



UNIVERSIDAD NACIONAL
AUTÓNOMA
DE MÉXICO

FACULTAD DE CIENCIAS

Aplicación de Algoritmos Genéticos para la
selección de variables en la clasificación de la
señal electroencefalográfica (EEG) P300

T E S I S

QUE PARA OBTENER EL TÍTULO DE:
Matemático

PRESENTA:

Jaime Arturo Licea Pinzón

TUTORA

Dra. Katya Rodríguez Vázquez

México, Ciudad de México, 2020





Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

A mi papa:

Quien con su paciencia y entusiasmo me adentró en el hermoso mundo de las matemáticas; y siempre me apoyó y enseñó a hacer lo que me gusta.

Índice general

Introducción	VII
Clasificación del fenómeno P300	VII
Propósito de la tesis	IX
Estructura de la tesis	X
Agradecimientos	XI
1. Clasificación de señales electroencefalográficas (EEG)	1
1.1. Funcionamiento del cerebro y la actividad neuronal	2
1.1.1. Potenciales relacionados a eventos	5
1.1.2. BCI P300	5
1.2. Principio general del procesamiento de las señales EEG	6
1.3. Discriminate lineal de Fisher	8
1.4. Validación cruzada	13
1.5. Extracción de características	15
2. Algoritmos genéticos	19
2.1. Uso de los algoritmos genéticos	20

2.2.	Funcionamiento general de los algoritmos genéticos	21
2.3.	Los operadores esenciales	23
2.3.1.	Selección	23
2.3.2.	Cruzamiento	25
2.3.3.	Mutación	25
2.3.4.	Elitismo	26
2.3.5.	Escalamiento	27
3.	Diseño e implementación	29
3.1.	Generación y obtención de los datos	29
3.2.	Implementación del algoritmo	34
3.2.1.	Estructuras primarias	35
3.2.2.	Algoritmo genético	36
3.2.3.	Función objetivo	37
3.2.4.	Formato de tablas	38
3.3.	Descripción del algoritmo	39
3.3.1.	La codificación de los individuos	39
3.3.2.	Obtención de la primera población	40
3.3.3.	Función objetivo	41
3.3.4.	Apareamiento de las soluciones	43
3.3.5.	Operadores adicionales para mejorar el AG estándar	44
4.	Resultados	47
4.1.	Funciones objetivo exploradas	48

4.2. Comparación estadística de distintos métodos de selección de variables	53
4.2.1. Usando todas las variables	55
4.2.2. Método univariados	56
4.2.3. Comparación de resultados	58
Conclusiones	63
BIBLIOGRAFÍA	64

Introducción

Clasificación del fenómeno P300

Las interfaces cerebro-máquina consisten en dispositivos que hacen posible que las señales producidas por el cerebro habiliten una actividad externa. El principio básico de la interfaz cerebro-máquina es que se pueden detectar las señales del cerebro y éstas se pueden convertir en acciones por medio de una computadora. Por ejemplo, las señales que transmite el cerebro podrían permitir controlar la dirección de una silla para discapacitados.

Un electroencefalograma registra la actividad eléctrica de tu cerebro a través de electrodos que se fijan en el cuero cabelludo, los cuales se colocan sobre las zonas correspondientes a las diferentes áreas del cerebro para así detectar y registrar patrones de actividad eléctrica. Éstas son llamadas señales electroencefalográficas (EEG).

Un tipo de señal EEG muy relevante son los potenciales relacionados a eventos, los cuales son generadas por el cerebro en respuesta a un estímulo determinado. La señal P300 es una señal relacionada a eventos cuya respuesta en la señal

EEG se manifiesta aproximadamente 300 milisegundos después del estímulo. Para el diseño de una interfaz máquina-cerebro basada en el fenómeno P300, es necesario que las señales se puedan identificar correctamente. Para lo anterior, se usa normalmente un algoritmo clasificador que en la presente tesis está dado por el discriminante lineal de Fisher (FLDA¹).

Generalmente, se ha encontrado que se puede obtener una buena clasificación usando diferentes clasificadores, pero lo que realmente más importa es la correcta selección de las características² que describen las señales EEG.

En el presente trabajo se hace uso de los datos experimentales que fueron proporcionados por el Dr. Omar Montoya durante su doctorado en la Freie Universität Berlín. Se recolectaron las señales EEG de ocho sujetos durante un experimento para detectar las señales P300. Después del pre-procesamiento de las señales EEG, se redujo la información a un vector³ cuyas posiciones contenían las amplitudes de las señales de cada electrodo a través del tiempo. Estas observaciones fueron etiquetadas para su posterior clasificación como positivas, en caso que se presentara el fenómeno P300, y negativas, en caso de no presentarse.

Para la selección de características se usan normalmente algoritmos univariados, los cuales evalúan el poder discriminante de cada característica individualmente. El Dr. Omar Montoya propuso la exploración de métodos multivariados para la selección de las variables pues estos evalúan un subconjunto de carac-

¹Del inglés: Fisher Linear Discriminant Analysis.

²Es la información temporal (temporalidad de la señal después del estímulo) y espacial (electrodos específicos) que se obtiene del electroencefalograma

³Ver sección 3.1

terísticas juntas y mantienen el conjunto con las mejores características. Esto permite a estos algoritmos considerar el impacto de redundancias y complementariedades entre las características. La desventaja de este método es que lleva a un alto requerimiento computacional por lo que normalmente solo se puede evaluar un pequeño subconjunto de todas las posibles combinaciones.

De lo anteriormente expuesto, los algoritmos genéticos surgen como una solución natural a este problema pues permite explorar los múltiples subconjuntos de características de manera heurística con gran eficacia.

Propósito de la tesis

El propósito de la presente tesis es hacer uso de algoritmos genéticos para poder seleccionar las mejores características de la información pre-procesada de las señales EEG con el objetivo de mejorar la precisión en la clasificación de la actividad P300. Se hace una comparación entre los métodos de selección de variables univariados y los algoritmos genéticos en esta tesis implementados para evaluar la mejora en la clasificación de las señales EEG. Se encontró que los algoritmos genéticos (AG) resulta en una mejora estadística significativa respecto del uso de la totalidad de atributos y los métodos univariados. La mejora porcentual respecto del uso de la totalidad de variables es de hasta de 43 % y respecto a los métodos univariados de hasta 25 %.

Estructura de la tesis

Capítulo 1 Se da un panorama general del funcionamiento del cerebro y la actividad neuronal con el objetivo de entender el fenómeno P300, la cual será el tipo de actividad cerebral central en la presente tesis. Se expone de una manera general el discriminante lineal de Fisher el cual se utilizará para la clasificación de las señales; así como la técnica de la validación cruzada cuyos errores de clasificación serán indispensables para determinar la variable objetivo del AG.

Capítulo 2 Se expone de manera general el funcionamiento de los algoritmos genéticos poniendo énfasis en las características que se usarán para implementarse en la presente investigación.

Capítulo 3 Se expone la obtención de los datos que se usarán para la clasificación que se pretende optimizar. Dado que los datos para el presente trabajo fueron recolectados experimentalmente y pre-procesados por una fuente externa, se dará solo un panorama general de la metodología del experimento y obtención de los datos. Por otra parte, se describirá de manera detallada la implementación del AG diseñado.

Capítulo 4 Se presentan los resultados del uso del AG implementado. Se hace una comparación entre los métodos de selección de variables univariados y los algoritmos genéticos para evaluar la mejora en la clasificación del fenómeno P300. Finalmente, se hace una comparación de diversas funciones objetivo en el AG.

Agradecimientos

Quiero hacer un especial reconocimiento al Dr. Omar Mendoza quien fue la persona que me introdujo al maravilloso estudio de las interfaces máquina-cerebro y quien me proporcionó los datos experimentales que hicieron posible el presente trabajo. Igualmente quiero agradecer a la Dra. Katya Rodríguez quien me ha dado su apoyo, conocimiento y paciencia para la realización de esta tesis y quien me mostró las infinitas posibilidades de los algoritmos genéticos.

Capítulo 1

Clasificación de señales electroencefalográficas (EEG)

En el presente capítulo se da un panorama general del funcionamiento del cerebro y la actividad neuronal para poder entender el fenómeno P300 que será la actividad cerebral que se busca clasificar de una manera óptima. Se expondrá de manera general el discriminante lineal de Fisher el cual se utilizará para clasificar las señales EEG. Finalmente se describirá la técnica de la validación cruzada cuyos errores de clasificación los cuales serán usados para hacer la validación del algoritmo genético a implementar.

1.1. Funcionamiento del cerebro y la actividad neuronal

El cerebro es el encargado de controlar y regular las funciones del cuerpo. Este órgano está formado por millones de células nerviosas que responden a diferentes estímulos que se envían desde el organismo y su exterior. Está relacionado con los sentidos, las emociones, los recuerdos, las reacciones. Se encarga de recibir señales y devolverlas en forma de respuesta.

La parte exterior del cerebro se llama corteza cerebral la cual se divide en dos hemisferios (izquierdo y derecho) y los cuales a su vez se dividen en cuatro lóbulos.

El cerebro se compone de millones de células especializadas llamadas neuronas que se caracterizan por su excitabilidad eléctrica y su habilidad de mandar señales eléctricas entre ellas.

El potencial eléctrico de una neurona es cambiado por iones que pasan de adentro hacia afuera de la membrana celular, fenómeno que se conoce como potencial de acción (ver figura 1.1¹). Este fenómeno es básicamente un cambio muy rápido en la polaridad de la membrana de negativo a positivo y vuelta a negativo, en un ciclo que dura unos milisegundos. Cada ciclo comprende una fase ascendente, una fase descendente y por último una fase hiperpolarizada. Cuando el potencial eléctrico alcanza un cierto umbral, un potencial viaja hacia lo largo del axón. Una vez que el potencial alcanza el final del axón, se manda una señal a

1

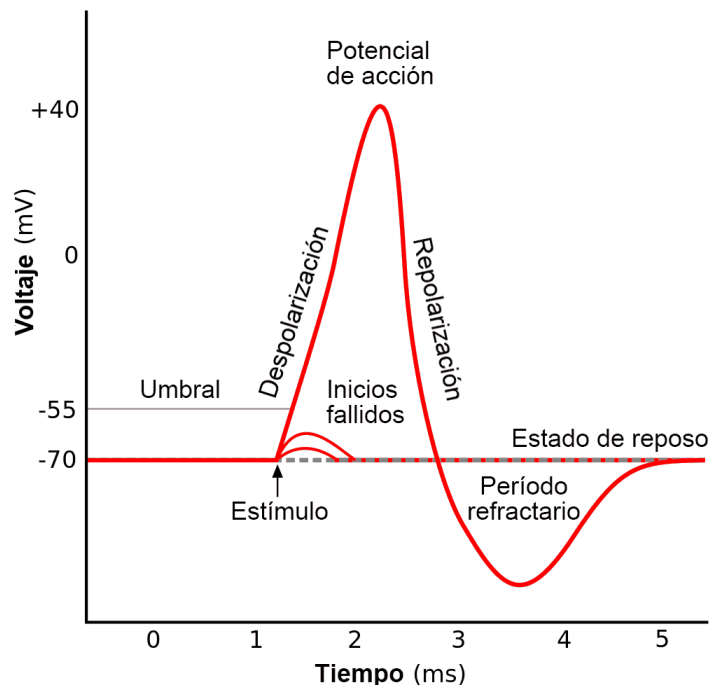


Figura 1.1: Potencial de acción de una neurona. [Fuente: <https://mejorconsalud.com/potencialdeaccion>]

través del espacio sináptico² a las dendritas de otra neurona a través de un proceso químico. La corteza cerebral es la estructura que más genera potencial eléctrico medible desde el cráneo.

Una de las formas de medir la actividad neuronal es la técnica de la electroencefalografía. Esta es un método no invasivo para grabar las señales eléctricas del cerebro por medio de poner varios electrodos (de 32 a 128) sobre el cráneo del individuo a estudiar.

Cuando el potencial eléctrico cambia debido a una oscilación neuronal, se pue-

²El espacio sináptico o hendidura sináptica es el pequeño espacio que existe entre el final de una neurona y el inicio de otra. Se trata de un espacio extracelular de entre 20 a 40 nanómetros

de detectar por medio de electrodos. Este cambio de voltaje causa el movimiento de las cargas que se encuentran en el metal con el que se construyen los electrodos, las cuales pueden ser grabadas. Esta actividad eléctrica se puede dividir en dos grandes categorías:

- Potenciales espontáneos: Se generan según el estado de excitación del sujeto, tales como los ritmos al dormir.
- Potenciales relacionados a eventos. Son generados por un estímulo externo tales como la luz o algún sonido.

Una interfaz cerebro computadora (BCI³) es un sistema de comunicación y control que crea un canal de salida no muscular para el cerebro. Las intenciones del usuario son transportadas por las señales del cerebro en lugar de los nervios y músculos. Estas señales no dependen de la actividad neuromuscular. Como cualquier otro sistema de control, la BCI establece una interacción de tiempo real entre el usuario y el mundo exterior [7]. El usuario codifica sus intenciones en señales cerebrales que la BCI detecta, analiza y traduce en comandos que serán ejecutados. Por ejemplo, si una persona usa la BCI para controlar los movimientos de un brazo robótico. Los campos eléctricos producidos por la actividad cerebral pueden ser recolectados del cráneo por medio del electroencefalograma. Cualquier cambio en el voltaje grabado puede constituir una variable o característica útil para la BCI. Mientras que el número de variables que provienen de las señales cerebrales son muchas, solo algunas han sido probadas para ese propósito [7].

