

Filtrado de potenciales sacádicos
presentes en señales
electroencefalográficas mediante
análisis de componentes
independientes y aprendizaje no
supervisado

JESÚS PEÑA RODRÍGUEZ

UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE INGENIERÍAS FÍSICO-MECÁNICAS
ESCUELA DE INGENIERÍA ELÉCTRICA, ELECTRÓNICA Y DE TELECOMUNICACIONES.
BUCARAMANGA, SANTANDER, COLOMBIA
2016

Filtrado de potenciales sacádicos presentes en señales electroencefalográficas mediante análisis de componentes independientes y aprendizaje no supervisado

JESÚS PEÑA RODRÍGUEZ

TRABAJO DE GRADO PARA OPTAR AL TÍTULO DE
MAGÍSTER EN INGENIERÍA ELECTRÓNICA

DIRECTOR
DANIEL A. SIERRA BUENO, PH.D.

UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE INGENIERÍAS FÍSICO-MECÁNICAS
ESCUELA DE INGENIERÍA ELÉCTRICA, ELECTRÓNICA Y DE TELECOMUNICACIONES.
BUCARAMANGA, SANTANDER, COLOMBIA
2016

ESTA TESIS ES DEDICADA A MIS PADRES JOSÉ DE JESÚS Y ELOÍNA, A MI HIJO
ALEJANDRO Y A TODAS LAS PERSONAS QUE HICIERON PARTE DE ESTA HERMOSA
EXPERIENCIA.

Agradecimientos

Agradezco a mi familia la cual siempre me ha apoyado en las metas que me propongo. Agradezco a mis tutores Daniel Sierra y Carlos Conde, los cuales con su experiencia y calidad humana pasaron sólo de tener una responsabilidad académica conmigo, a que formáramos una gran amistad. Agradezco a mis compañeros de trabajo en el grupo Halley y en especial a Luis Nuñez el cual me tendió la mano en un momento que fue crucial para seguir en el camino hacia la búsqueda del conocimiento. También agradezco a mis compañeros del grupo NYC, Yesith Toloza y Manuel Mejía, por su colaboración. Finalmente agradezco a todas esas personas que he conocido en la «Ciudad Bonita», con las cuales he compartido instantes de alegría y de esparcimiento que siempre estarán en mi recuerdo.

Contenido

Introducción.	18
1 Electroencefalografía y artefactos oculares	20
1.1 Electroencefalografía	21
1.2 Movimientos oculares	27
1.3 💡 Influencia de los artefactos oculares en la EEG	30
1.4 💡 Base de datos EEG y EOG.	40
2 Separación de fuentes independientes	44
2.1 Separación ciega de fuentes	45
2.2 Análisis de componentes independientes	47
2.3 ICA por neguentropía.	52
2.4 Algoritmo FastICA	54
2.5 💡 ICA en señales EEG contaminadas con artefactos oculares. .	58
3 Extracción de características: Complejidad	66
3.1 Extracción de características	67
3.2 Características de complejidad	71
3.3 💡 Extracción de características discriminates entre EEG y artefactos oculares	75
4 Aprendizaje supervisado: SVM balanceada	78
4.1 Aprendizaje supervisado	79
4.2 Máquinas de vectores soporte	81
4.3 Máquinas de vectores soporte balanceadas	88
4.4 Evaluación de modelos de aprendizaje supervisado	89

4.5	💡 Eliminación de artefactos oculares en señales EEG mediante una SVM balanceada.	92
5	Aprendizaje no-supervisado: Agrupamiento SVM	104
5.1	Agrupamiento.	105
5.2	Agrupamiento SVM	108
5.3	💡 Eliminación de artefactos oculares en señales EEG mediante agrupamiento SVM Externo Híbrido.	125
6	Conclusiones	136
	Bibliografía	139

Lista de figuras

1.1	Posicionamiento de electrodos subdurales (corticales) y electrodos profundos [19].	21
1.2	Ejemplo de la gran variedad de morfologías de células nerviosas encontradas en el sistema nervioso humano [51].	22
1.3	Estructura de la neo-corteza humana [36].	23
1.4	Morfología de la sinapsis [51].	24
1.5	Diferentes tipos de ondas EEG	25
1.6	Sistema internacional 10-20 [65].	26
1.7	Musculatura responsable de los movimientos oculares, [21].	27
1.8	Parámetros para la caracterización de los movimientos sacádicos. La gráfica muestra la diferencia de potencial en la posición inicial y final del ojo. Los movimientos sacádicos presentan amplitudes menores a 30° y duraciones entre 30 y 120 ms, [23]. La línea azul representa la señal cruda y la línea roja es la señal después de pasar por un filtro de media con una ventana de 500 ms de ancho.	29
1.9	Localización de los electrodos en los registros de movimientos oculares verticales y horizontales, [5].	30
1.10	Propagación de los artefactos oculares sobre el cuero cabelludo. Cada sección muestra el potencial registrado en cada uno de los 22 canales EEG dispuestos sobre el cuero cabelludo, además, se indica el instante de tiempo de la señal EOG en el que ocurre la variación. El experimento toma cuatro muestras cada $\Delta t = 4ms$ y estas son respectivamente: a). Propagación en el tiempo t , b). Propagación en $t + \Delta t$, c). Propagación en $t + 2\Delta t$, d). Propagación en $t + 3\Delta t$	32

1.11	Ubicación del sujeto durante la prueba.	33
1.12	Prueba de calibración del registro VEOG para encontrar la relación entre el ángulo de movimiento del ojo y la amplitud de la señal ocular registrada.	34
1.13	Prueba de calibración del registro HEOG.	34
1.14	Señal EOG registrada durante el proceso de calibración.	35
1.15	Curva resultante de la regresión lineal de los datos de calibración.	36
1.16	Estímulo usado para generar el nistagmo optocinético.	37
1.17	Señales EEG y EOG registradas durante el nistagmo optocinético.	37
1.18	Estímulo usado para evaluar la contaminación por artefactos sacádicos en la EEG durante los procesos de lectura, [41]	38
1.19	Señales EEG y EOG registradas durante el proceso de lectura.	38
1.20	Muestra de las imágenes proyectadas durante el proceso de estimulación por percepción.	39
1.21	Señales EEG y EOG registradas durante el proceso de estimulación por percepción.	40
1.22	Etapas de una sesión de registro. En este caso se muestra el canal Pz.	41
1.23	Esquema de tiempo de una sesión.	41
1.24	Izquierda: Montaje correspondiente al sistema 10-20. Derecha: Montaje de los tres canales EOG monopolares.	42
1.25	Muestra del registro EEG y HEOG perteneciente al sujeto A01E en el rango de tiempo de 280 a 288 s.	42
2.1	Señales observadas provenientes de la mezcla de las fuentes originales.	46
2.2	Señales originales estimadas mediante BSS.	47
2.3	Comparación entre una distribución de probabilidad de Laplace, línea roja sólida, la cual es un ejemplo de distribución supergausiana y una distribución gaussiana, línea azul punteada.	49
2.4	Comparación entre una distribución de probabilidad de uniforme, línea roja sólida, la cual es un ejemplo de distribución subgausiana y una distribución gaussiana, línea azul punteada.	50
2.5	Distribución de laplace con $\mu = 0$ y $b = 1$ sin presencia de valores atípicos.	51
2.6	Distribución de laplace con $\mu = 0$ y $b = 1$ con presencia de un valor atípico igual a 10.	52
2.7	Histograma de una señal EEG tomada del canal FC3 del sujeto 1.	60
2.8	Histograma de una señal EOG del sujeto 1.	61

2.9	Señales EEG y HEOG registradas durante un periodo de 12 segundos. La influencia de los artefactos oculares es más notoria en los canales cercanos a la parte frontal del cráneo que en la parte occipital.	62
2.10	Ubicación de los canales EEG registrados.	62
2.11	Descomposición de las señales EEG mediante ICA en las fuentes independientes estimadas.	63
2.12	Proyección de los coeficientes a_{ij} sobre el cráneo.	64
2.13	Proyección de las componentes IC1 e IC2 sobre el cráneo.	64
2.14	Vista tridimensional de la proyección de las componentes independientes pertenecientes a los artefactos oculares.	65
3.1	Esquema del procedimiento de extracción de características desde los registros EEG.	75
3.2	Matriz de relación entre las características extraídas desde las componentes independientes. Los patrones azules pertenecen a la clase EEG y los patrones rojos a los artefactos oculares. Las características son: curtosis, Análisis de Fluctuación sin Tendencia (DFA), Información de Fisher (FI), Dimensión Fractal de Higuchi (HFD), Movilidad de Hjorth (HJM), Complejidad de Hjorth (HJC), Exponente de Hurst (HE) y Dimensión Fractal de Petrosian (PFD).	76
4.1	Efecto del sesgo y la varianza en un clasificador. En la parte superior-izquierda vemos una buena compensación entre el sesgo y la varianza debido a que el límite de decisión del clasificador, <i>círculo azul</i> , se adapta correctamente a los datos de prueba. En la parte superior-derecha existe una alta varianza en el clasificador, debido a que el límite de decisión no se adapta a la totalidad de los datos generando un aumento del error de generalización. En la parte inferior-izquierda el límite de decisión está sesgado de los datos de prueba, causando que el error de generalización aumente. En la parte inferior-derecha se muestra el peor de los casos, es decir, cuando el clasificador tiene una varianza alta y un sesgo alto, en este caso el error de generalización será mayor a los dos casos expuestos anteriormente.	80
4.2	Error de entrenamiento y de prueba en función de la complejidad del modelo, [28]	81
4.3	Hiperplano de separación entre dos clases separables linealmente.	83
4.4	Hiperplano de separación y variables de holgura en el caso no separable.	86

4.5	Mapeo de las muestras del espacio de entrada al espacio de Hilbert mediante una función <i>kernel</i> $\Phi(\mathbf{x})$	87
4.6	Espacio de la curva ROC. La línea roja indica el desempeño de un clasificador aleatorio, a medida que la curva avanza hacia la parte superior izquierda mejora el desempeño del estimador. . . .	91
4.7	Ejemplo de una iteración de la validación cruzada con $K = 5$. En este caso las particiones 1, 2, 3 y 5 son empleadas para entrenar el modelo y la partición 4 para validarlo.	92
4.8	Evaluación de la exactitud de una SVM desbalanceada y balanceada para diferentes valores de C y γ	94
4.9	Evaluación de la sensibilidad de una SVM desbalanceada y balanceada para diferentes valores de C y γ	95
4.10	Curva ROC para la SVM desbalanceada.	98
4.11	Curva ROC para la SVM balanceada	98
4.12	Influencia de los artefactos aculares en el registro EEG tomado del sujeto A01E en el rango temporal de 280-288 s. En la parte superior se muestran los 22 canales EEG registrados y en la parte inferior el EOG el cual se toma como referencia para observar el instante de tiempo en que ocurrieron los artefactos. Las frajas de color gris indican los artefactos de mayor amplitud.	100
4.13	Descomposición de los canales EEG, registrados en el sujeto A01E, en componentes independientes mediante ICA.	101
4.14	Eliminación de las componentes independientes de los artefactos oculares.	101
4.15	Canales EEG libres de artefactos oculares.	102
5.1	Etapas de un algoritmo de agrupamiento, [32]	106
5.2	Clasificación de los algoritmos de agrupamiento.	107
5.3	Formulación del agrupamiento interno. Esfera envolvente con centro a y radio R en el espacio <i>Kernel</i>	109
5.4	Algoritmo de agrupamiento SVM para un conjunto de 100 datos que está formado por dos clases distribuidas concéntricamente. En la parte superior se muestra el resultado para los parámetros $q = 1$ y $C = 1$, en la parte inferior para $q = 0,1$ y $C = 1$. Por otra parte, en la parte izquierda se muestra el mapeo de la esfera envolvente en el espacio de entrada y en la derecha los grupos formados. Estos resultados se obtuvieron mediante una librería desarrollada en <i>python</i> por el autor.	112

5.5	Algoritmo de agrupamiento SVM para un conjunto de 100 datos provenientes de dos clases distribuidas en forma de semicírculo para los parámetros $q = 1$ y $C = 1$. En la parte superior-izquierda se pueden ver las muestras, las BSVs y las SVs. En la parte superior-derecha se muestra el mapeo de la esfera envolvente en el espacio de entrada. En la parte inferior-izquierda se indican los grupos formados en el espacio de entrada. En la parte inferior-derecha se muestra en número de muestras por grupo, indicando que el 95 % de las muestras pertenecen a los grupos 1 y 2.	113
5.6	Algoritmo de agrupamiento SVM Re-etiquetado. En la parte superior-izquierda se muestra el etiquetado aleatorio de las muestras y el hiperplano de decisión producto de la convergencia de la SVM. En la parte superior-derecha, se realiza el procedimiento de re-etiquetado de las muestras mal clasificadas. En la parte inferior-izquierda, se ve el hiperplano resultante de la convergencia de la SVM sobre las muestras re-etiquetadas. En la parte inferior-derecha, se muestra el hiperplano de separación final después algunas iteraciones.	114
5.7	Grupos formados por el algoritmo de agrupamiento SVM <i>Relabeler</i> en un conjunto de 1000 muestras después de 13 iteraciones, con los parámetros $C = 1$ y $\gamma = 0,01$	118
5.8	MSE entre la salida esperada y la salida estimada por el modelo de agrupamiento.	119
5.9	Reducción del SSE a medida que el algoritmo aumenta la compacidad de los grupos con cada iteración.	119
5.10	Grupos formados por el algoritmo de agrupamiento SVM <i>Relabeler</i> en un conjunto de 1000 muestras después de 18 iteraciones, con los parámetros $C = 1$ y $\gamma = 0,01$	120
5.11	MSE entre la salida esperada y la salida estimada por el modelo de agrupamiento.	121
5.12	Reducción del SSE a medida que el algoritmo aumenta la compacidad de los grupos con cada iteración.	121

5.13	Algoritmo de agrupamiento SVM <i>Relabeler</i> sobre un conjunto desbalanceado de 1000 muestras. En la parte superior-izquierda se muestra la distancia Euclidiana entre cada una de las muestras pertenecientes a un grupo y su centroide. En la parte superior-derecha, se pueden ver los grupos en el espacio de entrada formados por el algoritmo de agrupamiento. En la parte inferior-izquierda, se muestra la evolución del MSE con cada iteración. En la parte inferior-derecha, se indica la reducción del SSE a medida que el algoritmo converge.	122
5.14	Comparación entre el algoritmo de agrupamiento SVM <i>Relabeler</i> e híbrido sobre un conjunto desbalanceado de 1000 muestras con los parámetros $C = 1$ y $\gamma = 0,01$. En la parte superior-izquierda se muestran los grupos formados por el método <i>Relabeler</i> . En la parte superior-derecha, el valor SSE respecto al número de iteraciones para el método <i>Relabeler</i> . En la parte inferior-izquierda, se muestran los grupos formados por el método híbrido. En la parte inferior-derecha, se indica la reducción del SSE respecto al número de iteraciones para el método híbrido	124
5.15	SSE y eficiencia vs el número de muestras pertenecientes al grupo de artefactos oculares.	126
5.16	Grupos formados por el agrupamiento SVM híbrido para un conjunto de 1540 muestras. En rojo se muestran las ICs de los artefactos oculares y en azul las ICs de las señales EEG. La matriz muestra la relación diádica entre las 8 características extraídas, mientras su diagonal, indica la función de distribución de las muestras en cada característica.	127
5.17	Reducción del MSE durante el proceso de convergencia.	129
5.18	Aumento de la compacidad de los grupos formados a medida que el SSE se reduce.	129
5.19	Canales EEG registrados del sujeto A06T en el rango temporal de 288-296 s. En la parte superior se muestran los 22 canales EEG registrados y en la parte inferior un canal EOG el cual sirve de referencia para saber el momento en que ocurren los movimiento oculares. Las franjas de color gris indican los instantes de tiempo donde hay contaminación por artefactos.	130
5.20	Descomposición de los canales EEG en sus componentes independientes mediante ICA. las franjas de color gris indican las ICs pertenecientes a los artefactos oculares.	131
5.21	Eliminación de las ICs de los artefactos oculares.	131
5.22	Señales EEG reconstruidas y sin influencia de los artefactos oculares.	132

5.23	Canales EEG registrados del sujeto A08T en el rango temporal de 300-308 s. En la parte superior se muestran los 22 canales EEG registrados y en la parte inferior un canal EOG.	133
5.24	Descomposición de los canales EEG en sus componentes independientes mediante ICA. las franjas de color gris indican las ICs pertenecientes a los artefactos oculares.	133
5.25	Eliminación de las ICs de los artefactos oculares.	134
5.26	Señales EEG reconstruidas y libres de la influencia de los artefactos oculares.	134

Resumen

Título: Filtrado de potenciales sacádicos presentes en señales electroencefalográficas mediante análisis de componentes independientes y aprendizaje no supervisado.

Autor: Jesús Peña Rodríguez

Palabras clave: Artefactos oculares, Análisis de Componentes Independientes, Máquinas de Vectores Soporte, electroencefalografía.

Descripción:

La electroencefalografía (EEG) es la técnica más común para el estudio de la dinámica neuronal del cerebro humano [29]. La EEG es una herramienta esencial en la neurociencia cognitiva, ya que permite el registro no invasivo de las señales eléctricas provenientes del cerebro. Desafortunadamente, las señales neuronales en la EEG son altamente contaminadas por fuentes eléctricas no-cerebrales tales como los músculos del cuero cabelludo y la cara, el ritmo cardíaco y la respiración [35].

Sin embargo, los movimientos oculares son la mayor fuente de contaminación del electroencefalograma. Esto ocurre debido a que los movimientos oculares causan un cambio en el campo eléctrico que rodea los ojos [17]. Luego, las variaciones de potencial causadas se desplazan sobre el cuero cabelludo superponiéndose con los potenciales provenientes de los cúmulos neuronales [20], por lo tanto, esto influye de manera negativa en el registro y posterior análisis de las señales EEG.

En el presente trabajo de investigación se plantea una técnica de filtrado de los artefactos oculares mediante el análisis de componentes independientes y agrupamiento con máquinas de vectores soporte.

Con el ánimo de identificar fácilmente los aportes de la tesis, se optó por marcar las respectivas entradas de la tabla de contenido con .

1

¹Trabajo de grado
Facultad de Ingenierías Físico-Mecánicas. Escuela de Ingeniería Eléctrica, Electrónica y de Telecomunicaciones. Director: Daniel A. Sierra Bueno, Ph.D. en Ingeniería Biomédica.

Abstract

Title: Filtering of saccadic potentials from electroencephalographic signals by means of independent component analysis and non-supervised learning.

Author: Jesús Peña Rodríguez

Keywords: Ocular artifacts, Independent Component Analysis, Support Vector Machines, electroencephalography.

Description:

Electroencephalography (EEG) is the most common technique to study neuronal dynamics in the human brain [29]. EEG is an essential tool in cognitive neuroscience, because it allows the non-invasive recording of electrical signals provided by brain.

Unfortunately, neuronal signals in EEG are strongly contaminated by non-cerebral electrical sources such as: muscles of scalp and face, cardiac rhythm and breathing [35].

However, ocular movements are the main source of noise in EEG. This occurs because ocular movements cause changes in the electric field around the eyes [17]. Then, the potential variations travel over the scalp and they are superposed with the potentials provided by neuronal cumulus [20], therefore, this affects in negative way in the recording and analysis of EEG signals.

This work proposes a filtering technique of ocular artifacts by means of independent component analysis and clustering with support vector machines.

In order to identify the thesis contributions easily, the entries in the table of contents were marked with . In this way, a condensed reading can be realized, focused in the procedure and the results.

²

²Bachelor Thesis
Facultad de Ingenierías Físico-Mecánicas. Escuela de Ingeniería Eléctrica, Electrónica y de Telecomunicaciones. Director: Daniel A. Sierra Bueno, Ph.D. in Biomedical Engineering.

Introducción

Actualmente uno de los grandes objetivos de la ciencia es entender el comportamiento cerebral. Para hacer esto se necesita obtener información acerca de cómo el cerebro reacciona a diferentes estímulos, situaciones o patologías. Por esta razón, se han desarrollado diferentes herramientas para el registro y posterior estudio de la dinámica neuronal, tales como: rayos-x (CAT), resonancia magnética (MRI), emisión de positrones (PET), espectroscopia infrarroja (NIRS) y electroencefalografía (EEG).

La EEG es la técnica más usada debido a su gran simplicidad y a su naturaleza no invasiva. El registro de datos mediante EEG se hace a través de electrodos superficiales, los cuales son dispuestos sobre el cuero cabelludo formando una matriz la cual permite establecer una resolución espacial en relación al número de electrodos conectados.

Los electrodos registran la actividad eléctrica producto de la superposición de la actividad de los cúmulos neuronales cercanos al electrodo. Las señales registradas poseen amplitudes muy pequeñas, lo que las hace sensibles al ruido provocado por fuentes no-cerebrales tales como: la respiración, el ritmo cardíaco, señales musculares y movimientos oculares.

Los artefactos oculares causados por el movimiento de los ojos son las principales fuentes de contaminación de los registros EEG; el globo ocular actúa como un dipolo que está sujeto por pequeños músculos que controlan su movimiento durante las tareas visuales. Este movimiento genera diferencias de potencial que se transmiten a través del cuero cabelludo desde la parte frontal hasta la parte occipital de la cabeza, contaminando así todo el registro EEG. El grado de afectación disminuye a medida que la posición del electrodo se aleja de las cavidades oculares en donde se ubica la fuente de ruido.

Debido a lo anterior, en la última década se han venido desarrollando diversas herramientas para eliminar los artefactos oculares de las señales EEG. Estos métodos no se basan en técnicas de filtrado comunes en el tiempo, la frecuencia o el tiempo-frecuencia ya que es impredecible el instante en el que se presentan y además su espectro de frecuencia se superpone con el de las señales EEG. Por

esta razón, los métodos propuestos se basan en la separación ciega de fuentes (BSS). La BSS hace una búsqueda de las posibles fuentes que se mezclaron para dar origen a las señales EEG observadas, este proceso se hace mediante la maximización de la independencia estadística entre las fuentes. En este caso se supone que los artefactos oculares provienen de una fuente y las señales EEG de otras.

Por otra parte, se han aplicado técnicas de aprendizaje automático para evitar que la clasificación de las fuentes (componentes) independientes se hagan de manera manual. En este caso, dos métodos se han tenido en cuenta, el aprendizaje supervisado y el no-supervisado.

En las técnicas de aprendizaje supervisado el modelo de clasificación es entrenado mediante unas muestras de las cuales se tiene conocimiento acerca de a que clase pertenecen, luego el modelo es validado con una porción de la población de muestras y de esta manera se evalúa la efectividad del clasificador. En este caso en particular, una máquina de soporte vectorial (SVM) es utilizada como clasificador en vez de otras técnicas como redes neuronales ANN debido a su mejor desempeño en este tipo de problemas como se muestra en [57]. En [52], [27] y [56] se usa una SVM como herramienta de clasificación de artefactos oculares en EEG, pero en estos casos no se tiene en cuenta el origen desbalanceado de los datos, lo cual puede repercutir en el desempeño del modelo de clasificación, el cual tendrá mayor preferencia por la clase que contiene mayor número de muestras, en este caso por las componentes EEG. Por esta razón, se plantea el uso de una SVM normalizada la cual tiene en cuenta el desbalanceo de los datos. En [54] se muestra el uso una SVM balanceada para la eliminación de artefactos oculares de la EEG.

Por otro lado, el aprendizaje no supervisado busca agrupaciones que se forman en los datos de acuerdo a una medida de similaridad. En este caso el modelo no necesita de información a priori acerca de la cantidad de muestras por cada clase, ni cual sería la clase más probable para una muestra.

En [72], [47] y [49] se usan algoritmos de separación de artefactos oculares basados en agrupamiento jerárquico, K-means y difuso respectivamente, obteniendo resultados favorables y validando el potencial de la herramientas no supervisadas en este tipo de problemas.

En este trabajo se plantea un nuevo método de separación de artefactos oculares basado en agrupamiento por SVM, el cual ha mostrado ser una herramienta muy confiable según [68], [70], [10], [9], [15] y [66].

Las neuronas son células de formas delicadas y elegantes, las misteriosas mariposas del alma, cuyo batir de alas quién sabe si esclarecerá algún día el secreto de la vida mental.

Santiago Ramón y Cajal

1

Electroencefalografía y artefactos oculares

La electroencefalografía es una técnica que usa electrodos superficiales ubicados sobre el cráneo para el registro de las señales eléctricas que se generan durante la actividad cerebral. Debido a su bajo costo y poco riesgo, este método es muy usado; sin embargo, es muy susceptible a fuentes no-cerebrales de ruido. La principal fuente de contaminación en los registros EEG son los artefactos oculares causados por el movimiento de los ojos.

En este capítulo se muestran conceptos básicos del origen de las señales eléctricas del cerebro, de electroencefalografía, de los artefactos oculares y su influencia en la EEG, así como de la base de datos usada durante el desarrollo de este proyecto.

1.1. Electroencefalografía

El estudio del comportamiento cerebral es un campo relativamente nuevo. Su inicio fue en el siglo XIX cuando la actividad eléctrica de fondo del cerebro en animales sin anestesia fue descrita cualitativamente, luego esta fue analizada de manera sistemática por el psiquiatra alemán Hans Berger, quien introdujo el término *electroencefalograma* (EEG) para denotar el registro de fluctuaciones del potencial del cerebro.

Convencionalmente, la actividad eléctrica del cerebro es registrada con tres tipos de electrodos: superficiales, corticales y profundos, Figura 1.1.

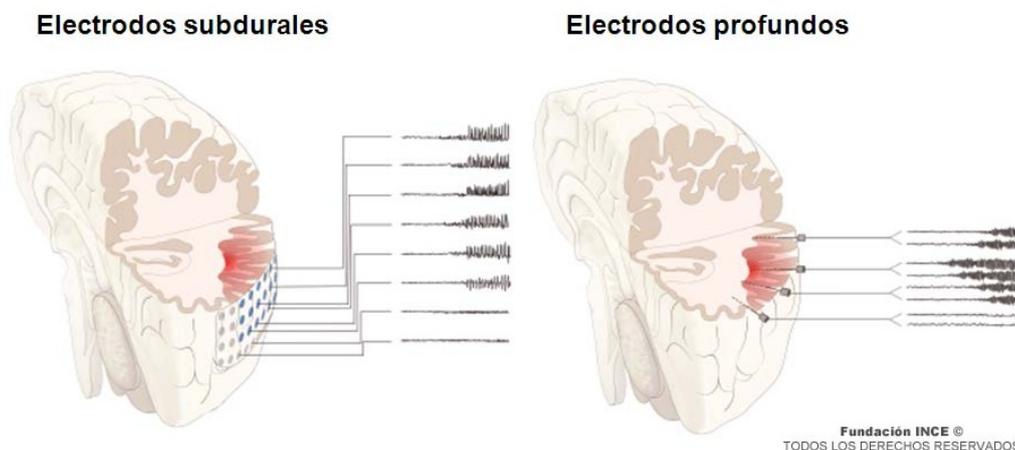


Figura 1.1: Posicionamiento de electrodos subdurales (corticales) y electrodos profundos [19].

Cuando los electrodos son situados sobre la superficie expuesta (corteza) del cerebro, los registros son llamados *electrocorticogramas* (ECoG). Por otra parte, finos electrodos de aguja con aislamiento de diversos diseños también pueden ser insertados en el tejido neural del cerebro, en tal caso el registro es llamado *registro de profundidad*. Ya sean obtenidos del cuero cabelludo, de la corteza cerebral o de la profundidad del cerebro, las fluctuaciones de potenciales registradas representan una superposición de los potenciales de campo producidos por una variedad de neuronas activas generadoras de corriente dentro del medio del volumen conductor [65].

Las fuentes que generan estos campos son elementos neuronales con interconexiones complejas. Los elementos neuronales mencionados son las dendritas,

el cuerpo celular (soma) y los axones de las células nerviosas. Además, la arquitectura del tejido neuronal cerebral no es uniforme de una localización a otra como se muestra en la Figura 1.2.

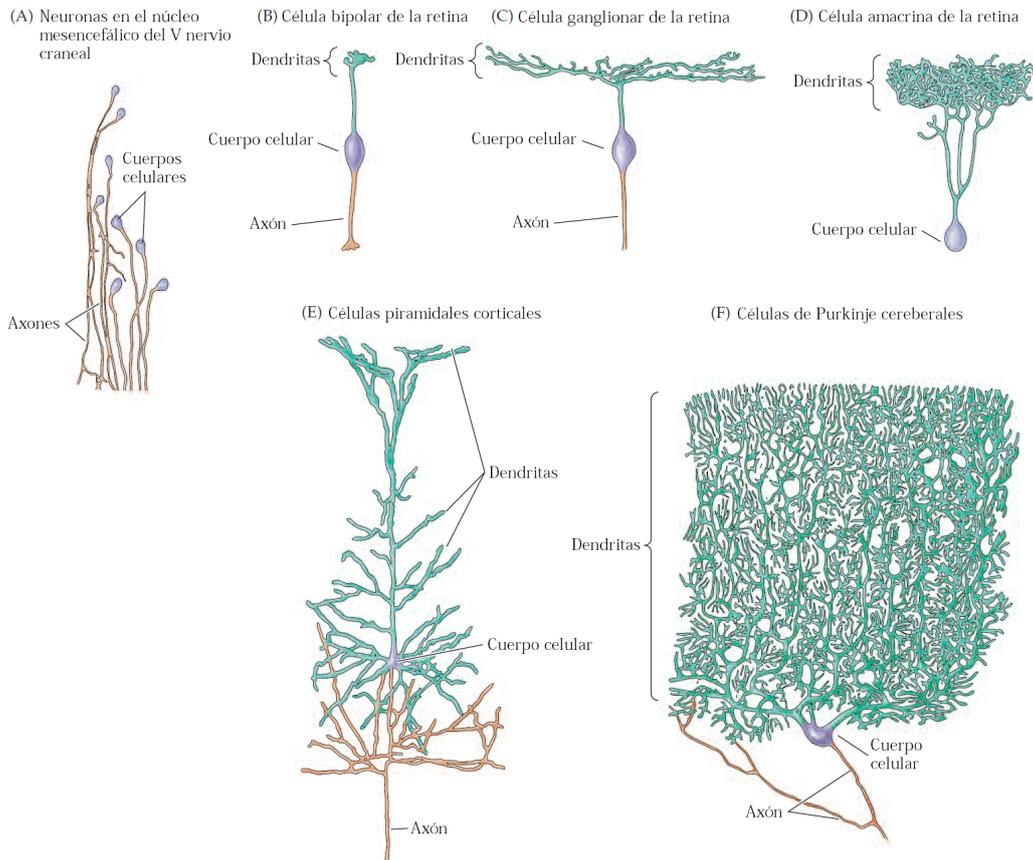


Figura 1.2: Ejemplo de la gran variedad de morfologías de células nerviosas encontradas en el sistema nervioso humano [51].

Debido a esta variabilidad, antes de emprender cualquier estudio detallado de EEG, primero se discutirá sobre la morfología cerebral.

1.1.1. Estructura de la corteza cerebral

La parte funcional del cerebro es la corteza cerebral, una capa relativamente delgada de materia gris (1.5 a 4.0 mm de espesor) que cubre la superficie externa del cerebro, incluyendo sus intrincadas convoluciones. Debido a que es la más reciente adquisición filogenética del cerebro, la corteza cerebral ha sido objeto

de un desarrollo relativamente mayor a otras partes del cerebro. El mayor avance en este crecimiento relativo ha sido la *neo-corteza*, la cual esta presente en la parte superior y lateral de los hemisferos cerebrales [65].



Figura 1.3: Estructura de la neo-corteza humana [36].

La corteza, claramente diferente, ubicada en la superficie media y en la base del cerebro es conocida como la *paleo-corteza*. Se usará la palabra *corteza* para referirnos específicamente a la neo-corteza.

Las arquitecturas corticales en los vertebrados comparten varias características comunes:

- Capas estratificadas que contienen los cuerpos celulares y haces de fibras.
- Una capa externa que carece de neuronas (capa I).
- Al menos una capa interior que contiene neuronas que dan lugar a grandes dendritas, que se elevan verticalmente a la capa I y viajan a través de esas capas formando múltiples ramas (arbolización).

La corteza cerebral humana está generalmente conformada por seis capas corticales, Figura 1.3. Las neuronas son principalmente de dos tipos: *piramidales* y *no-piramidales*. Hay también un gran número de capas de fibras nerviosas orientadas horizontalmente que se extienden entre regiones adyacentes de la corteza, así como haces de fibras orientados verticalmente que se extienden desde la corteza a las regiones más distantes de la corteza o hacia abajo para el tronco cerebral y la médula espinal [65].

1.1.2. Potenciales bioeléctricos del cerebro

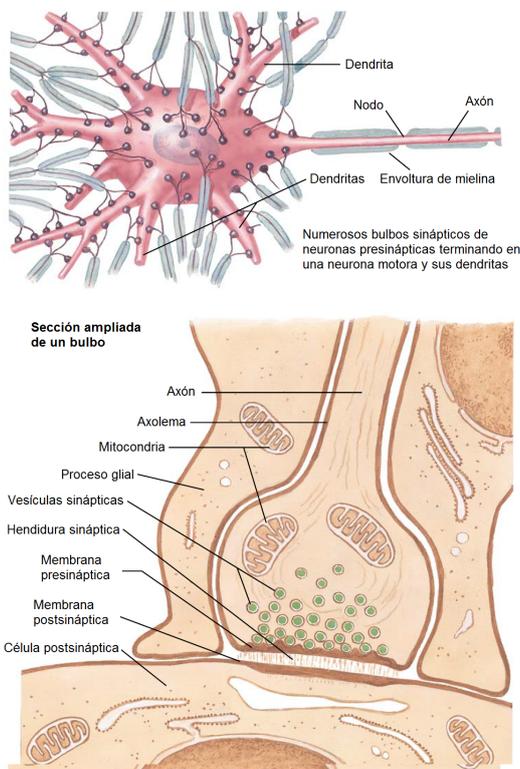


Figura 1.4: Morfología de la sinapsis [51].

La sinapsis es donde la transmisión de información de una neurona a otra toma lugar. La suma de las señales en el soma viaja al axón si este excede un cierto umbral en el cono axónico. En el final del axón (membrana presináptica), los potenciales eléctricos causan la difusión de químicos conocidos como *neurotransmisores* a través de la grieta sináptica, Figura 1.4; la cantidad de neurotransmisores que son difundidos a la membrana postsináptica causan un potencial eléctrico proporcional en la dendrita de la neurona conectada causando una señal de entrada al soma [48].

1.1.3. Ritmos cerebrales

Los registros eléctricos desde la superficie expuesta del cerebro o desde la superficie externa de la cabeza muestran oscilaciones continuas de actividad eléctrica dentro del cerebro. La intensidad y los patrones de esta actividad eléctrica son determinados en gran parte por la excitación general del cerebro resultante de funciones en el sistema de activación reticular del tronco cerebral (RAS). Las

ondulaciones en los potenciales eléctricos registrados son llamadas *ondas cerebrales* [65].

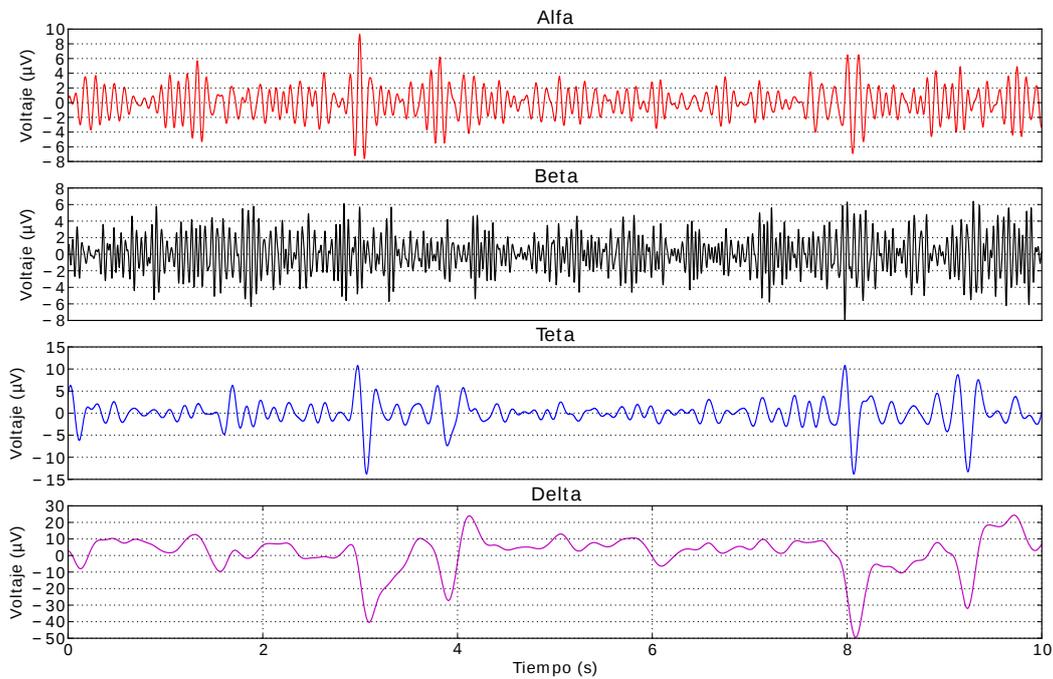


Figura 1.5: Diferentes tipos de ondas EEG

Las intensidades de las ondas cerebrales sobre la superficie del cerebro pueden llegar a $10mV$, mientras que los registrados en el cuero cabelludo tienen una amplitud más pequeña de aproximadamente $100\mu V$. Las frecuencias de estas ondas cerebrales van de 0.5 a 100 Hz, y su carácter es altamente dependiente del grado de actividad de la corteza cerebral.

Gran parte del tiempo, las ondas cerebrales son irregulares, y no pueden ser observados patrones generales. Sin embargo, en otras ocasiones, se producen distintos patrones. Algunos de estos son característicos de anomalías específicas del cerebro, tales como la epilepsia. Otros ocurren en personas normales y pueden ser clasificados como pertenecientes a uno de los cinco grupos de ondas (*alfa*, *beta*, *theta*, *delta* y *gamma*), Figura 1.5.

Las *ondas alfa* son ondas rítmicas que ocurren a una frecuencia entre 8 y 13 Hz. Se encuentran en el EEG de casi todas las personas normales cuando están despiertos en estado de reposo o descanso. Estas ondas ocurren con mayor intensidad en la región occipital pero también pueden ser registradas en las regiones parietal

y frontal del cuero cabelludo. Su valor es aproximadamente 2 a $20\mu V$, [65].

Cuando el sujeto está dormido, las ondas alfa desaparecen completamente. Cuando el sujeto despierta su atención es dirigida a algún tipo de actividad mental, entonces, las ondas alfa son remplazadas por señales asíncronas de mayor frecuencia pero menor amplitud.

Las *ondas beta* ocurren normalmente en el rango de frecuencia de 14 a 30 Hz, y algunas veces, particularmente durante actividad mental intensa llegan a 50 Hz. Estas son registradas frecuentemente en la región parietal y frontal, [65].

Las *ondas theta* tienen frecuencias entre 4 y 7 Hz. Estas ocurren principalmente en las regiones parietal y temporal en niños, pero también ocurren durante el estrés emocional en adultos, particularmente durante periodos de decepción y frustración, [65].

Las *ondas delta* incluyen todas las ondas en EEG por debajo de 3.5 Hz. Algunas veces estas ondas ocurren sólo una vez cada 2 o 3 segundos. Estas ocurren en el sueño profundo, en la infancia y en serias enfermedades cerebrales [65].

Las *ondas gama* poseen frecuencias entre 30 y 100 Hz. Estas se relacionan con los procesos cognitivos o de percepción, [53], [50].

1.1.4. El EEG clínico

El sistema comúnmente usado para situar los electrodos para monitorear la EEG clínica es el sistema de la Federación Internacional 10-20, Figura 1.6.

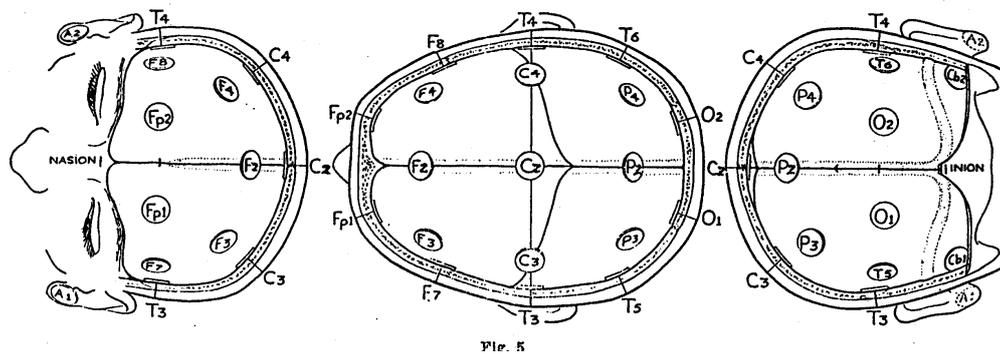


Figura 1.6: Sistema internacional 10-20 [65].

Este sistema usa ciertas referencias anatómicas para estandarizar el posicionamiento de los electrodos EEG. La representación de los canales EEG es referida como *montaje*.

En el montaje bipolar, cada canal mide la diferencia de potencial entre dos electrodos adyacentes.

En el montaje referencial, cada canal mide la diferencia de potencial entre un electrodo y un electrodo de referencia, dispuesto en la oreja.

En el montaje Laplaciano, cada canal mide la diferencia de potencial entre un electrodo y un promedio ponderado de los electrodos circundantes.

Los cambios de potencial registrados son amplificados y acondicionados para su posterior almacenamiento y análisis [65].

1.2. Movimientos oculares

Los movimientos oculares tiene como función principal orientar los ojos de forma que la imagen del objeto que interese al sujeto se proyecte en el área de la retina de mayor agudeza visual. Cada ojo puede moverse junto con la órbita en tres direcciones: verticalmente, horizontalmente y torsionalmente, [21]. En la Figura 1.7 se muestra el sistema de músculos que propician los movimientos oculares.

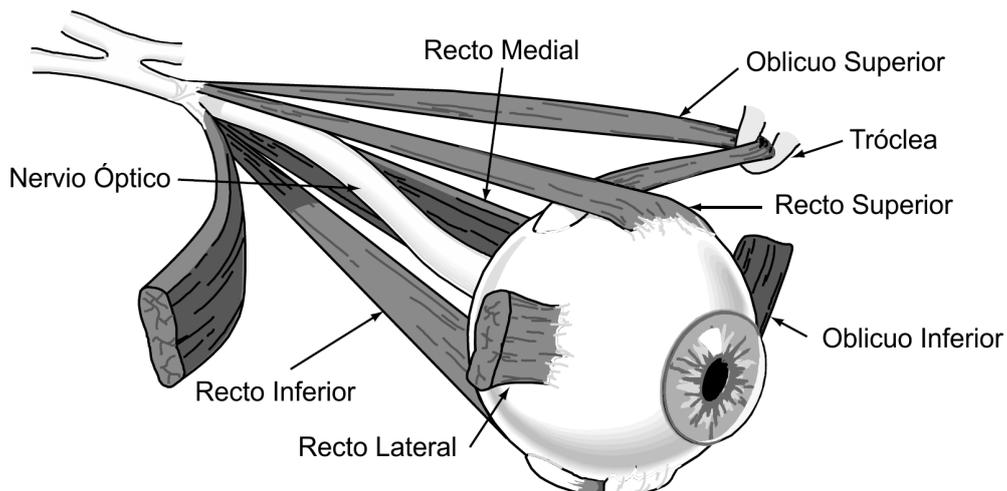


Figura 1.7: Musculatura responsable de los movimientos oculares, [21].

