



CIMAT

Centro de Investigación en Matemáticas, A.C.

MEJORA DEL DESEMPEÑO DE CLASIFICADORES FUNCIONALES

T E S I S

Que para obtener el grado de
Doctor en Ciencias
con Orientación en
Probabilidad y Estadística

Presenta

Jairo Arturo Ayala Godoy

Director de Tesis:

Dr. Rogelio Ramos Quiroga

Autorización de la versión final

A mis queridos padres, por el apoyo incondicional siempre.

Agradecimientos

Quiero expresar mis más sinceros agradecimientos a todas y cada una de las personas quienes con su apoyo y su ayuda hicieron posible mis estudios doctorales. A mi asesor Rogelio Ramos, por su apoyo constante, el tiempo compartido y la confianza durante estos años.

A mis queridos padres y hermanas porque a pesar de la distancia, siempre estuvieron para darme apoyo en los momentos difíciles.

A Lucy, por todo el apoyo brindado, sus consejos, la buena energía y el amor incondicional.

Al Centro de Investigación en matemáticas, A.C. CIMAT, y a todo el personal que ahí labora, por toda la ayuda recibida y hacer mi estancia placentera.

Al Consejo Nacional de Ciencia y Tecnología CONACyT, por la beca No. 292426 otorgada para realizar mi doctorado.

Al grupo de sinodales, Guillermina Eslava Gómez, Joaquín Ortega Sánchez, Johan Jozef Lode Van Horebeek y Rolando José Biscay Lirio, por el tiempo dedicado a la revisión de esta tesis y los comentarios brindados.

Finalmente quiero agradecer, a todo los profesores participes de mi educación, por su paciencia y excelente trabajo.

Resumen

Los avances tecnológicos han hecho de clasificación de datos funcionales una disciplina emergente con una gran variedad de publicaciones, y aunque en la literatura ya existen diferentes propuestas de clasificación, sigue siendo una línea de investigación relevante en la última década.

Nuestro trabajo se centró en clasificación supervisada de datos funcionales, particularmente en el mejoramiento de clasificadores funcionales, donde se pudieron presentar dos propuestas novedosas, las cuales son: Boosting con clasificadores funcionales y Boosting estructura funcional.

En el primer trabajo, adaptamos la forma de ponderar clasificadores funcionales para incorporarlos en las estructuras de ciertos algoritmos Boosting. Estas adaptaciones son ideas novedosas propuestas, las cuales fueron estudiadas analíticamente y desarrolladas computacionalmente. Para esto, tuvimos que estudiar el lugar exacto para realizar la ponderación, logrando obtener el mismo efecto que proponen los algoritmos Boosting. En la segunda investigación, modificamos algunos algoritmos Boosting extendiendo su aplicabilidad a datos funcionales. Estas modificaciones fueron estudiadas, buscando obtener la misma estructura del algoritmo para datos multivariados, pero ahora con datos funcionales.

Los resultados obtenidos con los dos trabajos de investigación muestran re-

sultados favorables. Observamos que cuando nos enfrentamos a datos difíciles de discriminar, nuestras propuestas siempre lograron mejoras significativas. Además, cuando se trabajó con datos reales, conseguimos siempre mejores resultados que los reportados en las referencias utilizadas.

Por otra parte, el estudio se inició un trabajo sobre Análisis de imágenes digitales de mamografías, donde se toma como dato funcional cada imagen, pudimos desarrollar un método nuevo, automático y sencillo de implementar computacionalmente para la etapa de pre-procesamiento, con resultados satisfactorios. Además, se encuentra en desarrollo una propuesta para las demás etapas, con el fin de obtener un método con un desempeño favorable en la clasificación automática de anomalías presentes en una mamografía.

Índice general

Resumen	v
Introducción	xi
1. Análisis de Datos Funcionales	1
1.1. Conceptos básicos	2
1.2. Singularidades y dificultades	7
1.3. Representación	8
1.4. Problemas importantes en FDA	9
1.5. Clasificación supervisada de datos funcionales	12
1.5.1. Métodos destacados	15
2. Boosting para datos finito-dimensionales	19
2.1. Métodos de optimización	22
2.1.1. <i>Forward stagewise additive modeling</i> (FSAM)	22
2.1.2. <i>Functional gradient descent</i> (FGD)	23
2.1.3. <i>Newton-like updates</i> (NU)	24
2.2. Funciones de pérdida	25
2.3. Algoritmos Boosting	26