³Del inglés: *Brain Computer Interface*.

1.1.1. Potenciales relacionados a eventos

La actividad cerebral se puede manifestar dado un estímulo sensorial lo cual se podrá detectar en el electroencefalograma en un periodo de tiempo acotado. Estos potenciales relacionados a un evento se podrán distinguir en diversas localizaciones del cerebro y a determinadas frecuencias. Mientras que muchas de estas señales pueden ser usadas para BCI, solo algunos tipos de estímulos han podido ser caracterizados correctamente. El que en mayor medida se ha estudiado es el potencial evocado por un estímulo visual. Este ocurre en los primeros 100 milisegundos después del estímulo visual; la frecuencia y las polaridades en las señales EEG se pueden identificar. Se presenta un componente negativo inicial a los 75 milisegundos seguido por uno positivo a los 100 milisegundos, seguido por otro negativo a los 145 milisegundos.

El potencial generado depende de las propiedades del estímulo visual, por ejemplo, diferentes frecuencias de flash. De esta manera una BCI diseñada con este método consistiría en que una persona vea diferentes objetivos (v. g. en una pantalla de ordenador) con diferentes frecuencias de flasheo y después de analizadas los potenciales emitidos, se pueda determinar cuál objetivo está viendo el sujeto.

1.1.2. BCI P300

Un estímulo visual o auditivo que está interespaciado por una frecuencia evoca normalmente un pico positivo aproximadamente 300 ms después de su aparición

en el centro parietal [7]. Este tipo de estímulo requiere que el sujeto ponga atención al objetivo. En el sistema BCI P300, al sujeto se le presenta con una serie de objetivos con diferentes tipos de estímulos (visuales por poner un ejemplo), cada uno de los cuales generará una respuesta de salida (v. g. el oprimir determinada tecla), lo cual dependerá de cuál estímulo el sujeto preste atención. La BCI reconocerá la señal P300 y ejecutará el comando especificado por el estímulo elegido. Esto solo requiere que el usuario module su atención. La BCI P300 tiene varias ventajas como el ser detectable en la mayoría de los usuarios y de relativa rapidez lo cual hace que sea un medio de comunicación rápido; además la calibración inicial requiere menos de una hora y se necesita poco entrenamiento por parte del sujeto [7].

1.2. Principio general del procesamiento de las señales EEG

Cuando se diseña una BCI, el procesamiento de las señales EEG tiene como objetivo traducir las señales EEG crudas en la clasificación de esas señales en un estado mental del usuario; que en el caso de la presente tesis, se clasificará la presencia del fenómeno P300 en las señales EEG. Esta traducción se logra usando técnicas de reconocimiento de patrones[12], la cual tiene los dos siguientes pasos, los cuales se esquematizan en la figura 1.2.

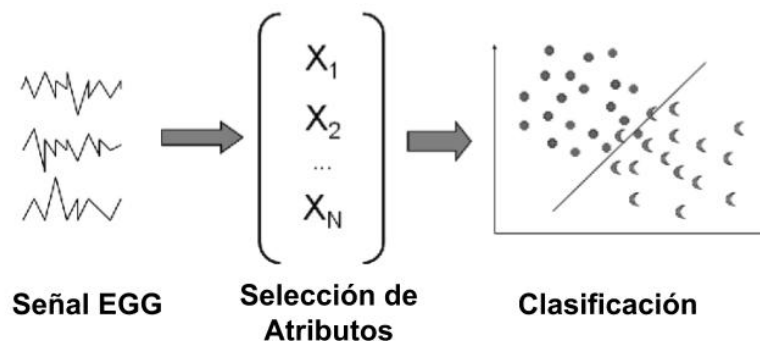


Figura 1.2: Diagrama del procesamiento de señales. [Fuente: Elaborada por el autor]

Extracción de Atributos. Este primer paso de procesamiento de la señal EEG tiene como objetivo describir las señales con pocos valores relevantes conocidas como *características*. Estas deberían captar la información metida en las señales EEG que es relevante para describir el estado mental a identificar mientras que se reduce el ruido y otra información no relevante. Estas *características* son arregladas en un vector conocido como vector de *características*.

Clasificación. Este segundo paso asigna un vector de *características* extraído de las señales EEG a una determinada clase que corresponde al estado mental a identificar (presencia de fenómeno P300); este paso se conoce también como la traducción de *características*. Para hacer la clasificación se usan algoritmos de clasificación entre los cuales se destaca el uso del discriminante lineal de Fisher (FLDA⁴) el cual se explicará más adelante.

Las señales EEG se procesan usando aprendizaje de maquina, por lo que las

⁴Del inglés: *Fisher Linear Discriminant Analysis*.

características son automáticamente clasificadas de acuerdo a observaciones de señales EEG del sujeto tomadas previamente. A este conjunto de señales les llamamos datos de entrenamiento y están etiquetadas según la clase a la que corresponda.

Basadas en estas muestras de entrenamiento el clasificador podrá reconocer apropiadamente la clase de las señales EEG. A esta etapa se llama de calibración y en la cual también se seleccionan *características* óptimas en cuanto a frecuencias y canales a considerar para una óptima clasificación [12].

1.3. Discriminate lineal de Fisher

Como se mencionó antes, la clasificación en las BCI tiene como objetivo traducir el vector de *características* en comandos. Para hacer esto se utiliza principalmente el FLDA como algoritmo de clasificación. Este clasificador es capaz de identificar la clase a la que pertenece un vector de *características* gracias a un conjunto de muestras de entrenamiento etiquetados extraídas de las señales EEG.

Para aprender a clasificar, los algoritmos de clasificación tratan de modelar los límites de las áreas cubiertas por los vectores de *características* de cada clase con ayuda de las muestras de entrenamiento.

En específico, el FLDA usa hiperplanos para separar los vectores de entrenamiento que representan las diferentes clases. En la etapa de prueba el vector de *características* se clasificará dependiendo en qué lado del hiperplano haya quedado. El algoritmo FLDA tiene requerimientos computacionales muy bajos lo cual

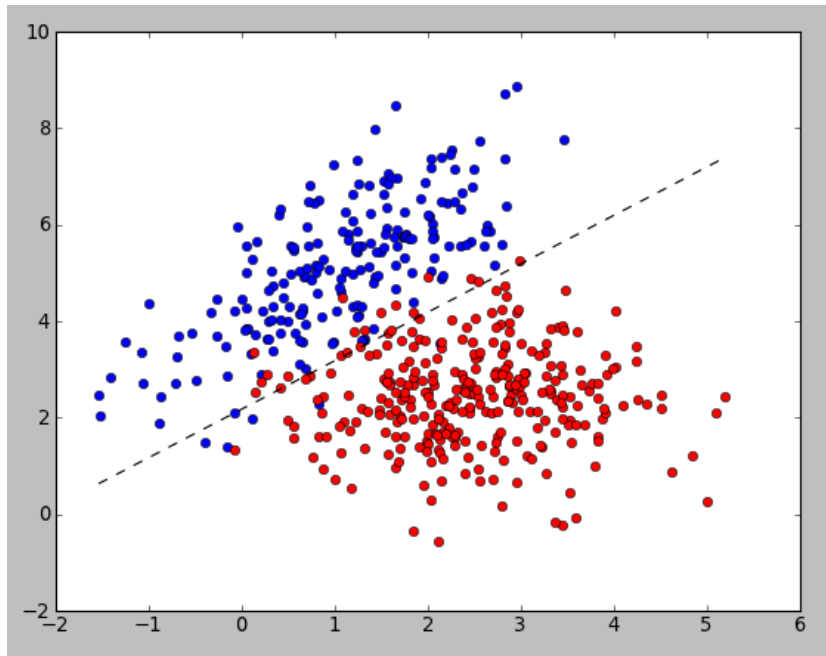


Figura 1.3: Discriminante Lineal. [Fuente: [1]]

lo hace ideal para BCI. Además, este algoritmo es simple y eficaz por lo que es el clasificador más usado para el diseño de BCI.

Un discriminante es una función que toma un vector como entrada y lo asigna a una determinada clase. Cuando es lineal la superficie de decisión es un hiperplano. La representación de un discriminante lineal tiene la siguiente forma: $y(x) = \vec{w}^T \vec{x} + w_0$ donde \vec{w} es un vector de coeficientes y \vec{x} es un vector de variables explicativas [1]. En la figura 1.3 se muestra un ejemplo de este tipo de discriminante, donde aparecen un conjunto de observaciones correspondiendo a dos clases (representadas por puntos rojos y azules respectivamente) que pueden ser separadas de una manera razonable por una línea recta.

Esta metodología se basa en la proyección de puntos en una dimensión infe-

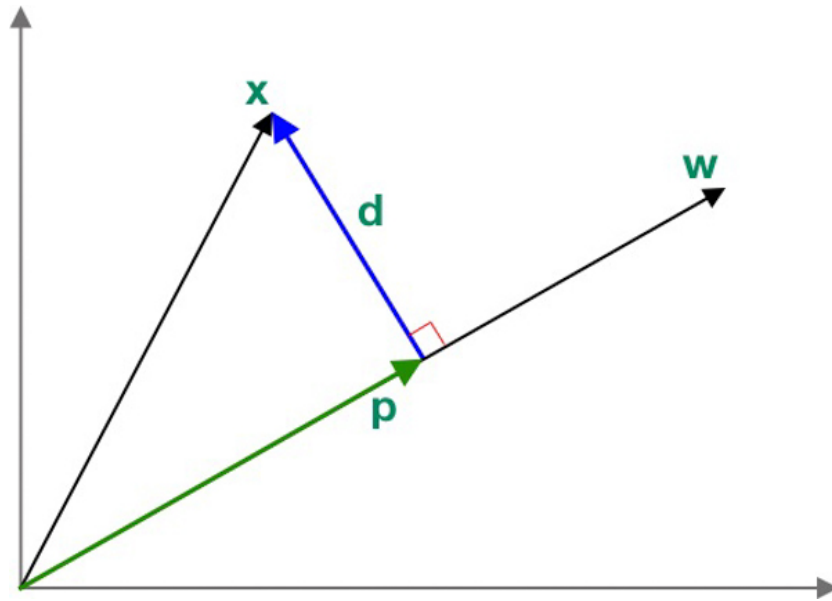


Figura 1.4: Projection ortogonal de un punto x sobre una línea w . [Fuente: [1]]

rior. En el caso de la figura 1.3, la proyección se hace en la línea punteada. Lo anterior se puede hacer para cada punto como se representa en la figura 1.4.

Una vez que los puntos son proyectados se pueden describir la dispersión de los puntos según su distribución. Lo anterior se muestra en la figura 1.5 donde se puede observar que dependiendo la línea que se usa para proyectar la distribución cambia y las clases se pueden diferenciar más claramente; en la Figura de la derecha las distribuciones se entrelapan mientras que en la derecha están completamente separadas.

De lo anterior podemos intuir que una forma de ver a un clasificador lineal es en términos de reducción de la dimensión. Para ver esto, consideremos el caso de dos clases, tomemos un vector \vec{x} y proyectemoslo en un vector \vec{w} usando la siguiente expresión.

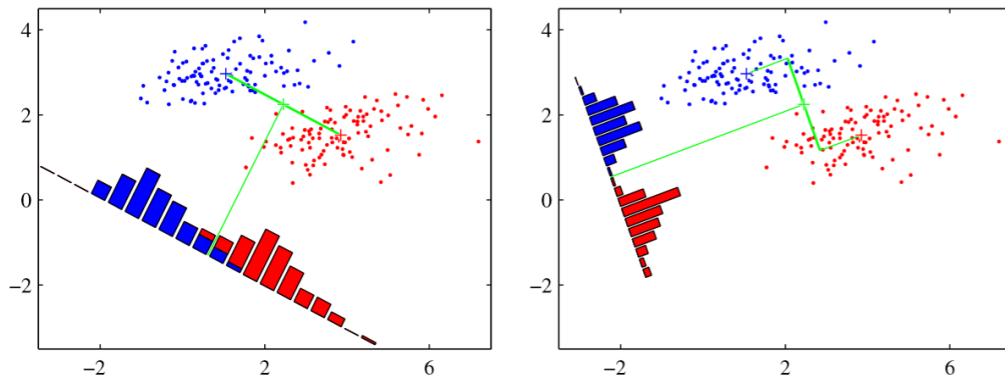


Figura 1.5: Distribución de la proyección ortogonal sobre dos diferentes líneas. [Fuente: [1]]

$$y = \vec{w}^T \vec{x}$$

Si pusieramos un limite y clasificamos $y > w_0$ como perteneciente a la clase C_1 y como C_2 en otro caso, tendríamos un clasificador lineal. La proyección explicada anterioremente provoca perdida de información pues clases que están muy bien separadas en un D-espacio puede ser traslapadas en una dimensión. Pero ajustando los componentes del vector de pesos \vec{w} podemos seleccionar una proyección que maximiza la separación de las medias de las clases mientras que minimiza la varianza dentro de las clases lo que minimizaría el traslape entre las clases [1]. Considere el problema de dos clases, en el cual hay N_1 puntos de la clase C_1 y N_2 puntos de la clase C_2 , en el que los vectores de las medias de las clases $(m_1 + m_2)$ están dados por las expresiones:

$$m_1 = \frac{1}{N_1} \sum_{n \in C_1} x_n$$

$$m_2 = \frac{1}{N_2} \sum_{n \in C_2} x_n$$

donde x_n son los puntos que se quieren clasificar; y la varianza dentro de cada clase (s_1^2 y s_2^2) está dada por las expresiones:

$$s_1^2 = \sum_{n \in C_1} (y_n - m_1)^2$$

$$s_2^2 = \sum_{n \in C_2} (y_n - m_2)^2$$

donde $y_n = \vec{w}^T x_n$.