Existen varios tipos de movimientos dependiendo de su velocidad y su función: los movimientos *rápidos*, hacen que el globo ocular alcance en su desplazamiento una velocidad de hasta 700 grados/segundo. Por otro lado, existen dos tipos de movimientos rápidos, [5]:

- Las *microsacadas* son pequeños saltos de 0.25 grados o menos que ocurren una o dos veces por segundo. Por ejemplo, si una imagen se proyecta en el mismo lugar de la retina durante unos segundos, deja de percibirse. Las microsacadas evitan esta posibilidad.
- Los movimientos *sacádicos* permiten un cambio veloz del punto de fijación de la mirada. Durante este desplazamiento no se produce percepción visual. Además, se trata de un movimiento balístico, es decir, no se puede detener una vez ha comenzado.

A menudo un primer movimiento sacádico puede no alcanzar correctamente el nuevo punto de fijación, por lo que se producen pequeñas sacadas de ajuste.

El sistema oculomotor responde a estímulos visuales, auditivos y vestibulares, [21]. Durante la sacada, el sistema oculomotor opera en un modo de lazo abierto, después el sistema opera en modo de lazo cerrado para asegurar que los ojos lleguen a la posición correcta.

Los movimientos sacádicos son conjugados y balísticos, con una duración típica de 30-120 ms y una latencia de 100-300 ms, [21] [23].

Las principales características de los movimientos sacádicos se muestran en la Figura 1.8.

1.2.1. El electrooculograma

El registro de los movimientos oculares se realiza mediante un electrooculograma, EOG, el cual capta la diferencia de potencial entre la cornea y la retina.

El EOG clínico mide este cambio indirectamente situando los electrodos a los costados del ojo. Además, el EOG es la prueba más extendida para el estudio del funcionamiento del sistema vestibular.

Para obtener resultados reproducibles y comparables, la Sociedad Internacional de Electrofisiología Clínica de la Visión (ISCEV) ha definido recomendaciones para estandarizar las pruebas EOG, [40].

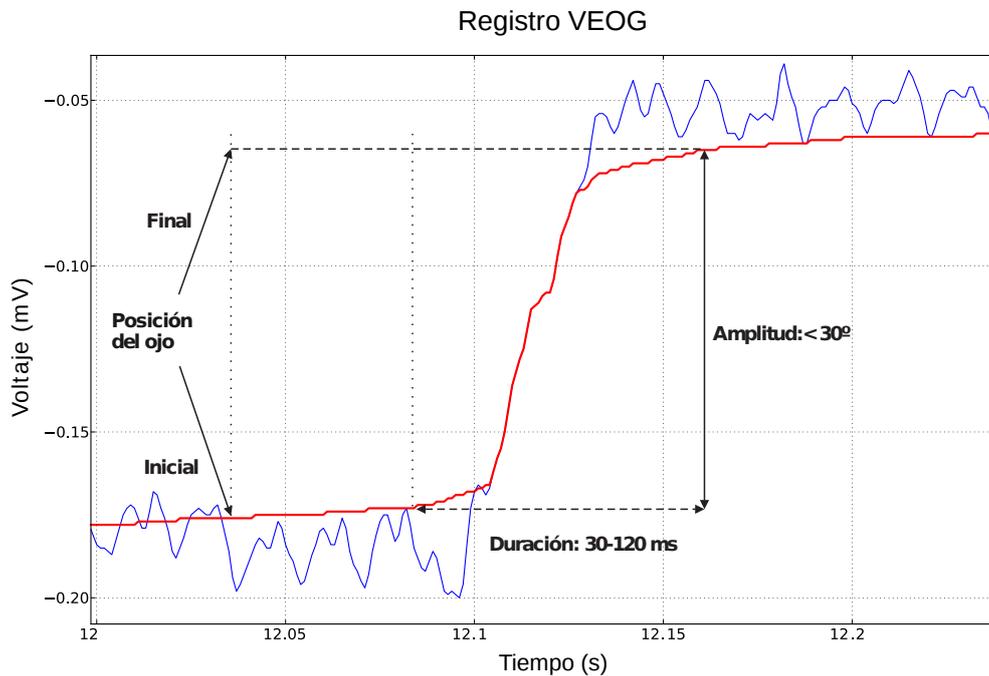


Figura 1.8: Parámetros para la caracterización de los movimientos sacádicos. La gráfica muestra la diferencia de potencial en la posición inicial y final del ojo. Los movimientos sacádicos presentan amplitudes menores a 30° y duraciones entre 30 y 120 ms, [23]. La línea azul representa la señal cruda y la línea roja es la señal después de pasar por un filtro de media con una ventana de 500 ms de ancho.

Por otro lado el EOG es ampliamente utilizado para tratar los artefactos oculares en el EEG. Otros métodos han sido empleados, tales como la reflexión retinal y las video cámaras, pero estas sólo son apropiadas para el rechazo de movimientos oculares grandes, por lo cual estos métodos son inadecuados para tratar los artefactos oculares.

Se cree que los diversos componentes de los artefactos oculares pueden ser registrados con las tres derivaciones EOG ortogonales: EOG vertical, horizontal y radial (*Elbert, 1985*). Para los movimientos oculares verticales se usa el VEOG. Este registra la diferencia de voltaje entre la parte superior e inferior del ojo. Este proceso de sustracción reduce los potenciales neuronales en el EOG, debido a que la cantidad de potencial neuronal alcanzando en E1 también alcanzará E3, y por lo tanto serán cancelados en el VEOG.

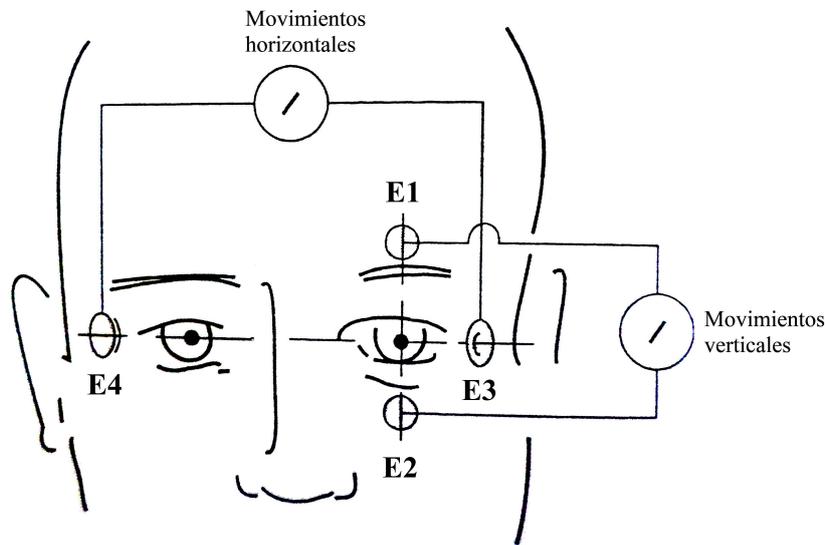


Figura 1.9: Localización de los electrodos en los registros de movimientos oculares verticales y horizontales, [5].

Los movimientos oculares horizontales son medidos con el EOG horizontal (HEOG), típicamente es la diferencia entre los voltajes en la parte exterior del canto izquierdo y derecho de los ojos, Figura 1.9. El componente radial es medido restando el voltaje promedio en los ojos del voltaje de el electrodo de referencia [17].

$$\frac{E_1 + E_2 + E_3 + E_4}{4} - referencia \quad (1.1)$$

Sin embargo, muchas variaciones y combinaciones de esas derivaciones son normalmente usadas, sin consenso en cuanto a la configuración óptima.

1.3. Influencia de los artefactos oculares en la EEG

Los movimientos oculares son la mayor fuente de contaminación del EEG. Esto es debido a que los movimientos oculares causan un cambio en los campos eléctricos que rodean los ojos, y eso distorsiona los campos eléctricos sobre el cuero cabelludo.

Los movimientos oculares no son sólo una preocupación teórica para los investigadores, ya que debido a la magnitud de los campos eléctricos que ellos provocan, estos afectan los resultados que son obtenidos. Por ejemplo, *Low, Borda, Frost y*

Kellaway (1966) notaron que en su estudio de variación negativa contingente (CNV), los potenciales evocados (ERP) en el cuero cabelludo fueron contaminados por artefactos oculares. Esta observación fue confirmada por *Hillyard y Galambos* (1970), quienes establecieron que en muchos de sus sujetos, una porción significativa de la CNV se debió al movimiento ocular, y no únicamente a potenciales neuronales como se había pensado previamente [17].

Con respecto a la EEG cuantitativa, en 1970 *Champan, Shelburne y Bragdon* ya estaban tratando de desentrañar los voltajes de los movimientos oculares de la banda alfa, mientras que más recientemente, el efecto de los movimientos oculares en EEG ha sido cuantificado rigurosamente.

Por ejemplo, *Gasser, Zieger y Gattaz* (1991) encontraron que los movimientos oculares incrementan la potencia de bandas de frecuencias bajas en pacientes esquizofrénicos, y también que disminuían la probabilidad de encontrar diferencias significativas entre grupos de control y grupos de esquizofrénicos.

Consecuentemente, los investigadores tomaron en cuenta el efecto que el movimiento ocular tiene sobre el EEG. Esto podría significar cualquier cosa, desde el rechazo de los datos obtenidos del sujeto, la reinterpretación de los datos, o aplicar los procedimientos de corrección necesarios. Además eso no puede tomarse a la ligera como muestra *Donchin* (1977) en las directrices para la publicación de investigaciones con ERP, donde se indica que «las medidas adoptadas para hacer frente a este problema deben ser consideradas en cualquier publicación» [17].

1.3.1. Patrón típico de los voltajes oculares a través del cuero cabelludo

Para describir el movimiento de los campos eléctricos generados por los movimientos oculares a través del cuero cabelludo, se usará el término *propagación*. El grado de propagación del potencial ocular a través del cuero cabelludo depende de un número de factores, tal como las propiedades de filtrado del cráneo, del cuero cabelludo y del tejido neuronal del sujeto. Sin embargo, como una aproximación, el voltaje en el cuero cabelludo tiende a ser inversamente proporcional al cuadrado de la distancia desde los ojos, de tal manera que los artefactos oculares decrecen rápidamente hacia localizaciones posteriores (*Rockstroh, Elbert, Birbaumer y Lutzenberger*, 1982), como se muestra en la Figura 1.10.

A menudo parece que no hay efecto de los movimientos oculares horizontales sobre sitios centrales, lo que sugiere que estos voltajes se cancelan entre sí y no

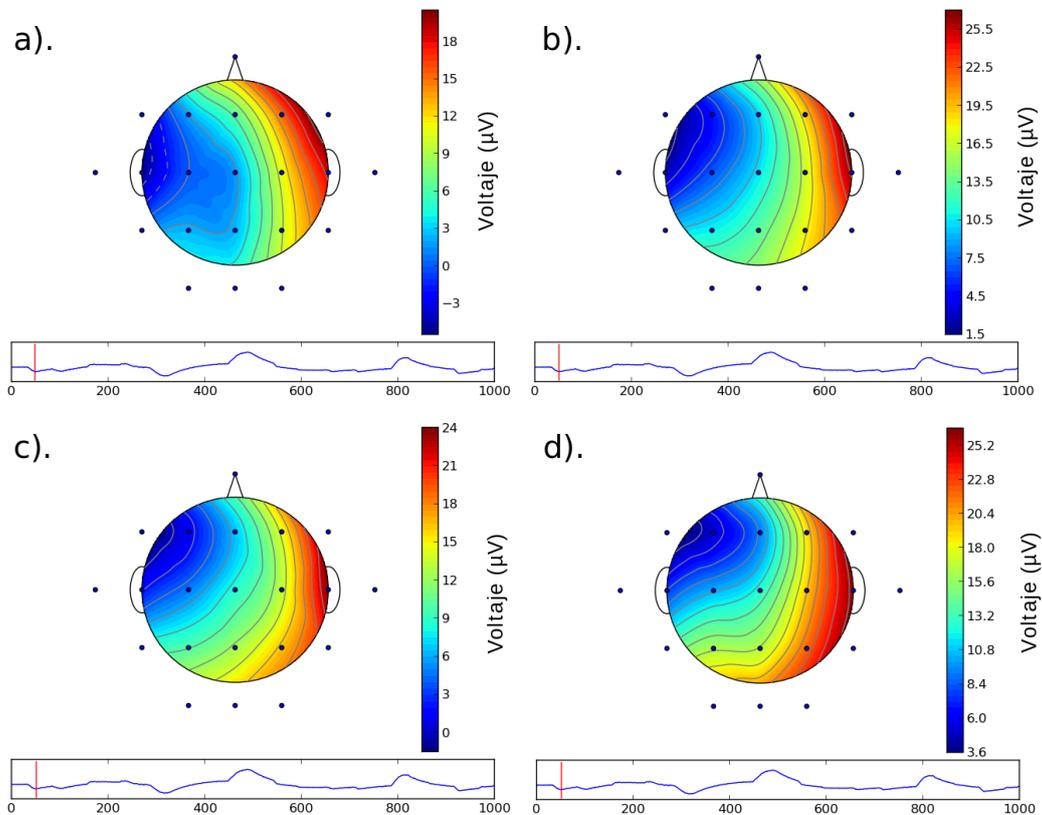


Figura 1.10: Propagación de los artefactos oculares sobre el cuero cabelludo. Cada sección muestra el potencial registrado en cada uno de los 22 canales EEG dispuestos sobre el cuero cabelludo, además, se indica el instante de tiempo de la señal EOG en el que ocurre la variación. El experimento toma cuatro muestras cada $\Delta t = 4ms$ y estas son respectivamente: a). Propagación en el tiempo t , b). Propagación en $t + \Delta t$, c). Propagación en $t + 2\Delta t$, d). Propagación en $t + 3\Delta t$.

afectan el EEG centralmente. Sin embargo contrario a los resultados que han sido reportados, no hay consenso al respecto en la actualidad. Una posibilidad es que las observaciones de artefactos de movimientos oculares horizontales centralmente pueden ser debido a errores en la estimación de la propagación, en lugar de las diferencias en la propagación.

Se sugiere esto porque diferentes métodos de estimación de la propagación aplicados al mismo conjunto de datos producen diferentes estimaciones de la contaminación de los movimientos oculares horizontales (simposio de Tilburg; *Brunia, Mocks y van den Berg-Lenssen* 1989), [18]. Así las diferencias pueden ser debido a

las metodologías de estimación en vez de la bio-electrodinámica de los diferentes sujetos.

1.3.2. Pruebas de laboratorio

Para comprobar el efecto de los movimientos oculares en los registros EEG se realizó una serie de pruebas en el laboratorio del Grupo de Investigación en Neurociencia y Comportamiento (NYC) de la Universidad Industrial de Santander, Colombia.

La prueba consistió en estimular los movimientos oculares de un sujeto mediante tareas visuales de seguimiento, lectura y exploración. De manera simultánea se registra las señales EEG provenientes de la zona frontal y occipital del cráneo, así como los potenciales oculares por medio de un electrooculograma vertical.

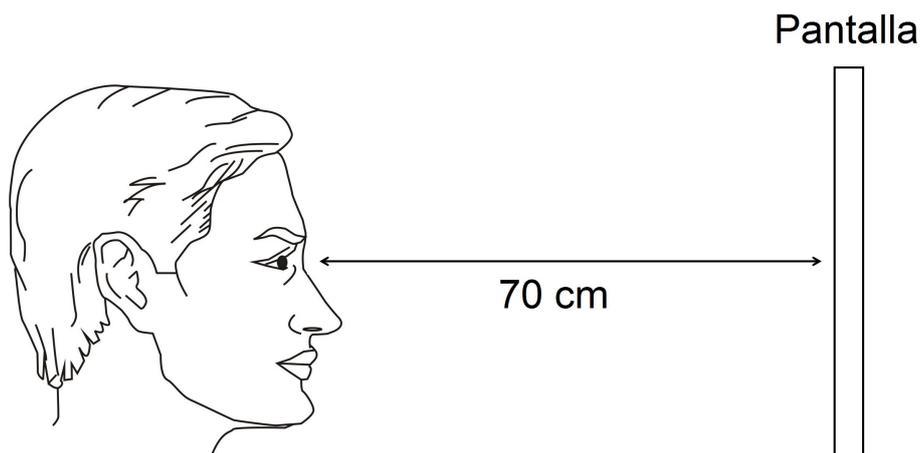


Figura 1.11: Ubicación del sujeto durante la prueba.

El equipo utilizado para el registro de las señales EEG y EOG fue el Powerlab 8/30 de ADInstruments. Por otra parte, la estimulación de los movimientos oculares se generó a través de una interfaz gráfica proyectada en una pantalla ubicada frente al sujeto a una distancia de 70 cm, Figura 1.11.

La prueba está constituida por dos etapas, la calibración y la estimulación.

La *calibración* consiste en la medición de la relación entre el ángulo del movimiento ocular y la amplitud de voltaje de la señal EOG registrada. Para esto se hace la proyección de un objetivo visual que cambia de posición de manera gra-

dual y que el sujeto debe seguir, Figura 1.12. El objeto visual tiene una posición inicial en el centro de la pantalla y luego realiza desplazamientos de 1.81, 5.44, 9.07 y 12.7 cm al lado derecho e izquierdo respectivamente, [7].

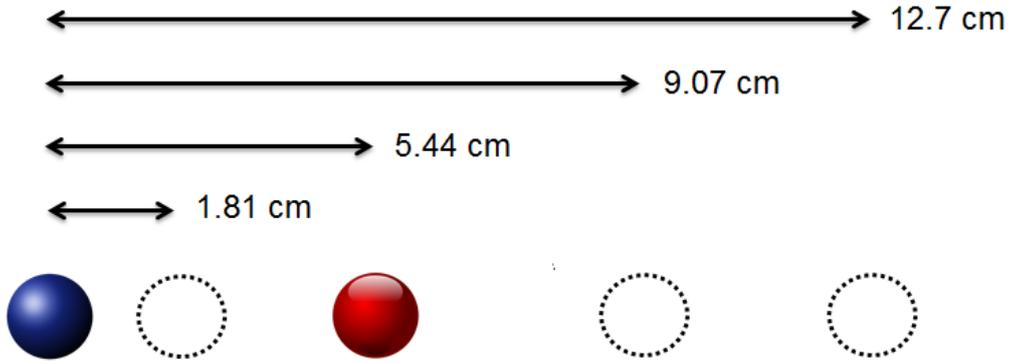


Figura 1.12: Prueba de calibración del registro VEOG para encontrar la relación entre el ángulo de movimiento del ojo y la amplitud de la señal ocular registrada.

El desplazamiento del objeto se puede calcular mediante la ecuación:

$$D_r = \frac{D_p T}{R} \quad (1.2)$$

donde, R es la resolución de la pantalla, T es el tamaño de la pantalla, D_p es la distancia en píxeles y D_r es la distancia real.

El ángulo α de la sacada, Figura 1.13, se puede determinar con:

$$\alpha = \arctan \frac{D_r}{D} \quad (1.3)$$

donde, D es la distancia desde los ojos del sujeto hasta el objetivo visual.

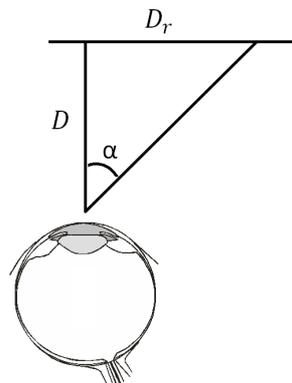


Figura 1.13: Prueba de calibración del registro HEOG.

La relación entre el desplazamiento del objeto y el ángulo de la sacada se muestra en la Tabla 1.1.

D_r (cm)	1.81	5.44	9.07	12.7
α (Grados)	1.48	4.44	7.38	10.28

Tabla 1.1: Relación entre D_r y α .

Las señales EOG registradas en el proceso de calibración se muestra en la Figura 1.14.

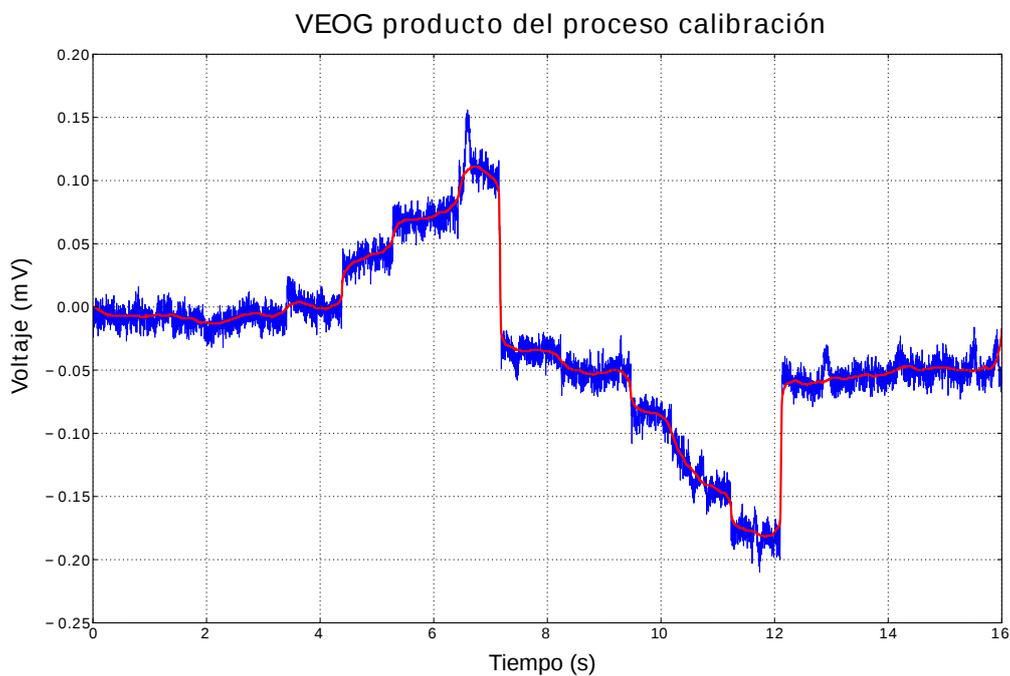


Figura 1.14: Señal EOG registrada durante el proceso de calibración.

Mediante esta información se puede determinar la relación entre el ángulo de la sacada y el voltaje registrado, Tabla 1.2.

α (Grados)	-10.28	-7.38	-4.44	-1.48	0	1.48	4.44	7.38	10.28
V_{sac} (V)	-0.178	-0.141	-0.084	-0.053	-0.007	0.004	0.03	0.069	0.111

Tabla 1.2: Relación entre el voltaje registrado en el EOG V_{sac} y el tamaño de la sacada α .

Luego se ajustaron los datos a una recta mediante regresión por mínimos cuadrados Figura 1.15 y se determinó la ecuación de calibración:

$$\alpha = \frac{V_{sac} - b}{m} \quad (1.4)$$

donde m es la pendiente y b el intercepto. En este caso m es 0.014 Voltios/Grado y b es 1.2 Voltios.

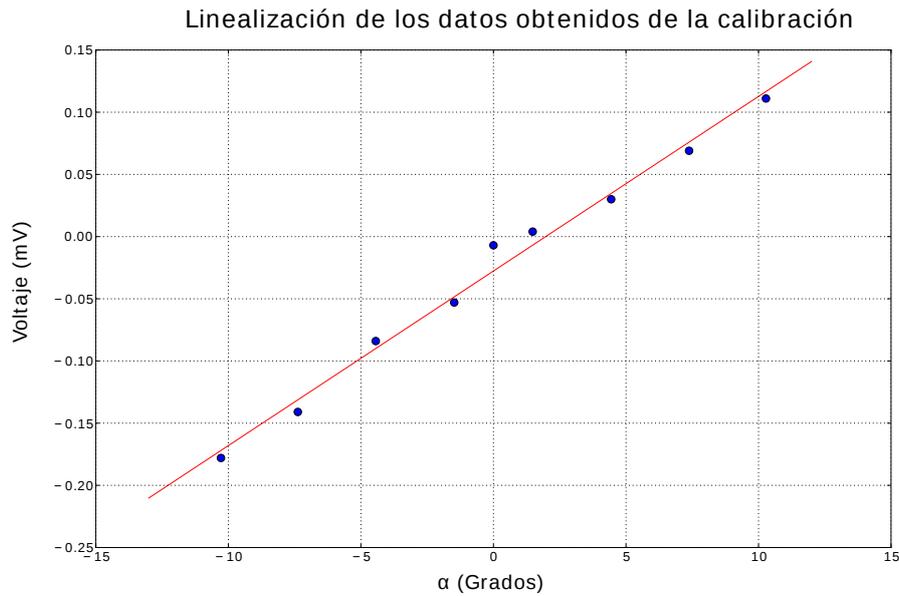


Figura 1.15: Curva resultante de la regresión lineal de los datos de calibración.

La etapa de *estimulación* se hizo de tres maneras, primero se generó un **nigstagmo optocinético** [51] el cual consiste en la alternación entre movimientos rápidos y lentos de los ojos. El estímulo consiste de una serie de barras verticales de color blanco y negro que se desplazan hacia la derecha a una velocidad de 60 pixeles/s, los ojos siguen automáticamente las franjas y el sujeto trata de fijar la mirada en el centro de la pantalla, este proceso genera sacadas rápidas en sentido contrario al movimiento de las barras. Las barras proyectadas en el experimento se puede ver en la Figura 1.16

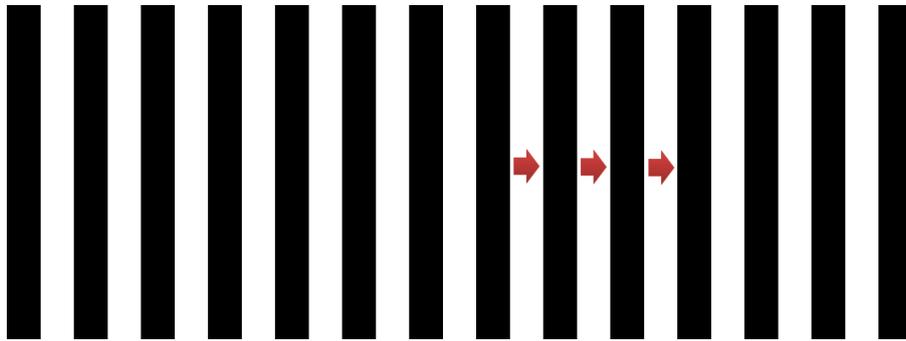


Figura 1.16: Estímulo usado para generar el nistagmo optocinético.

Las señales EEG y VEOG que se obtuvieron en esta prueba se muestran en la Figura 1.17. En esta gráfica se puede observar que en el instante que se generan las sacadas, el registro de la actividad cerebral en la parte frontal del cráneo se ve seriamente afectada, en cambio en la parte occipital la afectación es imperceptible. Con ello se concluye que cuando el sujeto realiza tareas de seguimiento se producen movimientos sacádicos que contaminan los registros EEG de manera gradual desde los ojos a partes posteriores del cráneo.

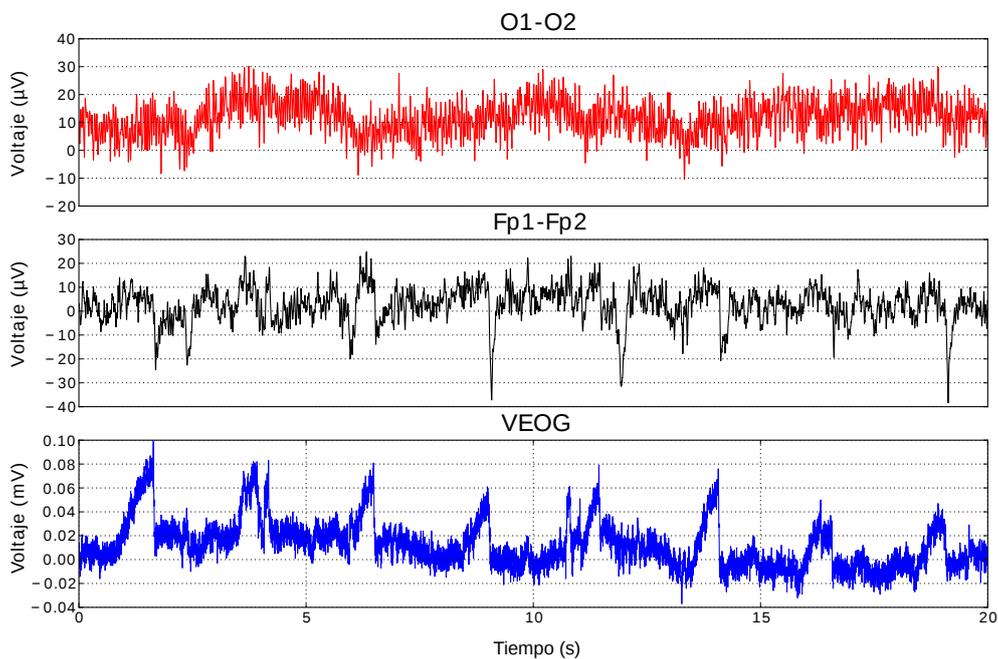


Figura 1.17: Señales EEG y EOG registradas durante el nistagmo optocinético.

En el segundo estímulo el sujeto debe leer un párrafo que es proyectado en la pantalla, Figura 1.17. El resultado de esta prueba muestra que cuando el sujeto

hace cambio de renglón durante la lectura se generan movimientos sacádicos amplios; estos movimientos afectan los registros EEG frontales generando picos de contaminación que alcanzan 3 veces la amplitud normal de las señales, Figura 1.19.

Muchos años después, frente al pelotón de fusilamiento, el coronel Aureliano Buendía había de recordar aquella tarde remota en que su padre lo llevó a conocer el hielo. Macondo era entonces una aldea de veinte casas de barro y cañabrava construidas a la orilla de un río de aguas diáfanas que se precipitaban por un lecho de piedras pulidas, blancas y enormes como huevos prehistóricos. El mundo era tan reciente, que muchas cosas carecían de nombre, y para mencionarlas había que señalarlas con el dedo. Todos los años, por el mes de marzo, una familia de gitanos desarrapados plantaba su carpa cerca de la aldea, y con un grande alboroto de pitos y timbales daban a conocer los nuevos inventos. Primero llevaron el imán. Un gitano corpulento, de barba montaraz y manos de gorrión, que se presentó con el nombre de Melquiades, hizo una truculenta demostración pública de lo que él mismo llamaba la octava maravilla de los sabios alquimistas de Macedonia. Fue de casa en casa arrastrando dos lingotes metálicos, y todo el mundo se espantó al ver que los calderos, las pailas, las tenazas y los anafes se caían de su sitio, y las maderas crujían por la desesperación de los clavos y los tornillos tratando de desenclavarse, y aun los objetos perdidos desde hacía mucho tiempo aparecían por donde más se les había buscado, y se arrastraban en desbandada turbulenta detrás de los fierros mágicos de Melquiades.

Figura 1.18: Estímulo usado para evaluar la contaminación por artefactos sacádicos en la EEG durante los procesos de lectura, [41]

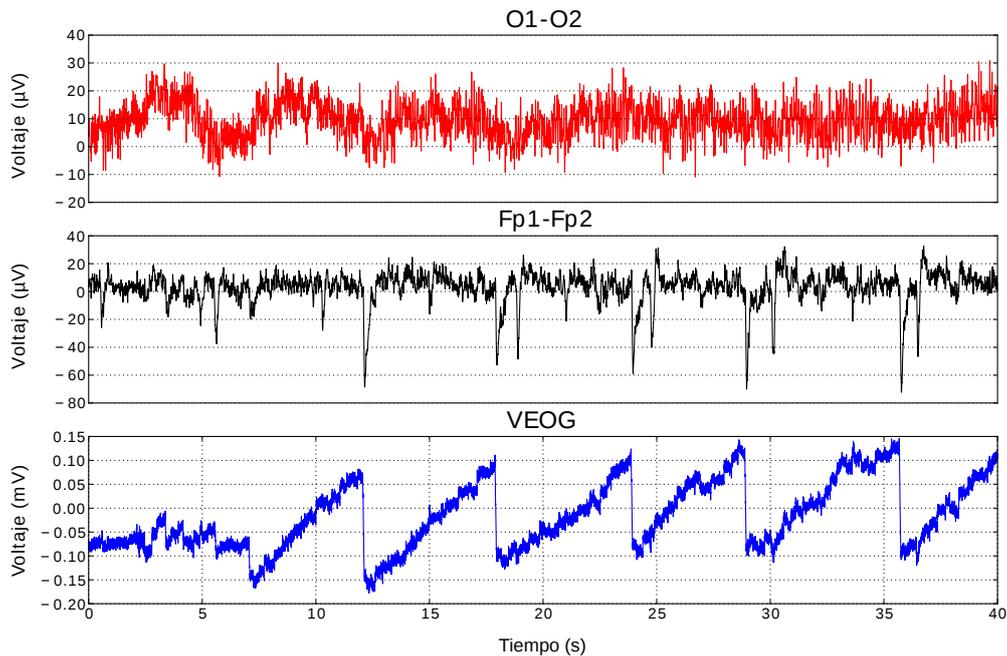


Figura 1.19: Señales EEG y EOG registradas durante el proceso de lectura.

Finalmente, se evaluó la generación de movimientos oculares durante tareas de percepción. Para esto, el sujeto es estimulado con una serie de imágenes que son proyectadas de manera aleatoria durante 3 segundos en la pantalla, Figura

1.20. Las imágenes usadas fueron tomadas de la base de datos ALOI creada para propósitos científicos [6]. Esta está conformada por 750 imágenes de objetos reales con un tamaño de 769 x 576 píxeles y con fondo negro.

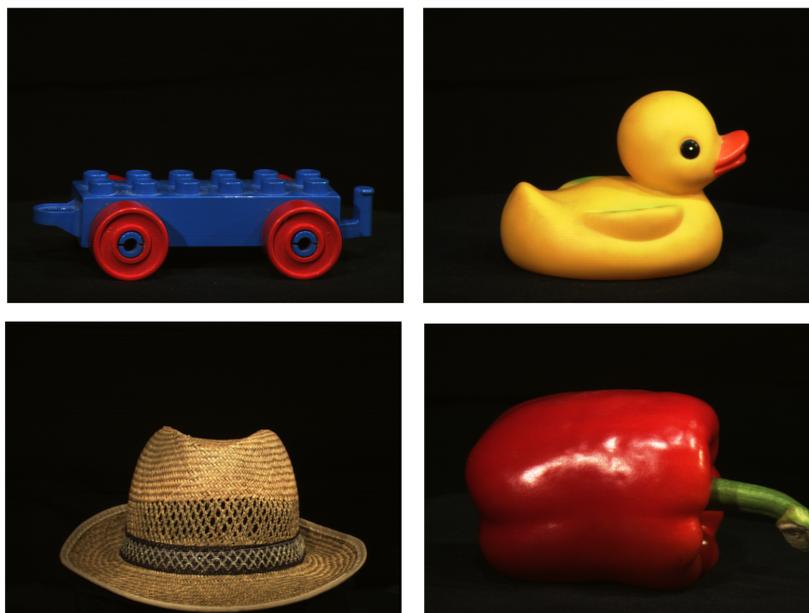


Figura 1.20: Muestra de las imágenes proyectadas durante el proceso de estimulación por percepción.

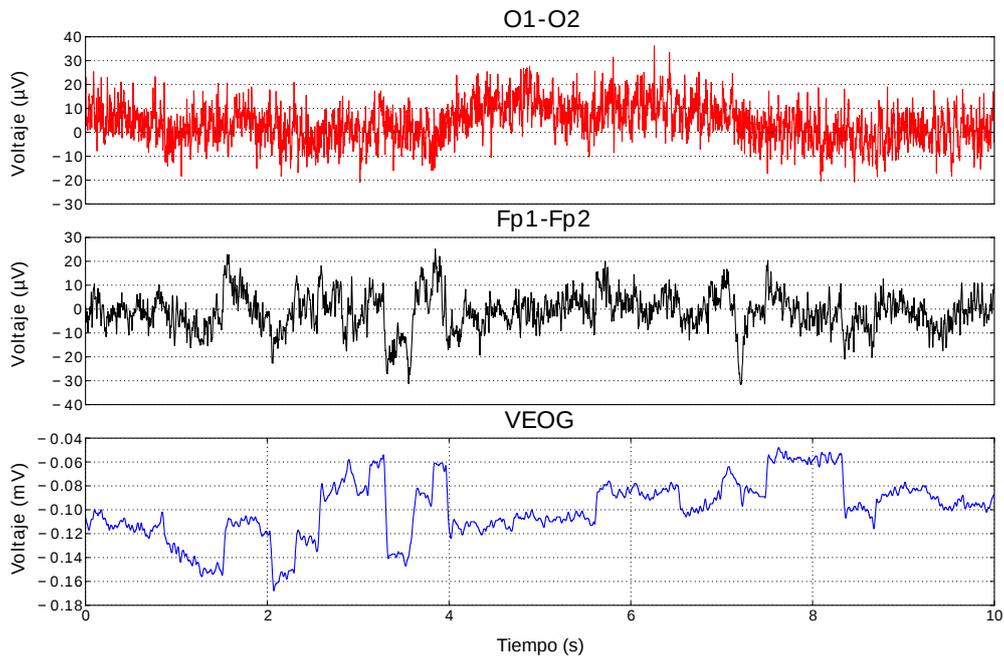


Figura 1.21: Señales EEG y EOG registradas durante el proceso de estimulación por percepción.

1.4. Base de datos EEG y EOG

Una vez el efecto de los artefactos oculares en los registros EEG fue comprobado, se presenta la base de datos con la cual se desarrolló el trabajo.

La base de datos usada fue suministrada por el *Institute for Knowledge Discovery* in the Graz University of Technology, Austria. Esta consiste del registro EEG de nueve sujetos mientras realizan cuatro diferentes tareas de movimientos imaginarios: la imaginación del movimiento de la mano izquierda, la mano derecha, ambos pies y la lengua. Dos sesiones en diferentes días fueron registradas para cada sujeto. Cada sesión esta compuesta de 6 registros separados por descansos cortos, Figura 1.22. Cada registro consiste de 48 ensayos (12 para cada una de las cuatro tareas) conformando un total de 288 ensayos por sesión, [14].

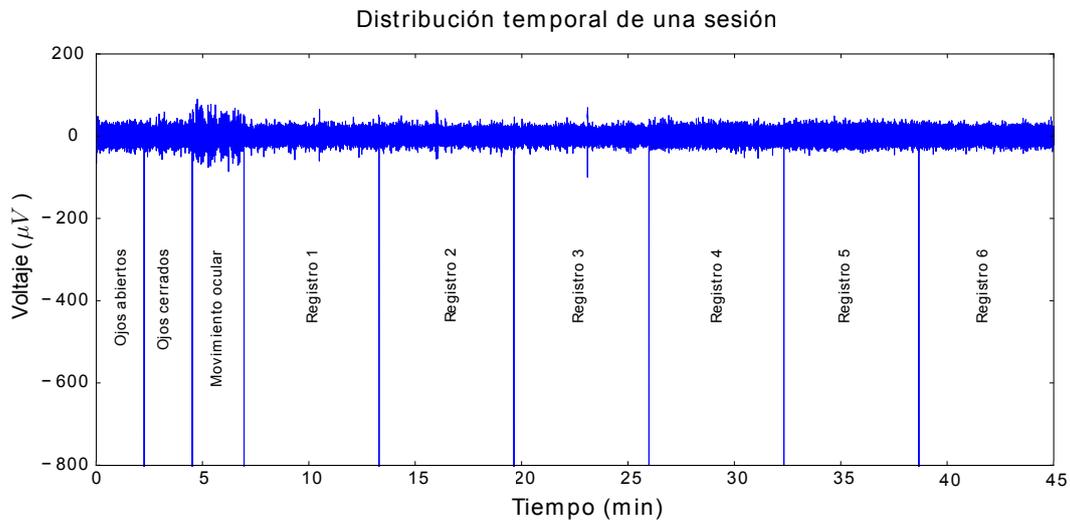


Figura 1.22: Etapas de una sesión de registro. En este caso se muestra el canal Pz.

Al inicio de cada sesión, un registro de aproximadamente 5 minutos fue hecho para determinar la influencia de los artefactos oculares. Estos registros a su vez se dividen en tres bloques: (1) dos minutos con los ojos abiertos mirando fijamente a la pantalla, (2) dos minutos con los ojos cerrados y (3) un minuto con movimiento ocular. El esquema de tiempo de una sesión se muestra en la Figura 1.23.

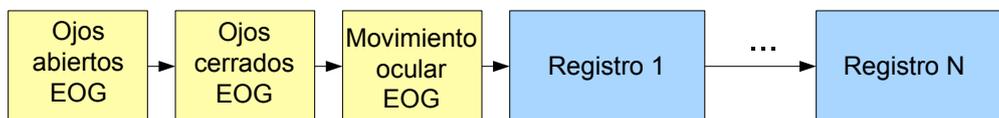


Figura 1.23: Esquema de tiempo de una sesión.

1.4.1. Registro de los datos

El registro EEG fue hecho mediante el uso de 22 electrodos con una distancia de separación entre ellos de 3.5 cm, el montaje se muestra en la Figura 1.24. Todas las señales fueron registradas de manera monopolar con el mastoides izquierdo como referencia y el mastoides derecho como tierra. Las señales fueron muestreadas a 250 Hz y filtradas entre 0.5 Hz y 100 Hz. La sensibilidad del amplificador fue configurada a $100 \mu\text{V}$. Adicionalmente se implementó un filtro notch de 50 Hz para eliminar el ruido de línea. Los canales EEG registrados son: Fz, FC3, FC1, FCz, FC2, FC4, C5, C3, C1, Cz, C2, C4, C6, CP3, CP1, CPz, CP2, CP4, P1, Pz, P2 y POz.

Además de los 22 canales EEG, 3 canales EOG monopolares fueron registrados con las mismas condiciones anteriormente mencionadas. En este caso la sensibilidad del amplificador fue configurada a 1 mV.

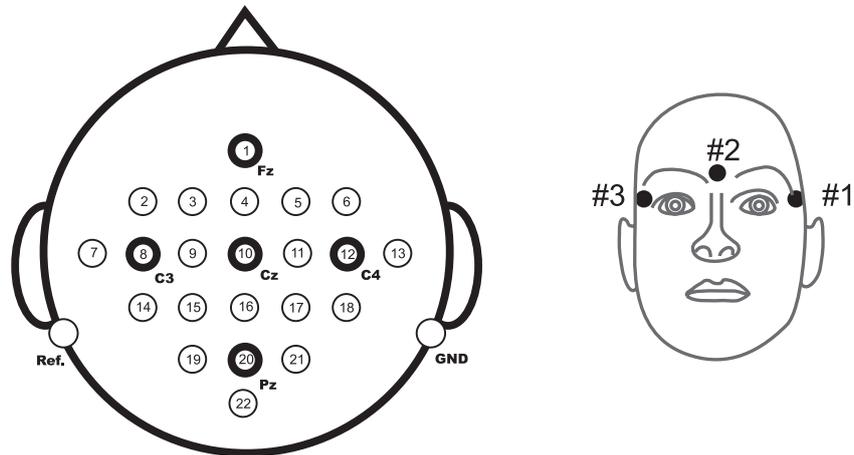


Figura 1.24: Izquierda: Montaje correspondiente al sistema 10-20. Derecha: Montaje de los tres canales EOG monopolares.