2.3.1. AdaBoost	27
2.3.2. LogitBoost	29
2.3.3. Binomial Boosting	31
2.3.4. Ponderación Alternativa	33
3. Boosting para datos infinito-dimensionales	37
3.1. Boosting con Clasificadores funcionales	38
3.1.1. Clasificadores funcionales ponderados	39
3.1.2. Resultados y discusión	44
3.1.3. Conclusiones	51
3.2. Boosting con estructura funcional	52
3.2.1. Modificaciones realizadas	52
3.2.2. Resultados y Discusión	54
3.2.3. Conclusiones	61
4. Análisis de imágenes digitales de mamografías	63
4.1. Anatomía de la mama	66
4.1.1. Tipo de severidad	66
4.1.2. Tipo de tejido	67
4.1.3. Tipo de anomalía	69
5. Clasificación automática de anomalías presentes en una mamografía	73
5.1. Pre-procesamiento	74
5.1.1. Metodología	75
5.1.2. Resultados y discusión	84
5.1.3. Conclusiones	94

5.2. Segmentación	95
5.2.1. Ideas trabajadas	96
6. Conclusiones y trabajos futuros	101
6.1. Conclusiones generales	101
6.2. Trabajos futuros	103
Bibliografía	105

Introducción

El gran avance en la tecnología computacional (desempeño de procesadores y la capacidad de almacenamiento) nos permite cada vez más recolectar y almacenar datos de alta dimensionalidad, los cuales naturalmente pueden ser curvas o imágenes e incluso vivir en un espacio infinito-dimensional. Esta clase de datos provienen de diversas áreas como, Medicina, Biomecánica, Química, Econometría, Genética, Geofísica, Meteorología, Geología, entre otras. En este contexto, cada dato observado corresponde a un individuo diferente y cada punto de él corresponde a una misma variable, medida a lo largo de su trayectoria.

En la práctica, en general, los datos se obtienen de manera discretizada, es decir, las mediciones están dadas por vectores finito-dimensionales los cuales pueden llegar a tener dimensiones del orden de cientos de miles. Sin embargo, podemos trasladar este problema al contexto de datos funcionales sabiendo que existe una curva subyacente la cual es observada sólo en instantes discretos de tiempo. Así, podemos explotar la naturaleza infinito-dimensional de dichas observaciones para inferir su estructura y obtener más información y como consecuencia, mejores resultados.

El conjunto de herramientas usadas para analizar esta clase de datos es conocido en la literatura como: **Análisis de Datos Funcionales (FDA)**, por sus siglas

en inglés). Esta área a pesar de ser una disciplina relativamente nueva, inició sus pasos con el trabajo pionero de [Deville \(1974\)](#) y más adelante con [Ramsay y Dantzell \(1991\)](#). Los autores que de alguna manera popularizaron esta área dentro de la Estadística son Ramsay, Silverman, Ferraty y Vieu, con libros como: *Functional Data Analysis* y *Nonparametric Functional Data Analysis*, donde presentan importantes resultados teóricos y prácticos.

En FDA se han propuesto versiones funcionales para métodos estadísticos tradicionales como, por ejemplo, Regresión, [Quin y Choi \(2011\)](#), Análisis de varianza, [Cuevas et al. \(2004\)](#); [Delicado \(2007\)](#), Modelos lineales generalizados, [Escabias et al. \(2004\)](#), Componentes principales [Li y Hsing \(2010\)](#) o Clasificación donde podemos nombrar diversos trabajos. Ésta última es la rama a la cual dedicamos nuestro trabajo.

En una primera etapa de esta tesis y basados en los trabajos de [Freund y Schapire \(1996\)](#), [Friedman et al. \(2000\)](#), quienes desarrollaron los métodos **Boosting**, analizamos y exploramos la extensión de estos métodos para aplicarlos a datos infinito-dimensionales, es decir, abordar los métodos Boosting con herramientas de la literatura de FDA. Aquí, logramos estudiar este problema desde dos puntos de vista: primero, adaptamos algunos clasificadores funcionales logrando incorporarlos en ciertos algoritmos Boosting, y segundo, modificamos algunos algoritmos Boosting extendiendo su aplicabilidad a datos funcionales. En este momento este trabajo se encuentra en revisión para ser enviado a una revista científica arbitrada.