El criterio de Fisher se define como la razón de la varianza entre las clases $(m_1 - m_2)^2$ y la varianza entre las clases ($s_1^2 + s_2^2$) de la siguiente forma:

$$J(w) = \frac{(m_1 - m_2)^2}{s_1^2 + s_2^2}$$

Lo anterior se puede reescribir de manera que quede explícitamente dependiente del vector \vec{w} de la siguiente forma:

$$J(\vec{w}) = \frac{\vec{w}^T S_B \vec{w}}{\vec{w}^T S_W \vec{w}} \quad (1.1)$$

donde:

$$S_B = (m_2 - m_1)(x_2 - m_1)$$

y

$$S_W = \sum_{n \in C_1} (x_n - m_1)(x_n - m_1)^T + \sum_{n \in C_2} (x_n - m_2)(x_n - m_2)^T$$

Diferenciando la ecuación 1.1 respecto de \vec{w} y despejando podemos encontrar que $J(\vec{w})$ queda maximizada cuando:

$$\vec{w} \propto S_w^{-1}(m_2 - m_1) \quad (1.2)$$

La ecuación 1.2 se conoce como el discriminante de Fisher [1] aunque estrictamente no sea un discriminante sino que especifique la dirección normal del hiperplano a proyectar los datos. A partir de esta expresión se construye el discriminante dando un punto de corte y_0 de forma que un vector será clasificado en la clase C_1 si $y(x) > \vec{w}^T x$ o en la C_2 en otro caso. Este punto y_0 será determinado donde haya el mínimo traslape entre las distribuciones de ambas clases proyectadas.

1.4. Validación cruzada

La validación cruzada es una técnica utilizada para evaluar los resultados de un análisis estadístico y garantizar que son independientes de la partición que se haga de los datos de entrenamiento y prueba. Consiste en repetir y calcular la media aritmética de las medidas de evaluación sobre diferentes particiones. Se utiliza en entornos donde el objetivo principal es la predicción y se quiere estimar la precisión de un modelo.

Existen diversas formas de realizar esta validación cruzada que dependen de

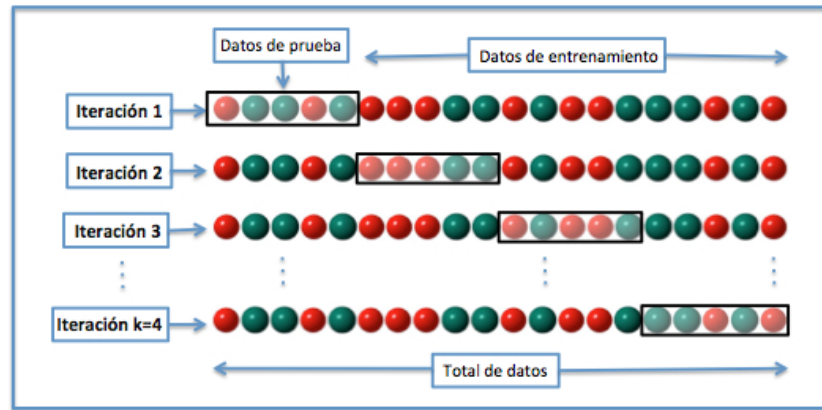


Figura 1.6: Diagrama de validación cruzada de k iteraciones. [Fuente: <https://commons.wikimedia.org/wiki/File:K-foldcrossvalidation.jpg>]

cómo se toman las diversas particiones. La que se utilizará en el presente trabajo se denomina validación cruzada de k iteraciones en la cual los datos de muestra se dividen en k subconjuntos. Uno de los subconjuntos se utiliza como datos de prueba y el resto ($k - 1$) como datos de entrenamiento. El proceso de validación cruzada es repetido durante k iteraciones, con cada uno de los posibles subconjuntos de datos de prueba. En la figura 1.6 se ejemplifica lo anterior tomando $k = 4$. En ésta se ponen en rectángulos los elementos que se tomarán como datos de validación, los cuales corresponden a las cuatro particiones que se harán para cada iteración.

Finalmente se calcula la media aritmética de los resultados de cada iteración para obtener un único resultado[9]. La evaluación del modelo usando la validación cruzada viene dada por el error obtenido en cada iteración. En cada una de las k iteraciones se realiza el cálculo de error, obteniendo el resultado final a partir de realizar la media aritmética de los k errores obtenidos.

1.5. Extracción de características

Llamamos *características* a la información temporal (temporalidad de la señal después del estímulo) y espacial (electrodos específicos) que se obtiene del electroencefalograma.

Hay tres fuentes de información que pueden ser usadas para extraer *características* de las señales EEG.

- Información espacial. Describe el lugar espacial de donde la señal EEG proviene. Se seleccionan canales específicos de lugares relevantes del cerebro.
- Información espectral. Esto consiste en describir sólo bandas de frecuencia relevantes de las señales EEG.
- Información temporal: Tales *características* describen qué tanto varía la señal respecto del tiempo. Esto se traduce en tomar las medidas en diferentes tiempos y con diferentes separaciones temporales.

Como se ha mencionado antes, la extracción de *características* tiene como objetivo representar las señales EEG crudas en un reducido número de valores relevantes, los cuales describen la información contenida en las señales.

Una primera forma de descartar *características* es usando el conocimiento que se tiene del funcionamiento del cerebro. Siguiendo esta idea; los potenciales relacionados a eventos como el P300, son respuestas a determinado estímulo que se presenta en un periodo de tiempo después de presentado éste. Por lo anterior,

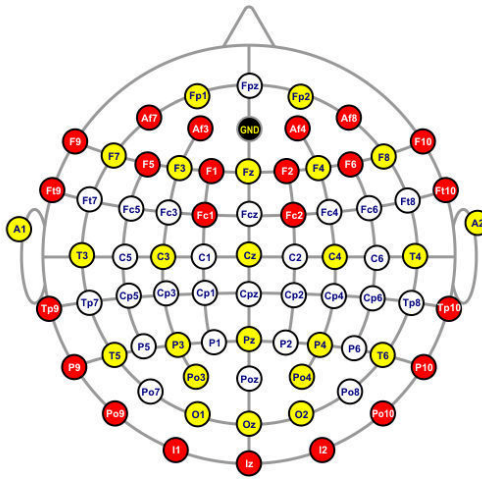


Figura 1.7: Posiciones de electrodos en electroencefalograma. [Fuente: <https://www.medicaexpo.es/prod/medical-computer-systems/product-69255-565055.html>]

resulta muy importante la variación temporal para la clasificación de este tipo de señales.

En lo que se refiere a la información espacial, las BCI basadas en P300 se enfocan más en los electrodos localizados sobre el lóbulo parietal donde se sabe que las señales P300 se originan. Por ejemplo, se recomienda el uso de un conjunto de ocho canales en las posiciones: Fz, Cz, P3, Pz, P4, PO7, Oz, PO8 [11]. En la figura 1.7 se muestran los electrodos usados con más frecuencia en los estudios de electroencefalograma para este tipo de fenómeno.

Se ha visto, que en general se puede obtener una buena clasificación usando diferentes clasificadores, pero lo que realmente parece más importante es la correcta selección de las *características* que describen las señales EEG [11]. Para la selección de *características* se emplean algoritmos usados ampliamente en

aprendizaje de máquina. Hay dos principales familias de metodologías:

- Algoritmos univariados: Tratan de evaluar el poder discriminante de cada *característica* individualmente. Se seleccionan las N mejores *características* individuales⁵, según alguna métrica que se pueda asociar a cada una, sin tomar en cuenta su interacción con las demás.

La utilidad de cada *característica* es evaluada por alguna métrica como la ganancia de información, la cual mide la diferencia del valor de la *característica* entre las dos clases y la correlación que tiene la *característica* dependiendo de la clase a la que pertenece [8]. Estos métodos son muy rápidos y computacionalmente eficientes, pero normalmente obtienen una precisión en la clasificación significativamente por debajo de la máxima posible. Esto es porque solo consideran la utilidad de una *característica* individualmente e ignoran las posibles redundancias y complementariedad entre distintas *características*. Esto puede ser porque las N mejores *características* sean altamente redundantes y medir la misma información y, por tanto, no aportar más poder discriminante al clasificador. Además, añadiendo otras *características* que individualmente no sean muy buenas pero, que midan diferente información de las que captan las N mejores, podría mejorar significativamente el poder discriminante.

- Algoritmos multivariados. Estos evalúan un subconjunto de *características* juntas y mantienen el conjunto con las mejores N *características*. Estos algo-

⁵ N se define por el diseñador de la BCI.

ritmos típicamente usan una medida del desempeño global del conjunto de *características* como, por ejemplo, el desempeño de clasificación en los datos de entrenamiento usando por lo general validación cruzada. Lo anterior, permite a estos algoritmos considerar el impacto de redundancias y complementariedades entre las *características*[8]. La desventaja de este método es que lleva a un alto requerimiento computacional. De hecho, la mayoría de las veces solo se puede evaluar un pequeño subconjunto las posibles combinaciones y, por tanto, en la práctica se hace uso de métodos heurísticos para reducir el número de conjuntos a evaluar.

Capítulo 2

Algoritmos genéticos

Según vimos en el capítulo anterior, para clasificar las señales EEG se necesita seleccionar de manera óptima las características a tomar en cuenta por el algoritmo. Para esto, se propone usar algoritmos multivariados pues estos permiten considerar el impacto de redundancias y complementariedades entre las características. En particular, los algoritmos genéticos (AG) surgen como una solución a este problema pues permiten explorar los múltiples subconjuntos de características de manera heurística con gran eficacia.

En el presente capítulo se expone el funcionamiento de los AG, poniendo énfasis, en las características que se usarán en su implementación en las secciones posteriores.

2.1. Uso de los algoritmos genéticos

Los AG son algoritmos de optimización inspirados por la selección natural y la genética. Este método es muy robusto y es capaz de ser aplicado a una variedad de problemas. A muy grandes rasgos un AG consiste en lo siguiente [3]:

- Un número (población) de candidatos de la solución del problema.
- Una forma de calcular que tan buena o mala es cada solución candidata.
- Un método para mezclar los fragmentos de las mejores soluciones para formar nuevas y, de ser posible, mejores soluciones.
- Un operador de mutación que evite que se pierda la diversidad entre las soluciones.

En optimización matemática se denomina *espacio de búsqueda* al conjunto de todos los posibles puntos que satisfacen las condiciones de la solución; esto es, son las soluciones candidatas al problema.

Los candidatos de solución son representados en los AG como un conjunto de *Cromosomas*. Estos típicamente toman la forma de cadenas de bits donde, cada posición en el *Cromosoma*, puede ser 1 o 0. Cada *Cromosoma* puede ser pensado como un punto en el *espacio de búsqueda* de las soluciones candidatas.

El AG requiere una función de *aptitud* que asigna un *puntaje* a cada *Cromosoma* de la población. Esta *aptitud* depende de que tan bien este *Cromosoma* resuelve el problema que se ha planteado; por ejemplo, en el caso del problema

de optimizar el máximo de una función, se daría una mayor *aptitud* a la solución que tenga una evaluación mayor. En este problema, las soluciones candidatas son valores que podrían codificar como cadenas de bits que podrían traducirse en un número real y después evaluarse en la función. La *aptitud* sería el valor de la función en ese punto.

Los AGs son inicializados en una población de candidatos que son usualmente esparcidos en todo el *espacio de búsqueda*. Después, el AG usa los tres operadores de selección, cruza y mutación para dirigir a la población hacia la convergencia de un óptimo global [3].

2.2. Funcionamiento general de los algoritmos genéticos

Un AG puede presentar diversas variaciones dependiendo de cómo se aplican los operadores genéticos, cómo se realiza la selección y cómo se decide el reemplazo de los individuos para formar la nueva población. El AG consiste de los siguientes pasos [3]:

- **Inicialización:** Se genera aleatoriamente la población inicial, la cual está constituida por un conjunto de *Cromosomas* que representan las posibles soluciones del problema.
- **Evaluación:** A cada uno de los *Cromosomas* de esta población se aplicará la función de aptitud para saber qué tan buena es la solución que se está

codificando.

- El AG se deberá detener cuando se alcance la solución óptima, se llegue a un número máximo de iteraciones (generaciones) o no haya cambios según alguna métrica que se haya establecido. Mientras no se cumpla la condición de término se hace lo siguiente:
 - Selección: Después de saber la aptitud de cada *Cromosoma* se procede a elegir los *Cromosomas* que serán cruzados en la siguiente generación. Los *Cromosomas* con mejor aptitud tienen mayor probabilidad de ser seleccionados.
 - Cruzamiento: Opera sobre dos *Cromosomas* a la vez para generar dos descendientes que combinan las características de ambos *Cromosomas* padres.
 - Mutación: Modifica al azar parte del *Cromosoma* de los individuos, lo cual permite alcanzar zonas del *espacio de búsqueda* que no estaban cubiertas por los individuos de la población actual.
 - Reemplazo: Una vez aplicados los operadores genéticos, se seleccionan los mejores individuos para conformar la población de la generación siguiente.