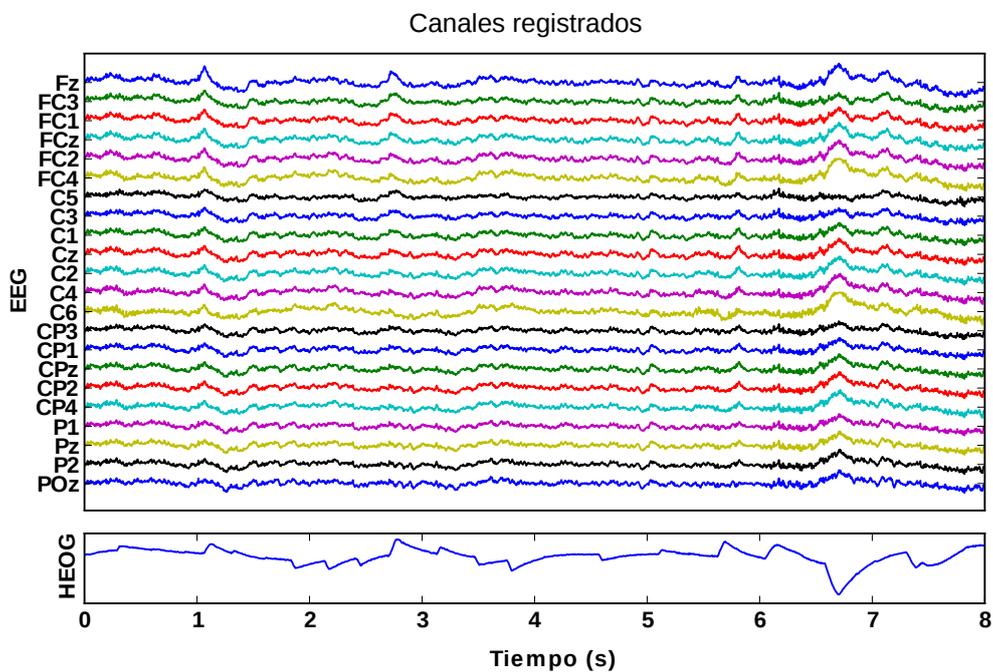


Figura 1.25: Muestra del registro EEG y HEOG perteneciente al sujeto A01E en el rango de tiempo de 280 a 288 s.

La señal del EOG horizontal fue calculada por medio de los canales EOG monopolares registrados de la siguiente manera:

$$HEOG = EOG_I - EOG_D \quad (1.5)$$

donde, EOG_I es el canal EOG registrado en el canto izquierdo y EOG_D es el canal EOG registrado en el canto derecho de los ojos . Finalmente, en la Figura 1.25 se muestra un ejemplo de las señales electroencefalográficas y electrooculográficas.

La ciencia no es sino una perversión de sí misma a menos que tenga como objetivo final el mejoramiento de la humanidad.

Nikola Tesla

2

Separación de fuentes independientes

La EEG es una técnica que registra la dinámica eléctrica cerebral mediante electrodos dispuestos en la superficie del cuero cabelludo; sin embargo, estos registros son afectados por señales provenientes de fuentes de origen no-cerebral dando como resultado una mezcla de señales neuronales y de artefactos generados por diversos fenómenos fisiológicos. Para poder eliminar la influencia de estos artefactos en la señales EEG se debe, como primer paso, descomponer las señales registradas en las fuentes originales. Por tal motivo, en este capítulo se establecen los conceptos básicos de la separación ciega de fuentes (BSS) y del análisis de componentes independientes (ICA), además, se muestra el modelo ICA mediante la maximización de la no-gaussianidad y el algoritmo FastICA. Final-

mente, se aplica ICA a registros EEG afectados por artefactos oculares.

2.1. Separación ciega de fuentes

La separación ciega de fuentes es un concepto muy usado para descomponer señales mezcladas que provienen de fuentes independientes. Para entender esto, se considera un número de señales generadas por algún fenómeno físico, por ejemplo, señales de voz, señales cerebrales o señales de radio. Al otro lado una serie de sensores dispuestos en diferentes ubicaciones registran una mezcla de las señales originales con un determinado peso.

Ahora bien, las señales observadas se denotan como $x_1(t), x_2(t), \dots, x_n(t)$ las cuales son resultado de la suma ponderada de las fuentes originales $s_1(t), s_2(t), \dots, s_m(t)$. Los coeficientes a_{ij} representan el peso de cada una de las observaciones $s_j(t)$ y es establecido por diferentes factores como por ejemplo la distancia entre los sensores. El modelo resultante es:

$$\begin{aligned} x_1(t) &= a_{11}s_1(t) + a_{12}s_2(t) + \dots + a_{1m}s_m(t) \\ x_2(t) &= a_{21}s_1(t) + a_{22}s_2(t) + \dots + a_{2m}s_m(t) \\ &\vdots \\ x_n(t) &= a_{n1}s_1(t) + a_{n2}s_2(t) + \dots + a_{nm}s_m(t) \end{aligned} \tag{2.1}$$

En este caso los coeficientes a_{ij} y las fuentes originales $s_j(t)$ son desconocidos debido a que no se tiene conocimiento del sistema físico de mezclado ni se pueden registrar las señales originales de manera directa.

Ahora se deben hallar las señales originales mediante las observaciones $x_1(t), x_2(t), \dots, x_n(t)$, para ello se asume que los coeficientes de mezcla a_{ij} crean una matriz invertible. Es decir, existe una matriz W con coeficientes w_{ji} , que pueden separar las observaciones $s_j(t)$ como:

$$\begin{aligned} s_1(t) &= w_{11}x_1(t) + w_{12}x_2(t) + \dots + w_{1n}x_n(t) \\ s_2(t) &= w_{21}x_1(t) + w_{22}x_2(t) + \dots + w_{2n}x_n(t) \\ &\vdots \\ s_m(t) &= w_{m1}x_1(t) + w_{m2}x_2(t) + \dots + w_{mn}x_n(t) \end{aligned} \tag{2.2}$$

Luego se hallan los coeficientes w_{ji} , en este caso se usará una solución sencilla basada en la independencia estadística de las señales. De hecho, si las señales son no-gaussianas, está será suficiente para determinar los coeficientes w_{ji} , entonces

se tienen las señales:

$$\begin{aligned}y_1(t) &= w_{11}x_1(t) + w_{12}x_2(t) + \cdots + w_{1n}x_n(t) \\y_2(t) &= w_{21}x_1(t) + w_{22}x_2(t) + \cdots + w_{2n}x_n(t) \\&\vdots \\y_m(t) &= w_{m1}x_1(t) + w_{m2}x_2(t) + \cdots + w_{mn}x_n(t)\end{aligned}\tag{2.3}$$

Si $y_1(t), y_2(t), \dots, y_m(t)$ son independientes, estas serán iguales a las fuentes originales $s_1(t), s_2(t), \dots, s_m(t)$ salvo por algún factor de escala.

En la Figura 2.1 se muestra un conjunto de señales observadas las cuales se asume son resultado de la mezcla de algunas fuentes. Luego, se aplicó el concepto expuesto anteriormente de BSS y se estimaron las fuentes originales de la Figura 2.2.

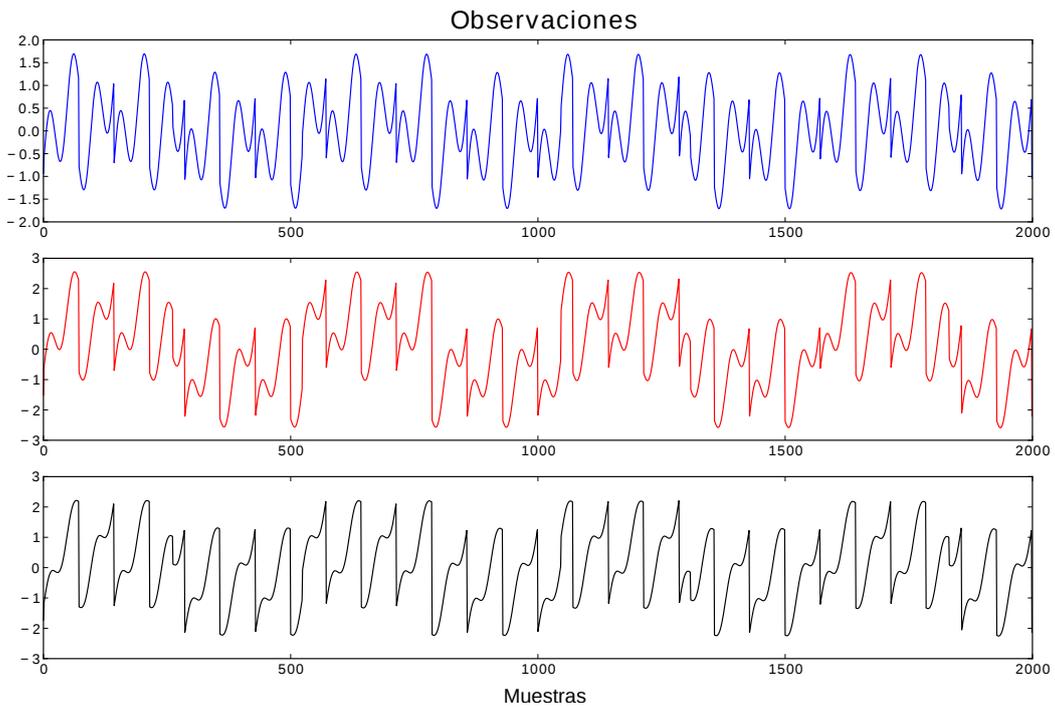


Figura 2.1: Señales observadas provenientes de la mezcla de las fuentes originales.

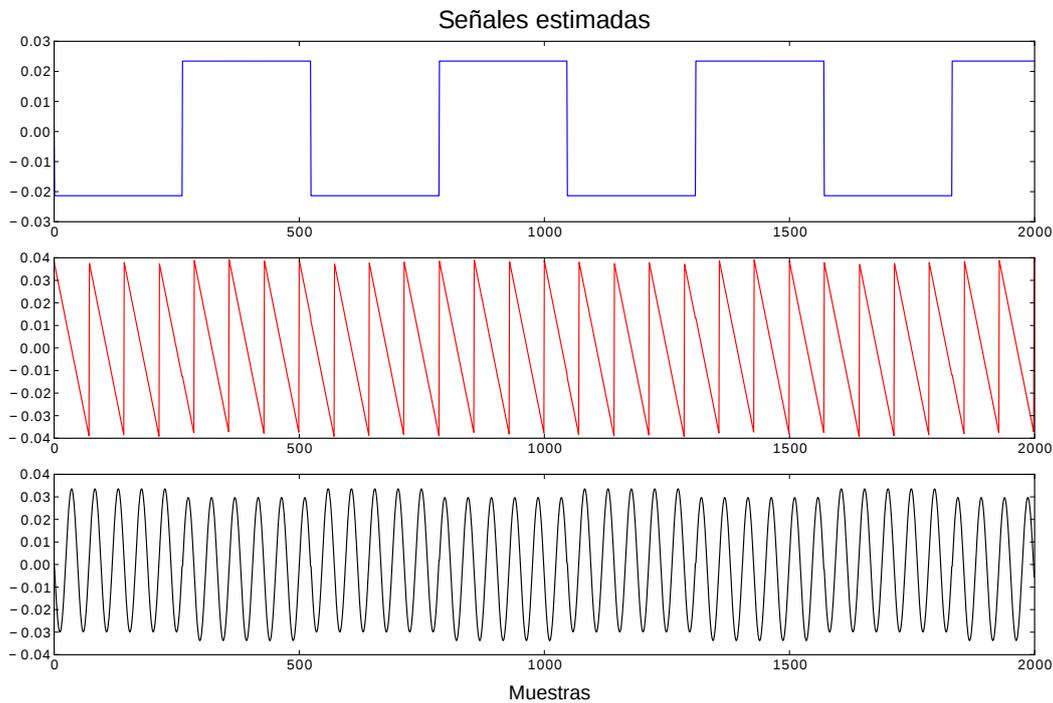


Figura 2.2: Señales originales estimadas mediante BSS.

2.2. Análisis de componentes independientes

El análisis de componentes independientes (ICA) es una técnica estadística y computacional que revela factores ocultos en una base de conjuntos de variables aleatorias, mediciones, o señales.

El ICA define un modelo generativo para los datos multivariantes observados, los cuales típicamente están dados como una gran base de datos de muestras. En el modelo, los datos son asumidos como una mezcla lineal o no lineal de algunas variables latentes desconocidas, y el sistema de mezclado también es desconocido. Las variables latentes son asumidas como no gaussianas y mutuamente independientes, y estas son llamadas componentes independientes de los datos observados. Estas componentes independientes, también llamadas fuentes, pueden ser encontradas por ICA [30]. Cabe resaltar que en situaciones prácticas las componentes no son totalmente independientes, pero se busca encontrar la mayor independencia posible.

Entonces ICA es definido así: dado un conjunto de observaciones de variables aleatorias $x_1(t), x_2(t), \dots, x_n(t)$ que se asume provienen de una mezcla lineal de

la siguiente forma:

$$\begin{pmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_n(t) \end{pmatrix} = A \cdot \begin{pmatrix} s_1(t) \\ s_2(t) \\ \vdots \\ s_m(t) \end{pmatrix} \quad (2.4)$$

donde A es la matriz de mezcla desconocida. Entonces mediante el análisis de componentes independientes se estima la matriz A y las fuentes independientes $s_j(t)$ a partir de las observaciones $x_i(t)$. Para simplificar el proceso matemático, en este caso, se asume que el número de observaciones n es igual al número de componentes independientes m .

Ahora se debe encontrar una matriz W tal que las variables estimadas $y_1(t), y_2(t), \dots, y_m(t)$ sean lo más independientes posible. En este caso W es la matriz inversa de A . Sin embargo, el modelo presentado en (3.4) sólo puede ser aplicado para variables no-gaussianas.

En la estimación de las componentes independientes se deben tener en cuenta dos principios básicos:

1. **Decorrelación no-lineal.** En el modelo ICA se tiene en cuenta la independencia entre las variables estimadas y no la decorrelación. Esto es debido a que la independencia es un factor más fuerte que la decorrelación, es decir que la independencia implica una decorrelación no-lineal. Por ejemplo, si s_1 y s_2 son variables independientes entonces las transformaciones no-lineales $g(s_1)$ y $h(s_2)$ son decorreladas (la covarianza es cero).
2. **Máxima no-gaussianidad.** Otro principio importante en la estimación de ICA es la máxima no-gaussianidad. Este se basa en el teorema del límite central, el cual dice que la suma de variables aleatorias no-gaussianas será más gaussiana que las variables originales. Por ejemplo, si se toma una combinación lineal $y = \sum_i b_i x_i$ de la mezcla de variables observadas, está será máximamente no-gaussiana si es igual a una de las componentes independiente. Cada máximo local representa una componente independiente.

2.2.1. Medición de la no-gaussianidad

Para medir la no-gaussianidad se puede usar, por ejemplo, la *curtosis*. La curtosis es un cúmulo de cuarto orden el cual es una generalización de la varianza usando polinomios de alto-orden.

La curtosis de y es definida como, [30]

$$kurt(y) = E\{y^4\} - 3(E\{y^2\})^2 \quad (2.5)$$

donde E es la esperanza y y es una variable aleatoria con media cero.

La curtosis puede tomar valores positivos y negativos. La curtosis de una variable aleatoria gaussiana es igual a cero. Si la curtosis es negativa se dice que la variable aleatoria es subgausiana y si la curtosis es positiva se dice que la variable aleatoria es supergausiana.

La función de densidad de probabilidad (PDF) de una variable supergausiana tiene forma puntiaguda debido a que la mayoría de sus valores son cero, Figura 2.3. Un ejemplo es la distribución de Laplace la cual es descrita por la función:

$$p(y) = \frac{1}{2b} \exp - \left(\frac{|y - \mu|}{b} \right) \quad (2.6)$$

donde μ es la posición del pico de la distribución y b es un parámetro de escala.

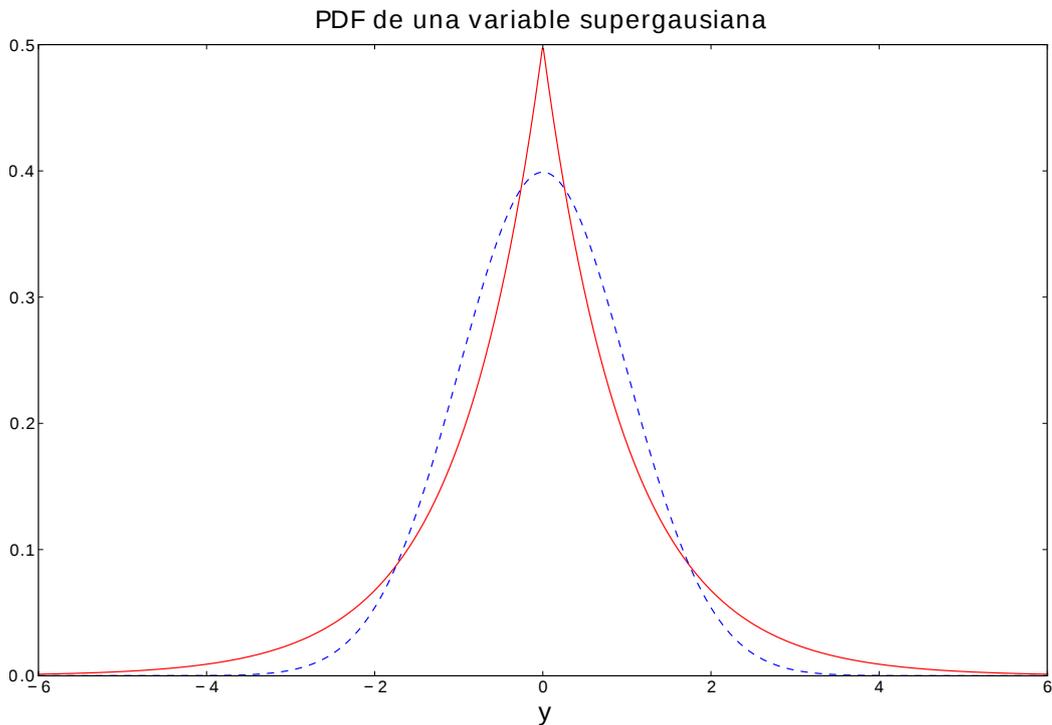


Figura 2.3: Comparación entre una distribución de probabilidad de Laplace, línea roja sólida, la cual es un ejemplo de distribución supergausiana y una distribución gaussiana, línea azul punteada.

Por otra parte la PDF de una variable subgausiana presenta un achatamiento en los valores cercanos a cero, Figura 2.3. Un ejemplo es la distribución uniforme cuya densidad está dada por:

$$p(y) = \begin{cases} \frac{1}{2\sqrt{3}} & \text{if } |y| \leq \sqrt{3} \\ 0 & \text{en otro caso} \end{cases} \quad (2.7)$$

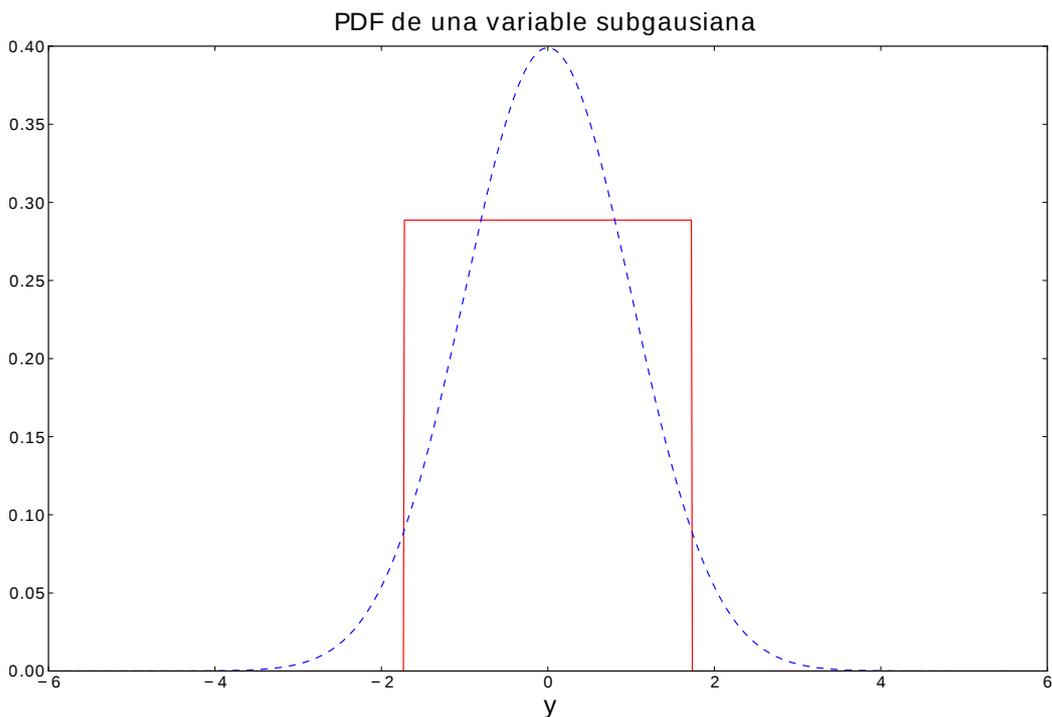


Figura 2.4: Comparación entre una distribución de probabilidad de uniforme, línea roja sólida, la cual es un ejemplo de distribución subgausiana y una distribución gaussiana, línea azul punteada.

Normalmente se usa el valor absoluto de la curtosis como medida de no-gaussianidad, con esto la curtosis será cero para variables aleatorias gaussianas y valores positivos para variable supergaussianas y subgaussianas.

Sin embargo, la curtosis tiene ciertas desventajas que la alejan de ser la candidata perfecta para implementar el modelo ICA. El principal de estos problemas es que la curtosis es muy sensible a los *outliers*, es decir a los valores atípicos dentro de la distribución de una variable aleatoria.

A continuación se presenta un ejemplo: se tiene una variable aleatoria de 1000

valores con distribución de laplace con $\mu = 0$ y $b = 1$, Figura 2.5. En este caso su curtosis es igual a 1.89.

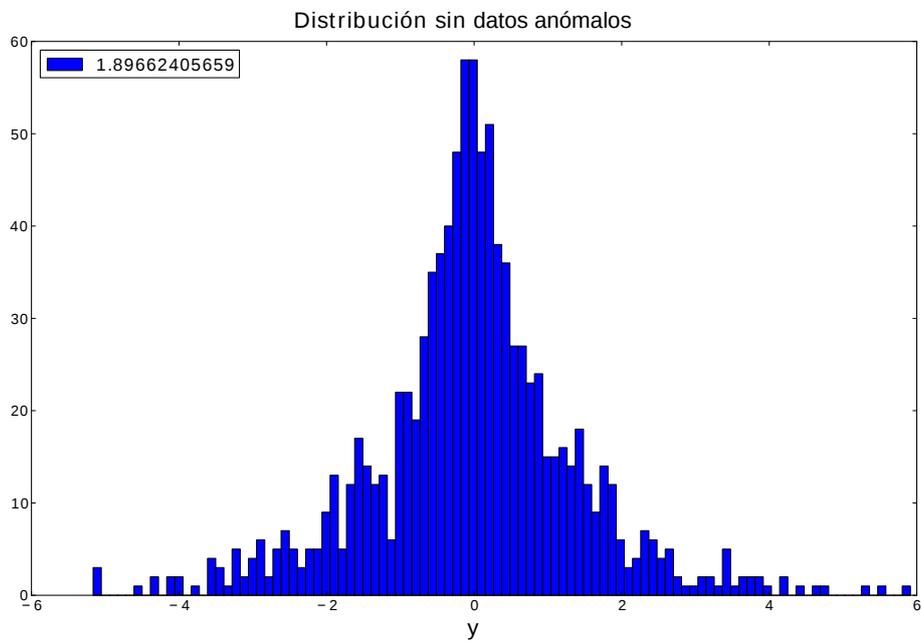


Figura 2.5: Distribución de laplace con $\mu = 0$ y $b = 1$ sin presencia de valores atípicos.

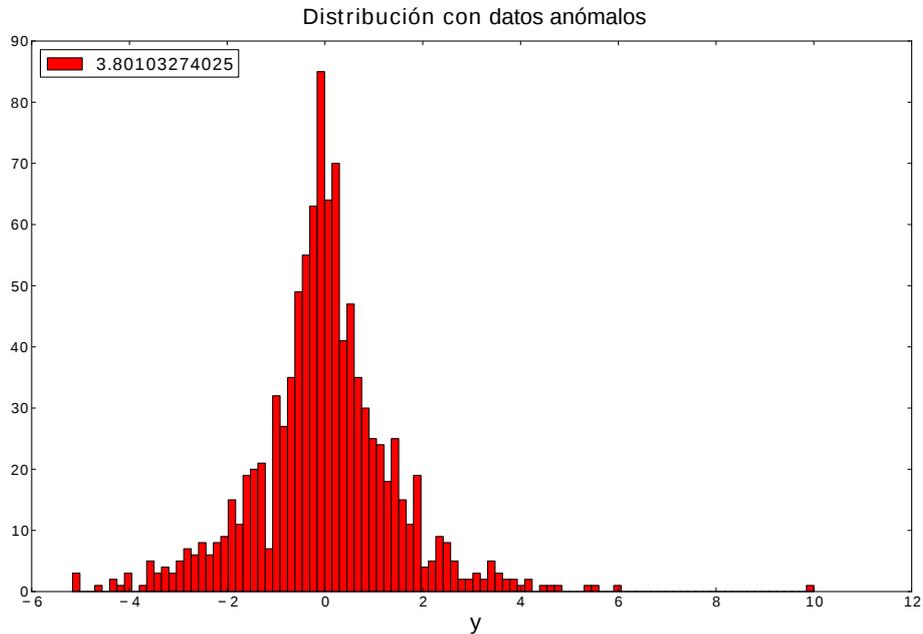


Figura 2.6: Distribución de laplace con $\mu = 0$ y $b = 1$ con presencia de un valor atípico igual a 10.

Ahora bien, cambiando la última de sus muestras por un valor igual a 10, Figura 2.6, la curtosis cambia a 3.8. Esto ejemplifica que la curtosis es poco efectiva como medida de no-gaussianidad debido a su sensibilidad a los outliers.

Por esta razón, en [30] se propone otra medida de no-gaussianidad más robusta basada en la teoría de la información, la *neguentropía*.

2.3. ICA por neguentropía

La neguentropía es basada en la cantidad de la información teórica de la entropía. La entropía de una variable aleatoria se relaciona con la cantidad de información que da la observación de la variable. Es decir, entre más aleatoria y desestructurada sea la variable, el valor de su entropía será mayor. La entropía H de una variable y con densidad $p_y(\eta)$ es, [30]:

$$H(y) = - \int p_y(\eta) \log p_y(\eta) d\eta \quad (2.8)$$

Una variable gaussiana tiene el valor de entropía más alto dentro de las variables aleatorias de igual varianza lo cual muestra que es la menos estructurada de todas las distribuciones.

Para obtener una medida de no gaussianidad la cual sea cero para las distribuciones gaussianas y tenga valores positivos para las otras distribuciones, se usa una versión normalizada de la entropía llamada negentropía. La negentropía J es definida como:

$$J(y) = H(y_{gauss}) - H(y) \quad (2.9)$$

donde y_{gauss} es una variable aleatoria gaussiana con la misma matriz de correlación y covarianza de la variable y .

La negentropía es considerada el estimador óptimo de no-gaussianidad, sin embargo, su implementación es costosa computacionalmente. Para resolver este problema se usan aproximaciones de la negentropía, [30].

2.3.1. Aproximación de la negentropía

El método clásico de aproximación de la negentropía es mediante cúmulos de alto orden, el cual dá como resultado, [30]:

$$J(y) \approx \frac{1}{12} E\{y^3\}^2 + \frac{1}{48} kurt\{y\}^2 \quad (2.10)$$

La variable aleatoria y se asume con media cero y varianza 1. En este caso la aproximación está expresada en términos de la curtosis heredando sus problemas de robustez. Además, para distribuciones simétricas, el término $E\{y^2\}$ es igual a cero, dando como resultado que la negentropía es una versión normalizada del cuadrado de la curtosis.

Para generalizar la aproximación por cúmulos de alto orden se usa la esperanza de funciones no-cuadráticas. Es decir, se rempazan las funciones polinomiales y^3 y y^4 por funciones G_i cuya esperanza es $E\{G_i(y)\}$.

Usando sólo una función no-cuadrática, la aproximación es:

$$J(y) \propto [E\{G(y)\} - E\{G(\nu)\}]^2 \quad (2.11)$$

donde ν es una variable aleatoria gaussiana con media cero y varianza unitaria. En este caso, si $G(y) = y^4$ obtenemos una aproximación basada en curtosis.

El principal objetivo es encontrar una mejor aproximación que la presentada en (3.10). Elegir una G que tenga un crecimiento lento permite obtener un estimador más robusto. Un ejemplo de funciones G son las siguientes, [30]:

$$G_1(y) = \frac{1}{a_1} \log \cosh(a_1 y) \quad (2.12)$$

$$G_2(y) = -\exp(-y^2/2) \quad (2.13)$$

donde $1 \leq a_1 \leq 2$ es una constante.

2.4. Algoritmo FastICA

2.4.1. Estimación de una componente independiente

Para estimar una de las componentes independientes se considera una combinación lineal de las observaciones x_i denotada por:

$$y = \mathbf{b}^T \mathbf{x} = \sum_i b_i x_i \quad (2.14)$$

donde \mathbf{b} es una de las filas de la inversa de la matriz de mezcla. Por otro lado, y en términos de las fuentes originales s_i puede ser definida como:

$$y = \mathbf{b}^T \mathbf{x} = \mathbf{b}^T \mathbf{A} \mathbf{s} = \mathbf{q}^T \mathbf{s} = \sum_i q_i s_i \quad (2.15)$$

Inicialmente, el vector aleatorio con media cero \mathbf{x} sufre una transformación lineal de blanqueo \mathbf{V} , es decir, los elementos del vector blanqueado \mathbf{z} son decorrelados y tienen varianza unitaria:

$$\mathbf{z} = \mathbf{V} \mathbf{x} \quad (2.16)$$

Entonces la combinación lineal queda:

$$y = \mathbf{w}^T \mathbf{z} \quad (2.17)$$

sujeto a la restricción $\|\mathbf{w}\| = 1$ para que las proyecciones queden normalizadas a una esfera unitaria.

El algoritmo FastICA encuentra un vector unitario \mathbf{w} tal que la proyección $\mathbf{w}^T \mathbf{z}$ maximice la no-gaussianidad. La no-gaussianidad es medida por la aproximación de la negentropía $J(\mathbf{w}^T \mathbf{z})$ dada en (3.11).

La función objetivo para el proceso de optimización es:

$$J(\mathbf{w}^T \mathbf{z}) \propto [E\{G(\mathbf{w}^T \mathbf{z})\} - E\{G(\nu)\}]^2 \quad (2.18)$$

La tasa de cambio del vector \mathbf{w} está dada por:

$$\Delta \mathbf{w} \propto \frac{\partial J(\mathbf{w}^T \mathbf{z})}{\partial \mathbf{w}} \quad (2.19)$$

donde:

$$\Delta \mathbf{w} \propto \gamma E\{\mathbf{z}g(\mathbf{w}^T \mathbf{z})\} \quad (2.20)$$

$$\mathbf{w} \leftarrow \mathbf{w}/\|\mathbf{w}\| \quad (2.21)$$

con $\gamma = E\{G(\mathbf{w}^T \mathbf{z})\} - E\{G(\nu)\}$. La normalización de \mathbf{w} es necesaria para que su proyección esté dentro de la esfera unitaria. La función g es la derivada de la función G usada en la aproximación de la negentropía.

De (3.20) podemos sugerir la siguiente iteración de punto fijo:

$$\mathbf{w} \leftarrow E\{\mathbf{z}g(\mathbf{w}^T \mathbf{z})\} \quad (2.22)$$

En este caso el coeficiente γ puede ser omitido debido a que este es eliminado por la normalización.

Debido a que la iteración de (3.22) no tiene buenas propiedades de convergencia, en [30] se plantea una modificación de esta la cual consiste en adicionar una constante α multiplicada por \mathbf{w} a ambos lados de la expresión:

$$(1 + \alpha)\mathbf{w} = E\{\mathbf{z}g(\mathbf{w}^T \mathbf{z})\} + \alpha\mathbf{w} \quad (2.23)$$

El coeficiente α , y por consiguiente, el algoritmo FastICA, puede ser encontrado usando un método de Newton aproximado. El método de Newton es un método rápido para resolver ecuaciones, es decir, converge en un número pequeño de pasos. Sin embargo, este usualmente requiere una inversión matricial en cada paso, por lo tanto, computacionalmente es más costoso que otros métodos de búsqueda como el gradiente descendente.

Para solucionar este problema en [30] se propone una aproximación del método de Newton que no necesita una inversión matricial pero que converge en el mismo número de pasos que el método original.

Primero se nota que los máximos de la aproximación de la negentropía de $\mathbf{w}^T \mathbf{z}$ se obtienen en ciertos óptimos de $E\{G(\mathbf{w}^T \mathbf{z})\}$.

De acuerdo a las condiciones de Lagrange, los óptimos de $E\{G(\mathbf{w}^T \mathbf{z})\}$, bajo la restricción $E\{(\mathbf{w}^T \mathbf{z})^2\} = \|\mathbf{w}\|^2 = 1$ son obtenidos en los puntos donde el gradiente del Lagrangiano es cero:

$$F = E\{\mathbf{z}g(\mathbf{w}^T \mathbf{z})\} + \beta\mathbf{w} = 0 \quad (2.24)$$

donde β es una constante y puede ser evaluada como:

$$\beta = E\{\mathbf{w}_0^T \mathbf{z}g(\mathbf{w}_0^T \mathbf{z})\} \quad (2.25)$$

donde \mathbf{w}_0 es el valor de \mathbf{w} en el óptimo.

Ahora se soluciona la ecuación (3.24) por el método de Newton obteniendo el gradiente el cual es la segunda derivada del Lagrangiano:

$$\frac{\partial F}{\partial \mathbf{w}} = E\{\mathbf{z}\mathbf{z}^T g'(\mathbf{w}^T \mathbf{z})\} + \beta \mathbf{I} \quad (2.26)$$

Para simplificar la inversión de esta matriz, se puede aproximar el primer término de la ecuación (3.26):

$$E\{\mathbf{z}\mathbf{z}^T g'(\mathbf{w}^T \mathbf{z})\} \approx E\{\mathbf{z}\mathbf{z}^T\} E\{g'(\mathbf{w}^T \mathbf{z})\} = E\{g'(\mathbf{w}^T \mathbf{z})\} \mathbf{I} \quad (2.27)$$

Con la anterior aproximación, el gradiente se convierte en una diagonal y puede ser invertida fácilmente. Entonces la aproximación de la iteración de Newton es:

$$\mathbf{w} \leftarrow \mathbf{w} - \frac{E\{\mathbf{z}g(\mathbf{w}^T \mathbf{z})\} + \beta \mathbf{w}}{E\{g'(\mathbf{w}^T \mathbf{z})\} + \beta} \quad (2.28)$$

Multiplicando (3.28) en ambos lados por $\beta + E\{g'(\mathbf{w}^T \mathbf{z})\}$ y simplificando, obtenemos la iteración básica de punto fijo en FastICA:

$$\mathbf{w} \leftarrow E\{\mathbf{z}g(\mathbf{w}^T \mathbf{z}) - E\{g'(\mathbf{w}^T \mathbf{z})\}\mathbf{w}\} \quad (2.29)$$

Las funciones g pueden ser calculadas como:

$$g_1(y) = \tanh(a_1 y) \quad (2.30)$$

$$g_2(y) = y \exp(-y^2/2) \quad (2.31)$$

$$g_3(y) = y^3 \quad (2.32)$$

Y por ende las funciones g' son:

$$g'_1(y) = a_1(1 - \tanh^2(a_1 y)) \quad (2.33)$$

$$g'_2(y) = (1 - y^2) \exp(-y^2/2) \quad (2.34)$$

$$g'_3(y) = 3y^2 \quad (2.35)$$

donde $1 \leq a_1 \leq 2$ es una constante a menudo igual a uno.

El método FastICA para encontrar una sola componente independiente se muestra en el Algoritmo 1.

Algoritmo 1 FastICA para estimar una componente independiente.

- 1: Centrar los datos \mathbf{x} para que tengan media cero
 - 2: Blanquear los datos centrados para obtener \mathbf{z}
 - 3: Elegir un vector inicial \mathbf{w} de norma unitaria
 - 4: $\mathbf{w} \leftarrow E\{\mathbf{z}g(\mathbf{w}^T\mathbf{z}) - E\{g'(\mathbf{w}^T\mathbf{z})\}\mathbf{w}\}$
 - 5: $\mathbf{w} \leftarrow \mathbf{w}/\|\mathbf{w}\|$
 - 6: **if** no hay convergencia **then**
 - 7: Regresar al paso 4
 - 8: **end if**
-

2.4.2. Estimación de varias componentes independientes

Para la estimación de varias componentes independientes se debe utilizar un método de máxima no-gaussianidad basado en que los vectores \mathbf{w}_i correspondientes a diferentes componentes independientes son ortogonales en el espacio blanqueado. Es decir, la independencia de las componentes requiere la decorrelación entre ellas, teniendo en el espacio blanqueado $E\{(\mathbf{w}_i^T\mathbf{z})(\mathbf{w}_j^T\mathbf{z})\} = \mathbf{w}_i^T\mathbf{w}_j$. Además, cabe resaltar que en el espacio blanqueado $A^{-1} = A^T$.

Entonces para hallar varias componentes independientes se debe correr el algoritmo varias veces para $\mathbf{w}_1, \dots, \mathbf{w}_n$, y después de cada iteración ortogonalizar los vectores para evitar que converjan al mismo máximo. Para lograr la ortogonalización se usan los siguientes algoritmos:

2.4.3. Ortogonalización deflacionaria

Un método de ortogonalización es la *ortogonalización deflacionaria* la cual usa el método de Gram-Schmit. Este método estima las componentes independientes una por una, es decir, si tenemos p componentes independientes, o p vectores $\mathbf{w}_1, \dots, \mathbf{w}_p$, se corre un algoritmo para \mathbf{w}_{p+1} , y después de cada iteración se restan de \mathbf{w}_{p+1} las proyecciones $(\mathbf{w}_{p+1}^T\mathbf{w}_j)\mathbf{w}_j$, $j = 1, \dots, p$ de los p vectores anteriormente estimados, y luego se renormaliza \mathbf{w}_{p+1} . El método FastICA con ortogonalización deflacionaria se muestra en el Algoritmo 2, [30].

2.4.4. Ortogonalización simétrica

En la *ortogonalización simétrica* los vectores \mathbf{w}_i no son estimados uno por uno, por el contrario, son estimados en paralelo. El método de estimación secuencial tiene la desventaja que los errores de estimación de los primeros vectores se acumulan en los vectores subsecuentes debido a la ortogonalización. Una ventaja de la ortogonalización simétrica es que abre la opción de paralelizar el cálculo de las

Algoritmo 2 FastICA con ortogonalización deflacionaria.

- 1: Centrar los datos \mathbf{x} para que tengan media cero
 - 2: Blanquear los datos centrados para obtener \mathbf{z}
 - 3: Elegir m , el número de componentes independientes
 - 4: Establecer $p \leftarrow 1$
 - 5: Inicializar \mathbf{w}_p aleatoriamente
 - 6: $\mathbf{w}_p \leftarrow E\{\mathbf{z}g(\mathbf{w}_p^T \mathbf{z}) - E\{g'(\mathbf{w}_p^T \mathbf{z})\}\mathbf{w}_p\}$
 - 7: $\mathbf{w}_p \leftarrow \mathbf{w}_p - \sum_{j=1}^{p-1} (\mathbf{w}_p^T \mathbf{w}_j) \mathbf{w}_j$
 - 8: $\mathbf{w}_p \leftarrow \mathbf{w}_p / \|\mathbf{w}_p\|$
 - 9: **if** \mathbf{w}_p no converge **then**
 - 10: Regresar al paso 6
 - 11: **end if**
 - 12: $p \leftarrow p + 1$
 - 13: **if** $p \leq m$ **then**
 - 14: Regresar al paso 5
 - 15: **end if**
-

componentes independientes.

La ortogonalización simétrica de \mathbf{W} puede ser hecha mediante el método que implica la matriz de raíces cuadradas:

$$\mathbf{W} \leftarrow (\mathbf{W}\mathbf{W}^T)^{-1/2} \mathbf{W} \quad (2.36)$$

La raíz cuadrada inversa $(\mathbf{W}\mathbf{W}^T)^{-1/2}$ es obtenida desde la descomposición en valores propios de $\mathbf{W}\mathbf{W}^T = \mathbf{E}\mathbf{D}\mathbf{E}^T$ como:

$$(\mathbf{W}\mathbf{W}^T)^{-1/2} = \mathbf{E} \text{diag}(d_1^{-1/2}, \dots, d_m^{-1/2}) \mathbf{E}^T \quad (2.37)$$

donde \mathbf{E} es la matriz ortogonal de vectores propios. El método FastICA con ortogonalización simétrica se muestra en el Algoritmo 3.

2.5. ICA en señales EEG contaminadas con artefactos oculares

El análisis de componentes independientes (ICA) ha jugado un papel muy importante en el procesamiento de señales cerebrales. ICA ha sido utilizado en diferentes tareas que van desde la separación de las señales EEG en los ritmos alfa, beta, delta, teta y gama, hasta la identificación de artefactos provenientes de diferentes fuentes como la respiración, el pulso cardiaco, el movimiento de

Algoritmo 3 FastICA con ortogonalización deflacionaria.

- 1: Centrar los datos \mathbf{x} para que tengan media cero
 - 2: Blanquear los datos centrados para obtener \mathbf{z}
 - 3: Elegir m , el número de componentes independientes
 - 4: Elegir los valores para \mathbf{w}_i , $i = 1, \dots, m$, con norma unitaria.
 - 5: Ortogonalizar la matriz W como en el paso 9
 - 6: **for** $i = 1, \dots, m$ **do**
 - 7: $\mathbf{w}_i \leftarrow E\{\mathbf{z}g(\mathbf{w}_i^T \mathbf{z})\} - E\{g'(\mathbf{w}_i^T \mathbf{z})\}\mathbf{w}_i$
 - 8: **end for**
 - 9: $\mathbf{W} \leftarrow (\mathbf{W}\mathbf{W}^T)^{-1/2}\mathbf{W}$
 - 10: **if** no hay convergencia **then**
 - 11: Regresar al paso 6
 - 12: **end if**
-

los músculos de la cara y los movimientos oculares, [39]. Además, teniendo en cuenta que los registros EEG son el resultado de una suma ponderada de la actividad eléctrica de millones de neuronas, ICA puede mostrar factores o información oculta que no puede ser observada en el análisis directo de los datos [61].