En una segunda etapa, iniciamos una investigación sobre el Análisis de imágenes digitales, particularmente en imágenes de Mamografías. La Mamografía es la herramienta más utilizada tanto para la identificación de anomalías como para la detección temprana del cáncer de mama. Sin embargo, un diagnóstico preciso sigue siendo una tarea clínica bastante difícil. Es por ello, que centramos nuestro

interés en este tipo de imágenes, donde el objetivo principal es aportar un método automático para la detección temprana de esta enfermedad, con ayuda de las herramientas de FDA. En un primer acercamiento a este problema, trabajamos en la eliminación automática del musculo pectoral en imágenes digitales de mamografías, el cual obstaculiza el análisis del tejido mamario. Este trabajo se inició durante una estancia de investigación que realicé en el departamento de Estadística de la Universidad Carlos III de Madrid, Madrid, España, bajo la supervisión del Dr. Juan Romo.

La tesis está organizada de la siguiente manera. En el [Capítulo 1](#) introducimos el tema de Análisis de datos funcionales y los problemas relevantes que se están trabajando recientemente. Además, al final de este capítulo, dedicamos una sección a Clasificación de datos funcionales, tema de nuestra investigación. En el [Capítulo 2](#), mostramos las ideas principales de Boosting para datos finito-dimensionales, y en el [Capítulo 3](#), extendemos su funcionamiento para poder aplicarlos a datos infinito-dimensionales. En el [Capítulo 4](#), hacemos un breve resumen sobre el tema de Análisis de imágenes digitales de Mamografías, y en el [Capítulo 5](#), presentamos el trabajo realizado sobre este tema, discutiendo los resultados obtenidos y las ideas trabajadas para culminar esta investigación. Finalmente, en el [Capítulo 6](#) resumimos los resultados generales y los posibles trabajos futuros.

Capítulo 1

Análisis de Datos Funcionales

El Análisis de Datos Funcionales (FDA) es una rama de la Estadística que estudia la información contenida en curvas, imágenes, o cualquier elemento que varía sobre un continuo. La tecnología ha hecho de FDA una disciplina emergente con una gran variedad de publicaciones y estudios. Prueba de este interés es la creciente presencia de publicaciones en revistas de primera línea como: **Journal of the American Statistical Association**, **Annals of Statistics**, **Statistical Science**, y la creación de grupos de trabajo específicos dentro de asociaciones importantes de investigación, como lo son: **European Research Consortium for Informatics and Mathematics** y la **Sociedad Española de Estadística e Investigación Operativa**.

De la literatura existente, resaltamos algunos libros por su relevancia en el tema, como por ejemplo, el libro de [Ramsay y Silverman \(2005\)](#), que debe citarse como un hito importante en la historia de FDA. Este libro tiene una orientación práctica, dirigida a un público científico general. El libro de [Ramsay y Silverman \(2002\)](#), se centra en la ilustración de las principales técnicas a través del estudio de casos concretos con datos reales. [Ferraty y Vieu \(2006\)](#) representaron un segundo punto de vista del tema, mostrando un lenguaje un poco más matemático, inclu-

yendo cuestiones asintóticas no triviales, junto con discusiones relevantes sobre el uso de semi-métricas, base de muchas dificultades teóricas. Otro libro que vale la pena mencionar es [Ferraty y Romain \(2011\)](#), que consta de 16 capítulos, de diferentes autores, con discusiones importantes de los principales temas en FDA.

Las particularidades de estos objetos, hacen que surja todo un conjunto de definiciones, teoremas y herramientas para poder estudiarlos correctamente. Es por ello, que el primer paso será definir y comentar los conceptos básicos y las características de este tipo de datos.

1.1. Conceptos básicos

El primer aspecto importante es definir qué se entiende por variable funcional. En general, cualquier observación que varíe sobre un continuo se puede tomar por un dato funcional. En la práctica, tenemos muestras de una determinada variable aleatoria en distintos instantes de tiempo dentro de un cierto rango (t_{\min}, t_{\max}) . De este modo, una observación puede expresarse mediante la familia $\{X(t_i)\}_{i=\{1,\dots,m\}}$. Realmente por su naturaleza funcional, podríamos considerar estas muestras como observaciones de la familia continua $\mathcal{X} = \{X(t) : t \in (t_{\min}, t_{\max})\}$.

Definición 1.1. *Una variable aleatoria \mathcal{X} se dice que es una **variable funcional** si toma valores en un espacio funcional \mathcal{F} (de dimensión infinita).*

Este sería el caso de un objeto unidimensional, que es el más común, pero la definición es general, $\{\mathcal{X} = X(t) : t \in T\}$, donde si $T = R$ será una curva, $T = R^2$ una imagen, u otras expresiones para casos más complejos.