Los anteriores pasos se esquematizan en la figura 2.1, donde se muestran los pasos de manera cíclica (*i*: inicialización, $f(X)$: evaluación, $?$: condición de término, *Se*: selección, *Cr*: cruzamiento, *Mu*: mutación, *Re*: reemplazo, X^* : mejor solución)

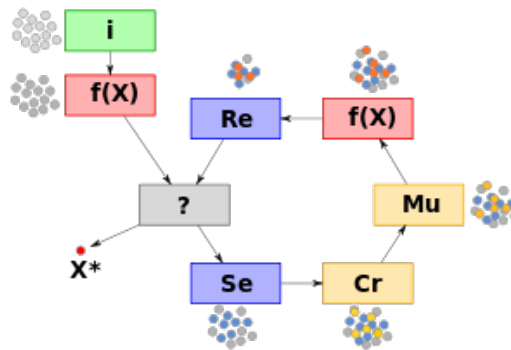


Figura 2.1: Diagrama de funcionamiento del AG. [Fuente: <https://es.wikipedia.org/wiki/Algoritmogenetico>]

2.3. Los operadores esenciales

En esta sección se profundizará más en los operadores que se mencionaron en la sección anterior, lo cual nos ayudará a entender los diferentes parámetros que se pueden ajustar en un AG cuando se aplica a un determinado problema en particular.

2.3.1. Selección

Este operador selecciona los *Cromosomas* en la población que se reproducirá, de manera que se escogen con mayor probabilidad los *Cromosomas* que tengan una mejor *aptitud*. Este operador tiene la intención de aplicar presión sobre la población de la misma manera que la selección natural en los sistemas biológicos. Entre mejor desempeño tenga un individuo, tendrá más oportunidad de pasar su información que contienen a la siguiente generación.

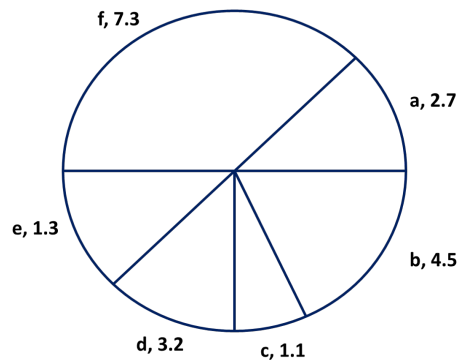


Figura 2.2: Ruleta de selección: [Fuente: Elaborada por el autor]

Con este método, la probabilidad de selección es proporcional a la *aptitud* de cada individuo. Se puede imaginar como una ruleta, siendo representado cada individuo con el tamaño de la ranura, la cual será proporcional a su *aptitud*. Esto se ilustra en la figura 2.2 donde se toma una población de seis individuos (a, b, c, d, e y f) con *aptitud* 2.7, 4.5, 1.1, 3.2, 1.3 y 7.3 respectivamente.

Además de ser ese método visualmente muy intuitivo, la implementación de esta estrategia resulta bastante simple:

- Se suma la *aptitud* de toda la población. Llamamos a esta suma: f_{sum} .
- Se escoge de manera aleatoria un número R_s entre 0 y f_{sum} .
- Se van sumando la *aptitud* de los individuos en un orden definido, deteniéndose inmediatamente cuando la suma es mayor a R_s .
- El último individuo que se ocupó en la suma, es el individuo seleccionado.

Este método de selección se aplica dos veces con el objetivo de seleccionar

una pareja de individuos para realizar la cruce. Este procedimiento de selección se continúa, hasta que se hayan obtenido tantos hijos como la población original.

2.3.2. Cruzamiento

Este operador escoge aleatoriamente una posición del *Cromosoma* e intercambia la secuencia siguiente y anterior entre los dos *Cromosomas* creando dos hijos. Por ejemplo, las cadenas 10000100 y 11111111, cruzadas en la tercera posición, producirían los hijos 10011111 y 11100100. La cruce tiene la función de intercambiar información de manera similar a la que usan los organismos biológicos en la reproducción sexual. El operador de cruce provee un método donde la información de diferentes soluciones se puede mezclar para permitir la exploración de nuevas partes del *espacio de búsqueda*.

Normalmente la cruce de los individuos seleccionadas solo se efectuará con una probabilidad P_c . Para hacer esto se escoge un número aleatorio R_c entre 0 a 1 y la cruce se efectúa sólo si $R_c \leq P_c$, de otra manera la cruce no se efectúa y los padres pasan intactos a la siguiente generación como hijos. Los valores típicos de este parámetro están entre 0.5 y 0.9 de manera que, en promedio, de un 10% a un 50% de los padres pasarán a la siguiente generación sin cruce.

2.3.3. Mutación

En el mundo natural, muchos procesos pueden causar mutación como el simple error en la duplicación de información genética en los seres vivos. Este ope-

rador cambia aleatoriamente algunos bits del *Cromosoma*. La mutación puede suceder en cada posición del *Cromosoma* con una probabilidad normalmente muy pequeña, siendo una típica probabilidad de mutación (P_m), del orden de 0.001. Sin embargo, la correcta configuración de la probabilidad de mutación dependerá del problema [3]. Normalmente es usado $P_m = 1/L$ o $P_m = N/L$ donde L es la longitud del *Cromosoma* y N es el tamaño de la población.

2.3.4. Elitismo

La selección basada proporcionalmente en la *aptitud* no siempre garantiza la selección de un individuo en particular, ni siquiera el que mejor solucione el problema. Si un individuo no se selecciona, éste desaparece. En otras palabras, con el método de selección proporcional según la *aptitud*, de manera regular se puede tirar la mejor solución obtenida hasta el momento. Esto no siempre es una desventaja pues hay veces que esto puede hacer que la convergencia sea más lenta, pero permitiendo explorar más en el *espacio de búsqueda*.

Sin embargo, también esto puede provocar que después de más exploración se haya perdido la posibilidad de encontrar algún individuo más apto. De esta manera el elitismo consiste en que el algoritmo siempre asegure seleccionar a los N individuos con mayor *aptitud* para la siguiente generación, lo cual se haría sin ningún tipo de mutación o cruzamiento.

2.3.5. Escalamiento

Es importante darnos cuenta que la velocidad de convergencia hacia la solución óptima en los AG no es constante en todas las etapas del algoritmo. En las etapas tempranas, un individuo con una *aptitud* particularmente alta, puede permitir que varias copias del mismo inunden rápidamente las subsecuentes generaciones. Aunque esto aumentaría rápidamente la convergencia, ésta podría ser errónea pues podría haberse caído solo en un óptimo local [3]. Por otra parte, en las etapas tardías del algoritmo, cuando ya hay individuos muy similares en *aptitud*, el método de selección proporcional a la *aptitud* escogería aproximadamente con igual probabilidad a todos los individuos, por lo que habría poca presión para distinguir individuos con buena *aptitud* de otros con mejor *aptitud*. Para corregir lo anterior, se necesita un método que permita evitar que individuos buenos se quiten en las primeras iteraciones del algoritmo y, que el mismo tiempo, resalte las pequeñas diferencias en las etapas más avanzadas.

Esto puede hacerse con un mecanismo llamado *escalamiento lineal*; el cual, funciona ajustando la *aptitud* de la población tomando como referencia la *aptitud* promedio de la misma. Esto permite que una proporción constante de los mejores individuos, comparados con el promedio, sean seleccionados. Un valor típico de esta constante estaría entre 1 y 2. Por ejemplo, si la constante fuera 2, aproximadamente 2 veces más se escogerían los mejores individuos a la siguiente generación que los individuos promedio.

Capítulo 3

Diseño e implementación

En el capítulo anterior se describió el funcionamiento de los algoritmos genéticos y los principales parámetros que los componen. En el presente capítulo se describirá la metodología de recolección y pre-procesamiento de los datos de entrada que serán utilizados por el clasificador de señales EEG. Posteriormente, se describirá de manera detallada la implementación del algoritmo genético diseñado para seleccionar las variables para la clasificación.

3.1. Generación y obtención de los datos

La información que se ocupó en el presente trabajo fue proporcionada por el Investigador Omar Mendoza mientras realizaba su doctorado en la Freie Universität Berlín. Dado que los datos para el presente trabajo fueron recolectados experimentalmente y pre procesados por el referido investigador, en la presente

sección solo expondré a grandes rasgos la metodología del experimento y obtención de los datos.

Para el presente trabajo se usó la información de ocho sujetos adultos. Las señales EEG fueron grabados con un gorro para electroencefalograma que contenía 16 electrodos (C3, C4, Cz, F3, F4, Fp1, Fp2, Fz, Oz, P3, P4, PO7, PO8, Pz, T7, T8). Las señales EEG fue amplificada y digitalizada.

El siguiente protocolo fue utilizado en cada una de las ejecuciones del experimento.

- El sujeto se sentó frente a una computadora y se le presentaron dos cuadros que destelleaban de forma independiente. Cada uno de los destellos sucedían de forma aleatoria pero con un tiempo de entre cada estímulo suficiente para que la señal de P300 fuera claramente identificable.
- Se pidió al sujeto que contara en silencio el número de veces que proyectaba el cuadro que se le indicaba.
- Cuando se indicaba el cuadro indicado para poner su atención se comenzaba a registrar las señales EEG.
- Cada ejecución consistió en un promedio de 35 ensayos positivos (presenta P300) y 35 ensayos de negativos (no presenta P300).

Del procedimiento anterior se obtuvieron un total de 70 muestras de señales EEG, de las cuales, 35 casos son positivos y 35 negativos. A estas señales se le aplicó el siguiente procesamiento de datos:

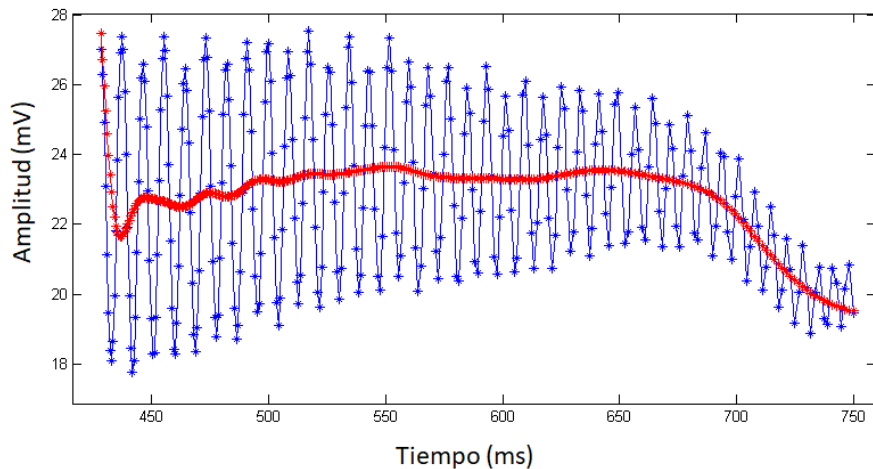


Figura 3.1: Efecto de un filtro de paso bajo en una señal. [Fuente: <https://stackoverflow.com/questions/14088658/matlab-low-pass-filter-output-initial-end-value>]

- Filtro: Se aplicó un filtro de paso bajo que elimina las frecuencias bajas de la señal grabada. Con esto se trata quitar el ruido de las señales EEG haciendo una curva suave. En la figura 3.1 se ejemplifica el efecto que tiene un filtro de paso bajo en una señal, donde se aprecia que se eliminan las altas frecuencias de la señal, lo cual resulta en una curva que capta la información general de la señal EEG.
- Submuestreo. Este procesamiento tiene como objetivo bajar la resolución de la señal guardada y así reducir la información que sería procesada por el clasificador. Dado un número entero, se conserva la parte de la señal cuya posición es múltiplo de ese número. Lo anterior se muestra gráficamente en la figura 3.2, donde la gráfica de abajo representa un submuestra de la señal

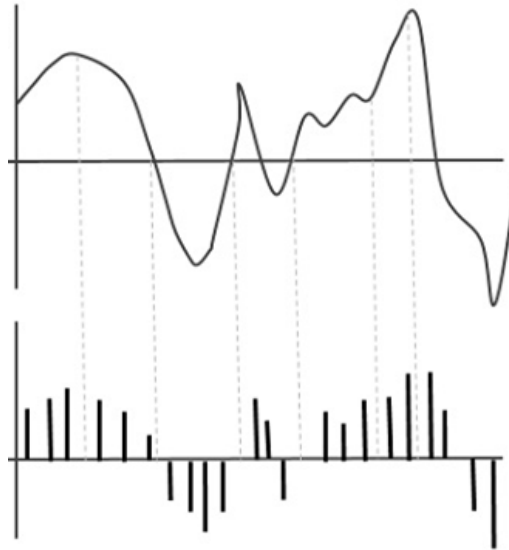


Figura 3.2: Submuestreo de la señal EEG. [Fuente: Elaborada por el autor]

original.

- Extracción de un segmento. La señal EEG completa que se obtiene del experimento se tiene que dividir en los varios fragmentos que se van a clasificarse. Se dividen en segmentos de 1000 ms. Cada uno de los segmentos inicia cuando se ha iniciado un estímulo; es decir, empieza al instante que uno de los cuadrados destellaba y terminaba 1000 ms después del estímulo. Dependiendo del tiempo de inter-estímulo, alguno de los siguientes estímulos podría suceder antes de pasados los 1000 ms; por tanto, varios de los segmentos estarían superpuestos temporalmente al siguiente corte.