Para aplicar ICA en el procesamiento de señales EEG se deben tener en cuenta dos aspectos: la primera es que las señales provengan de una suma ponderada de fuentes independientes y la segunda que las variables no sean gaussianas. En este caso las señales cerebrales son el resultado de la suma de muchas señales generadas por los cúmulos neuronales o zonas activas las cuales se pueden considerar como fuentes independientes. Por otra parte, los datos EEG presentan una distribución de probabilidad donde la mayoría de sus valores presentan valores cercanos a cero, lo cual se asimila al comportamiento de una variable supergausiana, Figura 2.7.

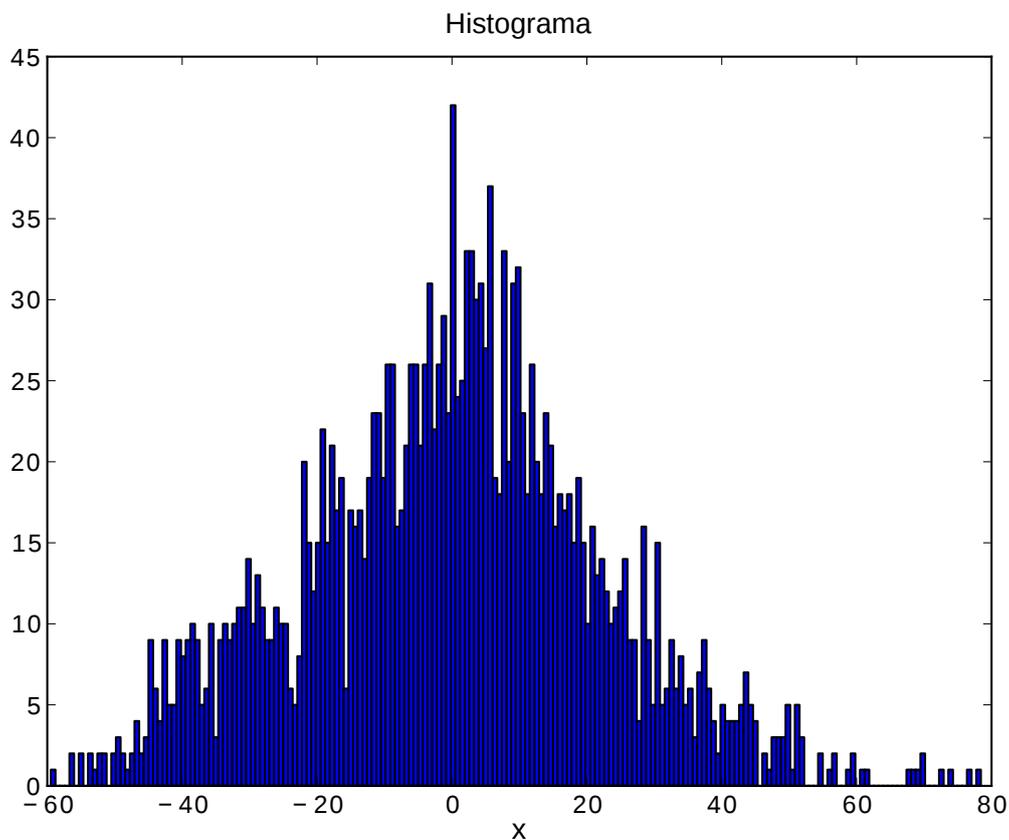


Figura 2.7: Histograma de una señal EEG tomada del canal FC3 del sujeto 1.

2.5.1. Detección de artefactos oculares en EEG mediante ICA

Como anteriormente se mencionó, una de las aplicaciones de ICA en señales EEG es la detección de artefactos o señales que provienen de fuentes no cerebrales. Este trabajo se enfocó en la detección de artefactos oculares los cuales son generados por el movimiento del globo ocular, es decir, de una fuente estadísticamente independiente a la que genera la actividad neuronal. Además, los artefactos oculares tienen una función de probabilidad puntiaguda centrada en cero, Figura 2.8, describiendo su naturaleza supergausiana.

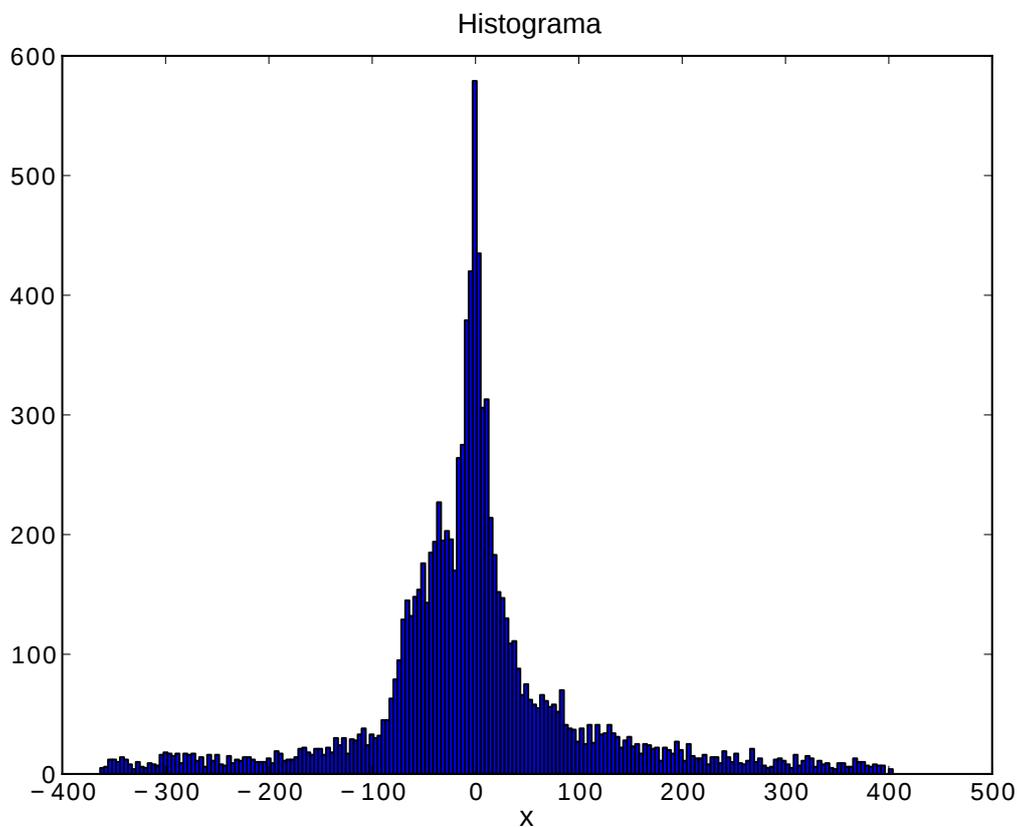


Figura 2.8: Histograma de una señal EOG del sujeto 1.

El ICA se aplicó a una porción de la base de datos la cual tiene una duración de 12 segundos, Figura 2.9, la cual está conformada por 22 canales los cuales fueron registrados mediante electrodos superficiales dispuestos sobre el cuero cabelludo como se muestra en la Figura 2.10. En este caso el algoritmo utilizado fue el FastICA con ortogonalización simétrica y con (3.31) como función g para la aproximación de la negentropía.

Por otra parte, los movimientos oculares son registrados mediante un HEOG los cuales indican el instante que se producen las sacadas. El efecto de estos artefactos oculares se propaga sobre todo el cuero cabelludo teniendo mayor significancia en los canales frontales debido a que los campos electromagnéticos generados por el dipolo ocular presentan amplitudes mayores cerca a las cuencas oculares, luego su amplitud se atenúa a medida que viajan hacia la parte posterior del cráneo mostrando efectos pequeños en los canales registrados en la región occipital.

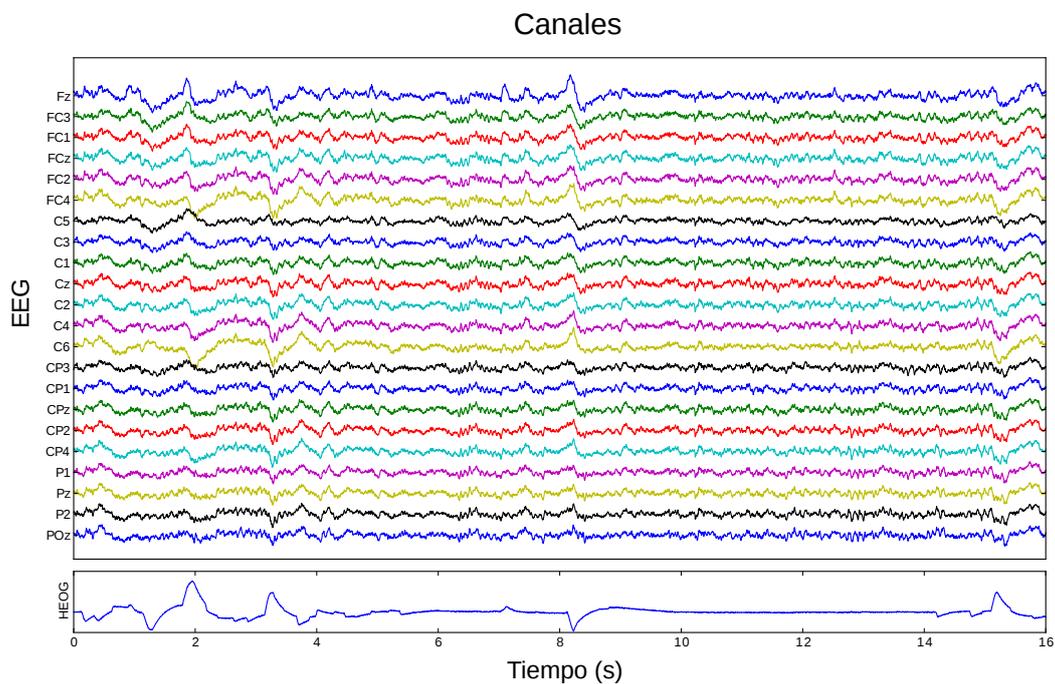


Figura 2.9: Señales EEG y HEOG registradas durante un periodo de 12 segundos. La influencia de los artefactos oculares es más notoria en los canales cercanos a la parte frontal del cráneo que en la parte occipital.

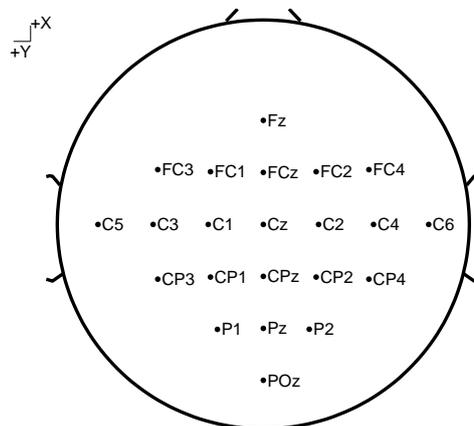


Figura 2.10: Ubicación de los canales EEG registrados.

En la Figura 2.11 podemos ver las componentes independientes estimadas por el algoritmo FastICA, las componentes IC1 y IC2 corresponden a los artefactos oculares, y las restantes a las diferentes componentes cerebrales.

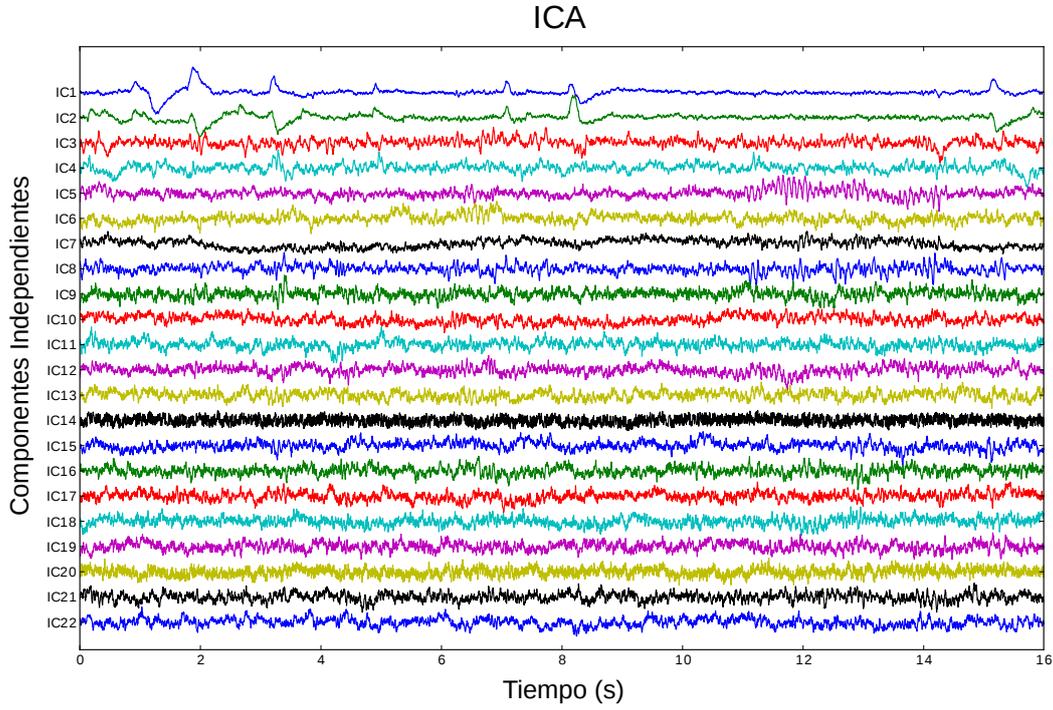


Figura 2.11: Descomposición de las señales EEG mediante ICA en las fuentes independientes estimadas.

Para ubicar espacialmente las fuentes estimadas se hace una proyección de los coeficientes a_{ij} de la matriz de mezcla $A = W^{-1}$ sobre el cráneo, Figura 2.12. La matriz de mezcla es definida como:

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{pmatrix} \quad (2.38)$$

donde, en este caso, $m = n = 22$.

De esta manera, los coeficientes pertenecientes a la componente estimada y_j son:

$$A(:, j) = \begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{nj} \end{pmatrix} \quad (2.39)$$

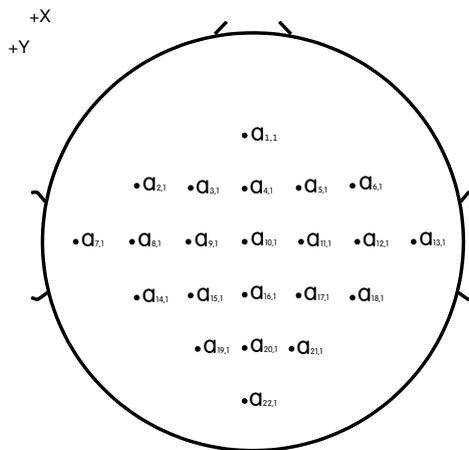


Figura 2.12: Proyección de los coeficientes a_{ij} sobre el cráneo.

En la Figura 2.13 se muestra la proyección de los coeficientes pertenecientes a las dos primeras componentes del ejemplo en cuestión. En este caso, el valor de los coeficientes cercanos a la zona frontal son mayores a los de otras regiones, debido a la cercanía con la fuente que origina los artefactos, en este caso los ojos.

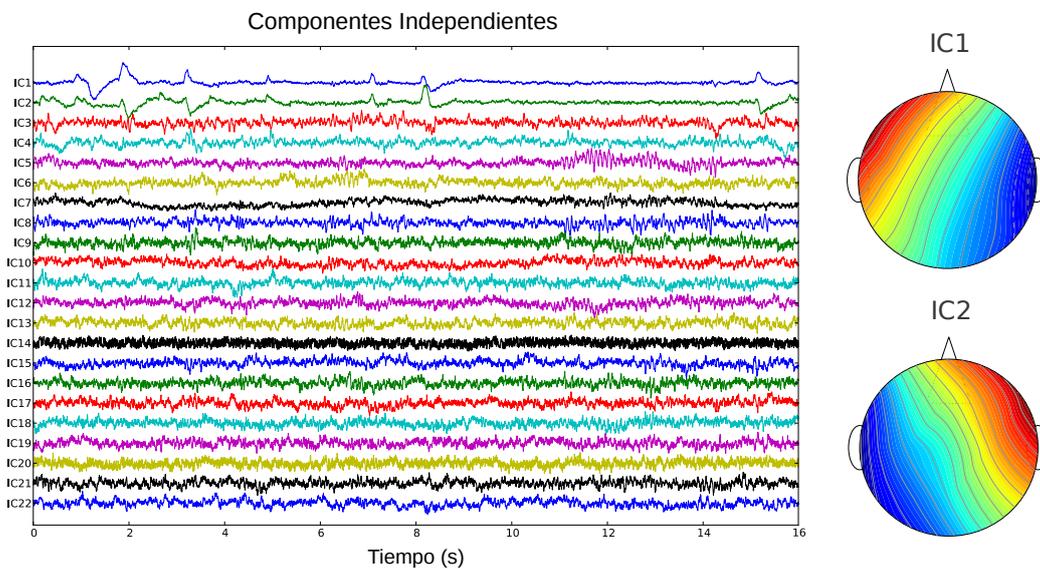


Figura 2.13: Proyección de las componentes IC1 e IC2 sobre el cráneo.

Para tener una mejor visión del efecto que causan las dos primeras fuentes independientes, en la Figura 2.14 podemos ver un modelo tridimensional de la proyección de los coeficientes a_{ij} , este modelo se generó por medio de la herramienta de análisis EEG llamada *EEGLab* de MatLab. En esta gráfica podemos ver

que la componente IC1 se ubica en el ojo izquierdo y la componente IC2 en el ojo derecho demostrando que el algoritmo FastICA hace una búsqueda eficiente de las fuentes independientes basado en la independencia estadística, teniendo en cuenta que los ojos tienen un comportamiento parecido pero no igual.

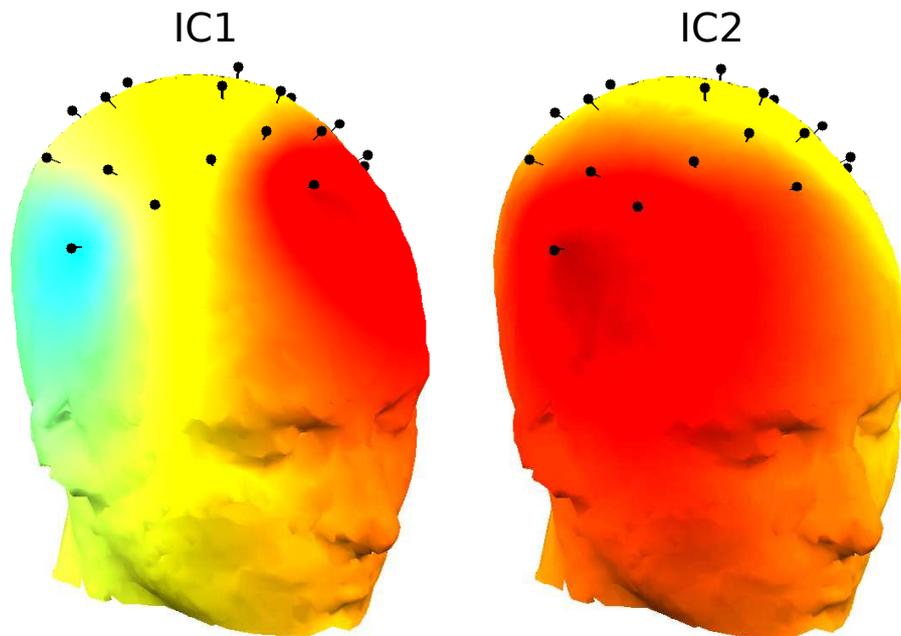


Figura 2.14: Vista tridimensional de la proyección de las componentes independientes pertenecientes a los artefactos oculares.

De esta forma se ha mostrado que el análisis de componentes independientes es una herramienta muy útil para la detección de artefactos oculares en los registros electroencefalográficos, basado en sus principales pilares conceptuales: la independencia entre las fuentes que generaron la mezcla observada, y la no-gaussianidad de las fuentes originales.

En el siguiente capítulo se muestra como la minería de datos nos proporciona un método para la caracterización de las componentes independientes halladas mediante ICA y así poder entrenar un modelo de reconocimiento de patrones que tenga la capacidad de discriminar los artefactos oculares de las señales EEG.

Cuando estás solucionando un problema, no tienes que preocuparte. Ahora, después de que has resuelto el problema, entonces sí es momento de preocuparse.

Richard Feynman

3

Extracción de características: Complejidad

En problemas de predicción y reconocimiento de patrones, la extracción de características juega un papel muy importante porque permite representar la información en un espacio dimensionalmente menor, reduciendo la complejidad para solucionar el problema y por ende el tiempo de cómputo; adicionalmente genera un crecimiento en el desempeño del modelo.

Para crear un sistema automático de filtrado de artefactos oculares en señales EEG se deben caracterizar las componentes independientes resultantes del ICA, estas características alimentarán un modelo de clasificación el cual discriminará las clases: artefactos y potenciales cerebrales.

En este capítulo se abordará la importancia de la minería de datos y de la extracción de características de complejidad en señales cerebrales para la eliminación de artefactos.

3.1. Extracción de características

La extracción de características se encarga de encontrar el conjunto de características más compacto y que aporte mayor información para mejorar la eficiencia en el almacenamiento y procesamiento de los datos [26]. El problema de extracción de características puede ser dividido en dos: construcción de características y selección de características.

3.1.1. Construcción de características

La construcción de características consiste en obtener una mejor representación de los datos crudos, esta puede estar integrada dentro de los procesos del modelo, por ejemplo en las capas ocultas de las redes neuronales donde se generan representaciones internas de los datos de entrada, o como una etapa de preprocesamiento.

El preprocesamiento puede ser definido así: Dado un vector de datos \mathbf{x} con una dimensión n , este puede sufrir una transformación donde se genera un vector de características \mathbf{x}' con dimensión n' .

Algunas de las transformaciones de preprocesamiento son, [26]:

1. *Estandarización*. Las características pueden tener diferentes escalas aunque se refieran a objetos comparables. Este sesgo en las escalas puede repercutir en el desempeño de los posteriores modelos de reconocimiento o predicción. Para evitar esto, se hace un proceso de estandarización, el cual es definido como:

$$\mathbf{x}' = \frac{\mathbf{x} - \mu}{\sigma} \quad (3.1)$$

donde μ es la media y σ la desviación estandar de \mathbf{x} .

2. *Normalización*. La normalización es un proceso mediante el cual se lleva la distribución de las características a una distribución normal eliminando la dependencia de la escala. La normalización es definida como:

$$\mathbf{x}' = \frac{\mathbf{x}}{\|\mathbf{x}\|} \quad (3.2)$$

donde $\|\mathbf{x}\|$ es la norma de \mathbf{x} .

3. *Mejoramiento de la señal.* La relación señal a ruido de la señal puede ser mejorada con filtros. Estas operaciones incluyen la corrección de la línea base, eliminación de ruido, suavizado y realce. Para esto normalmente se utiliza la transformada de Fourier o la transformada Wavelet.
4. *Extracción de características locales.* Para datos secuenciales, espaciales o estructurados se usan técnicas específicas parecidas a métodos convolucionales mediante kernels o métodos sintácticos. Estas técnicas codifican problemas de conocimiento específico en las características.
5. *Métodos espaciales embebidos lineales y no-lineales.* Cuando la dimensionalidad de los datos es muy grande, esta puede ser reducida a una dimensión menor mediante algunas técnicas, preservando la mayor información posible. Entre estas técnicas está el Análisis de Componentes Principales (PCA) y el Escalado Multidimensional (MDS).
6. *Expansiones no-lineales.* Aunque la reducción de la dimensionalidad es usada a menudo cuando se habla de la complejidad de los datos, algunas veces es mejor incrementar la dimensionalidad. Esto ocurre cuando el problema es muy complejo y las interacciones de primer orden no son suficientes para obtener buenos resultados. Este método consiste en calcular productos de las características originales para crear monomios.
7. *Discretización de características.* Algunos algoritmos no manejan bien los datos continuos siendo necesario discretizar los valores continuos en un conjunto discreto finito.

Algunos métodos no alteran el espacio dimensionalmente (mejoramiento de la señal, normalización, estandarización), otros lo aumentan (expansiones no-lineales, discretización de características), otros lo reducen (métodos espaciales intrínsecos) o pueden actuar en otra dirección (extracción de características locales), [26].

3.1.2. Selección de características

Uno de los principales objetivos de la selección de características es elegir las características más relevantes y que contengan mayor información, pero además de esto puede tener otros fines:

1. *Reducción de los datos.* Limitar los requerimientos de almacenamiento e incrementar la velocidad del algoritmo.

2. *Reducción del conjunto de características.* Para guardar recursos de la próxima etapa de recolección de los datos o durante su utilización.
3. *Mejoramiento del desempeño.* Para ganar precisión en los algoritmos de reconocimiento de patrones y predicción.
4. *Entendimiento de los datos.* Para obtener conocimiento acerca del proceso que genera los datos.

Los métodos para seleccionar características se pueden dividir en tres: métodos *Filters*, métodos *Wrappers* y métodos *Embedded*.

- *Métodos Filters.* Estos métodos hacen un escalafón de características dependiendo de un índice de relevancia, además seleccionan las características sin optimizar el desempeño del predictor.

Los métodos *Filters* pueden distinguir cuatro diferentes medidas de evaluación: distancia, información, dependencia y consistencia, [2].

- *Métodos Wrappers.* Involucran el predictor como parte del proceso de selección de características. Los *Wrappers* usan el algoritmo de aprendizaje como una “caja negra” para calificar subconjuntos de características dependiendo de la potencia del predictor. Por otra parte, estos métodos son más costosos computacionalmente que los *Filters* pero suelen obtener mejores resultados, [2]
- *Métodos Embedded.* Estos hacen la selección de características durante el proceso de entrenamiento y son usualmente específicos para el algoritmo de aprendizaje dado.

Tanto los métodos *Wrappers* y *Filters* pueden conseguir subconjuntos de características muy diferentes bajo pequeñas perturbaciones del conjunto de datos.

Por otra parte, la extracción de características tiene cuatro aspectos fundamentales:

- Construcción de características
- Generación del subconjunto de características, estrategia de búsqueda.
- Definición del criterio de evaluación
- Estimación del criterio de evaluación

Los últimos tres son relevantes en la selección de características.

La mayor diferencia entre *Filters* y *Wrappers* es en el criterio de evaluación, los primeros no involucran ningún algoritmo de aprendizaje mientras que los segundos usan el desempeño del predictor entrenado con un subconjunto de características dado. Ambos pueden usar estrategias de búsqueda para explorar todas las posibles combinaciones de características.

Existen métodos híbridos que combinan las mejores propiedades de los *Filters* y *Wrappers*, en estos un *Filter* genera una lista ordenada de características y luego un algoritmo de aprendizaje calcula y genera subconjuntos anidados de características.

La estimación del criterio puede ser hecha mediante dos estrategias: *in-sample* o *out-of-sample*. En la primera se usan todos los datos de entrenamiento para calcular una estimación empírica, luego es probada estadísticamente para evaluar su significancia. La segunda tiene el enfoque de “apredizaje automático”, esta divide los datos en dos: un conjunto de entrenamiento usado para estimar los parámetros del modelo predictivo y un conjunto de validación para estimar el desempeño del predictor. Para disminuir la varianza en el desempeño del predictor se usa el promedio de múltiples ensayos o validación cruzada.

3.1.3. Relevancia de las características

Un enfoque de la selección de características es ordenar las características de acuerdo a un índice de relevancia. Estos métodos son rápidos y efectivos, particularmente cuando el número de características es grande y el número de ejemplos de entrenamiento es pequeño. En estos casos, los métodos que intentan buscar un subconjunto de características para una predicción óptima pueden ser lentos y tender al *sobre-ajuste*, [26].

Sin embargo hay limitaciones para calificar las características de manera individual debido a que se hacen suposiciones de su independencia en los métodos univariados. Cuando hay combinación de características pueden ocurrir dos cosas:

- Características que individualmente son irrelevantes pueden ser relevantes en el contexto de otras.
- Características que individualmente son relevantes pueden generar redundancia en combinación con otras.

A partir de esto se crean los métodos multivariados que toman en cuenta la dependencia de las características. Estos logran mejores resultados porque no hacen suposiciones acerca de la independencia de las características.

Dentro de los métodos *Wrapper* y *Embedded*, los métodos *Greedy* (*selección hacia adelante* y la *eliminación hacia atrás*) son los más populares.

El método de selección hacia adelante empieza con un conjunto vacío al cual progresivamente se le adicionan características dependiendo del índice de desempeño.

El método de eliminación hacia atrás empieza el conjunto de todas las características y progresivamente elimina las menos útiles. Ambos procedimientos son rápidos y robustos contra el sobre-ajuste.

3.2. Características de complejidad

Como se mencionó en los capítulos anteriores el problema que se aborda en este trabajo es la eliminación de los artefactos oculares presentes en las señales EEG.

Caracterizar señales cerebrales en un trabajo difícil debido a su comportamiento caótico [59], [55], [31], y la extracción de características temporales y espectrales no son suficientemente discriminatorias debido a que los artefactos oculares ocurren en tiempos impredecibles y en rangos de frecuencia que se superponen al espectro EEG. Por esta razón se proponen características de complejidad que aplican conceptos desde teoría de la información (entropía) hasta la geometría computacional (dimensión fractal) como solución a este problema.

Las características seleccionadas son:

A. Curtosis

La curtosis es el cúmulo de cuarto orden de una variable aleatoria y es utilizada como medida de gaussianidad. La curtosis de x denotada por $kurt(x)$ se define como :

$$kurt(x) = E\{x\}^4 - 3(E\{x^2\})^2 \quad (3.3)$$

donde E es el operador esperanza. Una variable aleatoria gaussiana tiene la curtosis igual a cero, esto permite discriminar las ICs EEG debido a su gran aleatoriedad y por ende su similitud con las variables gaussianas. Las variables con

curtosis negativa son llamadas *subgaussianas* y las variables con curtosis positiva son llamadas *supergaussianas*, [30].

B. Información de Fisher (FI)

La información de Fisher es definida como la cantidad de información que es esperada en una variable observable x acerca de un punto de interés θ en una muestra de n observaciones independientes. El principio de Descripción de Longitud Mínima (MDL) muestra que FI puede comparar diversas señales y cuantificar su complejidad, [37].

La FI puede ser definida en el espectro singular normalizado σ , el cual se halla mediante la descomposición en valores singulares (SVD).

La SVD de x produce M valores singulares, $\sigma_1, \sigma_2, \dots, \sigma_m$ conocidos como el espectro singular [8], [33].

La información de Fisher puede ser definida como:

$$I = \sum_{i=1}^{M-1} \frac{(\bar{\sigma}_{i+1} - \bar{\sigma}_i)^2}{\bar{\sigma}_i} \quad (3.4)$$

donde M es el número de valores singulares normalizados y $\bar{\sigma}_i = \sigma_i / \sum_{j=1}^M \sigma_j$ es su normalización.

C. Análisis de fluctuación sin tendencia

El análisis de fluctuación sin tendencia (DFA) es un método de análisis de datos, el cual puede descubrir y cuantificar propiedades dependientes de la escala en sistemas complejos, y permite la detección de auto-similaridad embebida en señales no-estacionarias [16], [63].

El procedimiento para calcular el DFA de una serie de tiempo x es el siguiente, [8]:

1. Primero integrar x dentro de una nueva serie $y = [y(1), y(2), \dots, y(N)]$, donde $y(k) = \sum_{i=1}^k (x_i - \bar{x})$.
2. La serie integrada es cortada en porciones de longitud n . En cada segmento, una línea es ajustada por mínimos cuadrados a los datos, mostrando la *tendencia* en ese segmento. La coordenada y de los segmentos de línea recta es denotada por $y_n(k)$.

3. La fluctuación de la raíz cuadrática media de la serie integrada es calculada por $F(n) = \sqrt{(1/N) \sum_{k=1}^N [y(k) - y_n(k)]^2}$.
4. La fluctuación puede ser definida como la pendiente de la línea de la relación entre $\log F(n)$ y $\log n$.

D. Parámetros de Hjorth

Los parámetros de Hjorth son métodos para indicar las propiedades estadísticas de una señal x en el dominio del tiempo. El parámetro de movilidad es definido como la raíz cuadrada de la relación entre la varianza de la primera derivada de la señal y la varianza de la señal. El parámetro de complejidad muestra cuán similar es la señal a una señal senoidal. El valor de la complejidad converge a 1 cuando la señal es más similar a una onda seno [46].

Estos son definidos respectivamente como:

$$\text{Movilidad} = \sqrt{\frac{M_2}{T_P}} \quad (3.5)$$

$$\text{Complejidad} = \sqrt{\frac{M_4 \cdot T_P}{M_2 \cdot M_2}} \quad (3.6)$$

donde

$$T_P = \sum x_i / N \quad (3.7)$$

$$M_2 = \sum d_i / N \quad (3.8)$$

$$M_4 = \sum (d_i - d_{i-1})^2 / N \quad (3.9)$$

con $d_i = x_i - x_{i-1}$.

E. Dimensión Fractal de Higuchi (HFD)

HDF es un algoritmo para medir la dimensión fractal de una serie discreta de tiempo. Además, HDF puede ser usada para cuantificar la complejidad y la autosimilaridad de una señal [24].

Dada una serie de tiempo $X = x_1, x_2, \dots, x_N$, el algoritmo que calcula la HFD puede ser descrito como:

Primero, se generan k nuevas series de tiempo X_k^m definidas por:

$$x_m, x_{m+k}, x_{m+2k}, \dots, x_{m+\lfloor(N-m)/k\rfloor k} \quad (3.10)$$

donde $m=1,2,\dots,k$. La longitud de cada serie es calculada como:

$$L(m, k) = \frac{\sum_{i=2}^{\lfloor(N-m)/k\rfloor} |x_{m+ik} - x_{m+(i-1)k}| (N-1)}{\lfloor(N-m)/k\rfloor k} \quad (3.11)$$

donde N es la longitud de la serie de tiempo original X .

La longitud promedio es definida por:

$$L(k) = \frac{1}{k} \sum_{m=1}^k L(m, k) \quad (3.12)$$

Este procedimiento es repetido k_{max} veces para cada k desde 1 hasta k_{max} , luego mediante mínimos cuadrados se encuentra la línea que mejor se ajusta a la curva $\ln(L(k))$ versus $\ln(1/k)$ y la pendiente de esa línea es la HFD.

F. Exponente de Hurst(HE)

El exponente de Hurst es un método estadístico muy usado para inferir las propiedades de una serie de tiempo sin hacer suposiciones acerca de su estacionariedad [45]. El HE es la medida de la suavidad de una serie de tiempo fractal basada en el comportamiento asintótico del rango reescalado del proceso [34].

El HE de una serie de tiempo $X = x_1, x_2, \dots, x_N$, es calculado como:

El primer paso es calcular la desviación acumulada desde la media de la serie de tiempo dentro del rango T .

$$X(t, T) = \sum_{i=1}^t (x_i - \bar{x}) \quad (3.13)$$

donde

$$\bar{x} = \frac{1}{T} \sum_{i=1}^T x_i, \quad t \in [1, \dots, N] \quad (3.14)$$

Luego, el rango reescalado $R(T)/S(T)$ es definido por:

$$\frac{R(T)}{S(T)} = \frac{\text{máx}(X(t, T)) - \text{mín}(X(t, T))}{\sqrt{(1/T) \sum_{t=1}^T [x_t - \bar{x}]^2}} \quad (3.15)$$

El HE es la pendiente de la línea producida por $\ln(R(n)/S(n))$ versus $\ln(n)$ para $n \in [2, \dots, N]$.

G. Dimensión Fractal de Petrosian

El método de Petrosian calcula la dimensión fractal de una serie binaria de tiempo [22]. Primero se convierte la secuencia en una secuencia binaria y luego la PFD es calculada como:

$$PFD = \frac{\log_{10} N}{\log_{10} N + \log_{10}(N/(N + 0,4N_\delta))} \quad (3.16)$$

donde N es la longitud de la serie y N_δ es el número de cambios de signo en la derivada de la señal [8].

3.3. Extracción de características discriminates entre EEG y artefactos oculares

En el capítulo anterior se mostró como ICA permite separar las componentes de los artefactos oculares y las componentes de las señales EEG. En esta sección se muestra el procedimiento de extracción de características de las componentes independientes para su discriminación por el posterior modelo de predicción.

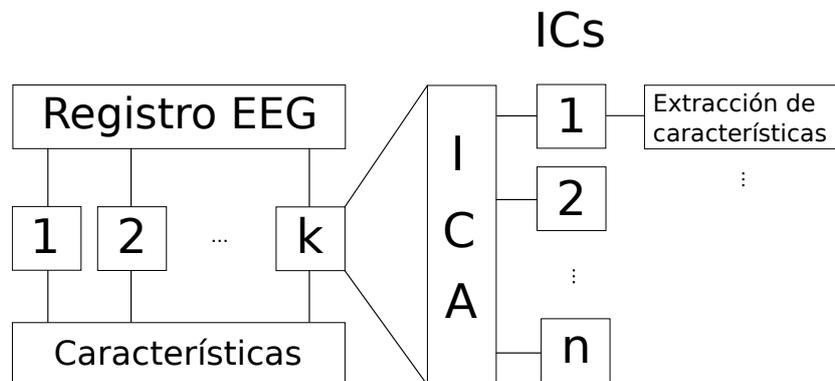


Figura 3.1: Esquema del procedimiento de extracción de características desde los registros EEG.

Como primer paso, los registros EEG fueron fragmentados en porciones de 8 segundos, luego a estas porciones se les aplicó ICA para obtener sus 22 componentes independientes, finalmente de cada una de ellas se extrajeron las 8 características de complejidad expuestas anteriormente mediante la librería *PyEEG*, [8],

Figura 3.2. Las características fueron estandarizadas mediante la ecuación (4.1) para favorecer el desempeño del predictor.

El vector de características extraído tiene un tamaño de 1540 patrones. En la Figura 3.2 se muestra las relaciones diádicas entre las características mediante su proyección bidimensional permitiendo reconocer agrupamientos emergentes de las componentes EEG y de los artefactos oculares, además se hace notar la naturaleza desbalanceada de los datos debido a que tan solo el 10% de los datos pertenecen a los artefactos oculares.

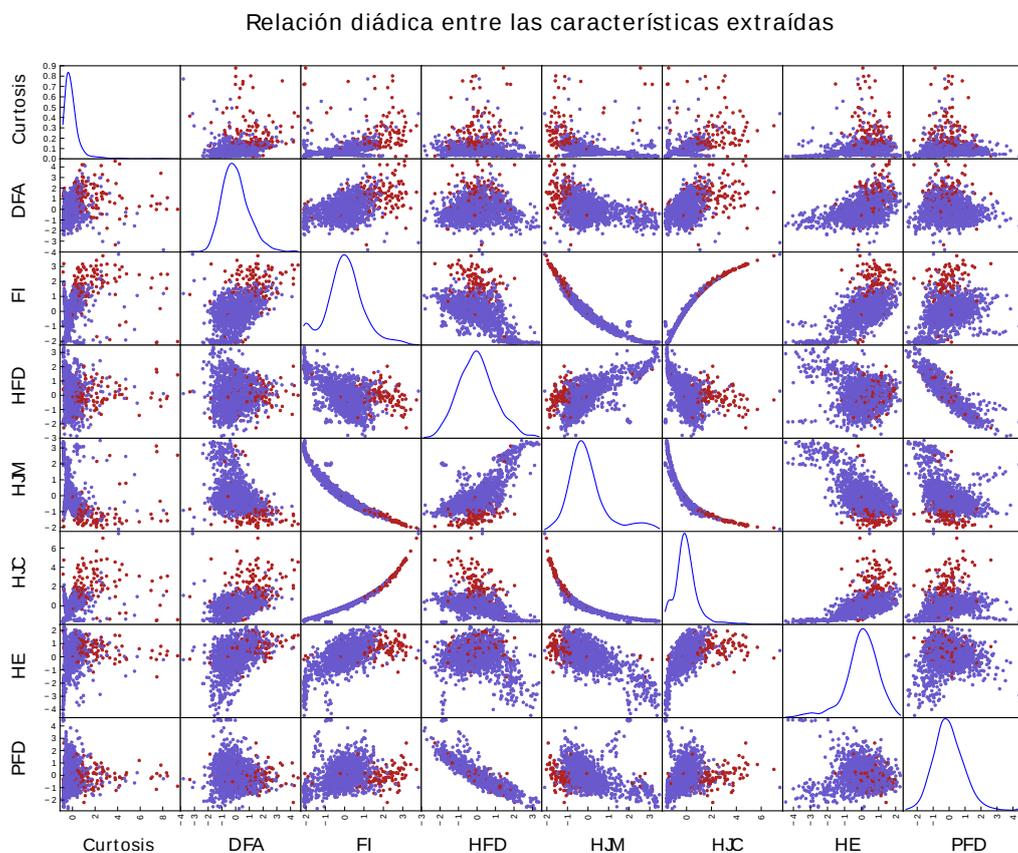


Figura 3.2: Matriz de relación entre las características extraídas desde las componentes independientes. Los patrones azules pertenecen a la clase EEG y los patrones rojos a los artefactos oculares. Las características son: curtosis, Análisis de Fluctuación sin Tendencia (DFA), Información de Fisher (FI), Dimensión Fractal de Higuchi (HFD), Movilidad de Hjorth (HJM), Complejidad de Hjorth (HJC), Exponente de Hurst (HE) y Dimensión Fractal de Petrosian (PFD).

Podemos ver como la combinación entre ciertas características generan complementaridad creando agrupamientos más definidos, por ejemplo: DFA vs HJM, FI vs HDF, FI vs HE, FI vs PFD, HFD vs HJM, HJM vs HE, HJM vs PFD, HJC vs HE o HJC vs PFD. Además, emergen relaciones interesantes entre las características, por ejemplo, logarítmicas entre FI vs HJM, FI vs HJC y HJM vs HJC o lineales entre las dimensiones fractales de Higuchi y Petrosian. Esto demuestra que características que individualmente parecen poco relevantes cuando interactúan con otras permiten ver información oculta de gran importancia.

Por otra parte, la mezcla de algunas características incrementa la varianza extraclase creando de esta manera un margen de separación más amplio para aumentar el desempeño del algoritmo de clasificación. Sin embargo, como podemos ver el margen de separación es complejo en un espacio de características de baja dimensión, por esta razón, un incremento de esta dimensión se reflejará en el mejoramiento de la eficiencia del predictor.

En el siguiente capítulo se muestra un método de clasificación de las ICs de señales EEG y de artefactos oculares basado en aprendizaje supervisado, más específicamente con una máquina de soporte vectorial balanceada.

Toda obra grande, en arte como en ciencia, es el resultado de una gran pasión puesta al servicio de una gran idea.

Santiago Ramón y Cajal

4

Aprendizaje supervisado: SVM balanceada

El problema del reconocimiento de patrones ha sido fundamental a lo largo de la historia del procesamiento de los datos. Este ha jugado un papel importante en diversos campos como: astronomía, astrofísica, medicina, economía, sociología, ingeniería, química, entre muchos otros. El campo de reconocimiento de patrones se basa en el descubrimiento automático de regularidades en los datos a través de algoritmos computacionales y mediante esas regularidades tomar acciones tales como clasificar los datos en diferentes categorías, [13].

En este capítulo se abordará el problema de clasificación de los artefactos oculares y señales EEG desde el punto de vista del aprendizaje supervisado mediante

una máquina de soporte vectorial balanceada.

4.1. Aprendizaje supervisado

El aprendizaje supervisado es la tarea de aprendizaje automático en la cual se infiere una función que relaciona las entradas con las salidas desde los datos de entrenamiento. Los datos de entrenamiento consisten de un conjunto de ejemplos de entrenamiento. Cada ejemplo está formado de un objeto de entrada y un valor de salida deseado. Un algoritmo de aprendizaje supervisado analiza los datos de entrenamiento y produce una función inferida, la cual es llamada *clasificador*, si la salida es cualitativa o *función de regresión* si la salida es cuantitativa. La función debe predecir el valor correcto de salida para cualquier entrada.

Un problema de aprendizaje supervisado puede ser definido como: Dado un vector \mathbf{x} con N muestras llamado *conjunto de entrenamiento*, el cual es usado para ajustar los parámetros de un modelo adaptativo y que está acompañado de un *vector objetivo* \mathbf{t} el cual representa la etiqueta o clase de la correspondiente muestra, se debe estimar una función de clasificación $\mathbf{y}(\mathbf{x})$ la cual toma una muestra x_i y genera una salida y_i que debe corresponder con el objetivo t_i .

La función $\mathbf{y}(\mathbf{x})$ es determinada durante la fase de entrenamiento, también conocida como fase de aprendizaje, basada en los datos de entrenamiento. Una vez el modelo es entrenado este puede determinar la identidad de una nueva muestra, la cual corresponde a un *conjunto de prueba*. La habilidad para categorizar correctamente nuevas muestras que difieren de las usadas en el entrenamiento es conocida como *generalización*, siendo esta un objetivo central en el reconocimiento de patrones [13].

4.1.1. Compensación sesgo-varianza

En el aprendizaje supervisado existe una relación entre el sesgo y la varianza del predictor que conforma el *error de generalización*. Supongamos que tenemos diferentes conjuntos de entrenamiento igualmente buenos. Un algoritmo de aprendizaje está sesgado para una entrada particular x_i , si cuando entrenamos con cada uno de esos conjuntos de datos, la predicción es sistemáticamente incorrecta. Por otra parte, un algoritmo de aprendizaje tiene una varianza alta para una entrada particular x_i si este produce diferentes valores de salida cuando entrenamos con diferentes conjuntos de datos. El efecto del sesgo y la varianza en un clasificador se muestra en la Figura 4.1.

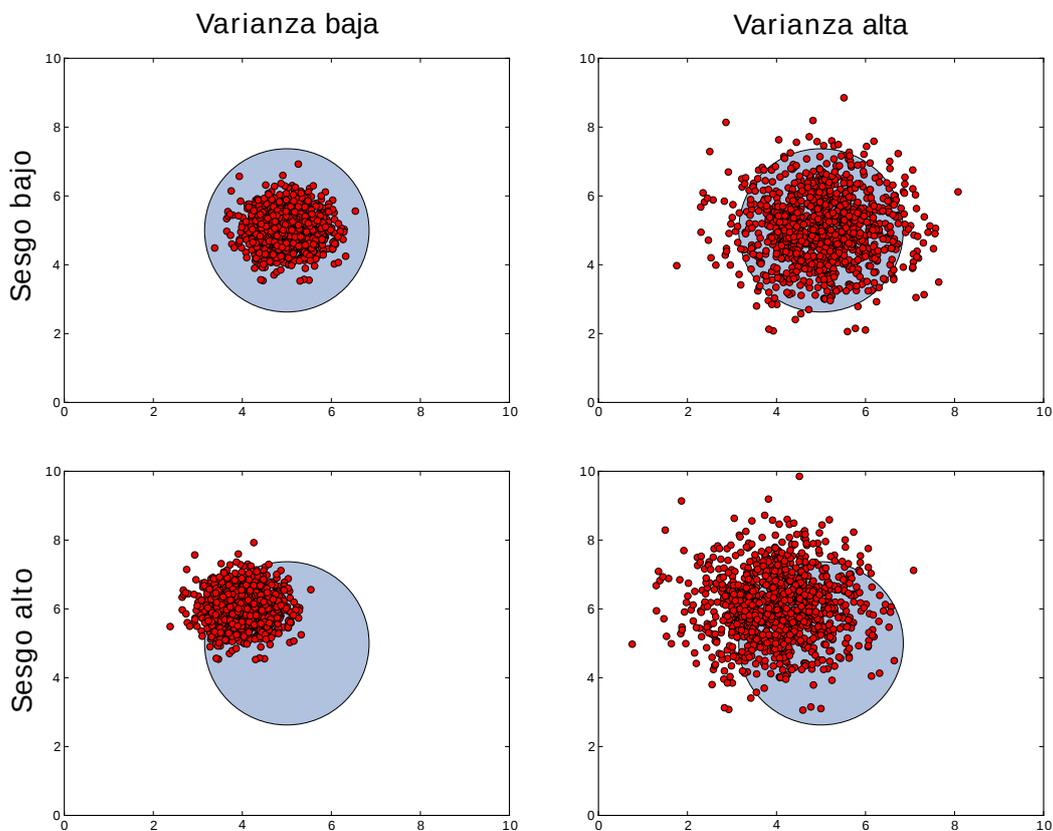


Figura 4.1: Efecto del sesgo y la varianza en un clasificador. En la parte superior-izquierda vemos una buena compensación entre el sesgo y la varianza debido a que el límite de decisión del clasificador, *círculo azul*, se adapta correctamente a los datos de prueba. En la parte superior-derecha existe una alta varianza en el clasificador, debido a que el límite de decisión no se adapta a la totalidad de los datos generando un aumento del error de generalización. En la parte inferior-izquierda el límite de decisión está sesgado de los datos de prueba, causando que el error de generalización aumente. En la parte inferior-derecha se muestra el peor de los casos, es decir, cuando el clasificador tiene una varianza alta y un sesgo alto, en este caso el error de generalización será mayor a los dos casos expuestos anteriormente.

El error de generalización es la suma del cuadrado del sesgo y la varianza del clasificador y es definido por:

$$E(x_i) = \sigma^2 + [\text{Sesgo}^2(y(x_i)) + \text{Var}(y(x_i))] \quad (4.1)$$

donde σ^2 es el error irreducible.

Más generalmente, a medida que incrementa la complejidad del modelo, la varianza tiende a incrementar y el cuadrado del sesgo a decrecer, si la complejidad

del modelo se reduce, la varianza decrece y el cuadrado del sesgo aumenta, Figura 4.2. Debido a esto se busca un modelo que compense la relación sesgo-varianza y que permita su generalización.

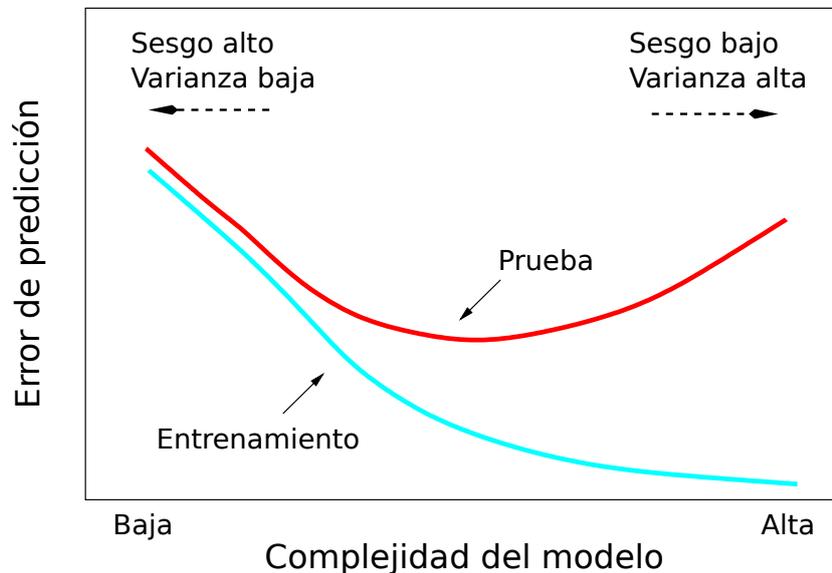


Figura 4.2: Error de entrenamiento y de prueba en función de la complejidad del modelo, [28]

Algunas de las acciones que se pueden tomar en cuenta para mejorar el error de generalización son:

a). Varianza alta

- Aumentar el número de datos de entrenamiento
- Disminuir el número de características o dimensión del espacio de entrada
- Incrementar el parámetro de regularización del modelo

b). Sesgo alto

- Aumentar el número de características o dimensión del espacio de entrada
- Disminuir el parámetro de regularización del modelo

4.2. Máquinas de vectores soporte

Las Máquinas de Vectores Soporte (SVM) son un conjunto de algoritmos de aprendizaje basados en la teoría de aprendizaje estadístico. La técnica SVM fue

desarrollada inicialmente como un clasificador lineal para problemas biclase con datos separables. Es decir, dado un problema de clasificación en el que los objetos a clasificar pertenecen a una de dos clases posibles, la máquina de aprendizaje implementada por SVM permite encontrar un modelo lineal que separa los elementos de ambas clases.

Posteriormente, SVM fue adaptado a problemas de clasificación con datos no separables. Actualmente, su funcionalidad se ha extendido para poder resolver problemas de regresión [43] y agrupamiento.

4.2.1. Caso separable

Para entender el funcionamiento de SVM se parte del caso más sencillo, la separación de dos clases. SVM busca encontrar el hiperplano de separación óptimo entre las dos clases, esto se logra mediante la maximización del *margin de separación* M en el cual se encuentran las muestras más cercanas entre las clases y por ende las más difíciles de separar, a estas muestras se les llama *vectores soporte*, Figura 4.3.

Partiendo del conjunto de entrenamiento formado por N pares $(\mathbf{x}_1, t_1), (\mathbf{x}_2, t_2), \dots, (\mathbf{x}_N, t_N)$, con $\mathbf{x}_i \in R^p$ y $t_i \in \{-1, 1\}$, se puede definir un hiperplano de la siguiente manera:

$$y(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + b = 0 \quad (4.2)$$

donde \mathbf{w} es un vector ortogonal al hiperplano con $\|\mathbf{w}\| = 1$ y b es un parámetro de sesgo. Los datos del conjunto de prueba serán clasificados por:

$$G(\mathbf{x}) = \text{sign}[y(\mathbf{x})] \quad (4.3)$$

Debido a que los datos son linealmente separables, existe al menos un valor de w y b que satisface $y(\mathbf{x}_i) > 0$ para las muestras con $t_i = +1$ y $y(\mathbf{x}_i) < 0$ para las muestras con $t_i = -1$; generalizando lo anterior, la condición es:

$$t_i y(\mathbf{x}_i) > 0, \quad i = 1, 2, \dots, N \quad (4.4)$$

La distancia perpendicular de un punto \mathbf{x} desde un hiperplano definido por $y(\mathbf{x}) = 0$ está dada por $|y(\mathbf{x})|/\|\mathbf{w}\|$. Entonces la distancia de un punto \mathbf{x}_i a la superficie de decisión está dada por:

$$\frac{t_i y(\mathbf{x}_i)}{\|\mathbf{w}\|} = \frac{t_i (\mathbf{w}^T \mathbf{x}_i + b)}{\|\mathbf{w}\|} \quad (4.5)$$

El margen está dado por la distancia perpendicular al punto \mathbf{x}_i más cercano desde el conjunto de datos, y se deben optimizar los parámetros \mathbf{w} y b para maximizar

dicha distancia. El máximo margen se encuentra resolviendo la siguiente expresión:

$$\arg \max_{\mathbf{w}, b} \left\{ \frac{1}{\|\mathbf{w}\|} \min_i [t_i(\mathbf{w}^T \mathbf{x}_i + b)] \right\} \quad (4.6)$$

donde se toma el factor $1/\|\mathbf{w}\|$ fuera de la optimización sobre n porque \mathbf{w} no depende de n .

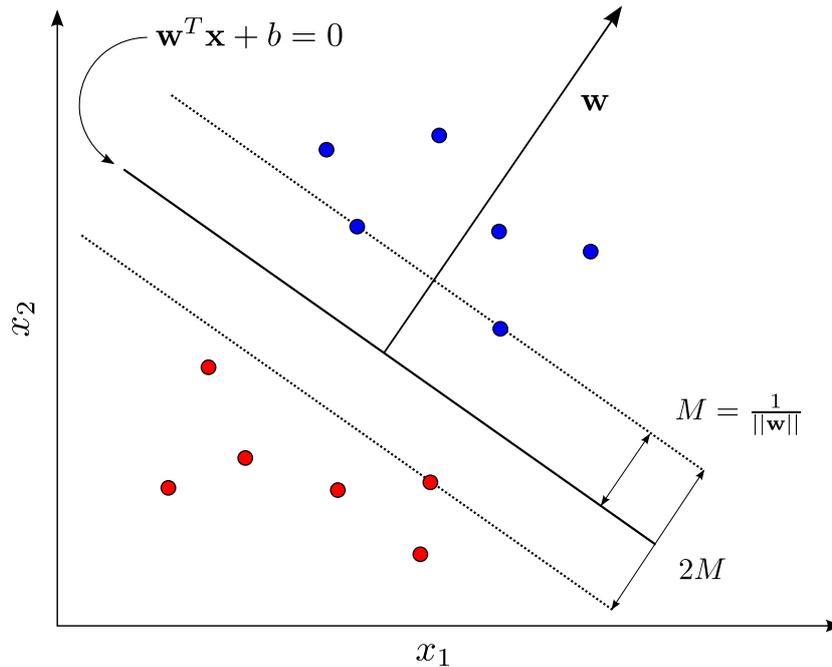


Figura 4.3: Hiperplano de separación entre dos clases separables linealmente.

La solución del problema de optimización puede resultar muy compleja entonces se opta por convertirlo en un problema equivalente más fácil de resolver. Se puede observar que si se hace un reescalamiento $\mathbf{w} \rightarrow \kappa \mathbf{w}$ y $b \rightarrow \kappa b$, entonces la distancia de cualquier punto \mathbf{x}_i a la superficie de decisión, dada por $t_i y(\mathbf{x}_i) / \|\mathbf{w}\|$ permanece igual. Entonces podemos hacer:

$$t_i(\mathbf{w}^T \mathbf{x}_i + b) = 1 \quad (4.7)$$

para los puntos más cercanos a la superficie de decisión. La restricción para todos los puntos es:

$$t_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1, \quad i = 1, 2, \dots, N \quad (4.8)$$

El problema de optimización requiere simplemente maximizar $1/\|\mathbf{w}\|$, lo que es equivalente a minimizar $\|\mathbf{w}\|^2$, entonces se tiene:

$$\arg \min_{\mathbf{w}, b} \frac{1}{2} \|\mathbf{w}\|^2 \quad (4.9)$$

El factor de $1/2$ es incluido por conveniencia. Este problema es un ejemplo de programación cuadrática en el que se busca minimizar una función cuadrática sujeta a restricciones lineales.

Para resolver el problema se plantea en base a multiplicadores de Lagrange de la siguiente manera:

$$L(\mathbf{w}, b, \alpha) = \frac{1}{2} \|\mathbf{w}\|^2 - \sum_{i=1}^N \alpha_i \{t_i(\mathbf{w}^T \mathbf{x}_i + b) - 1\} \quad (4.10)$$

donde $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_N)^T$ son los multiplicadores de Lagrange. Desarrollando las derivadas de $L(\mathbf{w}, b, \alpha)$ respecto a \mathbf{w} y b e igualandolas a cero, se obtienen las siguientes condiciones:

$$\mathbf{w} = \sum_{i=1}^N \alpha_i t_i \mathbf{x}_i \quad (4.11)$$

$$\sum_{i=1}^N \alpha_i t_i = 0 \quad (4.12)$$

Sustituyendo (5.11) y (5.12) en la función de costo (5.10), el problema es equivalente a maximizar:

$$\tilde{L}(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j t_i t_j (\mathbf{x}_i \cdot \mathbf{x}_j) \quad (4.13)$$

sujeto a las restricciones

$$\alpha_i \geq 0, \quad i = 1, 2, \dots, N \quad (4.14)$$

$$\sum_{i=1}^N \alpha_i t_i = 0 \quad (4.15)$$

El hiperplano de separación puede ser expresado en relación a los multiplicadores de Lagrange de la siguiente manera:

$$y(\mathbf{x}) = \sum_{i=1}^N \alpha_i t_i (\mathbf{x} \cdot \mathbf{x}_i) + b \quad (4.16)$$

Todas las muestras de entrenamiento tienen asociado un multiplicador de lagrange, sin embargo la mayoría de estos son cero, excepto los que pertenecen a los *vectores soporte*, [13].

Una vez resuelto el problema de optimización y hallado el valor de α , se determina el valor de b teniendo en cuenta que los vectores soporte x_i satisfacen $t_i y(\mathbf{x}_i) = 1$. Usando (5.16) tenemos

$$t_i \left(\sum_{m \in S} \alpha_m t_m (\mathbf{x}_i \cdot \mathbf{x}_m) + b \right) = 1 \quad (4.17)$$

donde S es el conjunto de índices de los vectores soporte. Aunque la ecuación puede ser resuelta eligiendo un número arbitrario de vectores soporte \mathbf{x}_i , una solución numéricamente más estable es la siguiente:

$$b = \frac{1}{N_S} \sum_{i \in S} \left(t_i - \sum_{m \in S} \alpha_m t_m (\mathbf{x}_i \cdot \mathbf{x}_m) \right) \quad (4.18)$$

donde N_S es el número total de los vectores soporte.

4.2.2. Caso no-separable

Existen dos casos donde las clases no están completamente separadas y el modelo SVM debe ser adaptado. La primera es cuando la clasificación se hace con una función de decisión lineal pero las muestras de entrenamiento no pueden separarse sin errores, la segunda, ocurre cuando las muestras no pueden clasificarse mediante una función lineal en el espacio de entrada y se debe aplicar una transformación a las muestras y llevarlas a una dimensión donde sean separables linealmente.

a.) Clasificación con errores

En este caso el hiperplano de separación es lineal pero existe al menos una muestra que se encuentra al lado opuesto de este. En la formulación del hiperplano, una muestra de entrenamiento es *no separable* cuando no cumple la ecuación (5.8), esto es debido a que la muestra está ubicada entre los márgenes o al lado equivocado de la frontera de decisión, [43].

La SVM debe ser modificada de tal manera que permita que algunas de las muestras de entrenamiento sean clasificadas de manera incorrecta. Para el problema de optimización se adiciona una penalización que aumenta con la distancia de separación entre la muestra mal clasificada y la frontera de decisión. Para esto se introducen las *variables de holgura*, $\xi_i \geq 0$ donde $i = 1, 2, \dots, N$. Cada muestra del conjunto de entrenamiento tiene una variable de holgura asociada. Para las

muestras que están clasificadas correctamente $\xi_i = 0$ y para las muestras mal clasificadas $\xi_i = |t_i - y(\mathbf{x}_i)|$. Entonces para las muestras que están sobre la frontera de decisión $y(\mathbf{x}_i) = 0$ tenemos $\xi_i = 1$, y para las muestras mal clasificadas $\xi_i > 1$, Figura 4.4.

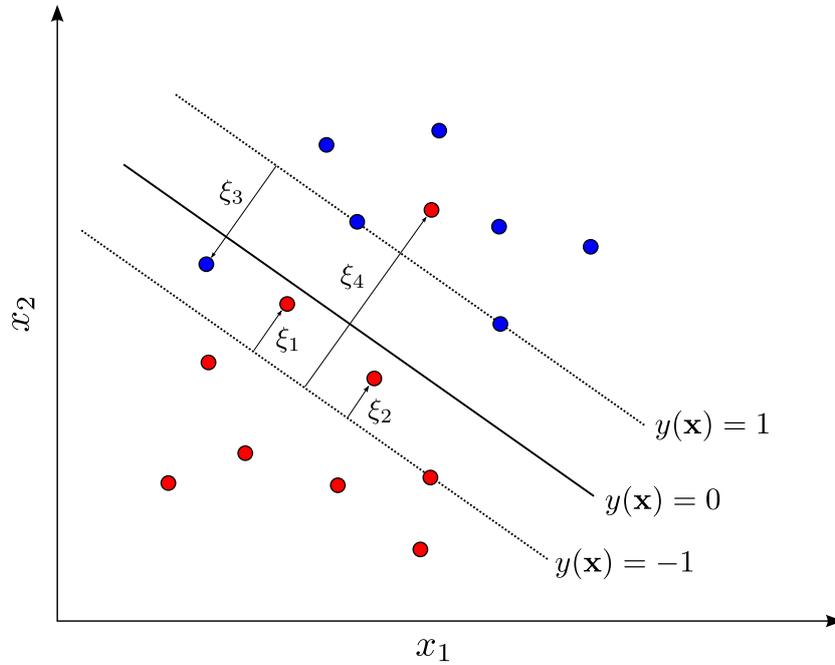


Figura 4.4: Hiperplano de separación y variables de holgura en el caso no separable.

El objetivo es maximizar el margen teniendo en cuenta la penalización de las muestras mal clasificadas. El problema de optimización es:

$$\arg \min_{\mathbf{w}, b, \xi} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^N \xi_i \quad (4.19)$$

sujeto a la restricción

$$t_i y(\mathbf{x}_i) \geq 1 - \xi_i, \quad i = 1, 2, \dots, N \quad (4.20)$$

donde $C > 0$ controla la compensación entre la penalización de la variable de holgura y el margen. El parámetro C es análogo a un coeficiente de regularización debido a que controla la relación entre minimizar los errores de entrenamiento y la complejidad del modelo.

El lagrangiano correspondiente está dado por

$$L(\mathbf{w}, b, \alpha) = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^N \xi_i - \sum_{i=1}^N \alpha_i \{t_i y(\mathbf{x}_i) - 1 + \xi_i\} - \sum_{i=1}^N \mu_i \xi_i \quad (4.21)$$

donde $\alpha_i \geq 0$ y $\mu_i \geq 0$ son los multiplicadores de Lagrange.

Derivando $L(\mathbf{w}, b, \alpha)$ respecto a \mathbf{w} , b y ξ_i e igualando a cero obtenemos tres expresiones

$$\mathbf{w} = \sum_{i=1}^N \alpha_i t_i \mathbf{x}_i \quad (4.22)$$

$$\sum_{i=1}^N \alpha_i t_i = 0 \quad (4.23)$$

$$\alpha_i = C - \mu_i \quad (4.24)$$

Ahora se reemplaza (5.22)(5.23) y (5.24) en (5.21) dando como resultado

$$\tilde{L}(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j t_i t_j (\mathbf{x}_i \cdot \mathbf{x}_j) \quad (4.25)$$

sueto a las restricciones

$$0 \leq \alpha_i \leq C \quad (4.26)$$

$$\sum_{i=1}^N \alpha_i t_i = 0 \quad (4.27)$$

La función de separación es igual al caso anterior (5.16).

b.) Clasificación no-lineal

Cuando las muestras de entrenamiento no son separables linealmente en el espacio de entrada, estas son transformadas mediante una función no-lineal Φ llamada *kernel* a un espacio transformado donde si pueden ser separadas con una función lineal, Figura 4.5.

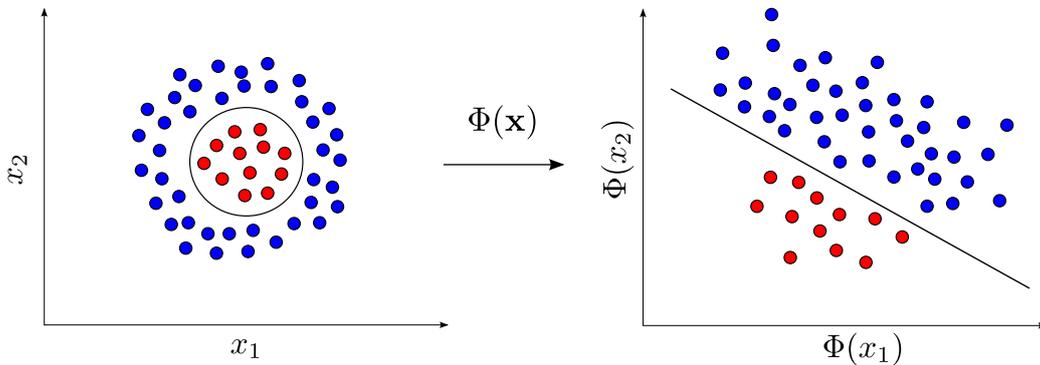


Figura 4.5: Mapeo de las muestras del espacio de entrada al espacio de Hilbert mediante una función *kernel* $\Phi(\mathbf{x})$.

El problema de optimización es el mismo que se muestra en la ecuación (5.19), sujeto a la restricción

$$t_i(\mathbf{w}^T \Phi(\mathbf{x}_i) + b) \geq 1 - \xi_i, \quad i = 1, 2, \dots, N \quad (4.28)$$

Siguiendo los pasos expuestos en los casos anteriores encontramos el funcional a maximizar

$$\tilde{L}(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j t_i t_j \Phi(\mathbf{x}_i, \mathbf{x}_j) \quad (4.29)$$

sujeto a las restricciones (5.26) y (5.27).

La función de decisión es

$$y(\mathbf{x}) = \sum_{i=1}^N \alpha_i t_i \Phi(\mathbf{x}, \mathbf{x}_i) + b \quad (4.30)$$

4.3. Máquinas de vectores soporte balanceadas

Los modelos anteriormente expuestos suponen un conjunto de entrenamiento balanceado, es decir, el número de muestras de ambas clases es igual. Sin embargo, en el mundo real los datos son desbalanceados. Recientemente, el problema del desbalanceo de clases ha recibido gran atención debido a que el desempeño de los algoritmos decae significativamente [64].

SVM ha sido usada en este tipo de problemas, pero puede ser sensible al alto desbalanceo de las clases, resultando en una deficiencia en la clasificación de los datos pertenecientes a la clase con menos muestras. Además, esto genera un gran sesgo en la clase mayoritaria [60].

La definición de una SVM balanceada parte del conjunto de entrenamiento el cual tiene N patrones. Para clases desbalanceadas l^+ (l^-) son el número de muestras para $t_i = +1$ ($t_i = -1$) y $l^+ + l^- = N$. Claramente, si las dos clases son balanceadas la relación será $l^+ = l^- = 1/2$, [4].

El problema de optimización resultante es:

$$\arg \min_{\mathbf{w}, b, \xi} \frac{1}{2} \|\mathbf{w}\|^2 + C^+ \sum_{i=1}^N \xi_i + C^- \sum_{i=1}^N \xi_i \quad (4.31)$$

sujeto a

$$t_i(\mathbf{w}^T \Phi(\mathbf{x}_i) + b) \geq 1 - \xi_i, \quad i = 1, 2, \dots, N \quad (4.32)$$

$$\xi_i \geq 0, \quad i = 1, 2, \dots, N \quad (4.33)$$

donde C es definido como $C^+ = \frac{C\sqrt{\lambda}}{l}$ ($C^- = \frac{C}{l\sqrt{\lambda}}$) con $\lambda = \frac{C^+}{C^-} = \frac{l^-}{l^+}$.

Este enfoque es normalizado en relación al número de patrones en ambas clases. Finalmente, la función a optimizar es la expuesta en (5.29) sujeta a las restricciones

$$0 \leq \alpha_i \leq C^+ \quad \text{para } t_i = +1 \quad (4.34)$$

$$0 \leq \alpha_i \leq C^- \quad \text{para } t_i = -1 \quad (4.35)$$

$$\sum_{i=1}^N \alpha_i t_i = 0 \quad (4.36)$$

Las funciones *kernel* más utilizadas son:

- *Polinomios de grado d*

$$\Phi(\mathbf{x}, \mathbf{x}') = (1 + (\mathbf{x} \cdot \mathbf{x}'))^d \quad (4.37)$$

- *Base radial*

$$\Phi(\mathbf{x}, \mathbf{x}') = \exp(-\gamma \|\mathbf{x} - \mathbf{x}'\|^2) \quad (4.38)$$

donde γ es el ancho del *kernel*, [1].

- *Red neuronal*

$$\Phi(\mathbf{x}, \mathbf{x}') = \tanh(\kappa_1 (\mathbf{x} \cdot \mathbf{x}') + \kappa_2) \quad (4.39)$$

4.4. Evaluación de modelos de aprendizaje supervisado

Para evaluar el desempeño del algoritmo de aprendizaje se hace uso de un grupo de muestras diferente al conjunto de entrenamiento, estos patrones son llamados *conjunto de prueba*.

Luego de que las muestras del conjunto de prueba sean clasificadas, el desempeño del modelo puede ser evaluado con una serie de medidas de desempeño como se muestra a continuación.

4.4.1. Precisión, sensibilidad, especificidad y exactitud

La *precisión* muestra cuantas de las muestras clasificadas como pertenecientes a la clase positiva son relevantes.

$$PPV = \frac{TP}{TP + FP} \quad (4.40)$$

La *sensibilidad* indica la capacidad del estimador para dar como casos positivos los casos realmente pertenecientes a la clase positiva.

$$TPR = \frac{TP}{TP + FN} \quad (4.41)$$

La *especificidad* indica la capacidad del estimador para dar como casos negativos los casos realmente pertenecientes a la clase negativa.

$$SPC = \frac{TN}{TN + FP} \quad (4.42)$$

La *exactitud* es la proximidad del valor estimado por el predictor al valor verdadero.

$$ACC = \frac{TP + TN}{TP + FP + TN + FN} \quad (4.43)$$

Estas medidas son expresadas en términos de TP verdaderos positivos, TN verdaderos negativos, FP falsos positivos y FN falsos negativos y pueden ser visualizadas mediante una *matriz de confusión*, Tabla 4.1.

	Clase estimada	
Clase verdadera	Clase A	Clase B
Clase A	TP	FN
Clase B	FP	TN

Tabla 4.1: Matriz de confusión.

4.4.2. Marcador F1

El marcador F1 mezcla la precisión y la sensibilidad del estimador, proporcionando una medida de desempeño en relación a la clase positiva. El marcador F1 es definido como:

$$F1 = 2 \frac{PPV \cdot TPR}{PPV + TPR} \quad (4.44)$$

En problemas de clases desbalanceadas el marcador F1 permite analizar la eficiencia del estimador respecto a la clase minoritaria.

4.4.3. Curva ROC

El desempeño del predictor binario puede también ser evaluado de manera gráfica a través de curvas ROC. Las curvas ROC tienen en cuenta la sensibilidad y la tasa de falsos positivos ($1 - \text{especificidad}$) del modelo. La estructura de una curva ROC se muestra en la Figura 4.6.

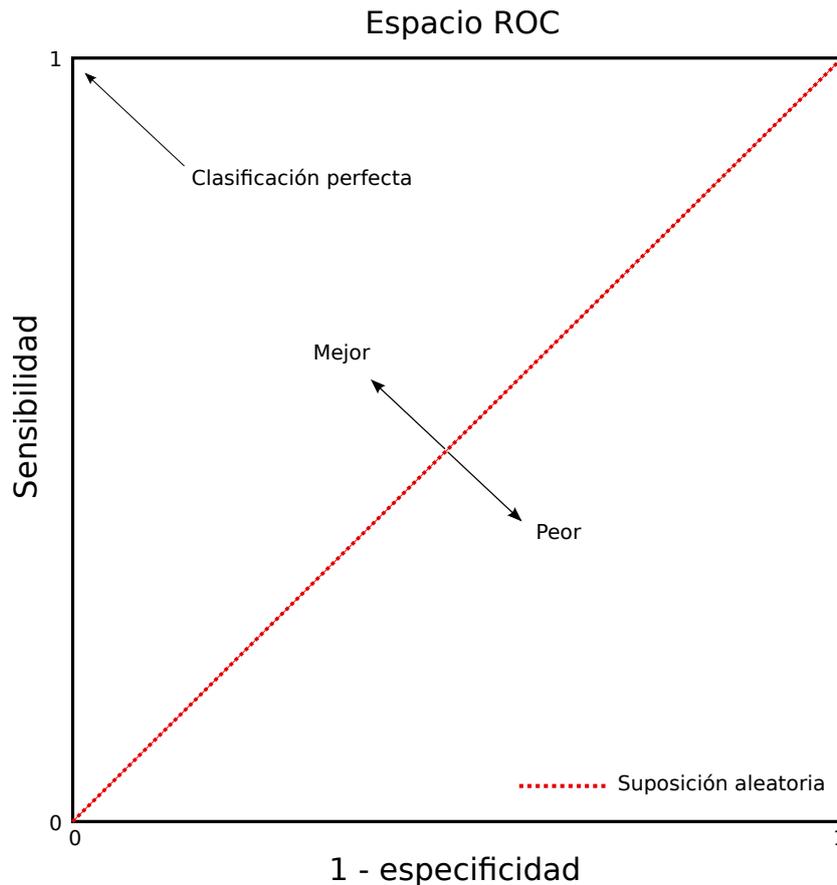


Figura 4.6: Espacio de la curva ROC. La línea roja indica el desempeño de un clasificador aleatorio, a medida que la curva avanza hacia la parte superior izquierda mejora el desempeño del estimador.

El área bajo la curva ROC es un indicador de desempeño muy utilizado. El AUC mide la exactitud del predictor, por ejemplo, una clasificación perfecta tiene una $AUC = 1$, en cambio una clasificación con $AUC = 0,5$ es totalmente aleatoria.

4.4.4. Validación cruzada

La validación cruzada es probablemente el método más simple y el más usado para estimar el error de predicción [28]. Idealmente se tuviese una gran cantidad de datos se conformaría un conjunto de validación y se evaluaría el desempeño del modelo de predicción. Sin embargo, en los caso donde los datos son escasos, esto no es posible. La validación cruzada de K particiones usa parte de los datos para entrenar el modelo, y una parte diferente para probarlo.

Las muestras son divididas en K partes de igual tamaño, Figura 4.7. Se elige una k -ésima parte de los datos y el modelo es entrenado con las $K - 1$ partes restantes, luego el error de predicción es calculado con la k -ésima parte. Este proceso se hace para $k = 1, 2, \dots, K$ y se combinan las K estimaciones del error de predicción.

Se debe elegir un número de particiones adecuado debido a que el sesgo y la varianza del modelo pueden depender de este. Si $K \rightarrow N$ el sesgo disminuye, pero la varianza aumenta debido a la similaridad de los N conjuntos de entrenamiento resultantes. Por otra parte, si K disminuye la varianza del predictor se reduce, pero el sesgo puede ser un problema dependiendo de cómo cambia el desempeño del predictor respecto al tamaño del conjunto de entrenamiento

1	2	3	4	5
Entrenamiento	Entrenamiento	Entrenamiento	Validación	Entrenamiento

Figura 4.7: Ejemplo de una iteración de la validación cruzada con $K = 5$. En este caso las particiones 1, 2, 3 y 5 son empleadas para entrenar el modelo y la partición 4 para validarlo.

4.5. Eliminación de artefactos oculares en señales EEG mediante una SVM balanceada.

Retomando el problema de interés de este libro, y teniendo en cuenta la naturaleza desbalanceada de los datos, la cual fue verificada en el capítulo anterior, en esta sección se exponen los resultados obtenidos de la clasificación y posterior eliminación de artefactos oculares presentes en señales EEG mediante una SVM balanceada.

En este caso se usó una SVM con una función *kernel* de base radial, (5.38). Además, se estimó la relación de desbalanceo $\lambda = 9,83$.

Para obtener una medida confiable del desempeño del estimador se usó validación cruzada con $K = 10$. En este caso, el desempeño del modelo SVM depende principalmente de dos parámetros: el parámetro de regularización C y el ancho del *kernel* γ . Para escoger los valores óptimos de C y γ se hizo una evaluación del modelo SVM con la mitad de los datos variando estos parámetros en el rango $0 < C, \gamma \leq 3$ y evitando el error de generalización mediante validación cruzada, es decir, por cada variación de C y γ se implementa la validación y se extraen las medidas de desempeño.

En la Figura 4.8 se muestra una comparación entre la exactitud de una SVM desbalanceada y una SVM balanceada en la tarea de clasificación de las componentes independientes de señales EEG y de artefactos oculares. La SVM desbalanceada presenta una exactitud por encima del 93 % para valores pequeños de $0,8 \leq C \leq 1,8$ y $\gamma \leq 0,1$, es decir, cuando la penalización de las muestras mal clasificadas es pequeña y el ancho de la función *kernel* es grande. A medida que γ aumenta, la exactitud del estimador tiende a decrecer.

Por otra parte, la SVM balanceada tiene valores de exactitud alrededor del 89 % para valores de $0,5 \leq C \leq 3$ y $\gamma \leq 0,2$. Finalmente, cabe recordar que para el caso de la SVM desbalanceada los valores de exactitud son mayores debido a que el origen desbalanceado de los datos favorece la clase con mayor número de muestras.

Para evaluar el desempeño del algoritmo respecto a la clase positiva, es decir, a la perteneciente a las componentes independientes de los artefactos oculares, y cuyo número de muestras es menor al de la clase negativa, en la Figura 4.9 se muestra la sensibilidad de la SVM desbalanceada y la SVM balanceada dependiendo de los parámetros C y γ .

Para la SVM desbalanceada, la máxima sensibilidad obtenida alrededor de 50 % para valores $\gamma \leq 0,5$ y $1 \leq C \leq 3$, a medida que su valor de γ aumenta la sensibilidad se reduce drásticamente por debajo del 10 %.

En la SVM balanceada, la sensibilidad del predictor alcanza valores del 90 % en valores bajos de C y γ , a medida que cualquiera de estos dos parámetros aumenta su valor, la sensibilidad del estimador disminuye.

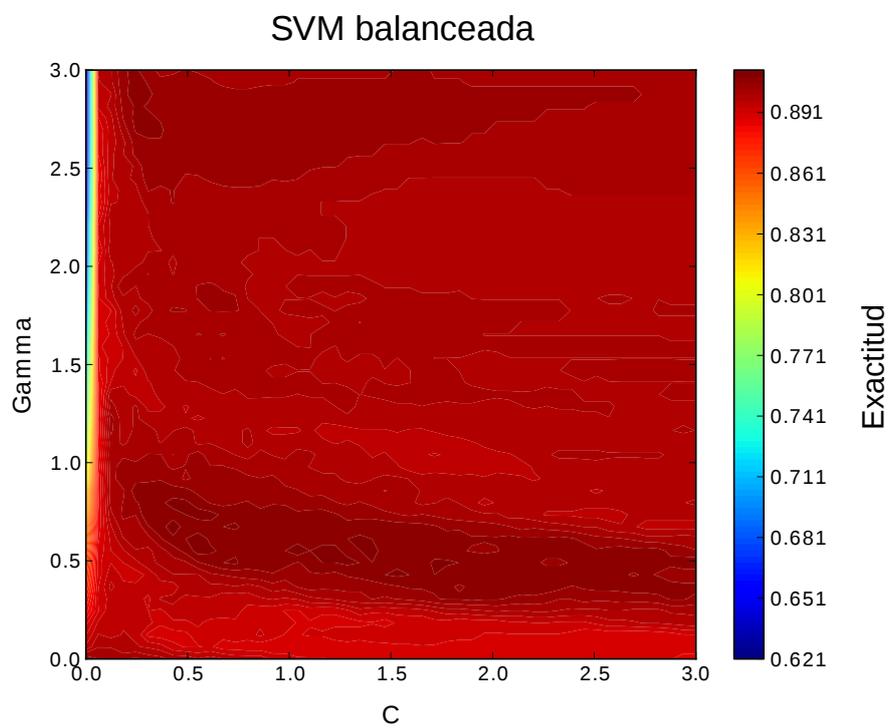
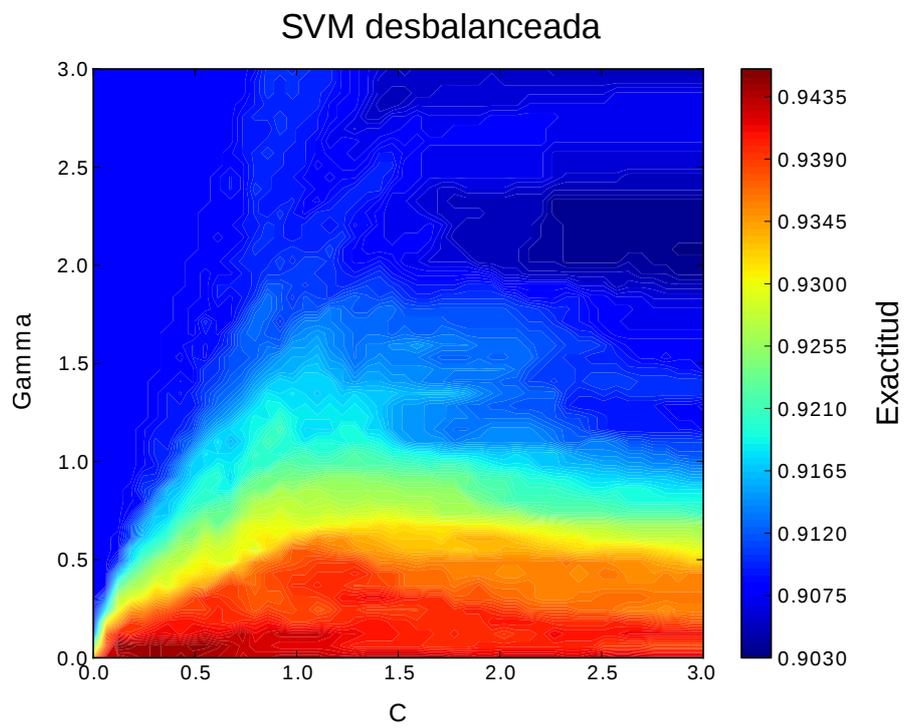


Figura 4.8: Evaluación de la exactitud de una SVM desbalanceada y balanceada para diferentes valores de C y γ .