La figura 3.2 esquematiza cómo se verían los distintos segmentos de las señales EEG. En cada gráfica se sobreponen una señal que presenta el fenómeno

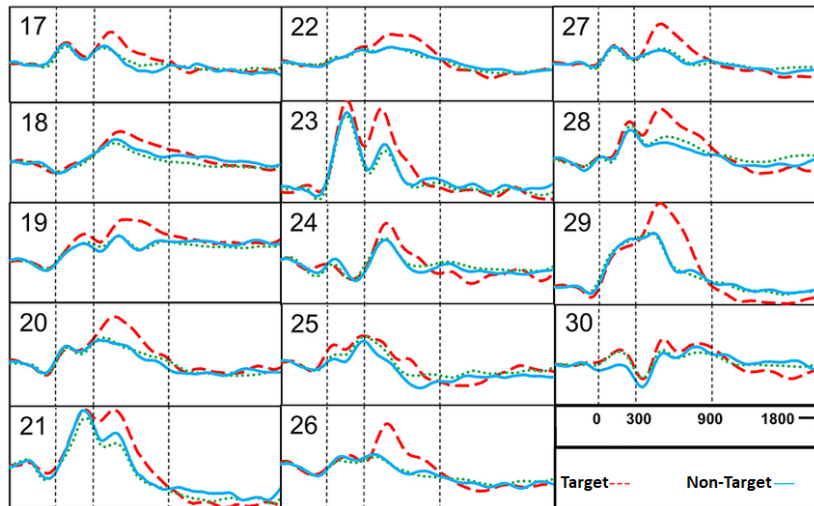


Figura 3.3: Ejemplificación de segmentos del fenómeno P300. [Fuente: <https://www.frontiersin.org/articles/10.3389/fnins.2014.00410/full>]

P300 (positivo) y la señal sin el fenómeno (negativo).

Los segmentos pre-procesados anteriormente descritos fueron codificados en un vector de 825 posiciones, las cuales, consisten la amplitud de la señal en un tiempo dado y de un determinado electrodo. En el lenguaje de análisis de datos, a cada uno de los siguientes vectores se identificará como una observación del experimento. Cada una de las observaciones corresponden a un segmento del fenómeno, los cuales, están etiquetados como positivas, en caso que se presentara el fenómeno P300 o, negativo en caso contrario.

Un extracto de una tabla para uno de los sujetos se encuentra en la figura 3.4, donde cada fila representa una observación del experimento y cada columna representa la amplitud de la señal de un electrodo y tiempo determinado.

		Electrodo													
		Cz - 1	F3 - 1	F4 - 1	C4 - 2	Cz - 2	F3 - 2	F4 - 2	Fp1 - 2	Fp2 - 2	Fz - 2	Oz - 2	C3 - 3	C4 - 3	...
Observaciones	1	-4.93	-5.49	-3.59	-0.63	-4.05	-5.29	-2.78	-7.89	-5.79	-4.67	2.03	-2.91	-0.61	...
	2	-0.31	-1.39	-0.56	0.08	-0.87	-1.20	-0.26	-3.20	-1.89	-0.85	-0.81	-2.20	-1.27	...
	3	-7.13	-2.67	-3.44	-6.17	-11.19	-7.30	-7.12	-1.14	-1.24	-8.42	-3.18	-13.71	-8.63	...
	4	-2.96	-9.86	-5.96	-0.20	-2.78	-9.66	-6.70	-9.17	-5.92	-8.40	2.63	-3.79	-1.44	...
	5	8.27	6.86	8.02	9.84	10.51	8.80	8.40	8.91	5.57	8.49	1.09	11.78	8.23	...
	6	-5.78	-3.39	-3.53	-5.10	-6.25	-3.02	-2.86	-2.50	-2.69	-1.89	-4.41	-5.71	-4.17	...
	7	2.24	-4.27	-0.73	2.22	1.27	-3.74	-0.38	-2.43	-2.66	-1.00	3.93	-0.37	1.32	...
	8	-2.39	-6.41	-6.55	-2.66	-1.72	-5.18	-5.65	-5.06	-7.05	-5.04	-1.87	-4.01	-3.39	...
	9	-8.17	-14.16	-14.41	-5.84	-9.04	-14.34	-13.29	-20.51	-24.41	-11.94	-3.57	-10.60	-7.22	...
	10	5.93	5.60	7.14	6.75	6.37	7.45	8.01	5.38	4.65	6.93	13.97	6.39	4.48	...
	11	1.60	-0.60	1.44	1.43	-0.50	-1.80	0.80	1.79	3.46	0.14	2.43	-1.49	-0.14	...
	12	14.90	16.35	14.36	15.62	14.70	16.97	13.49	46.41	38.81	14.81	11.55	13.44	14.56	...
	13	4.49	2.22	0.59	4.21	5.11	1.86	0.81	1.87	-1.48	2.33	-0.73	4.05	4.29	...
	14	3.32	5.60	6.32	0.10	1.16	4.31	4.34	3.21	3.94	4.50	-0.47	1.54	-0.69	...
	15	1.59	2.46	-1.02	4.08	3.50	3.94	0.68	2.80	-0.07	3.06	10.87	5.63	5.13	...
...	

Figura 3.4: Ejemplificación de la base de datos del fenómeno P300. [Fuente: Elaborado por el autor]

3.2. Implementación del algoritmo

Uno de los grandes retos de la presente tesis fue la implementación del AG pues no se usaron librerías especializadas, sino que el diseño de éste fue codificado manualmente para ajustarse a la medida al problema al resolver. Para lo anterior, se hizo principalmente uso de la librería Weka¹ que es una plataforma de software para el aprendizaje automático y la minería de datos escrito en Java y desarrollado en la Universidad de Waikato.

En las siguientes secciones se explicará brevemente la estructura de clases y las librerías más importantes que se usaron en la implementación. En esta parte no se detallará cómo funciona el algoritmo diseñado, lo cual, se dejará para la

¹Del inglés: Waikato Environment for Knowledge Analysis

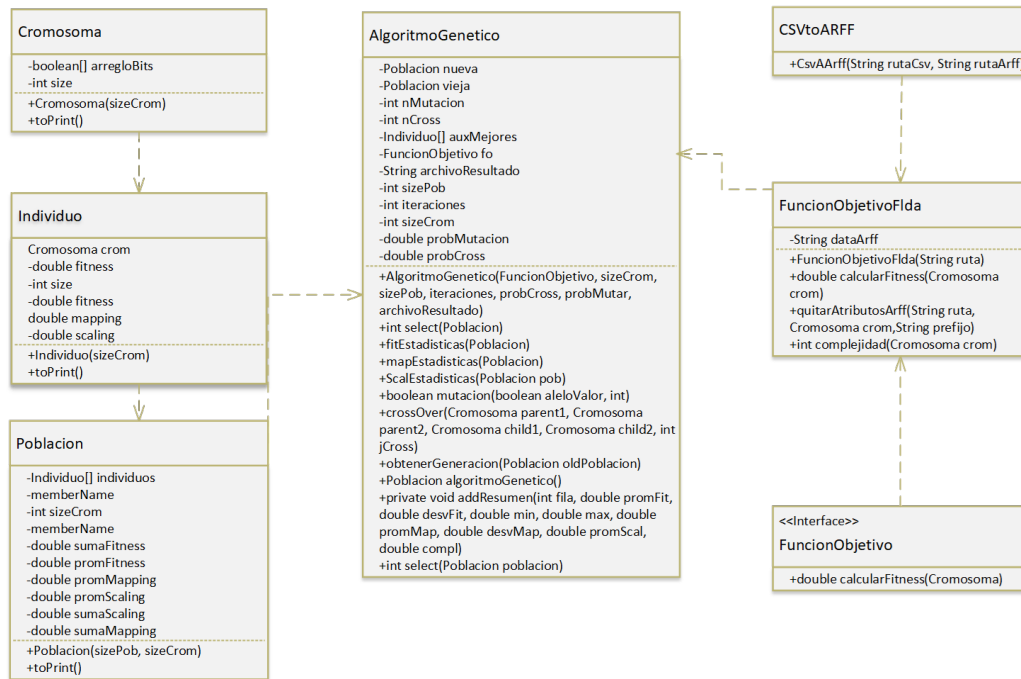


Figura 3.5: Diagrama de Clases. [Fuente: Elaborado por el autor]

sección 3.3. La figura 3.5 muestra un diagrama que esquematiza el diseño general de las clases implementadas, las cuales serán comentadas más adelante.

3.2.1. Estructuras primarias

A continuación, se mencionan las estructuras básicas del AG que fueron implementadas en forma de clases:

- *Cromosoma*. Consiste en un arreglo de bits. Su constructor las crea con bits aleatorios según el tamaño determinado.
- *Individuo*. Consiste principalmente de un *Cromosoma* y otras métricas pro-

pias como la *aptitud* original y escalada del *Individuo*.

- *Poblacion*. Es un arreglo de objetos *Individuo*. Tiene como atributos algunas métricas propias de cada *Poblacion* como la suma y la *aptitud* promedio de los *Individuos* que la conforman.

3.2.2. Algoritmo genético

La implementación del AG se creó dentro de una clase llamada *AlgoritmoGenetico*. Los atributos principales son los siguientes:

- Poblaciones auxiliares. Son poblaciones auxiliares que se ocuparán en las diferentes iteraciones del algoritmo para guardar la última *Poblacion* creada (*Poblacion nueva*) y la anterior a ésta (*Poblacion vieja*).
- Mejores *Individuos* (*Individuo[] auxMejores*). Es un arreglo que guarda los dos mejores *Individuos* que se han encontrado en las generaciones iteradas.
- Parámetros del algoritmo. Son atributos como el número de iteraciones (*iteraciones*), la probabilidad de mutación (*probMutacion*) y la probabilidad de cruzamiento (*probCross*).
- Función objetivo. Este es un objeto creado a partir de una interfaz llamada *FuncionObjetivoFlida*, la cual evalúa la función objetivo requerida.

Esta clase tiene varias funciones y métodos que sirven para generar las poblaciones en cada generación. Las funciones principales son las siguientes:

- *fitEstadisticas*. Recibe una *Poblacion* como parámetro y puebla las métricas de *aptitud* de ésta.
- *mapEstadisticas*. Recibe una *Poblacion* como parámetro y aplica la función inversa aditiva a la *aptitud*, con el objetivo de convertir el problema de reducción de errores a uno de optimización.
- *scalEstadisticas*. Recibe una *Poblacion* como parámetro y escala la *aptitud* de acuerdo a lo que se explica en la sección 3.3.5
- *crossOver*. Recibe los *Cromosomas* de dos *Individuos* padres y genera dos *Individuos* hijos por medio del cruzamiento en la posición que se pasa como parámetro.
- *mutacion*. Toma un bit y una probabilidad como parámetros. Cambia el bit ingresado según la probabilidad especificada.
- *obtenerGeneracion*. Toma como parámetro una *Poblacion* y obtiene otra nueva *Poblacion* realizando la cruce, mutación y selección de los *Individuos*.
- *select*. Toma una *Poblacion* y devuelve el *Individuo* de ésta que se selecciona para la cruce, según la metodología que se detalla en la sección 3.3.4.

3.2.3. Función objetivo

La implementación del AG se creó a través de una interfaz llamada *FuncionObjetivo*, la cual determinará el método esencial que tiene que cumplir toda fun-

ción objetivo para ser usada por el AG implementado. Este método se deberá llamar *calcularFitness*, tomará un *Cromosoma* como parámetro y devolverá un valor numérico que represente la *aptitud*.

Para la implementación de las funciones objetivo² se ocupan las bibliotecas de Weka relacionadas con el cálculo del discriminante lineal de Fisher³ y el filtrado de las variables⁴ que indica el *Cromosoma* que se toma como parámetro.

3.2.4. Formato de tablas

La información de las señales EEG pre-procesadas, que se refirió en la sección 3.1, fue usada como los datos base para la presente tesis. Esta fue proporcionada en archivos CSV⁵, el cual es un formato muy usado para representar datos en forma de tabla, donde las columnas se separan por comas y las filas por saltos de línea. Sin embargo, este formato no puede usarse en las funciones implementadas en Weka, por lo que se usaron diversas librerías para transformar los archivos CSV al formato ARFF, que es el apropiado en dichas librerías.

Los archivos ARFF tienen formato de texto plano en ASCII, por lo que pueden ser visualizados y modificados desde cualquier editor de texto. En un archivo ARFF se diferencian dos partes:

- Cabecera: Incluye la definición del nombre de la relación, la declaración de

²En el diagrama 3.5 se muestra como ejemplo la clase *FuncionObjetivoFlida*

³Las librerías usadas para el cálculo del discriminante lineal de Fisher son: `weka.classifiers.functions.FLDA` y `weka.classifiers.Evaluation`

⁴Las bibliotecas usadas para el filtro de las variables usadas son: `weka.filters.Filter` y `weka.filters.unsupervised.attribute.Remove`

⁵Del inglés: *Comma-separated values*

los atributos y su tipo.

- Datos: Cada línea representa una instancia. Dentro de cada instancia, los atributos se separan con comas. Estos atributos deben aparecer en el orden en que se declaran en la sección de la cabecera.

3.3. Descripción del algoritmo

En las siguientes secciones se darán los detalles relacionados al AG implementado. Esto se resolverá desarrollando los siguientes puntos:

- La codificación de los *Individuos*.
- Cómo se obtiene la primera *Poblacion*.
- Cómo se hace el apareamiento de las soluciones.
- Cómo se define la función objetivo.
- Qué implementaciones adicionales se usan para mejorar el AG estándar.

3.3.1. La codificación de los individuos

La materia prima que usan los AG es con los *Individuos* que están caracterizados principalmente por el *Cromosoma*. En éste, se codifica una solución específica del problema, la cual, vive en el conjunto de todas las posibles soluciones que tiene el problema y, de las cuales, se quiere optimizar.