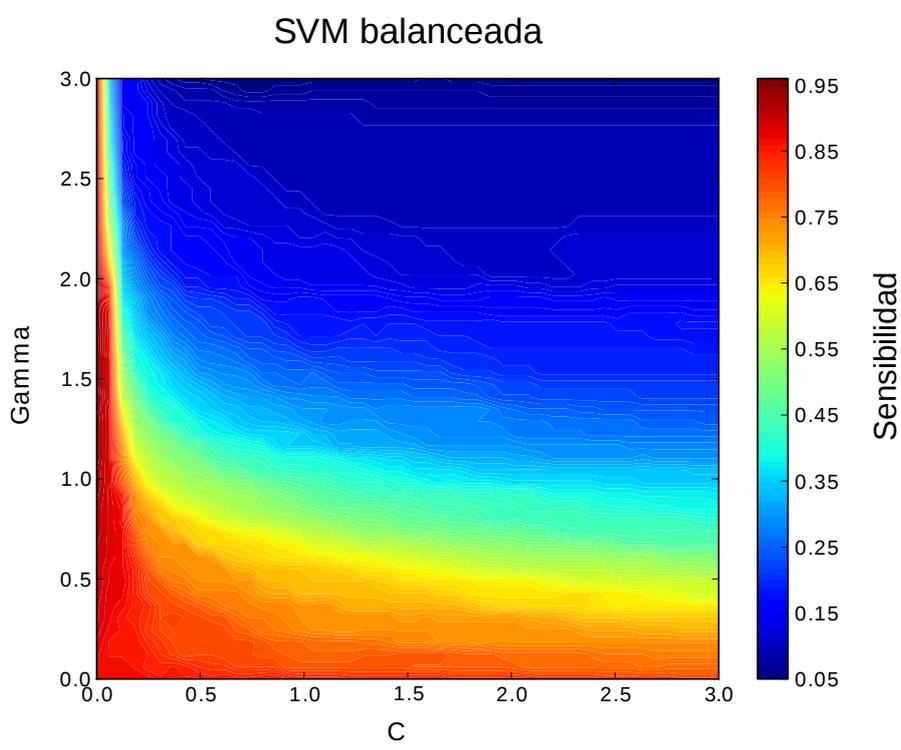
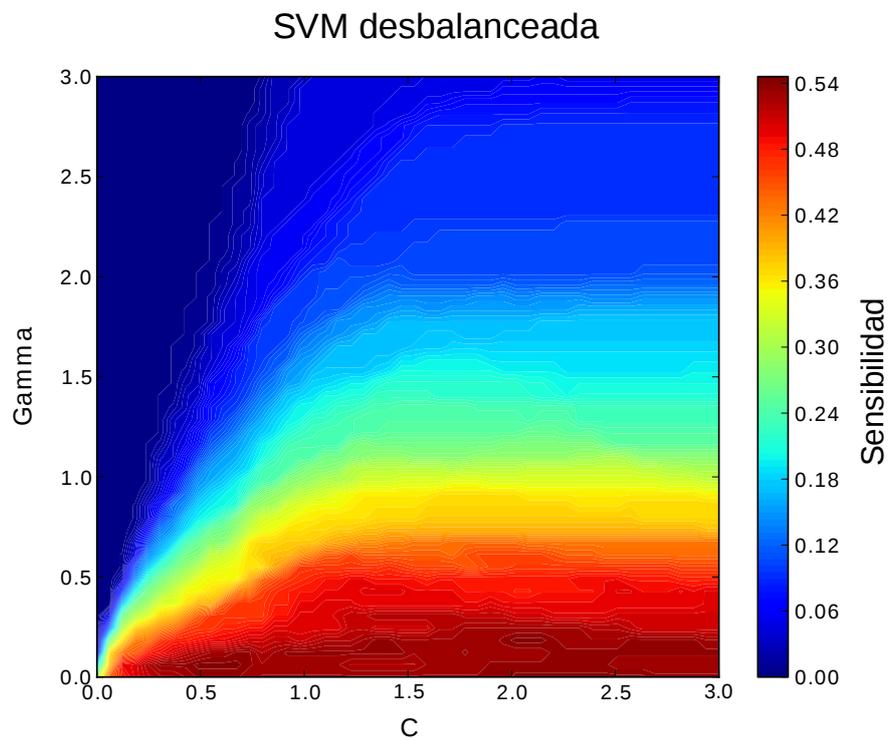


Figura 4.9: Evaluación de la sensibilidad de una SVM desbalanceada y balanceada para diferentes valores de C y γ .

Los parámetros de la SVM que mostraron el mejor desempeño del algoritmo teniendo en cuenta el desbalanceo de los datos fueron $C = 1$ y $\gamma = 0,0009$. En la Tabla 4.2 se muestra la matriz de confusión promedio para la SVM desbalanceada tanto para la etapa de entrenamiento como para la de validación. En este caso la clase g_1 es la perteneciente a las componentes independientes de los artefactos oculares y la clase g_2 son las componentes independientes de las señales EEG.

Clase verdadera	Clase estimada			
	Entrenamiento		Validación	
	g_1	g_2	g_1	g_2
g_1	52.7	47.2	50.7	49.3
g_2	1.0	99.0	1.2	98.7

Tabla 4.2: Matriz de confusión promedio para la SVM desbalanceada, $C = 1.0$ y $\gamma = 0.0009$.

Cuando no se tiene en cuenta el origen desbalanceado de los datos, el estimador tiene preferencia por la clase g_2 causando que sólo el 1,0 % de sus muestras sean mal clasificadas en el entrenamiento y 1,2 % en la validación.

Por otra parte, en la clase g_1 el porcentaje de muestras mal clasificadas es mucho mayor, teniendo un 47,2 % en entrenamiento y 49,2 % en la validación.

En el caso de la SVM balanceada, Tabla 4.3, la frontera de decisión es normalizada dependiendo del número de muestras de la clase g_1 y g_2 . El desempeño del modelo mejora drásticamente en comparación con el caso desbalanceado, el porcentaje de muestras mal clasificadas para la clase g_1 se reduce a 16,7 % en el entrenamiento y 19,7 % en la validación. En la clase g_2 este porcentaje aumenta a 9,61 % en el entrenamiento y 10,3 % en la validación, pero no influye en gran manera en el desempeño general del modelo.

Clase verdadera	Clase estimada			
	Entrenamiento		Validación	
	g_1	g_2	g_1	g_2
g_1	83.3	16.7	80.3	19.7
g_2	9.6	90.4	10.3	89.7

Tabla 4.3: Matriz de confusión promedio para la SVM balanceada, $C = 1.0$ y $\gamma = 0.0009$.

Otro método usado para evaluar el desempeño del predictor fue el marcador-F1. Este considera la precisión y el *recall* del modelo en la etapa de validación. En la Tabla 4.4 se evalúa la SVM desbalanceada y en Tabla 4.5 la SVM balanceada.

SVM desbalanceada			
Clase	Precisión	Recall	Marcador F1
g_1	83	50	59
g_2	95	99	97

Tabla 4.4: Evaluación del modelo, caso desbalanceado.

En este caso, el *recall* para la clase g_1 es de 50 % debido al gran número de falsos negativos y la precisión tiene un valor aceptable de 83 %. Sin embargo, cuando la SVM es normalizada, el *recall* de la clase g_1 se incrementa al 80 % y la precisión cae al 52 %. El desempeño total del modelo evaluado con el marcador-F1 muestra un incremento del 59 % hasta 60 % para la clase positiva g_1 y para la clase negativa g_2 varía levemente de 97 % a 93 %.

SVM balanceada			
Clase	Precisión	Recall	Marcador F1
g_1	52	80	60
g_2	98	90	93

Tabla 4.5: Evaluación del modelo, caso balanceado.

Finalmente, se hizo un análisis del predictor mediante curvas ROC, este toma en cuenta la especificidad y la sensibilidad. En la Figura 4.10 se muestra la curva ROC promedio de la SVM desbalanceada la cual presenta una AUC del 89 %, debido a su baja sensibilidad causada por el gran número de falsos negativos.

Por otra parte, en la Figura 4.11 podemos observar como la AUC aumenta hasta el 94 % para el caso balanceado, ya que en este caso el número de falsos negativos disminuye, sin embargo, la especificidad disminuye sin acarrear grandes consecuencias en el desempeño del predictor.

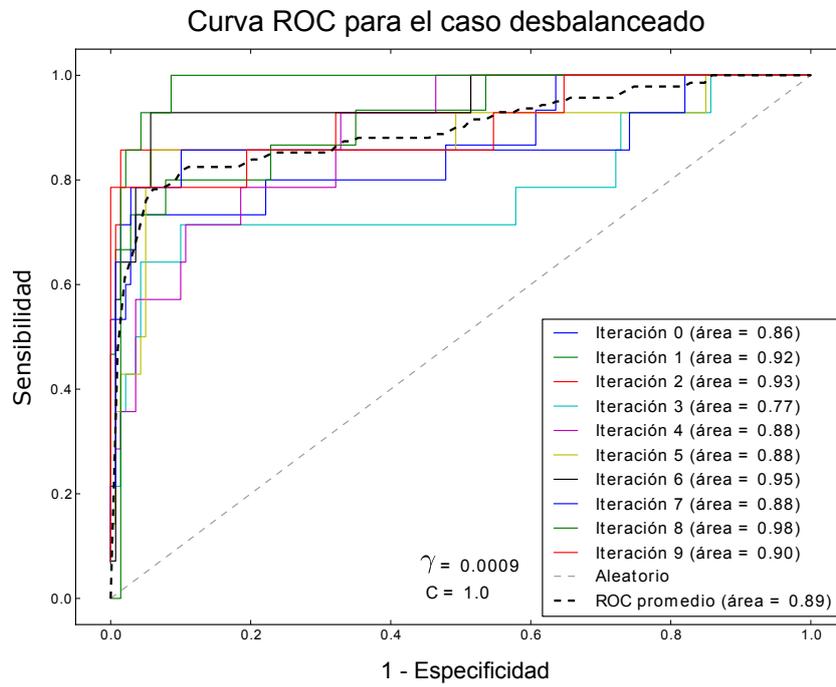


Figura 4.10: Curva ROC para la SVM desbalanceada.

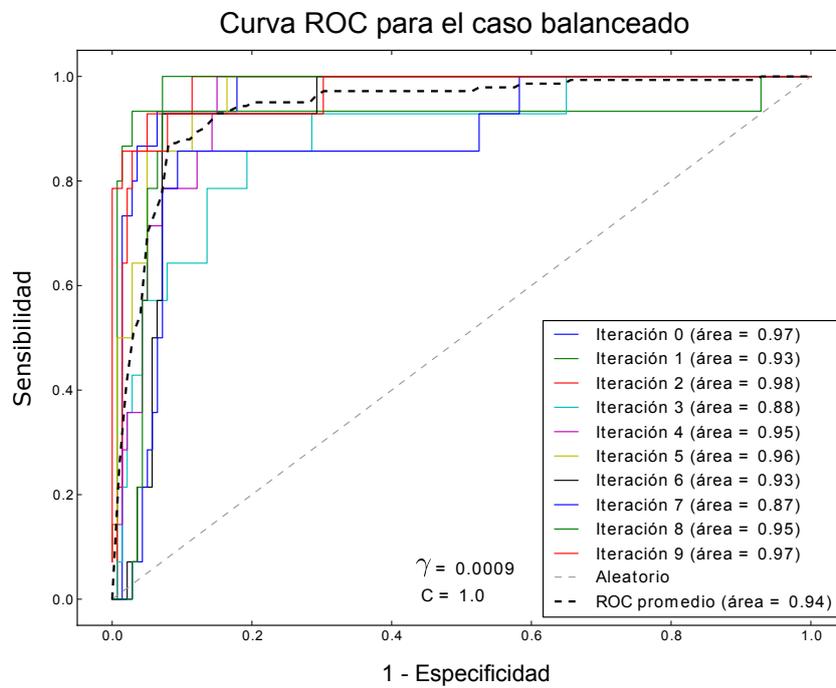


Figura 4.11: Curva ROC para la SVM balanceada

4.5.1. Reconstrucción de las señales EEG

La reconstrucción de las señales EEG sin artefactos oculares es hecha a través de la combinación lineal de la matriz de mezcla A y las componentes independientes modificadas \tilde{S} , de la siguiente manera:

$$\tilde{X} = A \cdot \tilde{S} \quad (4.45)$$

donde $\tilde{S} = [\tilde{s}_1, \tilde{s}_2, \dots, \tilde{s}_m]^T$, y \tilde{s}_i es definida como:

$$\tilde{s}_i = s_i \frac{-g_i + 1}{2} \quad (4.46)$$

donde g_i es la salida del estimador SVM y está dada por:

$$g_i = \begin{cases} 1 & \text{si } g_i = g_1 \\ -1 & \text{si } g_i = g_2 \end{cases} \quad (4.47)$$

4.5.2. Validación del sistema de eliminación de artefactos oculares mediante una SVM balanceada

A continuación se expone un ejemplo de prueba del sistema automático de eliminación de artefactos oculares. La Figura 4.12 muestra los 22 canales EEG pertenecientes al sujeto A01E en el intervalo de 280 – 288 s. Las señales fueron registradas desde los siguientes canales: Fz, FC3, FC1, FCz, FC2, FC4, C5, C3, C1, Cz, C2, C4, C6, CP3, CP1, CPz, CP2, CP4, P1, Pz, P2, y POz. Además, se registra un canal electrooculográfico para referenciar el momento en que ocurren los artefactos oculares.

La afectación de los artefactos oculares en los registros EEG muestra mayor influencia en la zona frontal debido a su cercanía a la fuente de ruido, en este caso los globos oculares.

En la Figura 4.13 se puede observar la descomposición de las señales EEG en componentes estadísticamente independientes por medio de ICA. Las componentes IC4 e IC9 pertenecen a los artefactos oculares y las componentes restantes a las señales EEG.

El resultado del proceso de eliminación de los artefactos oculares demuestra la efectividad del algoritmo de clasificación. En la Figura 4.14 se ve como las componentes IC4 e IC9 son suprimidas y las ICs EEG permanecen.

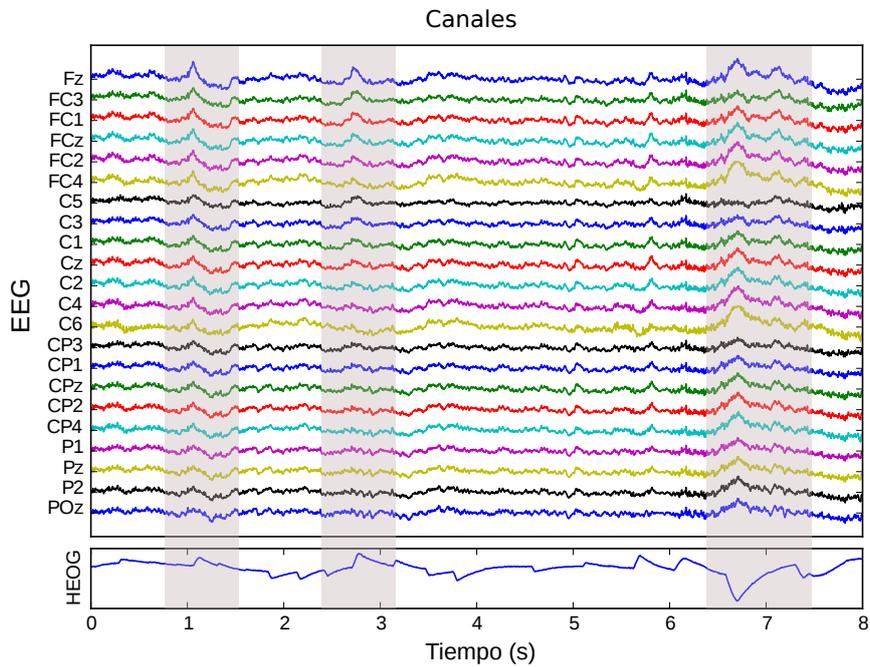


Figura 4.12: Influencia de los artefactos aculares en el registro EEG tomado del sujeto A01E en el rango temporal de 280-288 s. En la parte superior se muestran los 22 canales EEG registrados y en la parte inferior el EOG el cual se toma como referencia para observar el instante de tiempo en que ocurrieron los artefactos. Las frajas de color gris indican los artefactos de mayor amplitud.

La última etapa es la reconstrucción de las señales EEG, en este caso la restauración se lleva a cabo mediante la matriz de mezcla y las componentes independientes modificadas, es decir, después de pasar por el proceso de clasificación y posterior eliminación de los artefactos oculares. En la Figura 4.15 se observa el resultado de esta etapa.

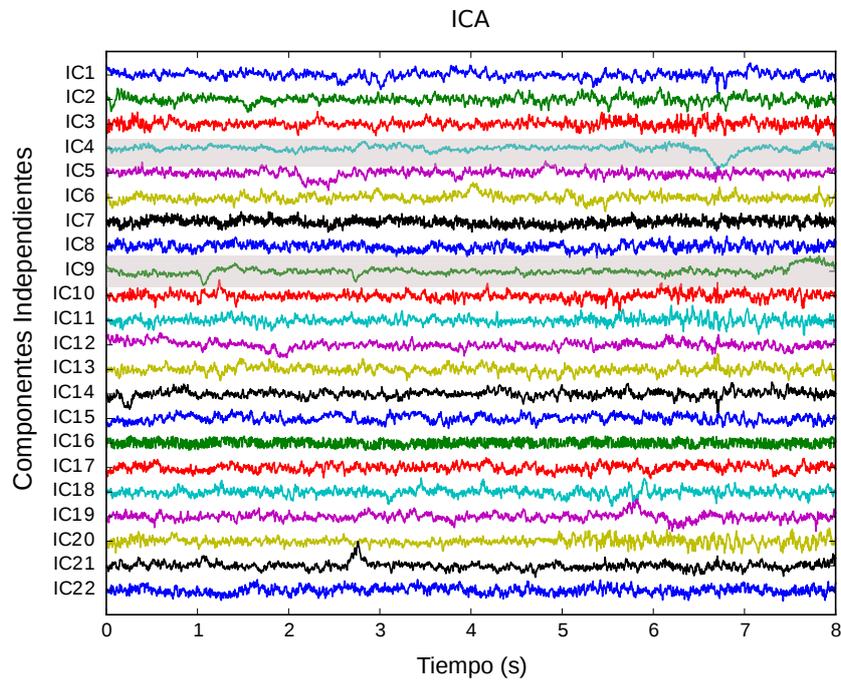


Figura 4.13: Descomposición de los canales EEG, registrados en el sujeto A01E, en componentes independientes mediante ICA.

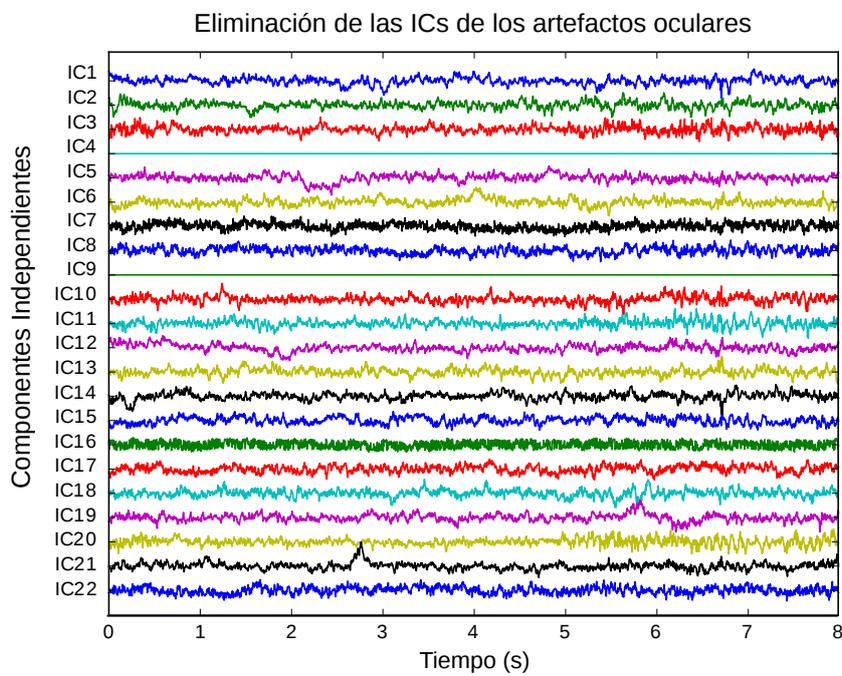


Figura 4.14: Eliminación de las componentes independientes de los artefactos oculares.

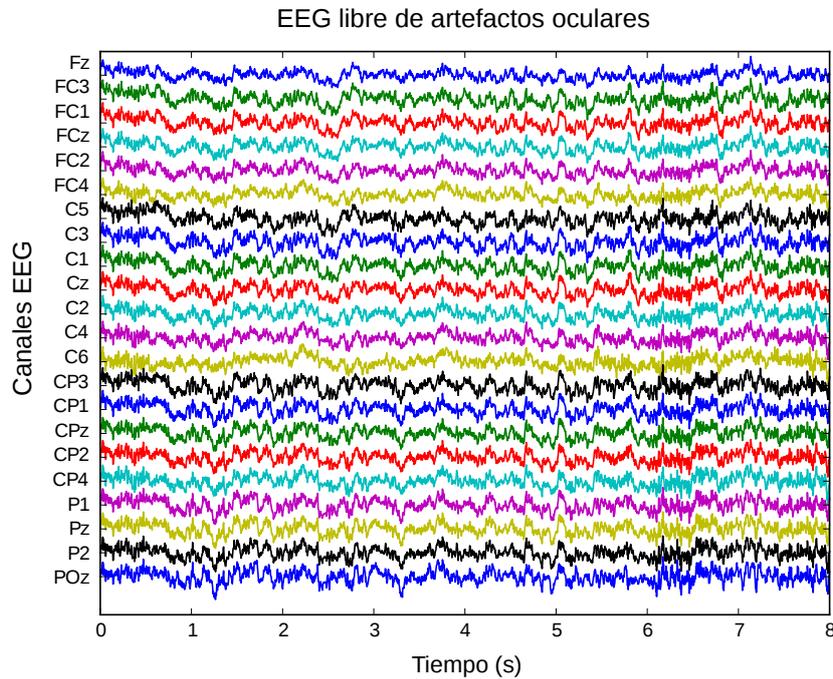


Figura 4.15: Canales EEG libres de artefactos oculares.

El método supervisado propuesto en este capítulo demostró un buen desempeño en la tarea de eliminación de artefactos oculares. Para evaluar esto, se realizó una comparación de la exactitud con métodos similares. En este caso se tomaron en cuenta los trabajos presentados en [20], [27] y [54].

En [20] no se proporciona ninguna medida de exactitud a pesar que se muestra la funcionalidad del método. En los otros dos métodos, los algoritmos también tienen la capacidad de eliminar otro tipo de artefactos como electromiográficos (EMG) y electrocardiográficos (ECG), por esta razón sólo se tomará en cuenta los resultados de la eliminación de artefactos EOG. En [27], no es tenida en cuenta la naturaleza desbalanceada de los datos, pero el método tiene una exactitud de 99 % en la clasificación de artefactos oculares. Por otra parte, en [54] la exactitud de la eliminación de los artefactos EOG es de 92 % y la preservación de la información EEG es de 88.4 %.

En nuestro caso, la exactitud de la eliminación de artefactos es del 80.3 % y la preservación de la información EEG es de 89.7 %. La comparación muestra que el método propuesto aquí tiene un menor desempeño, por esta razón, para mejorar esto se propone un aumento de los datos de entrenamiento o la búsqueda de otras características que permitan discriminar entre artefactos y EEG.

En este capítulo se mostraron los conceptos teóricos, la metodología y los resultados obtenidos de la implementación de un sistema automático de eliminación de artefactos oculares en registros EEG mediante una SVM balanceada. Este sistema probó tener bastante potencial para desarrollar la tarea de filtrado, sin embargo, depende de un conocimiento a-priori de la clase de donde provienen los datos para el entrenamiento del predictor.

Un sistema no supervisado puede ser usado como modelo de predicción sin la necesidad de poseer un conocimiento a-priori de la procedencia de los datos. En el siguiente capítulo se plantea un método de eliminación de artefactos basado en el agrupamiento de datos mediante SVM.

La ciencia no nos ha enseñado aún si la locura es o no lo más sublime de la inteligencia

Edgar Allan Poe

5

Aprendizaje no-supervisado: Agrupamiento SVM

El problema de agrupamiento ha sido usado en muchos contextos y en muchas disciplinas; esto refleja su amplio atractivo y utilidad como uno de los pasos en el análisis exploratorio de datos. El agrupamiento de datos surge de la necesidad de relacionar patrones por medio de medidas de similaridad entre características comunes sin tener conocimiento a priori de las clases a las que pertenecen los datos. Es así como en un mundo de datos abundantes, desbalanceados y sin etiquetas, el agrupamiento plantea una buena herramienta en el reconocimiento de patrones.

En este capítulo se presentan los conceptos, métodos y resultados del desarro-

llo de un modelo de eliminación de artefactos oculares en señales EEG a través del agrupamiento con máquinas de vectores soporte.

5.1. Agrupamiento

El análisis de agrupamiento, también llamado segmentación de datos, tiene una variedad de objetivos. Todos se refieren a la agrupación o la segmentación de una colección de objetos en subconjuntos o *clusters* de tal manera que los que están dentro de cada grupo tienen una relación más fuerte que con los objetos asignados a otros grupos. Un objeto puede ser descrito por un conjunto de mediciones, o por su relación con otros objetos. Además, el objetivo es algunas veces organizar los grupos en una jerarquía natural. Esto implica la integración sucesiva de grupos de manera que en cada nivel de la jerarquía, los grupos dentro del mismo grupo son más similares entre sí que los de diferentes grupos.

El agrupamiento también se utiliza para formar la estadística descriptiva que permite determinar si los datos consisten o no en un conjunto de distintos subgrupos, donde cada grupo representa objetos con propiedades sustancialmente diferentes. Este último objetivo requiere una evaluación del grado de diferencia entre los objetos asignados a los respectivos grupos [28].

5.1.1. Componentes de una tarea de agrupamiento

Típicamente las tareas de agrupamiento de patrones involucran los siguientes pasos, [32]:

1. Representación de patrones, incluyendo opcionalmente extracción y/o selección de características.
2. Definición de una medida de proximidad apropiada para el dominio de los datos.
3. Agrupamiento
4. Abstracción de los datos, si es necesario.
5. Evaluación de la salida, si es necesario.

La estructura de un algoritmo de agrupamiento se muestra en la Figura 5.1.

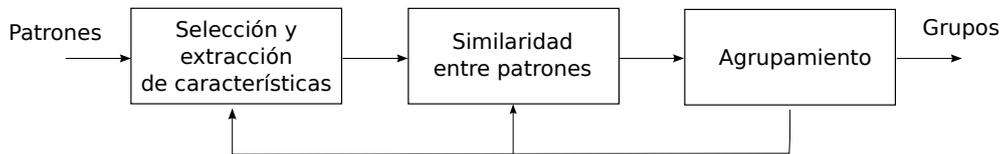


Figura 5.1: Etapas de un algoritmo de agrupamiento, [32]

La *representación de patrones* se refiere al número, tipo y escala de las características disponibles para el algoritmo de agrupamiento.

La *selección de características* es el proceso mediante el cual se identifica el subconjunto más efectivo para usar en el agrupamiento. La *extracción de características* es el uso de una o más transformaciones de las características de entrada para producir nuevas características. Ambas técnicas pueden ser usadas para obtener un conjunto de características apropiado, [32].

La *proximidad* es usualmente medida por una función de distancia definida entre pares de patrones. Una variedad de medidas de distancia son usadas, por ejemplo, la distancia Euclideana.

El *agrupamiento* de los datos puede ser hecho de diferentes maneras. Los algoritmos de agrupamiento puede ser clasificados en varios tipos [12], [38], [25], [3], [58] como se muestra en la Figura 5.2:

1. Los métodos *jerárquicos* producen una serie de particiones anidadas basadas en un criterio para mezclar y dividir grupos respecto a una medida de similaridad.
2. Los métodos *particionales* particionan el espacio de características en grupos e identifican la partición que optimiza un índice de costo.
3. Los métodos *Grid-based* particionan el espacio de características y buscan los objetos que pertenecen a cada celda resultante de la partición.
4. Los métodos *basados en densidad* agrupan los objetos de acuerdo a una función objetivo de densidad específica. La densidad es definida como el número de objetos en el vecindario particular de un objeto.
5. Los métodos *basados en modelos* buscan el valor de los parámetros, de un modelo matemático, que mejor representan el comportamiento de los datos.

6. Los métodos *basados en restricciones* hacen la tarea de agrupamiento teniendo en cuenta restricciones orientadas para obtener los resultados deseados.
7. Los métodos de *agrupamiento categórico* se especializan en el agrupamiento de datos cualitativos, es decir, donde las medidas de distancia no se pueden aplicar como criterio de similitud.

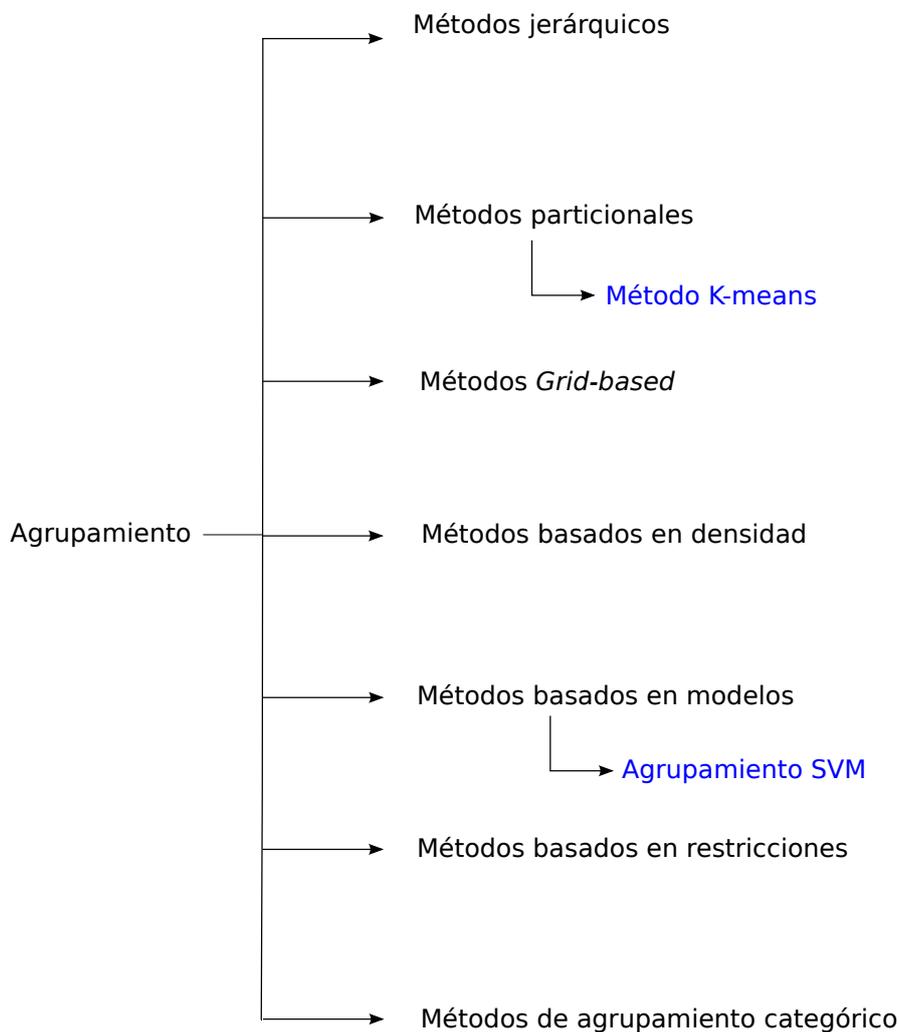


Figura 5.2: Clasificación de los algoritmos de agrupamiento.

La *validación de grupos* es una de las etapas más estudiadas debido a su naturaleza subjetiva, [3], [38]. Esta consiste en la evaluación de la salida de un proceso

de agrupamiento. A menudo este análisis usa un criterio de optimalidad; sin embargo, esos criterios usualmente son muy diversos. Esos criterios usualmente se dividen en dos categorías:

1. La validación *externa* compara la estructura recuperada con una estructura conocida.
2. La validación *interna* determina la compacidad de los grupos mediante una medida de similaridad. Esta evalúa la homogeneidad *intra-cluster*, la separabilidad *inter-cluster* o una combinación de las dos. Además, no usa ninguna información externa aparte de la proporcionada por los datos, [38].

5.2. Agrupamiento SVM

El agrupamiento SVM es una extensión de las cualidades del modelo SVM al aprendizaje no-supervisado, teniendo en cuenta que es considerado uno de los mejores discriminadores [71] debido a que es entrenado de manera rápida y fácil y ofrece una alta tolerancia al ruido. Actualmente, hay dos enfoques para implementar agrupamiento SVM. El primero llamado *agrupamiento interno* involucra solucionar un simple problema de optimización para todos los datos. El segundo llamado *agrupamiento externo* ordena varios clasificadores binarios en forma de árbol de decisión de tal manera que crean una simple función de decisión donde cada hoja pertenece a una clase, [71]. A continuación se hará una breve descripción de los dos métodos.

5.2.1. Agrupamiento interno

El agrupamiento interno usa aspectos intrínsecos de la formulación SVM para encontrar la esfera envolvente más pequeña. Dado un conjunto de datos \mathbf{x} de N muestras en \mathbb{R}^d , donde d es la dimensión del espacio de entrada, se usa una transformación no-lineal Φ que lleva a \mathbf{x} a un espacio de alta dimensión, *espacio Kernel*, en el cual se busca la esfera envolvente más pequeña con radio R . Esto es expresado como:

$$\|\Phi(\mathbf{x}_i) - a\|^2 \leq R^2, \quad \forall i \quad (5.1)$$

donde a es el centro de la esfera, Figura 5.3. Restricciones suaves son incorporadas adicionando variables de holgura ξ_j :

$$\|\Phi(\mathbf{x}_i) - a\|^2 \leq R^2 + \xi_i, \quad \forall i \quad (5.2)$$

sujeto a:

$$\xi_i \geq 0 \quad (5.3)$$

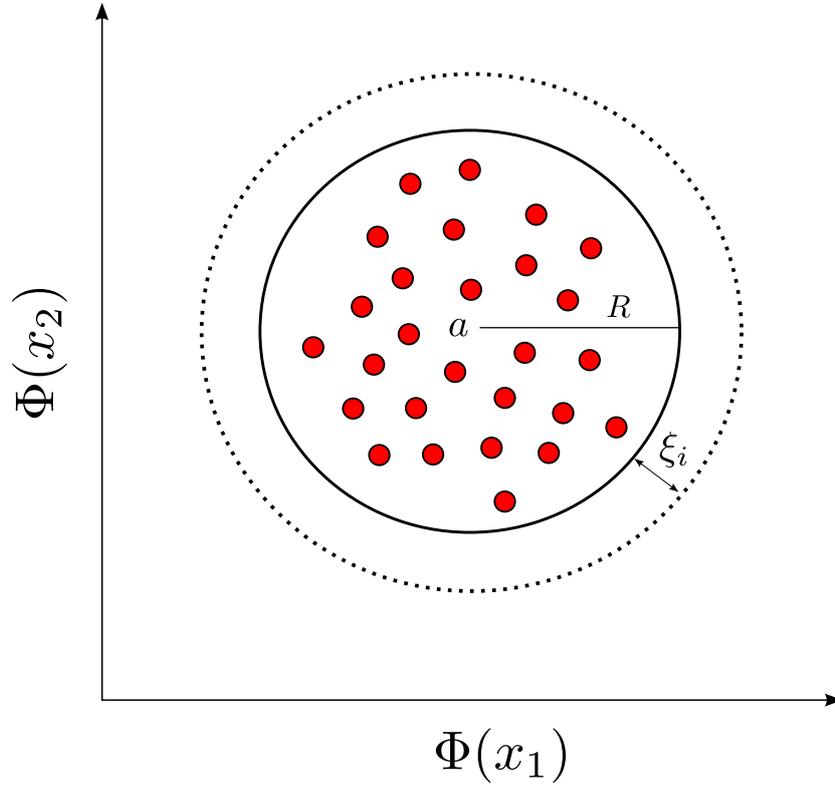


Figura 5.3: Formulación del agrupamiento interno. Esfera envolvente con centro a y radio R en el espacio *Kernel*.

Para solucionar el problema se presenta el Lagrangiano:

$$L(R, a, \xi) = R^2 - \sum_{i=1}^N (R^2 + \xi_i - \|\Phi(\mathbf{x}_i) - a\|^2) \beta_i - \sum_{i=1}^N \xi_i \mu_i + C \sum_{i=1}^N \xi_i \quad (5.4)$$

donde $\beta_i \geq 0$ y $\mu_i \geq 0$ son los multiplicadores de Lagrange, C es el costo para los *outliers* y $C \sum_{i=1}^N \xi_i$ es el término de penalización. Haciendo cero las derivadas de L respecto a R , a y ξ , respectivamente, obtenemos:

$$\sum_{i=1}^N \beta_i = 1 \quad (5.5)$$

$$a = \sum_{i=1}^N \beta_i \Phi(\mathbf{x}_i) \quad (5.6)$$

$$\beta_i = C - \mu_i \quad (5.7)$$

Las condiciones KKT complementarias son:

$$\xi_i \mu_i = 0 \quad (5.8)$$

$$R^2 + \xi_i - \|\Phi(\mathbf{x}_i) - a\|^2 \beta_i = 0 \quad (5.9)$$

La ecuación (6.9) dice que la imagen de un punto \mathbf{x}_i con $\xi_i > 0$ y $\beta_i > 0$ yace fuera de la esfera del espacio de características. La ecuación (6.8) establece que tal punto tiene $\mu_i = 0$, indicando que $\beta_i = C$. Entonces, este punto es llamado *vector soporte límite* o BSV. Un punto \mathbf{x}_i con $\xi_i = 0$ es mapeado dentro o en la superficie de la esfera. Si $0 < \beta_i < C$ entonces la ecuación (6.9) implica que su imagen $\Phi(\mathbf{x}_i)$ yace sobre la superficie de la esfera. Tal punto será referenciado como *vector soporte* o SV. Los SVs están en los límites del grupo y los BSVs están fuera de los límites, y todos los otros puntos yacen dentro de la esfera.

Usando esas relaciones se pueden eliminar las variables R , a y μ_i , convirtiendo el Lagrangiano en una forma dual de Wolfe la cual es una función de la variable β_i :

$$W = \sum_{i=1}^N \Phi(\mathbf{x}_i)^2 \beta_i - \sum_{i=1}^N \sum_{j=1}^N \beta_i \beta_j \Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j) \quad (5.10)$$

Ya que las variables μ_i no aparecen en el Lagrangiano, estas pueden ser reemplazadas con las restricciones:

$$0 \leq \beta_i \leq C, \quad \forall i \quad (5.11)$$

El producto punto $\Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j)$ puede ser representado como $\Phi(\mathbf{x}_i, \mathbf{x}_j)$, en este caso se usa un *Kernel* Gaussiano:

$$\Phi(\mathbf{x}_i, \mathbf{x}_j) = \exp(-q\|\mathbf{x}_i - \mathbf{x}_j\|^2) \quad (5.12)$$

donde q es el ancho del *Kernel*. El lagrangiano W es escrito ahora como:

$$W = \sum_{i=1}^N \Phi(\mathbf{x}_i, \mathbf{x}_i) \beta_i - \sum_{i=1}^N \sum_{j=1}^N \beta_i \beta_j \Phi(\mathbf{x}_i, \mathbf{x}_j) \quad (5.13)$$

En cada punto \mathbf{x} se define la distancia de su imagen en el espacio de características desde el centro de la esfera:

$$R^2(\mathbf{x}) = \|\Phi(\mathbf{x}) - a\|^2 \quad (5.14)$$

Reemplazando la ecuación (6.6) en (6.14) se tiene:

$$R^2(\mathbf{x}) = \Phi(\mathbf{x}, \mathbf{x}) - 2 \sum_{i=1}^N \beta_i \Phi(\mathbf{x}_i, \mathbf{x}) + \sum_{i=1}^N \sum_{j=1}^N \beta_i \beta_j \Phi(\mathbf{x}_i, \mathbf{x}_j) \quad (5.15)$$

El radio de esta esfera es:

$$R = \{R(\mathbf{x}_i) \mid \mathbf{x}_i \text{ es un vector soporte}\} \quad (5.16)$$

Los contornos que encierran los puntos en el espacio de datos son definidos por el conjunto:

$$\{\mathbf{x} \mid R(\mathbf{x}) = R\} \quad (5.17)$$

El algoritmo descrito anteriormente no diferencia entre los puntos que pertenecen a diferentes grupos. Para hacer esto se usa un enfoque geométrico involucrando $R(\mathbf{x})$, basado en la siguiente observación: dado un par de datos que pertenecen a diferentes grupos, debe existir una conexión entre ellos en la esfera del espacio de características. Esta conexión tiene un segmento de puntos y tal que $R(y) > R$. Esto conduce a la definición de la matriz de adyacencia A_{ij} entre pares de puntos \mathbf{x}_i y \mathbf{x}_j cuyas imágenes yacen en o sobre la esfera del espacio de características:

$$A_{ij} = \begin{cases} 1 & \text{si, } \forall y \in \overline{\mathbf{x}_i \mathbf{x}_j}, R(y) > R \\ 0 & \text{en otro caso} \end{cases} \quad (5.18)$$

Los grupos son definidos como los componentes conectados según A . El segmento de línea se implementa mediante un muestreo de un número dado de puntos.

Las BSVs no son clasificadas por el procedimiento ya que sus imágenes están fuera de la esfera envolvente. Se puede decidir si las BSVs se dejan sin clasificación, o se asignan al cluster más cercano, [11].

En la Figura 5.4 se muestra el comportamiento del algoritmo de agrupamiento SVM interno evaluado en dos casos. El primero con los parámetros $q = 1$ y $C = 1$ muestra que la esfera envolvente se adapta de manera adecuada a los datos formando dos grupos principales donde se encuentran la mayoría de las muestras. En el segundo caso con los parámetros $q = 0,1$ y $C = 1$ la esfera envolvente muestra poca sensibilidad a la estructura natural de los datos debido a que el parámetro q es inverso al ancho del *Kernel* de base radial generando así un solo grupo.

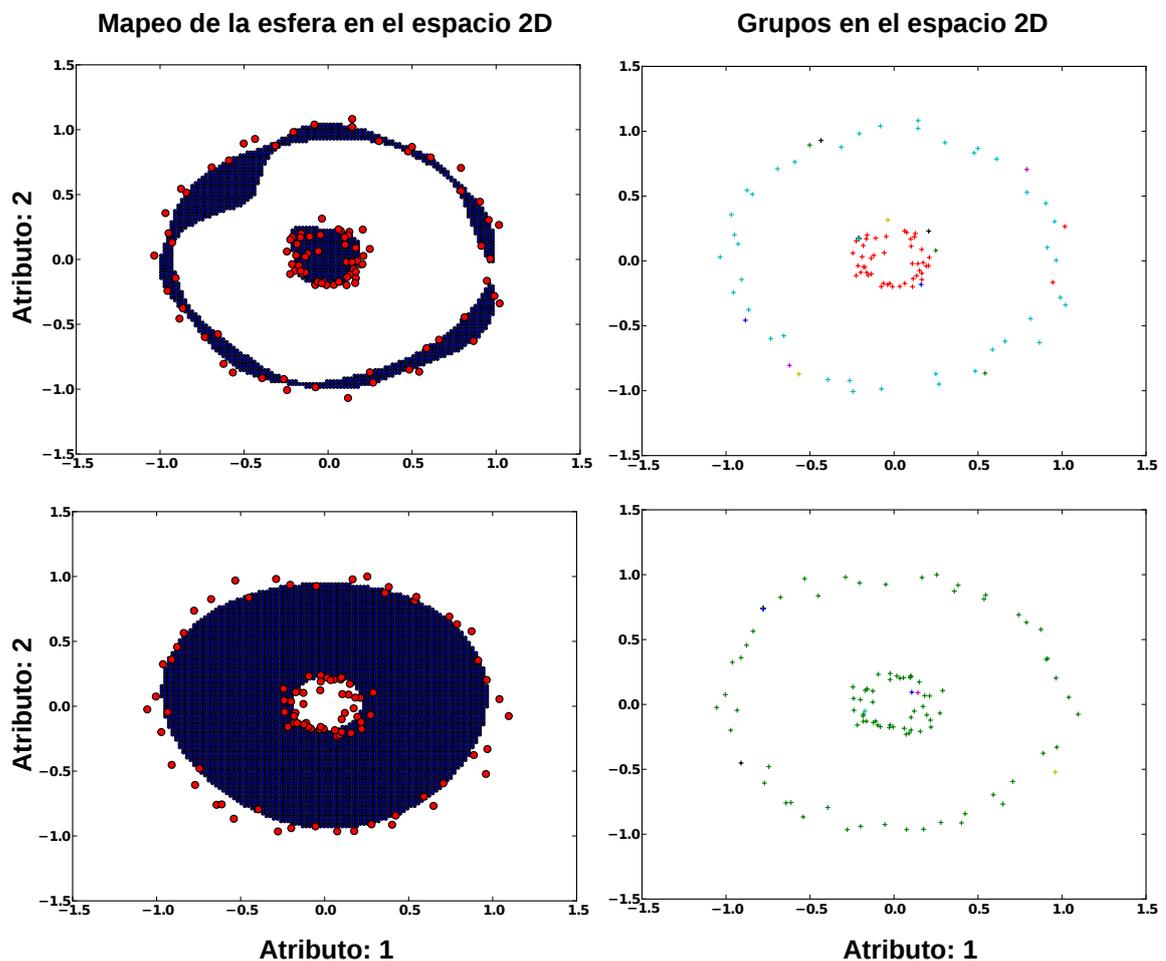


Figura 5.4: Algoritmo de agrupamiento SVM para un conjunto de 100 datos que está formado por dos clases distribuidas concéntricamente. En la parte superior se muestra el resultado para los parámetros $q = 1$ y $C = 1$, en la parte inferior para $q = 0,1$ y $C = 1$. Por otra parte, en la parte izquierda se muestra el mapeo de la esfera envolvente en el espacio de entrada y en la derecha los grupos formados. Estos resultados se obtuvieron mediante una librería desarrollada en *python* por el autor.

Un segundo ejemplo se muestra en la Figura 5.5, en este caso los datos forman dos semicírculos, cada uno de los cuales pertenece a una clase. Los parámetros del modelo de agrupamiento son $q = 1$ y $C = 1$. En la gráfica se indican los datos que yacen dentro de la esfera, los que son BSVs y los SVs, es decir, las muestras que están fuera y sobre la superficie de decisión respectivamente. La esfera mapeada en el espacio de entrada muestra una buena adaptación a los datos, así como la clara conformación de dos grupos principales a los cuales pertenecen la mayoría de las muestras.

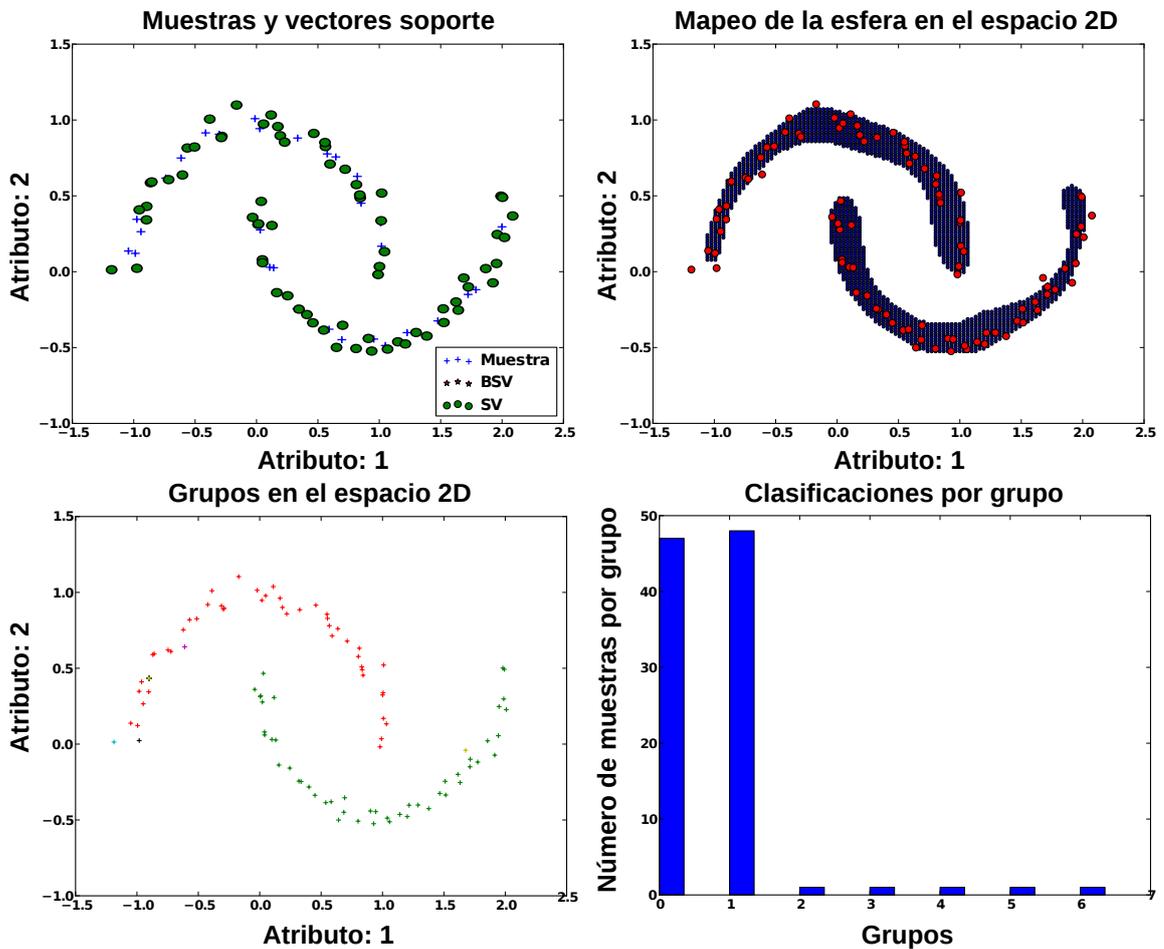


Figura 5.5: Algoritmo de agrupamiento SVM para un conjunto de 100 datos provenientes de dos clases distribuidas en forma de semicírculo para los parámetros $q = 1$ y $C = 1$. En la parte superior-izquierda se pueden ver las muestras, las BSVs y las SVs. En la parte superior-derecha se muestra el mapeo de la esfera envolvente en el espacio de entrada. En la parte inferior-izquierda se indican los grupos formados en el espacio de entrada. En la parte inferior-derecha se muestra en número de muestras por grupo, indicando que el 95 % de las muestras pertenecen a los grupos 1 y 2.

5.2.2. Agrupamiento externo

Aunque el enfoque interno para el agrupamiento SVM está débilmente sesgado con relación a la forma de los grupos en el espacio de características este carece de robustez. En el caso de problemas más reales donde los grupos están fuertemente superpuestos, el agrupamiento interno sólo puede delinear los núcleos de los grupos relativamente pequeños. Además, la implementación de la

formulación está firmemente ligada con la elección inicial del *kernel*, por esto, la naturaleza estática de la formulación y la implementación no se acomoda a algunas funciones *kernel*. Para remediar estas restricciones geométricas excesivas se crea un algoritmo de agrupamiento externo, [69].

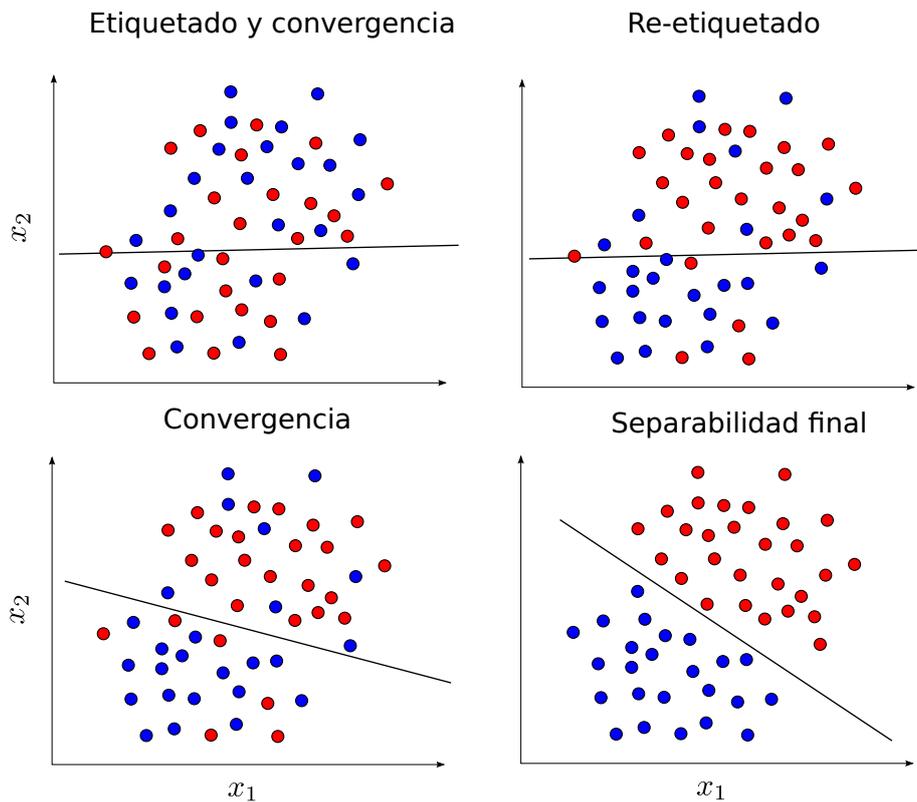


Figura 5.6: Algoritmo de agrupamiento SVM Re-etiquetado. En la parte superior-izquierda se muestra el etiquetado aleatorio de las muestras y el hiperplano de decisión producto de la convergencia de la SVM. En la parte superior-derecha, se realiza el procedimiento de re-etiquetado de las muestras mal clasificadas. En la parte inferior-izquierda, se ve el hiperplano resultante de la convergencia de la SVM sobre las muestras re-etiquetadas. En la parte inferior-derecha, se muestra el hiperplano de separación final después algunas iteraciones.

El agrupamiento externo trabaja como un árbol de decisión optimizado. El algoritmo primero inicializa ejecutando un clasificador SVM binario sobre el conjunto de datos cuyas muestras son etiquetadas de manera aleatoria, esto se repite hasta que ocurre una convergencia inicial. Una vez el paso de *inicialización* es completado, los parámetros de confianza para la clasificación son evaluados.

El algoritmo ahora mejora este resultado invirtiendo las etiquetas de los datos mal clasificados, los cuales tienen un valor de confianza por encima de un umbral establecido, y ejecutando de nuevo la SVM con el conjunto de datos re-etiquetado. Este proceso continúa hasta que ya no se pueda hacer algún progreso, es decir, que ya no haya incremento en el valor de la efectividad. Después de este proceso, un alto porcentaje de las etiquetas de los datos anteriormente desconocidas serán conocidas.

Este método ofrece una manera para agrupar datos sin tener conocimiento a priori de sus características de agrupamiento, o el número de grupos. Por otra parte, el algoritmo puede ser implementado sobre los grupos identificados para encontrar subgrupos, [69], [71].

El agrupamiento externo no está sesgado en relación a la forma de los grupos, además que su formulación no depende únicamente de una función *kernel*. El anterior algoritmo es llamado *agrupamiento SVM externo re-etiquetado* y es mostrado en el Algoritmo 4.

Algoritmo 4 Agrupamiento SVM externo re-etiquetado.

```

1:  $\hat{t} \leftarrow$  Asignación aleatoria desde  $\{-1, +1\}$ 
2: repeat
3:    $\theta \leftarrow doSVM(\mathbf{x}, \hat{t})$ 
4:    $\hat{t} \leftarrow doRelabel(\mathbf{x}, \theta)$ 
5: until  $\hat{t}$  permanezca constante

```

El procedimiento *doSVM* se realiza mediante un algoritmo de clasificación SVM estándar, este a su vez proporciona el hiperplano de decisión θ y etiquetas estimadas \hat{t} como entradas al procedimiento *doRelabel*, el cual es mostrado en el Algoritmo 5 y se encarga de reasignar las muestras mal clasificadas al otro grupo, es decir invierte la etiqueta de la muestra.

La reasignación se hace mediante la comparación de la distancia $D(\mathbf{x}_i, \theta)$ entre una muestra mal clasificada \mathbf{x}_i y el hiperplano de separación, y la distancia promedio de las muestras mal clasificadas pertenecientes a la misma etiqueta. En este caso la distancia utilizada es la *Euclídeana*.

El principio de funcionamiento del algoritmo se muestra en la Figura 5.6, en este caso el agrupamiento SVM Re-etiquetado es aplicado en un conjunto de muestras en dos dimensiones.

Algoritmo 5 Procedimiento *doRelabel*, [44], [67], [69]

Require:Vector de entrada: \mathbf{x} Grupo etiquetado: \hat{t} Modelo SVM: θ Factor de confianza: α

- 1: $\mathbf{x}'^+ \leftarrow K$ características mal clasificadas con $\hat{t} = +1$
 - 2: $\mathbf{x}'^- \leftarrow L$ características mal clasificadas con $\hat{t} = -1$
 - 3: **for all** i -ésima muestra de \mathbf{x}'^+ **do**
 - 4: **if** $1/K \sum_{j=1}^K D(\mathbf{x}'_j^+, \theta) < \alpha D(\mathbf{x}'_i^+, \theta)$ **then**
 - 5: $\hat{t}_i^+ \leftarrow -1$
 - 6: **end if**
 - 7: **end for**
 - 8: **for all** i -ésima muestra de \mathbf{x}'^- **do**
 - 9: **if** $1/L \sum_{j=1}^K D(\mathbf{x}'_j^-, \theta) < \alpha D(\mathbf{x}'_i^-, \theta)$ **then**
 - 10: $\hat{t}_i^- \leftarrow +1$
 - 11: **end if**
 - 12: **end for**
-

5.2.3. Evaluación del algoritmo de agrupamiento

A diferencia de los modelos de aprendizaje supervisado, en los métodos de agrupamiento no se cuenta con información a priori acerca de la clase a la cual pertenecen los datos, por esta razón no se puede hacer una evaluación supervisada comparando las etiquetas estimadas con las esperadas. En este caso, la validación del agrupamiento se hace por medio de la información intrínseca de las muestras, es decir, no depende de información externa.

Suma de Errores Cuadráticos (SSE)

La SSE mide la compacidad o varianza intra-grupos mientras que otras medidas calculan la separación entre grupos o varianza inter-grupos. La SSE es definida como, [38], [44].

$$SSE = \sum_{k=1}^K \sum_{\forall x_i \in C_k} \|x_i - \mu_k\|^2 \quad (5.19)$$

donde C_k es el conjunto de muestras en el grupo k , μ_k es la media del grupo k . Los componentes de μ_k son calculados como:

$$\mu_{k,j} = \frac{1}{N_k} \sum_{\forall x_i \in C_k} x_{i,j} \quad (5.20)$$

donde $N_k = |C_k|$ es el número de muestras pertenecientes al grupo k .

Los métodos de agrupamiento que minimizan el criterio SSE son llamados *particiones de varianza mínima*.

Para comprobar el desempeño del algoritmo de desarrollo una librería en *python* para método de agrupamiento SVM Externo. En este caso, se usó un factor de confianza $\alpha = 1$, una función *kernel* de base radial y como medida de convergencia el *error cuadrático medio* (MSE) entre la salida esperada y la salida estimada por el modelo SVM en cada iteración. El MSE es definido como:

$$MSE = \frac{1}{N} \sqrt{(\hat{t} - t)^2} \quad (5.21)$$

donde N es la longitud del vector t .

En la Figura 5.7 se muestran los grupos formados por el algoritmo de agrupamiento SVM Externo implementado sobre un conjunto de 1000 muestras claramente separables. En este caso, los parámetros de la SVM son $C = 1$ y $\gamma = 0,01$. El algoritmo converge en 12 iteraciones en las cuales logra obtener un $MSE < 1e - 6$, Figura 5.8. Por otro lado, el valor del SSE disminuye con cada iteración, es decir, la compacidad de los grupos aumenta con cada iteración, Figura 5.8.

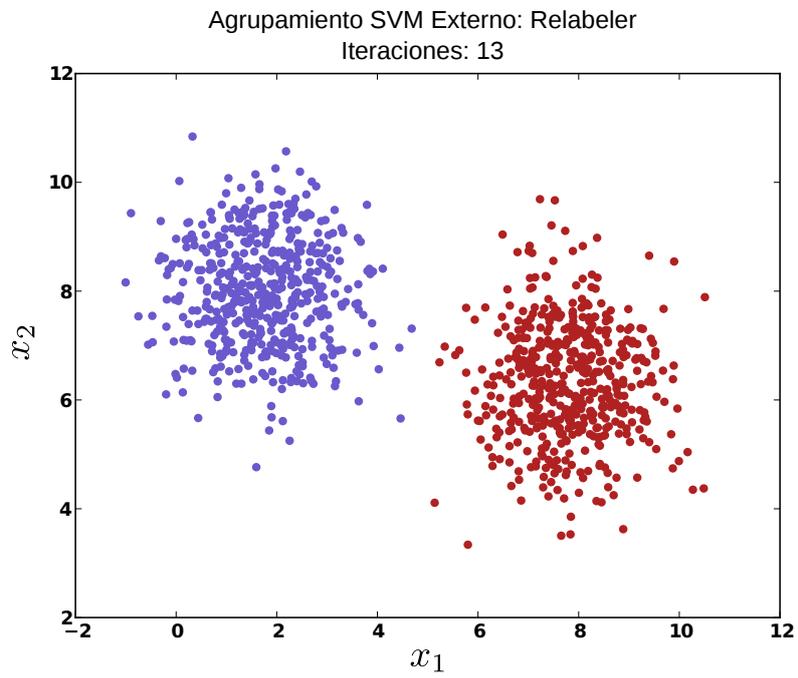


Figura 5.7: Grupos formados por el algoritmo de agrupamiento SVM *Relabeler* en un conjunto de 1000 muestras después de 13 iteraciones, con los parámetros $C = 1$ y $\gamma = 0,01$.

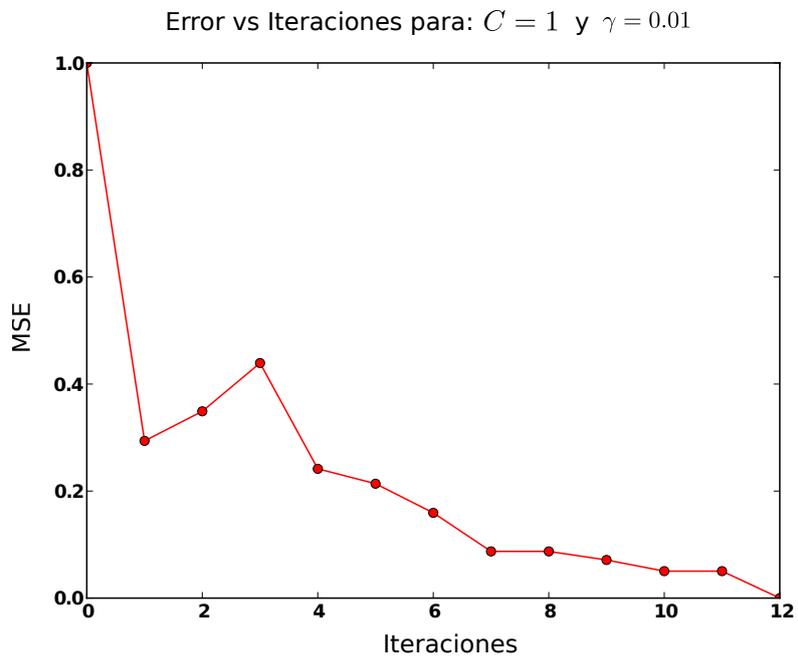


Figura 5.8: MSE entre la salida esperada y la salida estimada por el modelo de agrupamiento.

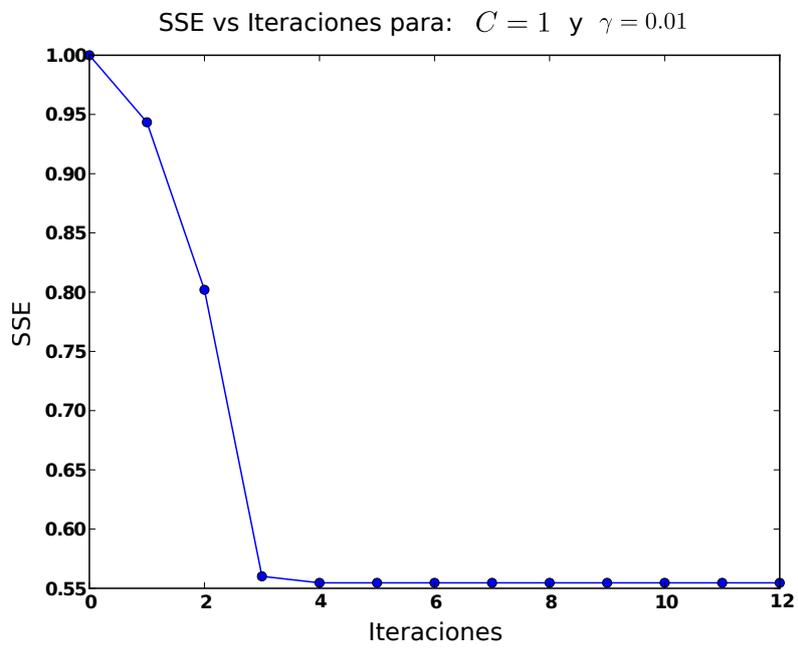


Figura 5.9: Reducción del SSE a medida que el algoritmo aumenta la compacidad de los grupos con cada iteración.

En la Figura 5.10 se muestra otro ejemplo, en este caso, los datos están superpuestos dificultando la búsqueda del hiperplano de separación. Los parámetros de la SVM son $C = 1$ y $\gamma = 0,01$. El algoritmo converge en 18 iteraciones, Figura 5.8 y el valor del SSE normalizado disminuye hasta 0,8, Figura 5.8.

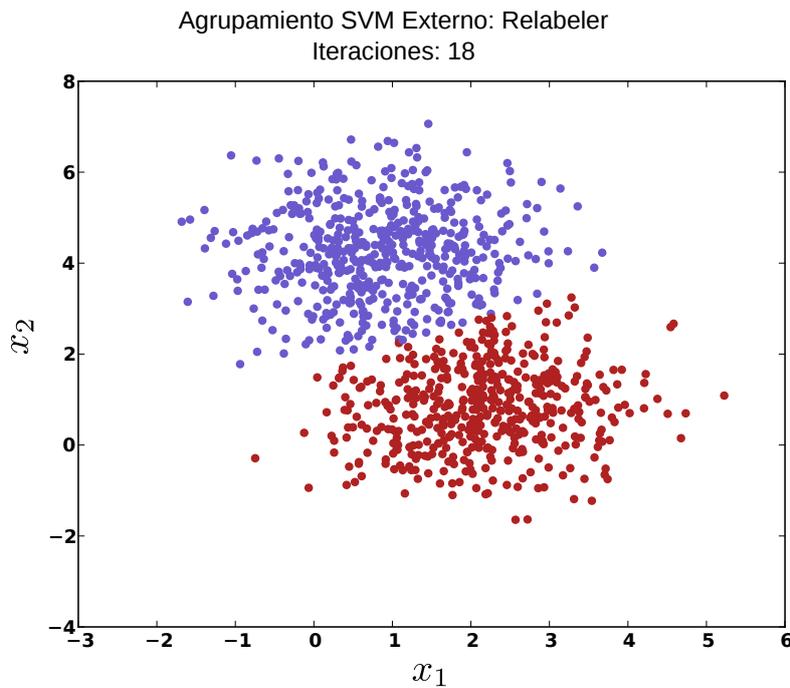


Figura 5.10: Grupos formados por el algoritmo de agrupamiento SVM *Relabeler* en un conjunto de 1000 muestras después de 18 iteraciones, con los parámetros $C = 1$ y $\gamma = 0,01$.

En la Figura 5.13 se puede ver el resultado de la tarea de agrupamiento mediante el algoritmo SVM Externo en un conjunto de 1000 muestras cuyos grupos están desbalanceados en una relación de 2 a 1. En este caso, se muestra la distancia de cada muestra perteneciente a un grupo respecto al centroide del grupo, esto demuestra cómo el algoritmo asigna las muestras al grupo más cercano. Por otra parte, se presenta la reducción del MSE en cada iteración a medida que el algoritmo converge en una solución indicando que la salida esperada y la estimada aumentan su similitud. Finalmente, el valor SSE decrece con cada iteración denotando que la compacidad de los grupos aumenta.

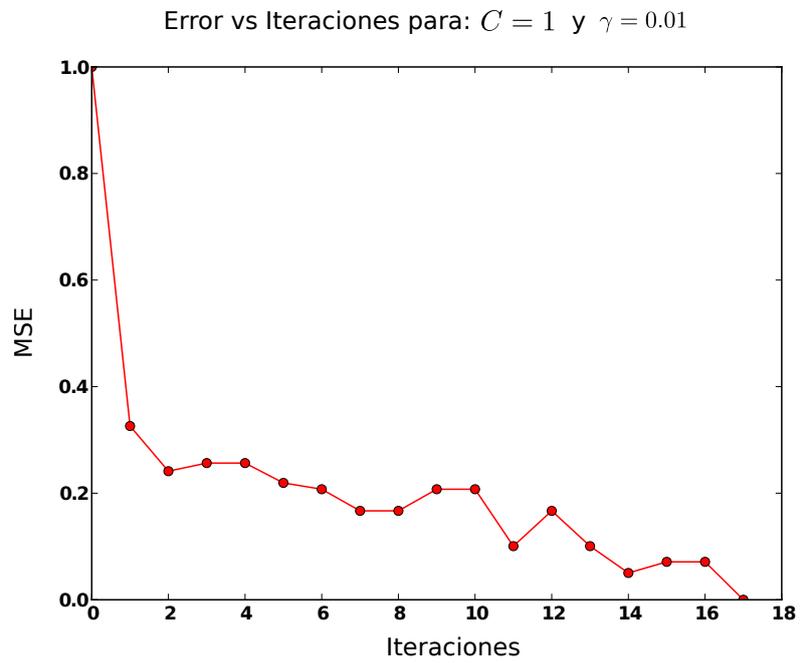


Figura 5.11: MSE entre la salida esperada y la salida estimada por el modelo de agrupamiento.

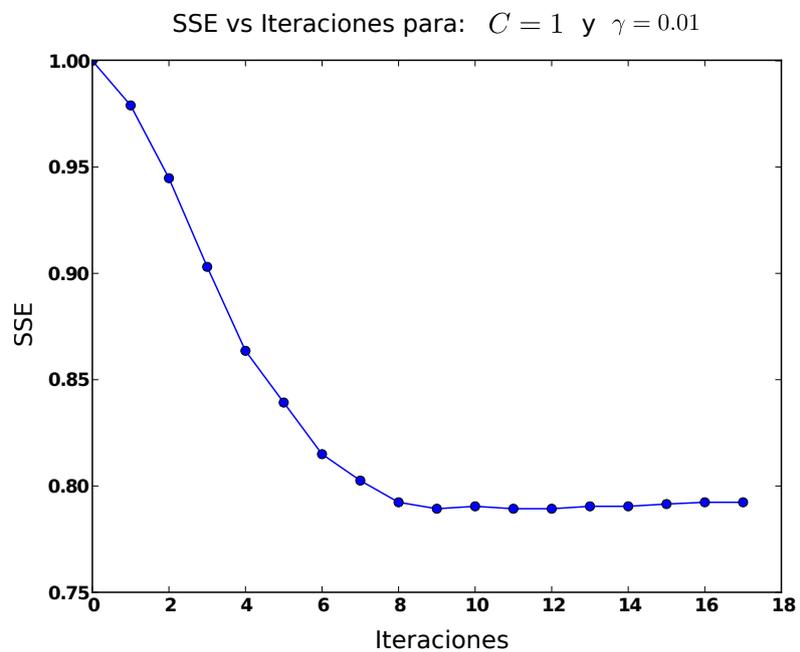


Figura 5.12: Reducción del SSE a medida que el algoritmo aumenta la compacidad de los grupos con cada iteración.

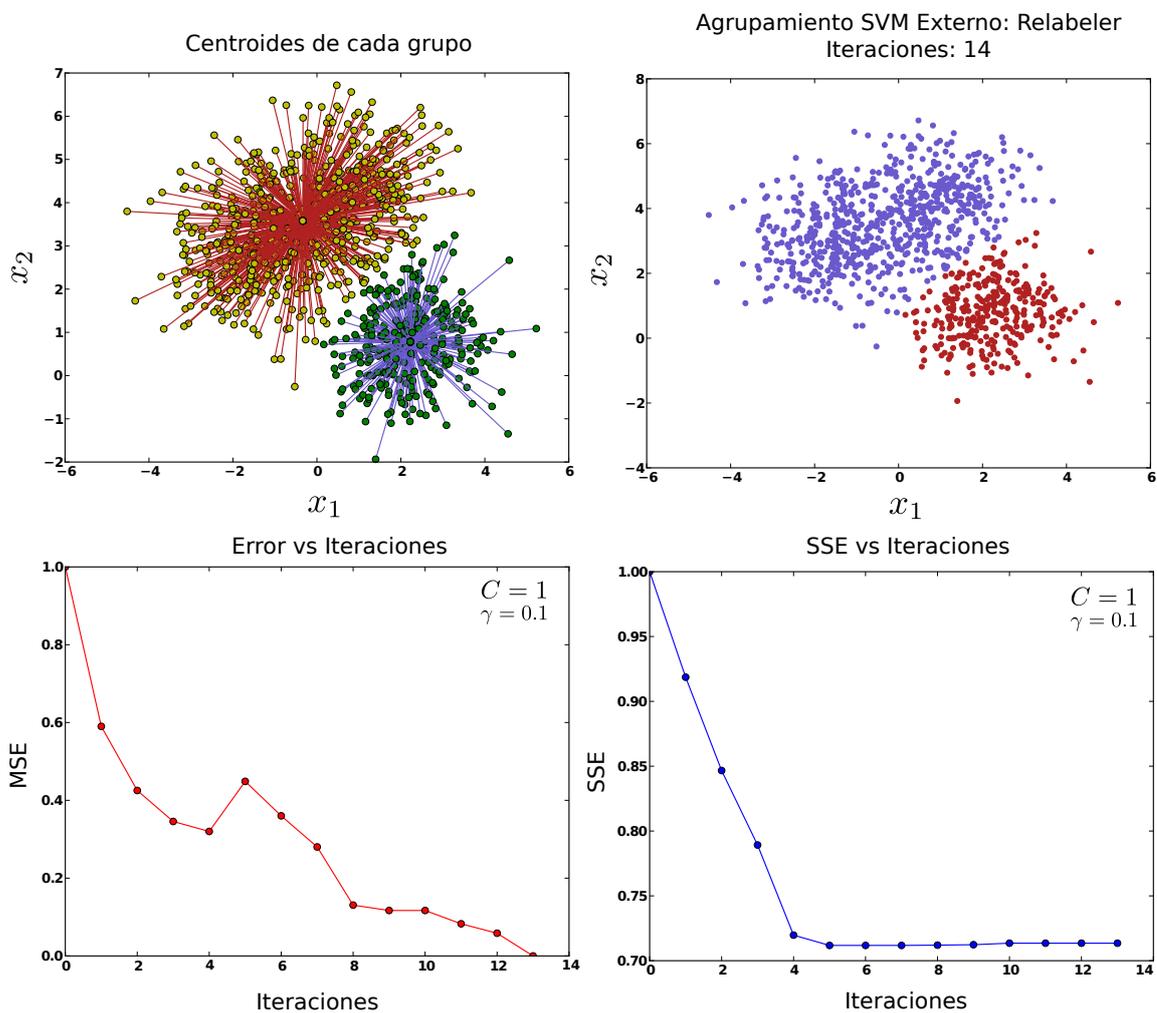


Figura 5.13: Algoritmo de agrupamiento SVM *Relabeler* sobre un conjunto desbalanceado de 1000 muestras. En la parte superior-izquierda se muestra la distancia Euclidiana entre cada una de las muestras pertenecientes a un grupo y su centroide. En la parte superior-derecha, se pueden ver los grupos en el espacio de entrada formados por el algoritmo de agrupamiento. En la parte inferior-izquierda, se muestra la evolución del MSE con cada iteración. En la parte inferior-derecha, se indica la reducción del SSE a medida que el algoritmo converge.

5.2.4. Agrupamiento híbrido

El agrupamiento SVM *Relabeler* encuentra un hiperplano de separación que está fuertemente ligado al hiperplano conseguido en la convergencia inicial, lo cual puede repercutir en su desempeño. Por esta razón, un enfoque un agrupa-

miento externo híbrido se presenta en [69], [67], el cual reemplaza el paso inicial de etiquetado aleatorio con un agrupamiento K-medias. El agrupamiento SVM Externo inicial debe ser levemente perturbado para inicializar apropiadamente el paso de re-etiquetado; de otra manera el algoritmo retornará a la solución original del método K-medias.

Agrupamiento K-medias

El algoritmo *K-medias* divide los datos en K grupos (C_1, C_2, \dots, C_K) , representados por sus centroides. El centroide de cada grupo es calculado como la media de todas las muestras que lo componen.

El algoritmo inicia con un conjunto inicial de los centroides de los grupos, eligiendolos de manera aleatoria. En cada iteración, cada muestra es asignada al centroide más cercano de acuerdo a la distancia Euclideana entre los dos. Después de este paso, el centroide de los grupos es recalculado.

El centroide de cada grupo es definido como la media de sus muestras:

$$\mu_k = \frac{1}{N_k} \sum_{i=1}^{N_k} x_i \quad (5.22)$$

donde N_k es el número de muestras pertenecientes al grupo k .

La convergencia se lleva a cabo cuando el error no se reduce con la re-localización del centroide del grupo, indicando que la partición actual esta en un óptimo local. Por otra parte, se puede definir también un número máximo de iteraciones como criterio de parada. El agrupamiento K-medias es mostrado en el Algoritmo 6.

Algoritmo 6 Agrupamiento por K-medias.

Require:

Vector de entrada: \mathbf{x}

Número de grupos: K

- 1: Inicialización de los K centroides de los grupos
 - 2: **while** la condición de parada no se satisfaga **do**
 - 3: Asignar las muestras al grupo más cercano
 - 4: Actualizar los centroides de los grupos
 - 5: **end while**
-

El algoritmo de agrupamiento SVM Externo híbrido, el cual reemplaza el paso inicial de asignación aleatoria de las etiquetas por una asignación mediante K-medias, se muestra en el Algoritmo 7.

Algoritmo 7 Agrupamiento SVM externo híbrido

- 1: Etiquetado inicial por *K-medias*
 - 2: **repeat**
 - 3: $\theta \leftarrow doSVM(\mathbf{x}, \hat{t})$
 - 4: $\hat{t} \leftarrow doRelabel(\mathbf{x}, \theta)$
 - 5: **until** \hat{t} permanezca constante
-

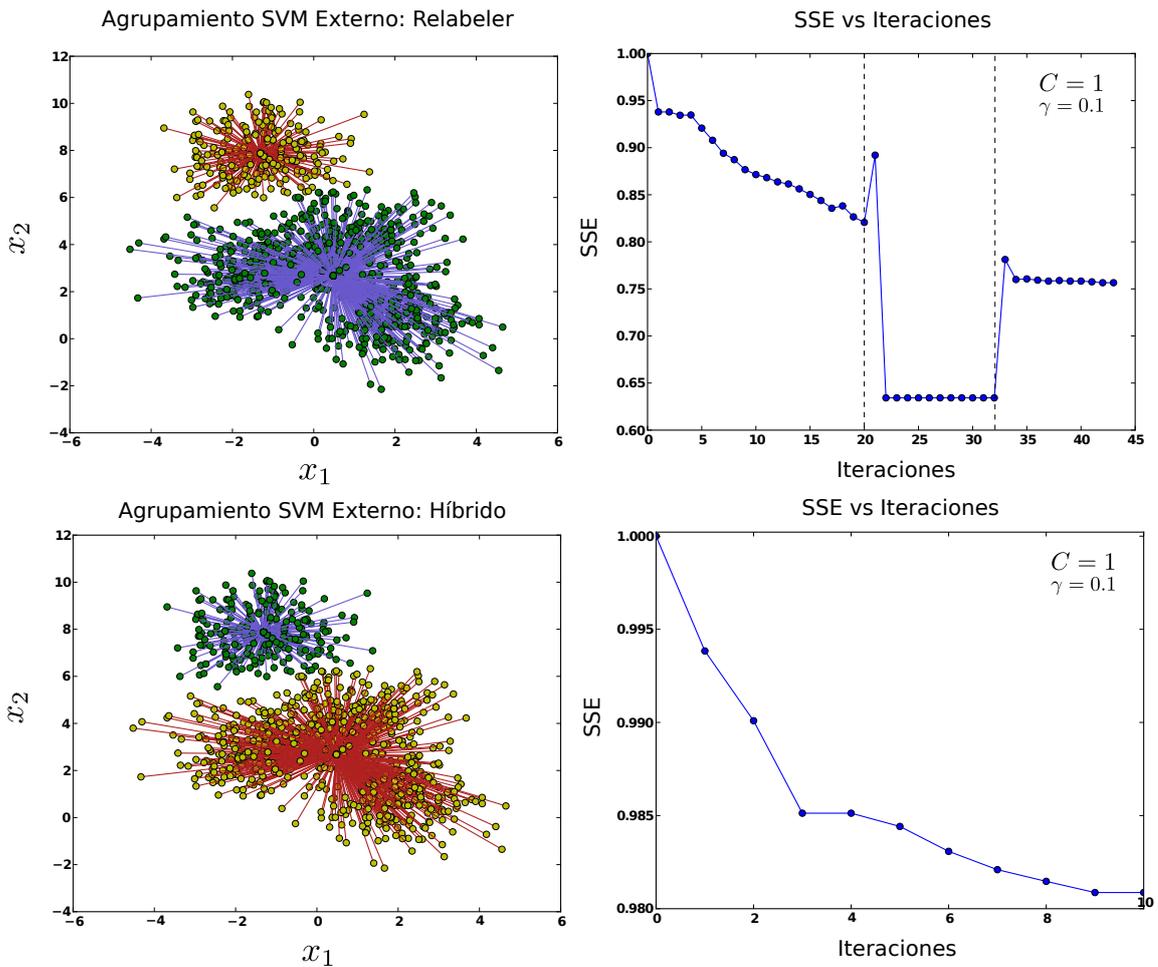


Figura 5.14: Comparación entre el algoritmo de agrupamiento SVM *Relabeler* e híbrido sobre un conjunto desbalanceado de 1000 muestras con los parámetros $C = 1$ y $\gamma = 0,01$. En la parte superior-izquierda se muestran los grupos formados por el método *Relabeler*. En la parte superior-derecha, el valor SSE respecto al número de iteraciones para el método *Relabeler*. En la parte inferior-izquierda, se muestran los grupos formados por el método híbrido. En la parte inferior-derecha, se indica la reducción del SSE respecto al número de iteraciones para el método híbrido

El enfoque híbrido del agrupamiento SVM externo mejora en gran medida el desempeño evitando que el algoritmo converga en un sólo grupo y disminuyendo el número de iteraciones para encontrar la solución. En la Figura 5.14 se muestra una comparación entre el algoritmo *Relabeler* y el híbrido, en este caso se muestran los grupos formados por ambos métodos.

En la parte superior-derecha de la gráfica se indica el SSE de los grupos formados por el método *Relabeler*. Aquí el algoritmo realiza tres ejecuciones las cuales se indican mediante la línea punteada, en la primera se hace una asignación errónea del número de muestras a cada grupo lo cual lleva al que el algoritmo asigne todos los datos a un solo grupo; en la segunda, el algoritmo inicialmente asigna todas las muestras a un grupo y el algoritmo no converge; en la tercera, la asignación se hace de manera correcta y el algoritmo logra converger a la solución. En la parte inferior-derecha, se muestra el SSE del método híbrido, en este caso, el algoritmo converge rápidamente en una sólo ejecución debido a que este involucra una asignación inicial de las muestras por medio de K-medias.

5.3. Eliminación de artefactos oculares en señales EEG mediante agrupamiento SVM Externo Híbrido

En esta sección se muestran los resultados obtenidos de la implementación de un sistema de detección y eliminación de artefactos oculares en EEG basado en el método de agrupamiento SVM Externo híbrido. El sistema está dividido en dos partes, la primera es la etapa de entrenamiento del modelo de agrupamiento y la segunda es la etapa de prueba. Los Algoritmos 8 y 9 muestran los procedimientos respectivamente.

Algoritmo 8 Etapa de entrenamiento

Require:

Vector de características: \mathbf{f}

Parámetro de penalización: C

Ancho del *kernel*: γ

Criterio de parada: MSE

1: $\theta \leftarrow \text{Agrupamiento}(\mathbf{f}, C, \gamma, MSE)$

Siguiendo el enfoque del aprendizaje no-supervisado, el algoritmo en este caso no recibe ninguna información acerca de a cuál grupo, en este caso componentes independientes de artefactos oculares o de señales EEG, pertenecen las

muestras. El algoritmo tiene como entrada sólo el vector de características \mathbf{f} y los parámetros del modelo C , γ y MSE .

La primera etapa proporciona un modelo entrenado, representado por la frontera de decisión θ , el cual es usado para asignar cada componente independiente a un grupo. El modelo se entrenó con 1540 muestras cada una de las cuales tiene una dimensión $d = 8$, debido a que esta compuesta por 8 características extraídas de las componentes independientes resultantes de la descomposición de las señales EEG mediante ICA. Las dimensiones representan la curtosis, el Análisis de Fluctuación sin Tendencia, la Información de Fisher, la Dimensión Fractal de Higuchi, la Movilidad y la Complejidad de Hjorth, el Exponente de Hurst y la Dimensión Fractal de Petrosian.