Variable	Cz - 1	F3 - 1	F4 - 1	C4 - 2	Cz - 2	F3 - 2	F4 - 2	Fp1 - 2	Fp2 - 2	Fz - 2	Oz - 2	C3 - 3	C4 - 3	...
Vector	{1,	0,	1,	0,	1,	1,	1,	0,	0,	0,	0,	1,	0,	...}

Figura 3.6: Ejemplo de vector que codifica la variable que se usará por el clasificador. [Fuente: Elaborado por el autor]

En la sección 3.1 se mencionó que las variables para hacer la clasificación consisten en las amplitudes de las señales EEG previamente pre-procesadas (filtradas y digitalizadas) y que se resumieron en un vector de 825 elementos, donde cada posición representa un canal (electrodo) en un determinado tiempo.

En el presente trabajo, el objetivo es encontrar el subconjunto de esas 825 variables que se requieren para optimizar el error de clasificación que proviene de la validación cruzada, usando el discriminador de Fisher. Dado que el *Cromosoma* debe ser una representación de este conjunto de soluciones, se decidió que, cada *Cromosoma* fuera un vector de 825 posiciones; la i -ésima posición tendría 1, si la variable i se utiliza en la clasificación, y 0 si no se utiliza. Lo anterior se puede ver gráficamente en la figura 3.6.

3.3.2. Obtención de la primera población

Una vez que está establecido el conjunto de soluciones en la que el AG buscará la solución óptima, se debe encontrar un conjunto de soluciones iniciales que serán evaluadas a través de la función. Para la presente implementación se decidió usar una *Poblacion* inicial de 50 *Individuos*, los cuales se construyeron con vectores de 825 posiciones usando una distribución aleatoria Bernoulli con parámetro $p = 0.5$.

3.3.3. Función objetivo

Una vez que la *Poblacion* inicial fue creada, cada una de las posibles soluciones se evaluó con la función objetivo correspondiente a los errores de la validación cruzada de la clasificación del discriminante de Fisher. Lo anterior se describirá continuación.

1. Se usó la técnica de la validación cruzada de cinco iteraciones; es decir, se dividió la *Poblacion* en cinco subconjuntos de las observaciones de manera aleatoria.
2. Para la primera iteración se tomó el primer subconjunto como datos de validación (datos de prueba) y se tomaron los restantes como datos de entrenamiento.
3. Con los datos de entrenamiento, se entrenó el discriminante lineal de Fisher y se validó en los datos de prueba (valores que no ha analizado antes).
4. Lo anterior, se hizo para cada uno de los cinco subconjuntos. En cada una de las cinco iteraciones, se realiza el cálculo del error. El resultado final se obtiene a partir de realizar el promedio de los cinco valores de errores obtenidos. El resultado anterior de los errores se referirá de aquí en adelante como el *Error*.
5. Con el resultado de la clasificación se construyó la tabla de confusión y, a partir de ella, la medida de Fisher que se usará en la creación de la variable

		True	
		Positivo	Negativo
Predichos	Total	Verdadero positivo	Falso Positivo
		Falso negativo	Verdadeto negativo

Figura 3.7: Matriz de confusión. [Fuente: Elaborado por el autor]

objetivo final. Lo anterior se puede ver gráficamente en la figura 3.7, de donde se obtienen las siguientes expresiones:

- $Error = Falso\ positivo / Falso\ negativo$
- $Recall = Verdaderos\ positivos / Positivos$
- $Precision = Verdaderos\ Positivos / Positivos\ predichos$
- $Medida\ de\ Fisher = (2\ Recall * Precision) / (Recall + Precision)$

6. Una vez obtenido el *Error* y la medida de Fisher, la función objetivo se definió de la siguiente manera:

$$Variable\ Objetivo = Error - Medida\ de\ Fisher$$

Podemos observar de la definición de medida de Fisher que esta varía de 0 a 1, de manera que, entre más cercano a 1 sea esta cantidad, se obtendría una mejor clasificación. De esta manera, damos preponderancia primero a minimizar el número

	Individuo									
	1	2	3	4	5	6	7	8	9	10
Fitness	19	21	10	11	15	26	18	16	12	18
Suma	19	40	50	61	76	102	120	136	148	S
Rango	[Suma,19)	[19,40)	[40,50)	[50,61)	[61,76)	[76,102)	[102,120)	[120,136)	[136,148)	[148,S)

↑
s = 45

Figura 3.8: Esquemática de la selección de un *Individuo* con probabilidad proporcional a la *aptitud*. [Fuente: Elaborado por el autor]

de *Errores*, pero cuando los *Errores* son similares entre la *Poblacion*, por medio del escalamiento del AG la medida de Fisher cobrará mayor importancia.

3.3.4. Apareamiento de las soluciones

Con la función objetivo descrita anteriormente, se hace la evaluación de cada uno de los *Individuos* de la *Poblacion* usando las variables que indica el correspondiente *Cromosoma*; el resultado de esta evaluación se designa como *aptitud*. Con esta información se procede a realizar el apareamiento de la siguiente manera:

1. Se seleccionan dos *Individuos* de la *Poblacion* con una probabilidad que sea proporcional a la *aptitud* que tengan respecto de los otros *Individuos*. Para esto, se ordenan los *Individuos* y se hace la suma parcial de su *aptitud*. Sea S la suma total de la *aptitud* se tiene que, para cualquier número s que cumpla que $0 < s < S$, éste se podrá corresponder inequívocamente con un *Individuo* según el intervalo que define la suma parcial respecto del anterior *Individuo*. Lo anterior se representa esquemáticamente en la figura 3.8, donde la $s = 45$ indica que el *Individuo* 3 se seleccionará para el cruzamiento.

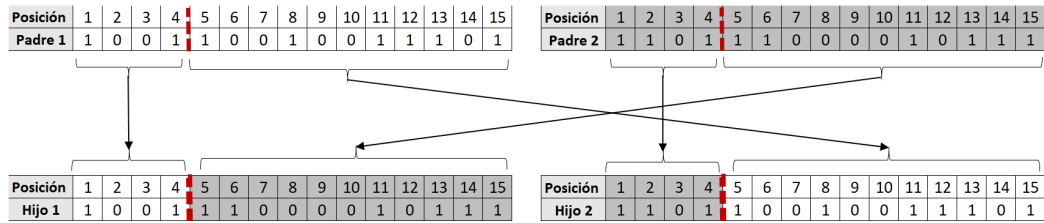


Figura 3.9: Esquemática de la cruce de dos *Individuos*. [Fuente: Elaborado por el autor]

2. Posteriormente se eligen dos valores aleatorios entre $[0, S]$, los cuales se determinarán según el método anterior cuáles serán los *Individuos* que se aparearán.
3. Posteriormente se elige un número aleatorio entre 0 y 825 para elegir la posición donde se dividirán los vectores.
4. Dado el punto del vector del primer padre, todos los datos más allá de este punto, se intercambiarán entre los dos *Individuos* padres y, de esta manera, se obtendrán dos hijos. Lo anterior se esquematiza en la figura 3.9, donde el punto de corte resultó ser en la posición 4.

3.3.5. Operadores adicionales para mejorar el AG estándar

Para mantener una constante presión en la selección de los mejores *Individuos* que se aparearán con el AG, se hace el escalamiento de la *aptitud* de los *Individuos* que se aparearán. Esto es importante pues en las etapas finales del algoritmo puede haber poca dispersión entre los *Individuos*, por lo cual, la selección con la *aptitud* normal tratará similarmente a todos los *Individuos* sin que las pequeñas mejoras

puedan trascender. Esto logró una mejora muy importante en el diseño del AG diseñado.

Se usó un escalamiento lineal, lo cual resulta en una relación lineal con la *aptitud* original y manteniendo una cierta relación entre el máximo de la *aptitud* individual y el promedio de la *Poblacion*. Para lo siguiente se le aplicó una transformación de estandarización, haciendo que la *aptitud* de la *Poblacion* tuviera una varianza 1 y media 1, y topándolo hacia abajo a 0. Lo anterior se expresa en la siguiente fórmula:

$$Fit_{escal} = \max[(Fit_{orig} - Fit_{prom})/Fit_{desv}, 0]$$

Donde:

- Fit_{escal} es la *aptitud* que resulta del escalamiento.
- Fit_{orig} es la *aptitud* original.
- Fit_{prom} es el promedio de la *aptitud* de la *Poblacion* de la generación.
- Fit_{desv} es la desviación estándar de la *aptitud* de la *Poblacion* en la generación.

Vemos que la anterior estandarización mantiene una distribución siempre constante entre los diferentes *Individuos* de las poblaciones.

A veces, por cuestiones de azar, algunos buenos *Individuos* de la *Poblacion* se pierden en el apareamiento o en la mutación, pues los hijos pueden tener menor *aptitud* que los *Individuos* padres. Por esta razón se decidió utilizar el mecanismo

del elitismo, el cual consiste en transmitir de generación a generación los mejores *Individuos*. Esto tuvo un gran impacto en la implementación del AG diseñado permitiendo una convergencia monótona y constante. Se decidió mantener los dos mejores *Individuos* de la generación.

Capítulo 4

Resultados

En el presente capítulo se hará una presentación de los resultados obtenidos con la aplicación de los AG al problema de la selección de variables presentada en los capítulos anteriores.

Se decidió hacer dos formas distintas de análisis con el objetivo de poder observar de una manera más clara los resultados y las ventajas de usar los AG en la selección de variables para un clasificador. Los siguientes análisis son los que se desarrollarán:

- Comparación de las distintas funciones objetivo exploradas y su impacto en los resultados obtenidos.
- Evaluación estadística de la mejora realizada con los AG en el uso de la selección de variables en la clasificación. Para esto se aplicará el AG a los ocho sujetos de los que se tienen datos y se evaluará la mejora estadística con una prueba de hipótesis.

4.1. Funciones objetivo exploradas

Es importante señalar que el AG tiene un componente aleatorio, lo que implica un resultado distinto en cada ejecución del algoritmo usando el mismo sujeto y parámetros. Por esta razón se reportarán los valores promedios de 10 ejecuciones del AG aplicadas a cada sujeto. Esto se ejemplifica en la figura 4.1, donde se muestran los *Errores* de los individuos con mayor *aptitud* de 10 ejecuciones del AG aplicado a un sujeto determinado. Por cada ejecución se obtuvieron 30 generaciones. Se puede ver que en la sexta ejecución se obtuvo el error mínimo.

De la anterior forma se usaron diferentes funciones objetivo que serían minimizadas por el AG, con el fin de encontrar un mínimo de errores en la clasificación. Se probaron las siguientes funciones objetivo:

- El número de *Errores*.
- La medida de Fisher.
- El número de *Errores* más el inverso aditivo de la medida de Fisher.

Los resultados obtenidos con las tres funciones objetivo se muestran en las figuras 4.2, 4.3 y 4.4, donde se muestra el promedio de los *Errores* que se obtuvieron en cada generación, el límite inferior y límite superior (definidos a partir de la desviación estándar de las 10 iteraciones).

Podemos apreciar de las gráficas en las figuras 4.2, 4.3 y 4.4, que las tres funciones objetivo usadas por el AG resultan muy útiles para encontrar el mínimo número de *Errores*. Podemos observar, que cuando usamos la función objetivo de

		Ejecución									
		Corrida 1	Corrida 2	Corrida 3	Corrida 4	Corrida 5	Corrida 6	Corrida 7	Corrida 8	Corrida 9	Corrida 10
G e n e r a l i ó n	1	17	14	15	18	15	18	19	16	16	16
	2	16	14	14	16	15	17	16	16	16	15
	3	16	14	14	16	14	16	15	15	16	15
	4	16	13	14	16	14	16	14	14	15	14
	5	15	13	14	16	14	15	14	12	15	14
	6	15	13	13	15	14	15	14	12	15	13
	7	13	13	13	14	14	14	14	12	14	13
	8	12	13	12	13	14	14	14	12	14	12
	9	12	12	12	13	13	14	14	11	14	12
	10	12	12	12	13	12	12	13	11	14	12
	11	11	12	11	13	12	12	13	11	14	12
	12	11	11	11	13	12	12	13	11	14	11
	13	11	11	11	13	11	12	13	11	14	11
	14	11	11	11	12	10	11	13	10	14	11
	15	10	11	11	12	10	11	13	10	14	10
	16	10	11	10	12	10	10	13	10	14	10
	17	10	11	10	12	10	10	13	10	14	10
	18	10	11	10	12	10	9	13	10	14	10
	19	10	11	10	12	10	9	13	10	13	10
	20	10	11	10	12	10	8	12	10	13	10
	21	10	11	10	12	10	8	12	10	13	10
	22	10	11	10	12	10	8	12	10	13	10
	23	10	10	10	12	10	8	12	10	12	10
	24	9	10	10	12	10	8	12	9	12	10
	25	9	10	10	12	9	7	12	9	12	10
	26	9	10	9	12	9	7	12	9	11	10
	27	9	10	9	11	9	7	12	9	11	10
	28	9	10	9	11	9	7	12	9	11	9
	29	9	10	9	11	9	7	11	9	11	9
	30	9	9	8	11	9	7	11	9	11	9

Figura 4.1: Ejecuciones de AG aplicada a un sujeto determinado. [Fuente: Elaborado por el autor]