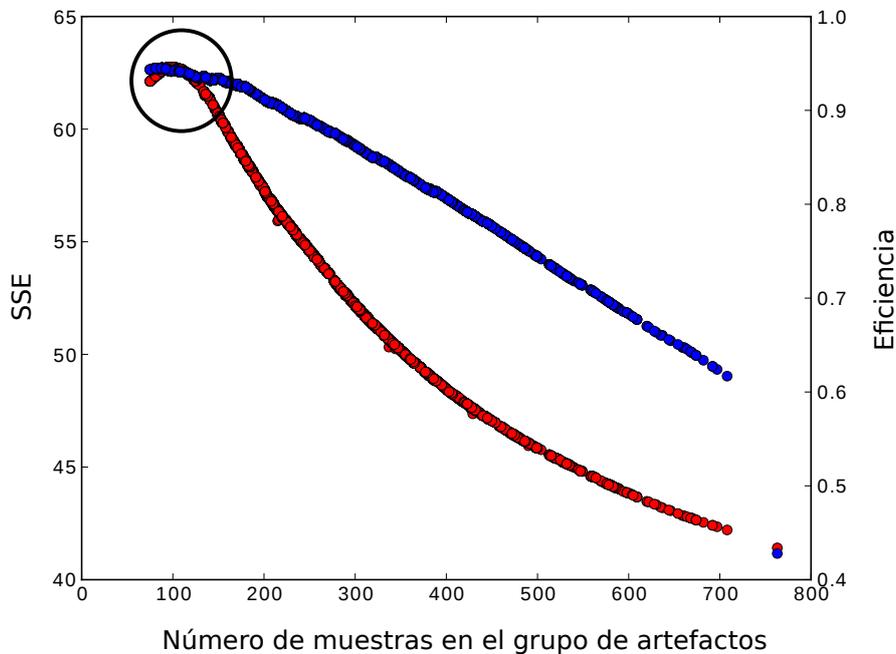


Figura 5.15: SSE y eficiencia vs el número de muestras pertenecientes al grupo de artefactos oculares.

El desempeño del método de agrupamiento fue validado usando el SSE y la efectividad para 10.000 ejecuciones. En la Figura 5.15 se muestra el SSE y la efectividad del modelo de agrupamiento vs el número de muestras pertenecientes al grupo de los artefactos oculares. El valor de efectividad más alto se produce cuando el grupo positivo tiene un tamaño alrededor de 100 muestras, sin embargo, el valor del SSE es grande indicando que existen grupos con mayor compacidad pero menor efectividad.

Los grupos generados en la etapa de entrenamiento del modelo con los parámetros elegidos para el sistema de eliminación de artefactos, $C = 2,5$, $\gamma = 0,009$ y $MSE = 1e - 6$, son mostrados en la Figura 5.16.

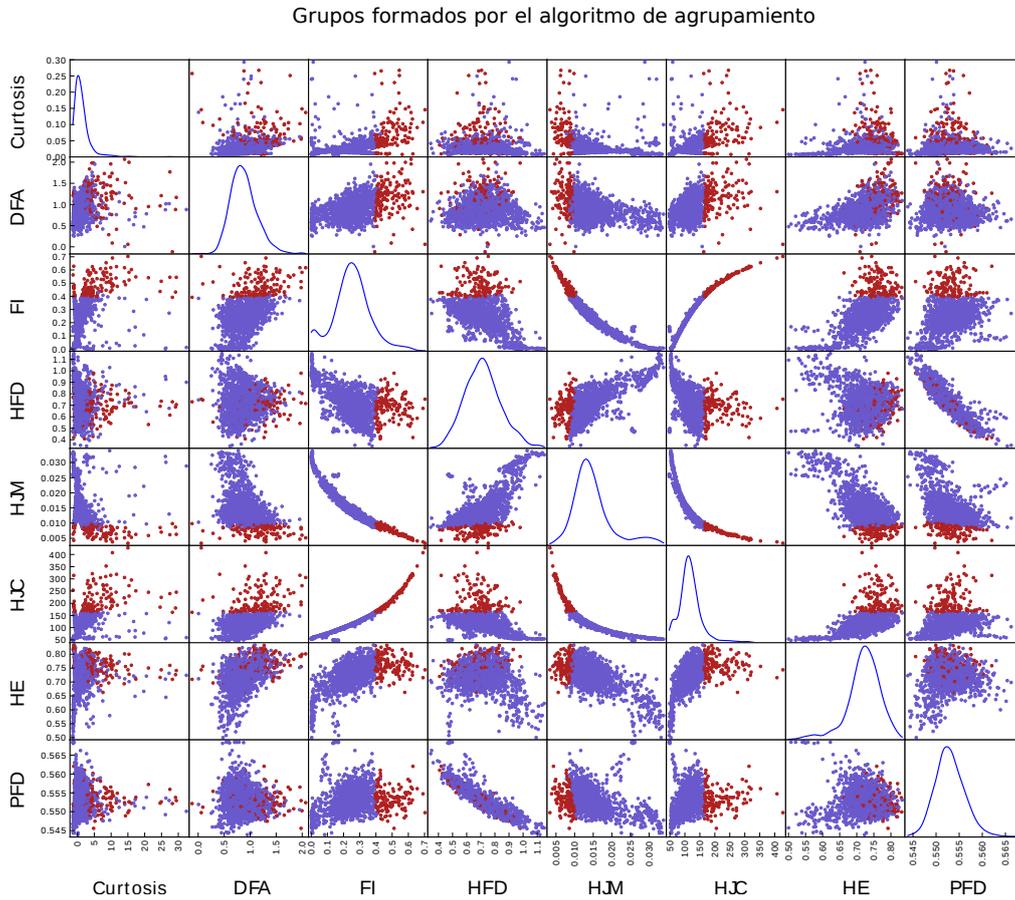


Figura 5.16: Grupos formados por el agrupamiento SVM híbrido para un conjunto de 1540 muestras. En rojo se muestran las ICs de los artefactos oculares y en azul las ICs de las señales EEG. La matriz muestra la relación diádica entre las 8 características extraídas, mientras su diagonal, indica la función de distribución de las muestras en cada característica.

Aquí las muestras de color rojo pertenecen a las ICs de los artefactos oculares y las de color azul a las ICs de las señales EEG. El algoritmo hace uso de la información intrínseca de las muestras para formar grupos los cuales revelan su naturaleza desbalanceada sin haber tenido conocimiento a priori de esto. La convergencia del algoritmo ocurre en tres iteraciones en las cuales el MSE se reduce hasta cumplir el criterio de parada, Figura 5.17. Por otro lado, la compacidad de los grupos, que es inversa al SSE, se incrementa con cada iteración, Figura 5.18.

En la segunda etapa del algoritmo, el registro EEG a procesar es cargado, estandarizado y descompuesto en sus componentes independientes mediante FastICA. A cada componente se le extrae el vector de características el cual a través la función de decisión θ indicará si es un artefacto ocular o una señal EEG. Finalmente se hace el procedimiento de eliminación de las ICs de los artefactos y la reconstrucción del registro EEG a través de las ecuaciones (5.45) y (5.46) expuestas en el capítulo anterior.

Algoritmo 9 Etapa de prueba

Require:

Datos EEG: X
Modelo entrenado: θ
1: $X \leftarrow$ Estandarización de X
2: $S, A \leftarrow \text{FastICA}(X)$
3: **for** cada $s_i \in S$ **do**
4: $f_i \leftarrow$ Extracción de características de s_i
5: $g_i \leftarrow \text{Predicción}(\theta, f_i)$
6: **end for**
7: $X \leftarrow \text{Reconstrucción}(S, A, \mathbf{g})$

A continuación se muestran dos ejemplos de la aplicación del sistema de eliminación de artefactos en dos porciones de registros EEG.

En el primer caso, el algoritmo se aplicó en el registro EEG del sujeto A06T en el rango temporal de 288 a 296 s. En la Figura 5.19 se muestran los canales EEG y HEOG pertenecientes al registro, además la influencia de los artefactos oculares en las señales EEG se indica a través de una franja gris.

En la Figura 5.20 se presentan las componentes independientes resultantes del proceso de descomposición a través del algoritmo *FastICA*. Aquí las ICs correspondientes a los artefactos se resaltan a través de una franja gris. En este caso, la IC9 describe claramente el comportamiento del artefacto en el registro EEG.

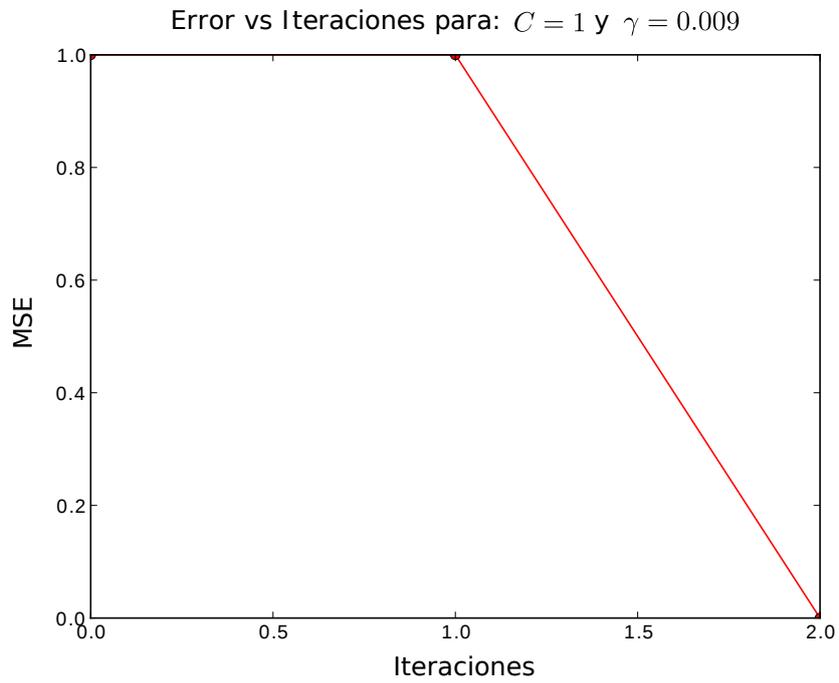


Figura 5.17: Reducción del MSE durante el proceso de convergencia.

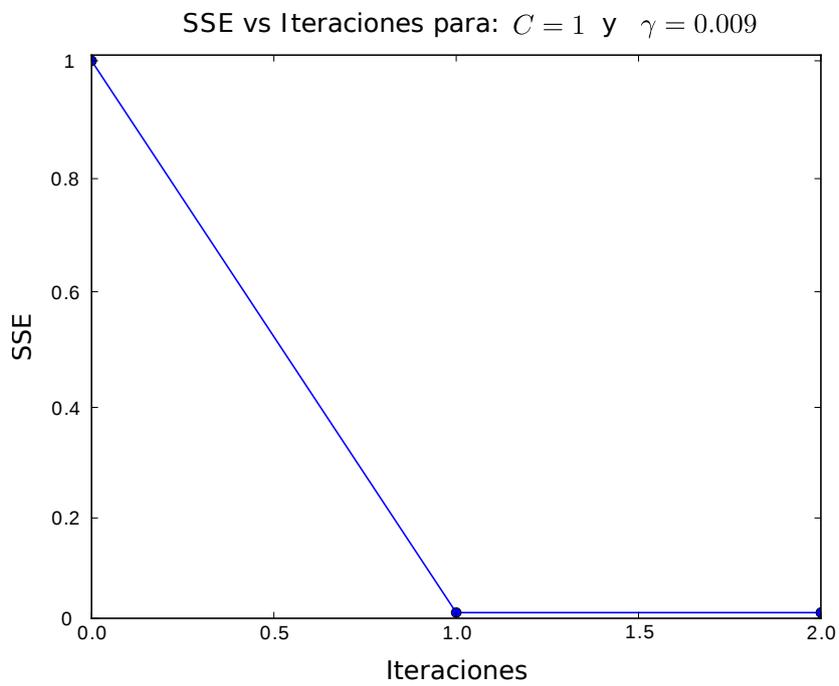


Figura 5.18: Aumento de la compacidad de los grupos formados a medida que el SSE se reduce.

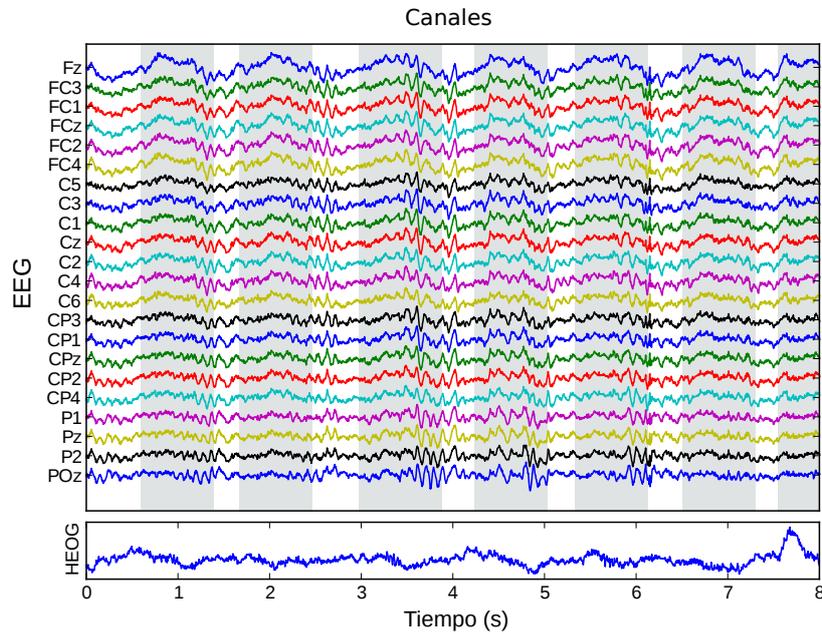


Figura 5.19: Canales EEG registrados del sujeto A06T en el rango temporal de 288-296 s. En la parte superior se muestran los 22 canales EEG registrados y en la parte inferior un canal EOG el cual sirve de referencia para saber el momento en que ocurren los movimientos oculares. Las franjas de color gris indican los instantes de tiempo donde hay contaminación por artefactos.

Una vez que el algoritmo identifica las ICs de los artefactos por medio de las características extraídas y del modelo de agrupamiento entrenado, este procede a eliminarlas, es decir, a atenuar su influencia sobre el registro EEG, Figura 5.21.

Finalmente se realiza el proceso de reconstrucción del registro EEG, para ello, se hace uso de la matriz de mezcla hallada mediante el método *FastICA* y de las ICs obtenidas anteriormente. Las señales EEG reconstruidas para este ejemplo se muestran en la Figura 5.22 donde se puede ver información importante en la zona occipital que anteriormente parecía oculta debido a la contaminación por los potenciales sacádicos provenientes del movimiento ocular.

El primer ejemplo demuestra la capacidad del sistema de eliminación automática de artefactos en un caso donde se presentan movimientos sacádicos sucesivos los cuales son comunes en los estudios del comportamiento de la corteza visual en tareas de seguimiento, reconocimiento y percepción [42].

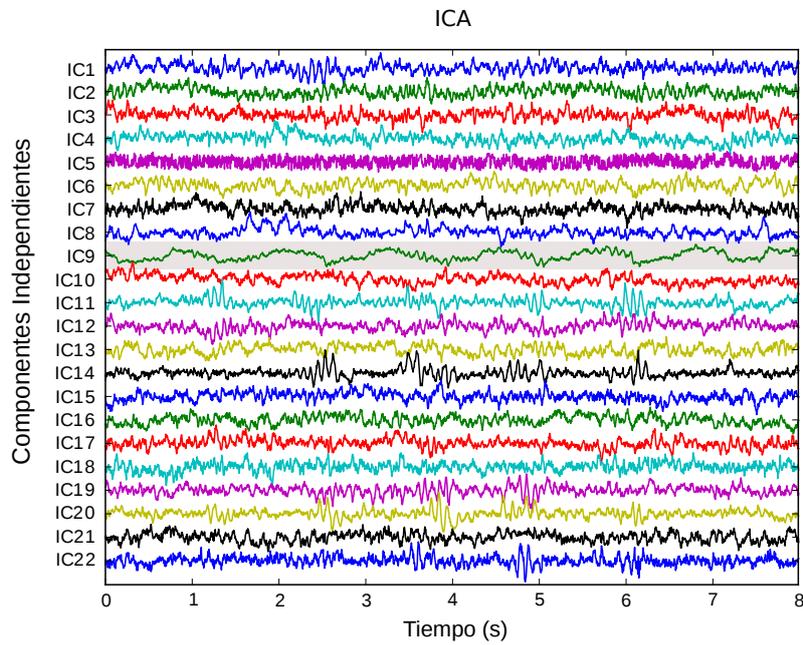


Figura 5.20: Descomposición de los canales EEG en sus componentes independientes mediante ICA. las franjas de color gris indican las ICs pertenecientes a los artefactos oculares.

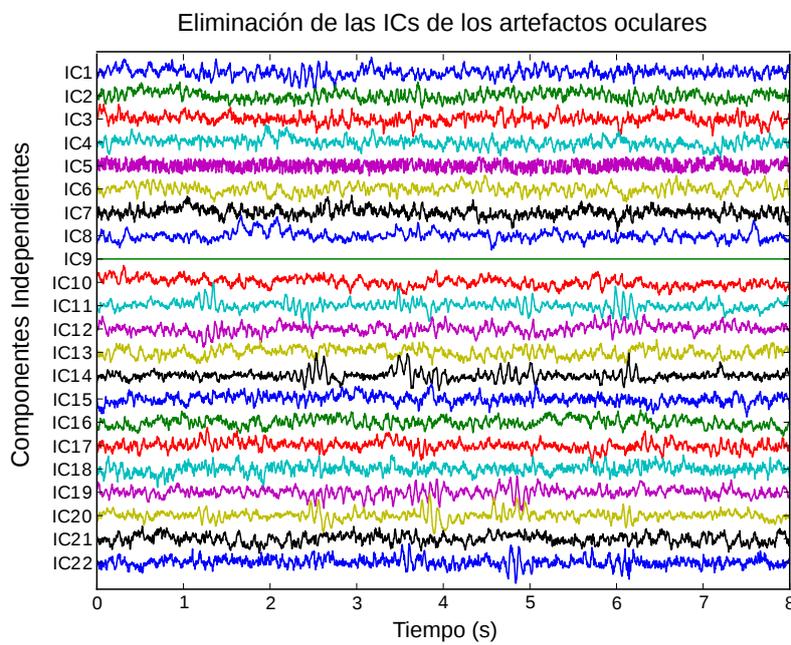


Figura 5.21: Eliminación de las ICs de los artefactos oculares.

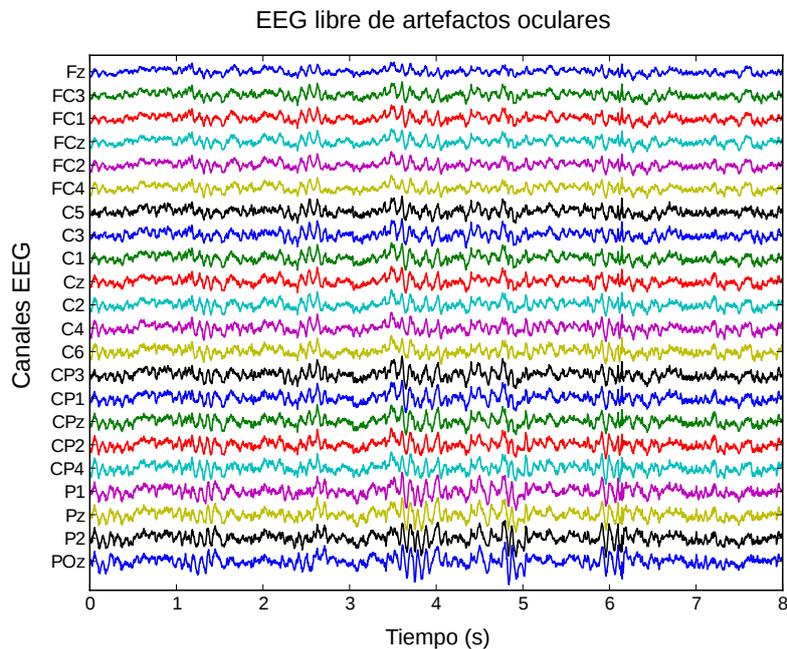


Figura 5.22: Señales EEG reconstruidas y sin influencia de los artefactos oculares.

En el segundo ejemplo se propone eliminar un artefacto ocular aislado en un registro EEG, es decir, el movimiento ocular ocurre en un instante donde no lo antecede ni lo procede otro.

En la Figura 5.23 se presenta el registro EEG que pertenece al sujeto A08T en el rango de tiempo entre 300 y 308 s. Aquí las señales EEG son afectadas por una sacada que ocurre entre los segundos de registro 5 y 6, su influencia se extiende a lo largo de todos los canales EEG.

La descomposición de las señales cerebrales en sus componentes independientes se muestran en la Figura 5.24, en este caso las ICs 5 y 13 conforman las fuentes que original el artefacto ocular. En la Figura 5.25 se puede observar como el algoritmo detecta y elimina de manera automática estas componentes.

La reconstrucción del registro EEG libre de artefactos oculares se expone en la Figura 5.26. Aquí se puede ver como la influencia del movimiento ocular ha sido removida de señales cerebrales proporcionando de esta manera una fuente de información libre de ruido que pueda influenciar un posterior estudio en base a estos datos.

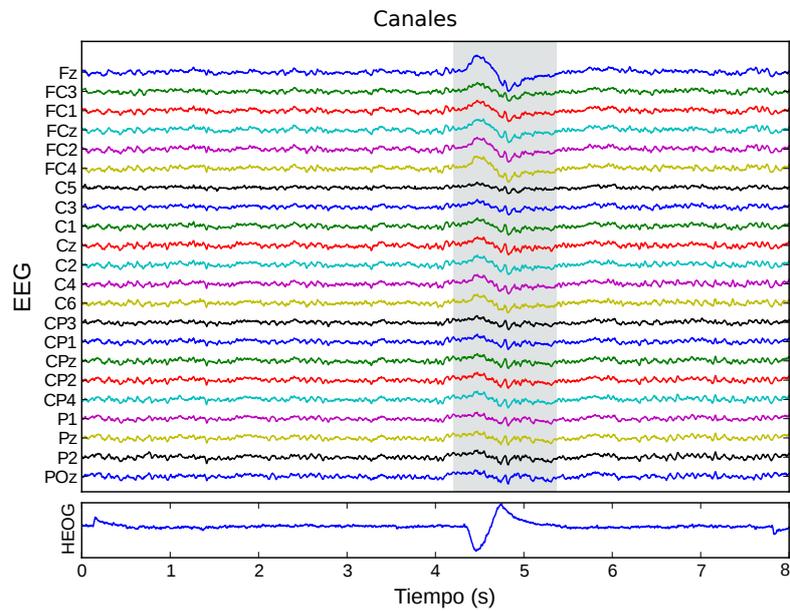


Figura 5.23: Canales EEG registrados del sujeto A08T en el rango temporal de 300-308 s. En la parte superior se muestran los 22 canales EEG registrados y en la parte inferior un canal EOG.

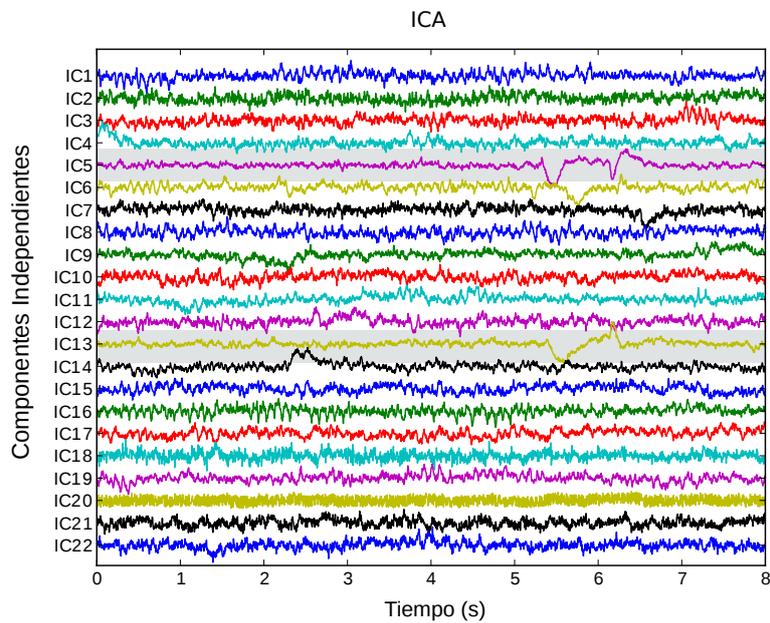


Figura 5.24: Descomposición de los canales EEG en sus componentes independientes mediante ICA. las franjas de color gris indican las ICs pertenecientes a los artefactos oculares.

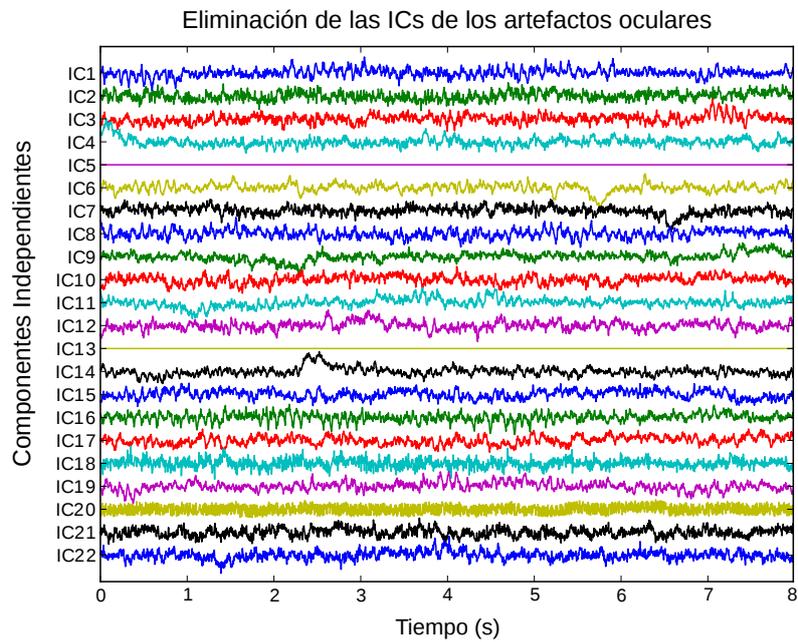


Figura 5.25: Eliminación de las ICs de los artefactos oculares.

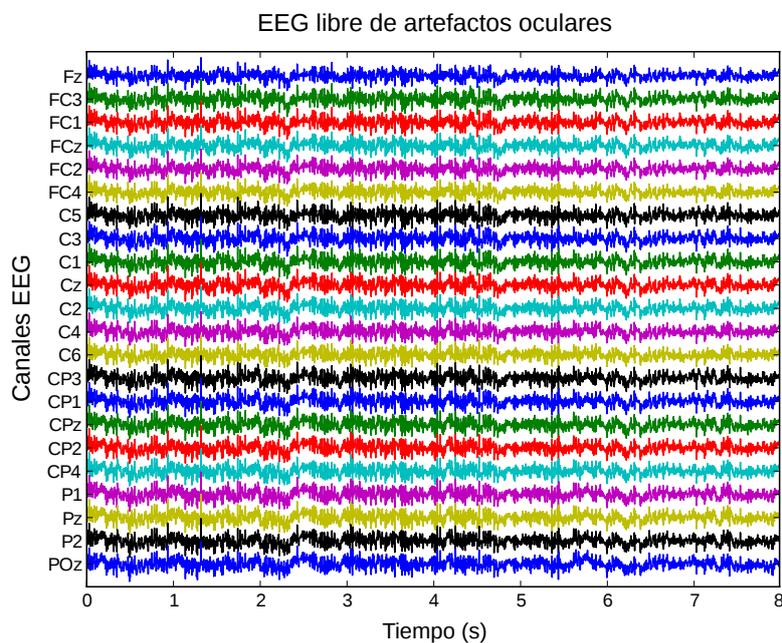


Figura 5.26: Señales EEG reconstruidas y libres de la influencia de los artefactos oculares.

En esta sección, se presentó un sistema automático para la eliminación de

artefactos oculares en EEG usando ICA y agrupamiento SVM. Esta herramienta provee un nuevo método para eliminar artefactos evitando su detección y eliminación manual en las señales EEG y de esta manera reduciendo el tiempo y costo del proceso. Algunos métodos actuales eliminan las secciones con ruido, sin embargo, estas secciones también tienen información del comportamiento neuronal.

El sistema presentado toma en cuenta estas porciones de la señal después de hacer el respectivo filtrado de los artefactos. Por otra parte, el método provee señales sin ruido para las etapas posteriores de procesamiento e.g. en el diagnóstico de enfermedades u otras aplicaciones que usen EEG.

La técnica se basó en el concepto del aprendizaje no-supervisado, este solo usa la información intrínseca de los datos para formar grupos, i.e. este no necesita de información a priori acerca de la clase a la que pertenecen de los datos.

Este método automático y dependiente solo de los datos, provó ser una herramienta útil en la tarea de eliminación de artefactos, además que aumenta las posibilidades de desarrollar una técnica robusta para procesar datos EEG antes de la etapa de análisis la cual tiene como resultado la toma de decisiones en procesos terapéuticos y farmacológicos.

La verdadera alegría está en el hallazgo más allá que en el saber.

Isaac Asimov

6

Conclusiones

En el presente trabajo se presentaron dos herramientas automáticas de detección y eliminación de artefactos oculares en registros EEG, una basada en el enfoque del aprendizaje supervisado y otra en el agrupamiento. Los dos métodos parten de la descomposición de las señales EEG mediante el Análisis de Componentes Independientes debido que esta ha probado ser una herramienta eficaz como se muestra en [62] donde se hace una revisión extensiva sobre los algoritmos de eliminación de artefactos en EEG. Aquí se determina que los enfoques clásicos son basados en: regresión, corrección de artefactos oculares, filtrado y la separación ciega de fuentes (BSS). Además, se concluye que el 45 % de los algoritmos actuales de eliminación de artefactos en EEG usan ICA y que la proyección de este campo de estudio apunta a la mejora de las técnicas existentes mediante la combinación de varios métodos, la automatización del proceso de filtrado y la

validación mediante métricas de desempeño objetivas [62].

En el primer enfoque, se debió tener en cuenta el origen desbalanceado del vector de características extraídas de las ICs producto del ICA. Esta relación de desproporcionalidad era aproximadamente de 10 a 1, es decir por cada muestra perteneciente a una componente de un artefacto ocular se tenían 10 pertenecientes a componentes de señales EEG. Por esta razón, se eligió usar como modelo de clasificación una Máquina de Vectores Soporte balanceada, la cual hace una normalización del hiperplano de decisión en base a la relación de desbalanceo, de esta manera no se ve afectado su desempeño. En la mayoría de los métodos de eliminación de artefactos con SVM no se ha tenido en cuenta este factor, el cual puede afectar de manera significativa su efectividad como se demuestra en [54].

Los resultados de esta primera metodología demostraron ser alentadores ya que la precisión del modelo alcanzó valores por encima del 80 % en la tarea de eliminación de los artefactos en comparación con un 50 % para el caso donde no se tiene en cuenta la naturaleza desbalanceada de los datos.

El segundo método, se basó en el concepto del aprendizaje no-supervisado, explorado sólo en los últimos años en este tipo de problemas [72], [47]. Este hace uso sólo de la información intrínseca de los datos para formar grupos o clases, es decir, no necesita de información a priori de las etiquetas de los datos. En este caso, se utilizó un algoritmo de agrupamiento basado en las bondades de SVM, llamado agrupamiento SVM Externo híbrido el cual funciona como un árbol de decisión conformado por una serie de SVMs binarias las cuales van modificando el hiperplano de decisión a medida que el algoritmo busca los grupos más compactos.

El modelo fue entrenado sobre un vector de características de 8 dimensiones. Estas correspondían a mediciones de fractalidad, cantidad de información y aleatoriedad las cuales permiten discriminar las ICs tanto de potenciales sacádicos como de señales cerebrales. Los datos formaron grupos naturalmente que eran similares a las clases formadas por las muestras usadas en el método supervisado, demostrando una relación interna entre las muestras de cada clase.

Esta herramienta, automática y dependiente sólo de los datos de entrada, mostró ser muy efectiva en el proceso de identificación de los artefactos oculares; se evaluaron varias pruebas en registros de diferentes sujetos y se obtuvieron resultados mejores que los logrados con el método de eliminación de artefactos supervisado.

Los métodos propuestos en este trabajo abren la puerta a la creación de una herramienta robusta que permita remover el efecto que causa los movimientos oculares en los registros EEG, el cual puede repercutir en procesos posteriores que tengan como base las señales EEG adquiridas y de esta manera evitar conclusiones erróneas de trabajos e investigaciones que se basen en datos contaminados.

Bibliografía

- [1] *Support Vector Machines*. Springer New York, 2008.
- [2] Jesús S Aguilar-Ruiz and Norberto Diaz-Diaz. Selección de atributos relevantes basada en bootstrapping. pages 21–30, 2005.
- [3] Periklis Andritsos. Data clustering techniques. Technical report, University of Toronto, 2002.
- [4] Davide Anguita, Andrea Boni, Ro Ridella, Fabio Riveccio, and Dario Sterpi. Theoretical and practical model selection methods for support vector classifiers. In *in L. Wang (Ed.), Support Vector Machines: Theory and Applications*. Springer Verlag, 2005.
- [5] L.C. Arangüena and J.I. Dorado. *Psicofisiología: fundamentos metodológicos*. Colección Psicología/Ediciones Pirámide Series. Pirámide, 1997.
- [6] Varios Autores. Amsterdam library of object images (aloi). <http://aloi.science.uva.nl/>, 2014. Accedido 05-03-2014.
- [7] Varios Autores. The mcgill physiology virtual laboratory. <http://www.medicine.mcgill.ca/physio/vlab/default.htm>, 2014. Accedido 29-03-2014.
- [8] Forrest Sheng Bao, Xin Liu, and Christina Zhang. PyEEG: An open source python module for EEG/MEG feature extraction. *Computational Intelligence and Neuroscience*, 2011, 2011.
- [9] Asa Ben-hur, David Horn, Hava Siegelmann, and Vladimir Vapnik. A support vector method for hierarchical clustering. In *In Advances in Neural Information Processing Systems 13*, pages 367–373. MIT Press, 2001.

- [10] Asa Ben-hur, David Horn, and Vladimir Vapnik. Support Vector Clustering. 2:125–137, 2001.
- [11] Asa Ben-hur, David Horn, and Vladimir Vapnik. Support Vector Clustering. 2:125–137, 2001.
- [12] Pavel Berkhin. Survey of clustering data mining techniques. Technical report, 2002.
- [13] Christopher M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.
- [14] C. Brunner, R. Leeb, G. R. Müller-Putz, A. Schlögl, and G. Pfurtscheller. Bci competition 2008 - graz data set a, 2008.
- [15] Tian-tian Chang, Horace H S Ip, and Jun Feng. Large scale image clustering with support vector machine based on visual keywords. In *Proceedings of the Tenth International Workshop on Multimedia Data Mining - MDMKDD '10*, pages 1–8, New York, New York, USA, 2010. ACM Press.
- [16] Jing Cheng and Guangyuan Liu. Detrended Fluctuation Analysis Based on the Affective ECG Theory Background and Related Work The Affective Detection Theory Based on Physiological Signals. 1(March):91–102, 2014.
- [17] R.J. Croft and R.J. Barry. Removal of ocular artifact from the EEG: a review. *Neurophysiologie Clinique/Clinical Neurophysiology*, 30(1):5–19, February 2000.
- [18] R.J. Croft, University of Wollongong, and University of Wollongong. Dept. of Psychology. *The Removal of Ocular Artifact from the EEG*. University of Wollongong, 1999.
- [19] Servicio de Neurología. Cirugía de la epilepsia: Los electrodos profundos mejoran la localización del foco epiléptico y reducen los riesgos. <http://www.neurologiaclinica.es/cirugia-epilepsia-electrodosprofundos-mejoran-localizacion-foco-epileptico-reducen-riesgo>, 2014. Accedido 25-04-2014.
- [20] Jie Dong, Tao Wang, Ai-tao Zhang, and Hong-ya Dai. The removal of blink and saccade artifact in EEG recordings by Independent Component Analysis. *2010 3rd International Conference on Biomedical Engineering and Informatics*, 1(Bmei):1071–1075, October 2010.

- [21] John Enderle, Susan Blanchard, and Joseph Bronzino. *Introduction to biomedical engineering*. 2001.
- [22] Samira Vafaye Eslahi and Nader Jafarnia Dabanloo. Fuzzy support vector machine analysis in EEG classification. 4(12):4062–4066, 2013.
- [23] L Gila, a Villanueva, and R Cabeza. [Physiopathology and recording techniques of the ocular movements]. *Anales del sistema sanitario de Navarra*, 32 Suppl 3:9–26, 2009.
- [24] Carlos Gómez, Ángela Mediavilla, Roberto Hornero, Daniel Abásolo, and Alberto Fernández. Use of the Higuchi’s fractal dimension for the analysis of MEG recordings from Alzheimer’s disease patients. *Medical Engineering & Physics*, 31(3):306–313, April 2009.
- [25] N. S. Gupta, B. S. Agrawal, and R. M. Chauhan. Survey on clustering techniques of data mining. *American International Journal of Research in Science, Technology, Engineering and Mathematics*, pages 206–211, February 2015.
- [26] I. Guyon. *Feature Extraction: Foundations and Applications*. Studies in Fuzziness and Soft Computing. Springer, 2006.
- [27] Sebastian Halder, Michael Bensch, Jürgen Mellinger, Martin Bogdan, Andrea Kübler, Niels Birbaumer, and Wolfgang Rosenstiel. Online artifact removal for brain-computer interfaces using support vector machines and blind source separation. *Computational intelligence and neuroscience*, 2007:82069, January 2007.
- [28] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition*. Springer Series in Statistics. Springer, 2009.
- [29] Joerg F Hipp and Markus Siegel. Dissociating neuronal gamma-band activity from cranial and ocular muscle activity in EEG. *Frontiers in human neuroscience*, 7(July):338, January 2013.
- [30] A. Hyvärinen, J. Karhunen, and E. Oja. *Independent Component Analysis*. Adaptive and Cognitive Dynamic Systems: Signal Processing, Learning, Communications and Control. Wiley, 2004.
- [31] M.G.M.P.S.F. Institute and C.T.P.B.C.P. Research. *Nonextensive Entropy : Interdisciplinary Applications: Interdisciplinary Applications*. Oxford University Press, USA, 2004.

- [46] Seung-Hyeon Oh, Yu-Ri Lee, and Hyoung-Nam Kim. A Novel EEG Feature Extraction Method Using Hjorth Parameter. *International Journal of Electronics and Electrical Engineering*, 2(2):106–110, 2014.
- [47] J. Otero-Millan, Jose L Alba Castro, Stephen L Macknik, and S. Martinez-Conde. Unsupervised clustering method to detect microsaccades. *Journal of Vision*, 14(2):18–18, February 2014.
- [48] Ramaswamy Palaniappan. *Biological signal analysis*. 2010.
- [49] Udit Patidar and George Zouridakis. A hybrid algorithm for artifact rejection in EEG recordings based on iterative ICA and fuzzy clustering. *Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference*, 2008:50–3, January 2008.
- [50] Pedro Ponce, Arturo Molina, David C., and Dimitra Grammatikou. Brain computer interfaces for cerebral palsy. In *Cerebral Palsy - Challenges for the Future*. InTech, mar 2014.
- [51] Dale Purves, George J. Augustine, David Fitzpatrick, William C. Hall, Anthony-Samuel LaMantia, James O. McNamara, and S. Mark Williams. *Neuroscience*. 2007.
- [52] Chunchu Rambabu and B. Rama Murthy. EEG Signal with Feature Extraction using SVM and ICA Classifiers. *International Journal of Computer Applications*, 85(3):1–7, January 2014.
- [53] Tomas Ros, Bernard J. Baars, Ruth A. Lanius, and Patrik Vuilleumier. Tuning pathological brain oscillations with neurofeedback: a systems neuroscience framework. *Frontiers in Human Neuroscience*, 8, dec 2014.
- [54] Shi-yun Shao, Kai-quan Shen, Chong Jin Ong, Einar P V Wilder-Smith, and Xiao-ping Li. Automatic EEG artifact removal: a weighted support vector machine approach with error correction. *IEEE transactions on bio-medical engineering*, 56(2):336–44, March 2009.
- [55] Minfen Shen, Chunhao Lin, Jialiang Huang, and Yanxun Li. Modeling and Prediction of EEG Signal Using Support Vector Machine. *2008 2nd International Conference on Bioinformatics and Biomedical Engineering*, pages 1988–1991, 2008.
- [56] Leor Shoker, Saeid Sanei, and Jonathon Chambers. Artifact removal from electroencephalograms using a hybrid BSS-SVM algorithm. *IEEE Signal Processing Letters*, 12(10):721–724, October 2005.

- [57] Rajesh Singla, Brijil Chambayil, Arun Khosla, and Jayashree Santosh. Comparison of SVM and ANN for classification of eye events in EEG. *Journal of Biomedical Science and Engineering*, 04(01):62–69, 2011.
- [58] Ignacio Javier Benítez Sánchez. Técnicas de agrupamiento para el análisis de datos cuantitativos y cualitativos, 2005.
- [59] Soroosh Solhjo. Classification of chaotic signals using HMM classifiers: EEG-based mental task classification. *Proc. European Signal ...*, pages 18–21, 2005.
- [60] Yuchun Tang, Yan-Qing Zhang, N.V. Chawla, and S. Krasser. SVMs modeling for highly imbalanced classification. *IEEE Trans. Syst., Man, Cybern. B*, 39(1):281–288, feb 2009.
- [61] M Ungureanu, C Bigan, R Strungaru, and V Laurescu. Independent Component Analysis Applied in Biomedical Signal Processing Politehnica. *Journal of the Institute of Measurement Science SAS*, 4(2):1–8, 2004.
- [62] Jose Antonio Urigüen and Begoña Garcia-Zapirain. EEG artifact removal—state-of-the-art and guidelines. *J. Neural Eng.*, 12(3):031001, apr 2015.
- [63] Frederi G. Viens, Maria C. Mariani, and Ionuț Florescu, editors. *Handbook of Modeling High-Frequency Data in Finance*. John Wiley & Sons, Inc., Hoboken, NJ, USA, December 2011.
- [64] Benjamin X. Wang and Nathalie Japkowicz. Boosting support vector machines for imbalanced data sets. *Knowledge and Information Systems*, 25(1):1–20, mar 2009.
- [65] John G. Webster. *Medical Instrumentation Application and Design*. 2009.
- [66] S. Winters-Hilt. *MacHine-Learning Based Sequence Analysis, Bioinformatics and Nanopore Transduction Detection*. Lulu Enterprises Incorporated, 2011.
- [67] Stephen Winters-Hilt. *MacHine-Learning Based Sequence Analysis, Bioinformatics and Nanopore Transduction Detection* -. Lulu.com, Raleigh, North Carolina, 2011.
- [68] Stephen Winters-Hilt and Sam Merat. SVM clustering. *BMC bioinformatics*, 8 Suppl 7:S18, January 2007.
- [69] Stephen Winters-Hilt and Sam Merat. SVM clustering. *BMC bioinformatics*, 8 Suppl 7:S18, January 2007.

- [70] Stephen Winters-Hilt, Anil Yelundur, Charlie McChesney, and Matthew Landry. Support vector machine implementations for classification & clustering. *BMC bioinformatics*, 7 Suppl 2:S4, January 2006.
- [71] Stephen Winters-Hilt, Anil Yelundur, Charlie McChesney, and Matthew Landry. Support vector machine implementations for classification & clustering. *BMC bioinformatics*, 7 Suppl 2:S4, January 2006.
- [72] Yuan Zou, John Hart, and Roozbeh Jafari. Automatic EEG artifact removal based on ICA and Hierarchical Clustering. In *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, volume 5, pages 649–652. IEEE, March 2012.