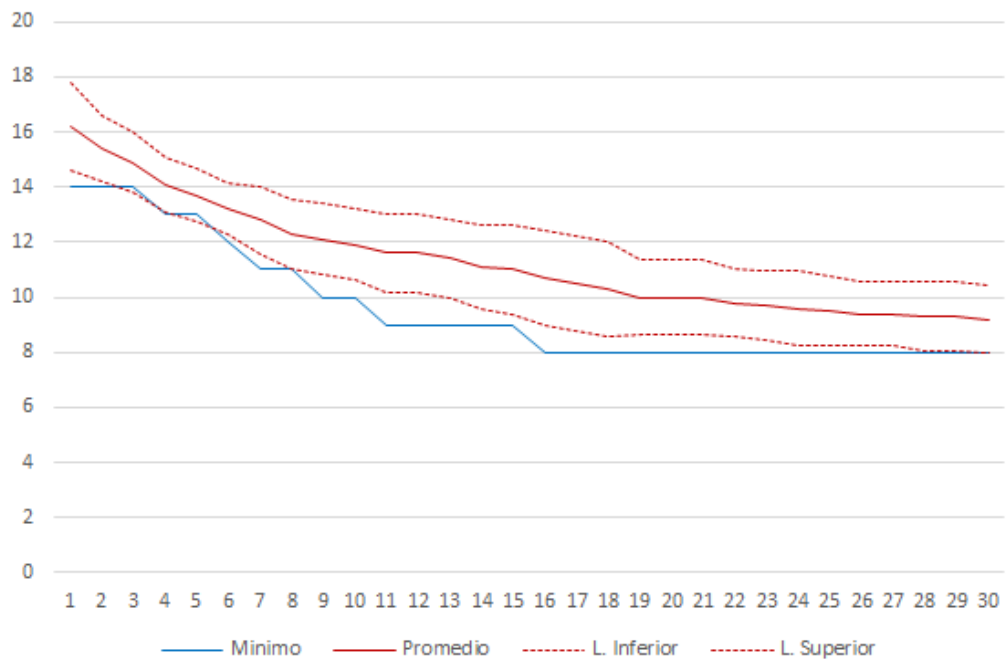


Figura 4.2: Promedio, límite inferior y límite superior de los *Errores*, usando como función objetivo el número de *Errores* únicamente. [Fuente: Elaborado por el autor]

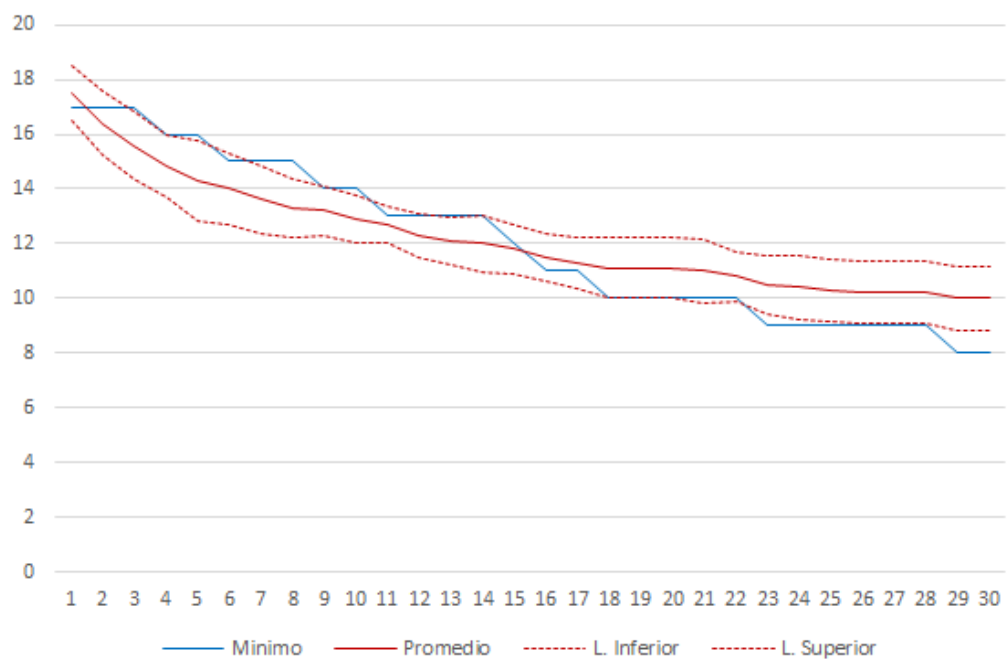


Figura 4.3: Promedio, límite inferior y límite superior de los *Errores*, usando como función objetivo la medida de Fisher. [Fuente: Elaborado por el autor]

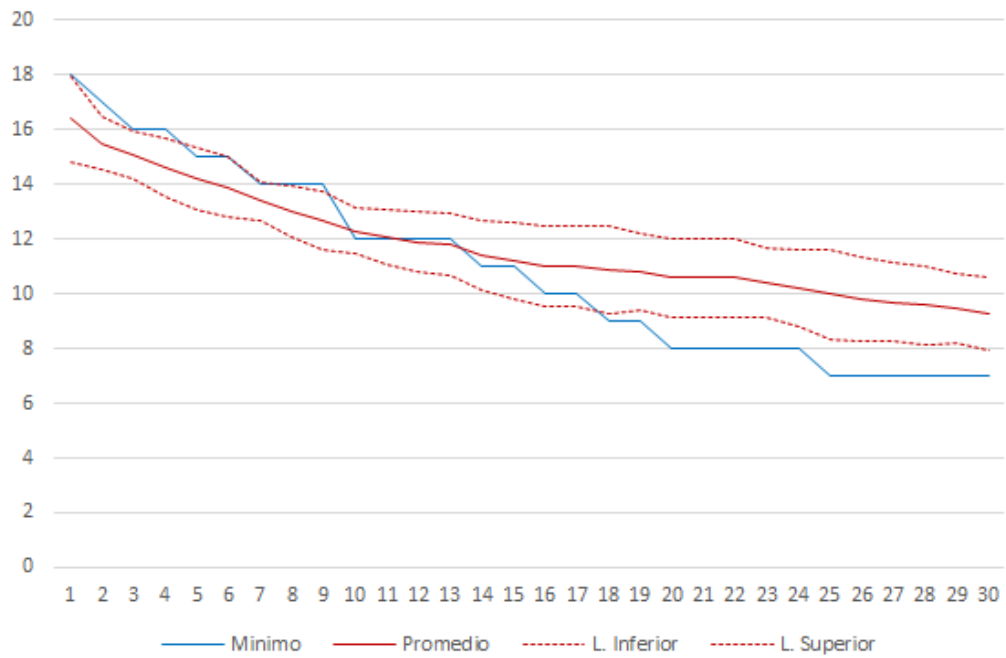


Figura 4.4: Promedio, límite inferior y límite superior de los *Errores*, usando como función objetivo el número de *Errores* más la medida de Fisher. [Fuente: Elaborado por el autor]

los *Errores* únicamente, la *aptitud* de la población se homogeniza y se estanca la mejora en las siguientes generaciones. Usando la medida de Fisher como una función objetivo menos discretizada a comparación de los errores, encontramos que se desempeñó también de forma similar encontrando un mínimo de ocho errores. Podemos ver de las anteriores gráficas que el mínimo de errores se encontró considerando la medida de Fisher junto con el número de errores en la función objetivo. Un análisis de la varianza de los resultados del AG usando las distintas funciones objetivo demostró que no hay alguna diferencia estadística entre ellas; sin embargo, se encontró el individuo con menor *Error* (7 errores) considerando la combinación de los *Errores* con la medida de Fisher. Esto se puede deber a que, a pesar de que los errores son la variable principal, cuando la población se vuelve muy homogénea la medida de Fisher lograba que hubiera una ligera diferencia entre individuos con igual número de errores, lo cual se amplificaba por el escalamiento implementado en el AG.

4.2. Comparación estadística de distintos métodos de selección de variables

Para los análisis siguientes sólo se utilizará la función objetivo considerando los *Errores* junto con la medida de Fisher, pues fue la que más precisión de clasificación obtuvo según la sección anterior. Se realizará la comparación de los distintos métodos de selección de variables para determinar la mejora estadística del uso de los AG.

Todas la variables					Algoritmo genético				
		Valor en la realidad					Valor en la realidad		
		Positivo	Negativo				Positivo	Negativo	
Valor en la predicción	Positivo	27	10	37	Valor en la predicción	Positivo	31	2	33
	Negativo	9	27	36		Negativo	5	35	40
		36	37	Totales			36	37	Totales

Figura 4.5: Comparación de las matrices de confusión. [Fuente: Elaborado por el autor]

	Todas la variables	Algoritmo genético
Sensibilidad	0.75	0.86
Especificidad	0.73	0.95
Exactitud	0.74	0.90
Medida de Fisher	0.74	0.90
AUC	0.74	0.90

Figura 4.6: Comparación de las métricas de la clasificación. [Fuente: Elaborado por el autor]

En las figuras 4.5 y 4.6 se comparan las matrices de confusión y las métricas correspondientes a la clasificación hecha con todas las variables y con la clasificación tomando en cuenta solo las variables que resultan de la aplicación del AG.

Vemos de las tablas anteriores que, haciendo una comparación entre el uso de todas las variables y las seleccionadas por el AG, resulta en una mejora considerable incrementando la precisión del clasificador más de 15 puntos porcentuales. Sin embargo, esto podría ser solo un caso aislado del uso del algoritmo por lo que, a continuación, se incluirá el análisis de otros siete experimentos usados en sujetos diferentes; así mismo, se incluirá la comparación con otros métodos univariados de selección de variables.

Para poder evaluar el desempeño del AG en la selección de variables en la clasificación de las señales EEG, haremos una comparación con las siguientes metodologías:

- Usando todas las variables.
- Método univariado usando el F Score¹.
- Método univariado usando el coeficiente de correlación.

Para la comparación entre los diferentes métodos de selección de variables se realizará un análisis de la varianza, el cual permite determinar si, diferentes tratamientos muestran diferencias significativas o, por el contrario, puede suponerse que las medias poblacionales no difieren entre sí estadísticamente.

4.2.1. Usando todas las variables

El caso de selección de variables más básico que se debe considerar es el uso de todas las variables disponibles para la clasificación. Este es el punto de partida para poder evaluar cualquier método de selección de variables, pues tendría que haber una mejora respecto de este caso base, ya sea en el poder discriminatorio del clasificador o la reducción de la complejidad sin comprometer el poder discriminatorio.

En la figura 4.7 se muestra el porcentaje de éxito de la clasificación usando la totalidad de las variables para los ocho sujetos de muestra.

¹Se define con la siguiente expresión: $F_{Score} = 2 \cdot \left(\frac{exhaustividad \cdot precision}{exhaustividad + precision} \right)$

Sujetos	Normal
1	74%
2	70%
3	67%
4	60%
5	63%
6	64%
7	63%
8	55%
Promedio	65%

Figura 4.7: Precisión usando la totalidad de los atributos disponibles para los ocho sujetos diferentes. [Fuente: Elaborado por el autor]

Se puede notar que el porcentaje de precisión de los diferentes sujetos varían ampliamente pues cada uno de los sujetos tienen características particulares en el experimento donde se recolectaron los datos; por tanto, podemos suponer razonablemente que la mejora de los demás métodos será relativa al sujeto que se trate.

4.2.2. Método univariados

Los algoritmos univariados tratan de evaluar el poder discriminante de cada característica individualmente. Se seleccionan los N mejores atributos individuales, donde N se define por el diseñador del modelo. Estos métodos son computacionalmente eficientes pero también subóptimos. Esto es porque solo consideran la utilidad de una característica individualmente e ignoran las posibles redundancias y complementariedad entre distintas características. En el presente trabajo se exploran dos métodos univariados: el F Score y el coeficiente de correlación.

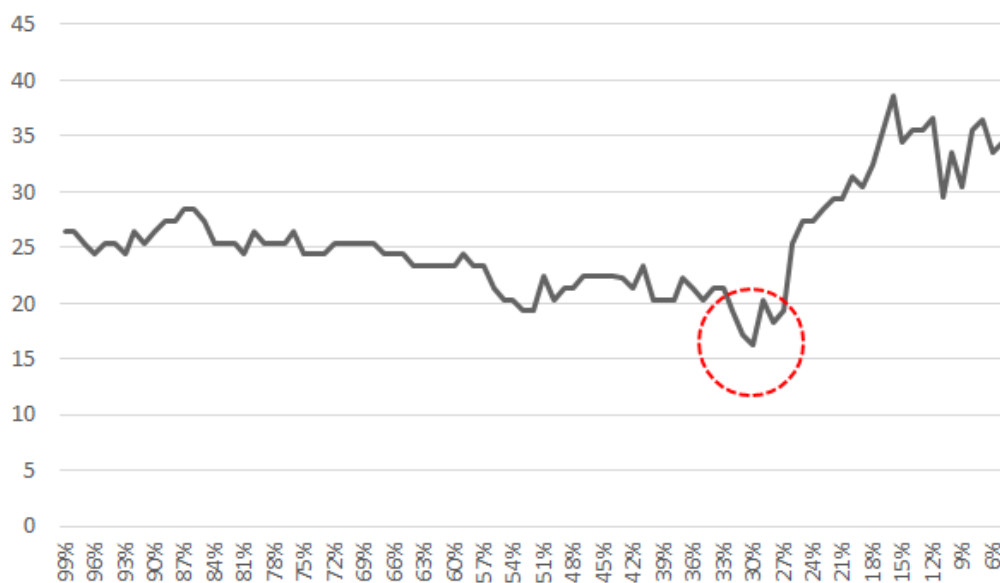


Figura 4.8: Error de clasificación contra porcentaje de mejores atributos según la F Score. [Fuente: Elaborado por el autor]

Para encontrar el error mínimo que puede obtenerse con este método de selección de variables se eligieron un determinado porcentaje de los mejores atributos según su F Score o su coeficiente de correlación. Dependiendo de este porcentaje se corría el algoritmo y se obtenía el error de la clasificación con los atributos escogidos. Esto se hizo para los diferentes percentiles de mejores atributos hasta encontrar el óptimo número de atributos que resultaron en el *Error* mínimo. La metodología anterior se muestra en las figuras 4.8 y 4.9, donde se grafica el *Error* de clasificación contra el porcentaje de mejores atributos de cada metrica. El círculo representa el porcentaje donde se encuentra el mínimo error obtenido.

Lo anterior se hizo para los ocho sujetos que fueron motivo de estudio, obteniéndose los resultados de sus *Errores* mínimos, como se muestran en la figura

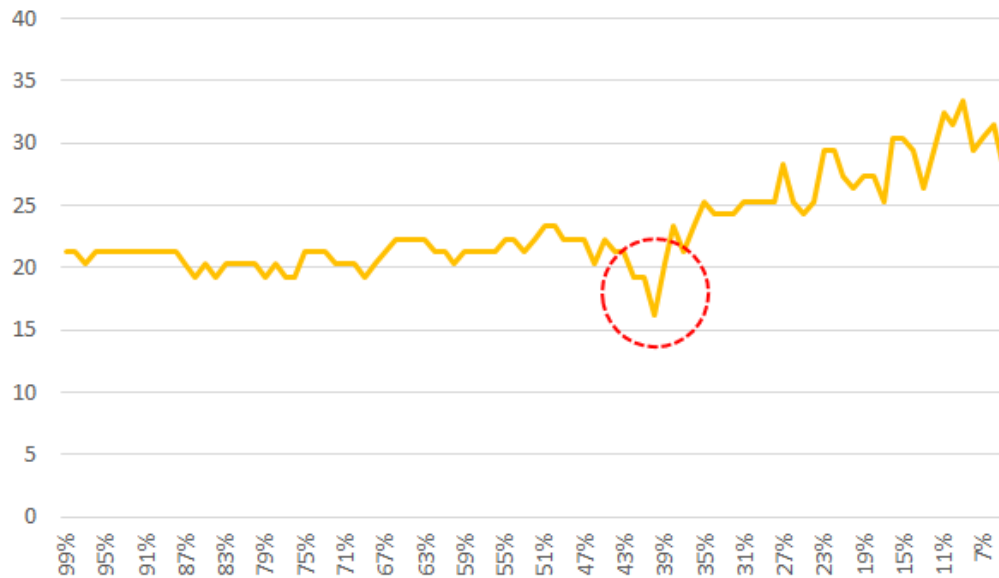


Figura 4.9: Error de clasificación contra porcentaje de mejores atributos según el coeficiente de correlación. [Fuente: Elaborado por el autor]

4.10.

4.2.3. Comparación de resultados

Tomando solo en cuenta la función objetivo que considera los errores y la medida de Fisher, se aplicó el AG para seleccionar las variables que resultaron en el error mínimo del clasificador para cada uno de los ocho sujetos del experimento.

En las figuras 4.11, 4.12 y 4.13 se muestran de diversas maneras la comparación de los resultados del AG contra los métodos univariados y la totalidad de los atributos.

De los anteriores resultados se puede ver claramente que en todos los casos los métodos univariados resultan en una mayor precisión del clasificador. Asimismo,

Sujetos	F Socre	Correlación
1	77%	78%
2	73%	77%
3	77%	74%
4	71%	66%
5	77%	71%
6	77%	81%
7	78%	75%
8	68%	67%

Figura 4.10: Precisión del clasificador usando la totalidad de los atributos disponibles, para los ocho sujetos diferentes. [Fuente: Elaborado por el autor]

Sujetos	Normal	F Socre	Correlación	Univariado	Génético
1	74%	77%	78%	78%	89%
2	70%	73%	77%	77%	86%
3	67%	77%	74%	77%	81%
4	60%	71%	66%	71%	82%
5	63%	77%	71%	77%	84%
6	64%	77%	81%	81%	92%
7	63%	78%	75%	78%	85%
8	55%	68%	67%	68%	75%
Promedio	65%	75%	74%	76%	84%
Desviación	6%	3%	5%	4%	5%

Figura 4.11: Precisión del clasificador usando los diversos métodos de selección de atributos, para los ocho sujetos diferentes. [Fuente: Elaborado por el autor]

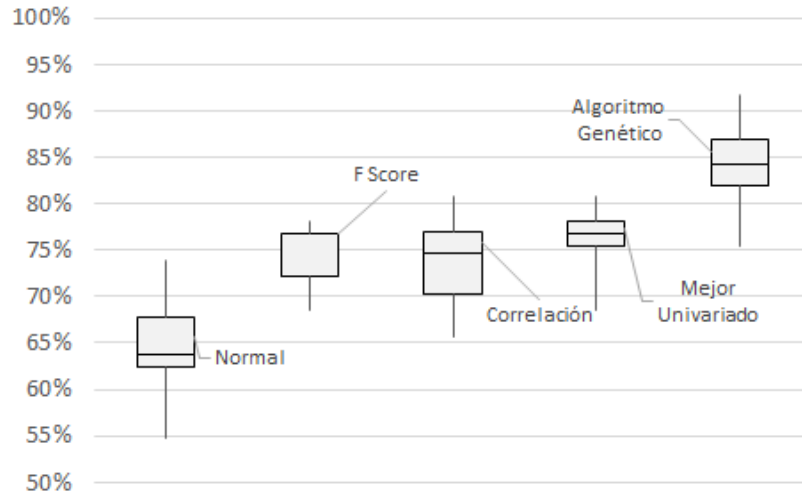


Figura 4.12: Gráfica de cajas de la precisión del clasificador de los ocho sujetos, usando los diferentes métodos de selección de atributos. [Fuente: Elaborado por el autor]

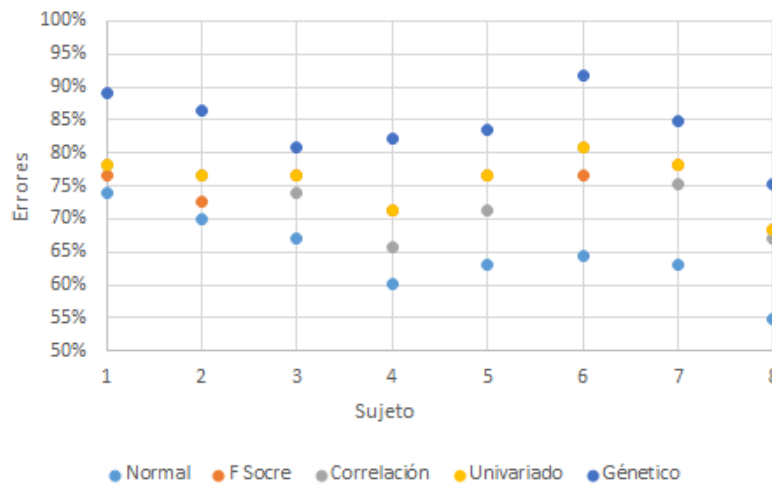


Figura 4.13: Precisión usando los diferentes métodos de selección de variables, para los ocho sujetos diferentes. [Fuente: Elaborado por el autor]

Sujetos	Normal	F Score	Correlación
1	20%	16%	14%
2	24%	19%	13%
3	20%	5%	9%
4	36%	15%	25%
5	33%	9%	17%
6	43%	20%	14%
7	35%	9%	13%
8	38%	10%	12%

Figura 4.14: Mejora porcentual de la precisión respecto del AG, para los ocho sujetos diferentes. [Fuente: Elaborado por el autor]

el AG resulta en una mejora muy importante respecto del uso de la totalidad de atributos y los métodos univariados. En la figura 4.14 se puede ver que la mejora porcentual respecto del uso de la totalidad de variables es de hasta de 43% y, respecto a los métodos univariados, de hasta 25%.

Sin embargo, es necesario corroborar esta intuición visual de alguna manera estadística. Para esto, usaremos un análisis de la varianza (ANOVA), comparando los resultados para los ocho sujetos usando los métodos de selección de variables de interés. En esta prueba la hipótesis nula es que los dos métodos arrojan resultados que no son estadísticamente diferentes.

En la figura 4.15 se puede observar que el p-valor del análisis de viarianza entre los métodos de la correlación y el F Score es de 0.651. De esta manera vemos que, a pesar de que en cinco de ocho sujetos el F Score dio mejores resultados de clasificación, estos no son estadísticamente significativos con un nivel de confianza del 95%. Por tanto, se concluye que ambos métodos univariados tienen un

ANÁLISIS DE VARIANZA (Normal vs F Score)		
<i>F</i>	<i>Probabilidad</i>	<i>Valor crítico para F</i>
17.619	0.001	4.600

ANÁLISIS DE VARIANZA (Correlación vs F Score)		
<i>F</i>	<i>Probabilidad</i>	<i>Valor crítico para F</i>
0.214	0.651	4.600

ANÁLISIS DE VARIANZA (Minimo Univariado vs Genético)		
<i>F</i>	<i>Probabilidad</i>	<i>Valor crítico para F</i>
13.478	0.003	4.600

Figura 4.15: Pruebas de hipótesis entre parejas de métodos de selección de variables usados en la clasificación. [Fuente: Elaborado por el autor]

desempeño similar en la selección de variables.

En el caso del AG, vemos que los p-valores respecto de los métodos univariados y usando todas las variables, es de 0.003 y 0.001 respectivamente; por tanto, se rechaza la hipótesis de que la distribuciones son las mismas y se concluye que la mejora con el uso de los AG, en comparación con usar todas las variables o usar la selección por métodos univariados, es estadísticamente significativa.

Conclusiones

Las interfaces cerebro-máquina basadas en P300 tienen muchas potenciales aplicaciones las cuales necesitan que la clasificación de las señales se pueda hacer de manera más precisa y rápida. Generalmente se encuentra que se puede obtener una buena clasificación usando diferentes clasificadores, pero lo que realmente parece más importante es el diseño de la correcta selección de las *características* que describen las señales EEG.

Las distintas metodologías univariadas para la selección de variables son computacionalmente económicas pero difícilmente pueden llegar a un óptimo. Las metodologías multivariadas que evalúan un subconjunto de características juntas y mantienen el conjunto con las mejores características necesitan un alto requerimiento computacional, por lo que solo es posible evaluarse un pequeño subconjunto de todas las posibles combinaciones. Los AG surgen como una solución natural a este problema pues permiten explorar los múltiples subconjuntos de características de manera heurística con gran eficacia.

En la presente tesis se hizo una comparación de diversas funciones objetivo probadas en el AG. Se observa que, aunque las distintas funciones objetivo son

similares estadísticamente, la combinación de los *errores* de clasificación con la medida de Fisher logró encontrar una mayor solución.

Asimismo, se hizo una comparación entre los métodos de selección de variables univariados y los AG para evaluar la mejora en la correcta clasificación del fenómeno P300. Se encontró que el AG resulta en una mejora estadísticamente significativa en comparación con el uso de la totalidad de atributos y los métodos univariados. La mejora porcentual respecto del uso de la totalidad de variables es de hasta de 43 % y, respecto a los métodos univariados, de hasta 25 %.

A pesar de que los resultados obtenidos en la presente tesis fueron importantes, se puede mejorar aún la implementación del AG propuesto. Principalmente resultará interesante en futuras investigaciones analizar una optimización multiobjetivo que minimice además de los errores la complejidad del clasificador. Además, se pueden usar técnicas más avanzadas de los AG, como el uso de diploides y técnicas que mantengan una correcta diversificación en la población de *Cromosomas*.

Bibliografía

- [1] C. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.
- [2] C.A. Coello *Introducción a la Computación Evolutiva*. CINVESTAV-IPN, 2014.
- [3] D.A. Coley. *An Introduction to Genetic Algorithms for Scientists and Engineers*. World Scientific Publishing , 1999.
- [4] G. Dornhege, J.D.R. Millán, T. Hinterberger, D.J. McFarland, K. Müller, and T.J. Sejnowski. *Toward Brain-Computer Interfacing*. MIT Press, 2007.
- [5] A.E. Eiben, J.E. Smith. *Introduction to Evolutionary Computing*. Springer, 2007.
- [6] D.E. Goldberg *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley, 1989.
- [7] B. Graimann, B. Allison, G. Pfurtscheller. *Brain-Computer Interfaces*. Springer, 2010.

- [8] I. Guyon, A. Elisseeff. *An Introduction to Variable and Feature Selection*. Journal of Machine Learning Research 3, 2003.
- [9] T. Hastie, R. Tibshirani, J. Friedman. *The Elements of Statistical Learning Data Mining, Inference, and Prediction*. Springer, 2008.
- [10] J.R. Koza. *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. MIT Press, 1992.
- [11] D.J. Krusienski, E.W. Sellers, F. Sellers, S. Bayouth, D.J. McFarland, T.M. McFarland, J.R. Wolpaw. *A Comparison of Classification Techniques for the P300 Speller*. Journal of Neural Engineering. 2006.
- [12] F. Lotte. *A Tutorial on EEG Signal Processing Techniques for Mental State Recognition in Brain-Computer Interfaces*. Springer, 2014.
- [13] M. Mitchell. *An Introduction to Genetic Algorithms*. MIT Press, 1996.
- [14] P.L. Nunez, R. Srinivasan. *Electric Field of the Brain*. Oxford University Press, 2006.
- [15] T. Vaughan, D. McFarland, G. Schalk, W. Sarnacki, D. Krusienski, E. Sellers, J. Wolpaw. *The wadsworth BCI research and development program: at home with BCI*. IEEE Trans Neural Syst Rehabil Eng, 2006.