Por el momento, tenemos un solo dato, pero si deseamos tener un conjunto de ellos, se pueden definir como:

Definición 1.2. Un conjunto de datos funcionales X_1, \dots, X_n es la observación de n variables aleatorias $\mathcal{X}_1, \dots, \mathcal{X}_n$ idénticamente distribuidas como \mathcal{X} .

Podemos observar algunos conjuntos con gran reconocimiento en la literatura de FDA, llamados datos Benchmark, como ejemplos de datos funcionales. Un primer ejemplo se muestra en la [Figura 1.1](#). Estos datos consisten en el ángulo que se forma en el plano sagital por la cadera y la rodilla en un ciclo de marcha de un niño. El tiempo se mide en términos del ciclo de la marcha individual, de manera que cada curva toma valores de t en $[0, 1]$. El ciclo comienza y termina en el punto en que el talón de la extremidad bajo observación golpea el suelo.

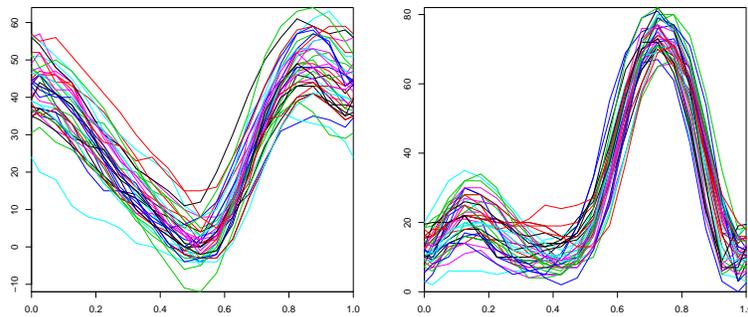


Figura 1.1: Ángulos en el plano sagital formados por la cadera (gráfica izquierda) y por la rodilla (gráfica derecha) de 39 niños.

En la [Figura 1.2](#) observamos las letras “*fda*”. Estos datos se obtuvieron al registrar la posición de la pluma por medio de un sistema de Optotrak, que da la posición de un diodo emisor de infrarrojos en el espacio tridimensional 600 veces por segundo con un error de alrededor de 0.5 mm, para más detalles de estos ejemplos se puede ver [Ramsay et al. \(2009\)](#).

En FDA, como los datos son variables funcionales, las observaciones son varia-

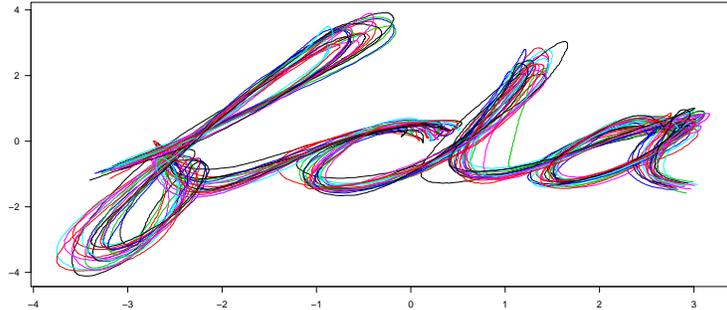


Figura 1.2: Muestras de escritura a mano.

bles aleatorias de procesos estocásticos, es decir, elementos aleatorios que toman valores en un espacio funcional. Por lo tanto, es necesario tener herramientas para su análisis. Estas herramientas vendrán dadas por el espacio en el que habite la observación y determinarán la dificultad del problema. Así, la primera preocupación es determinar el espacio funcional en el que se trabajará. Por ello, la métrica que se elija debe ser coherente con los datos que se describan.

En la literatura se suele utilizar $L^2[a, b]$ por su generalidad y sus buenas propiedades, pero se podrían considerar otros espacios funcionales con distintas métricas o semi-métricas que en algunos casos pueden ser más apropiadas.

Definición 1.3. Sea \mathcal{F} un espacio funcional, \mathcal{F} es un **espacio métrico** si existe una distancia $d : \mathcal{F} \times \mathcal{F} \rightarrow \mathbb{R}$ donde para cualesquiera elementos $f, g, h \in \mathcal{F}$ se cumple:

1. $d(f, g) \geq 0$ y $d(f, g) = 0 \Leftrightarrow f = g$,
2. $d(f, g) = d(g, f)$,
3. $d(f, g) \leq d(f, h) + d(h, g)$.

Definición 1.4. Sea \mathcal{F} un espacio vectorial funcional, \mathcal{F} es un **espacio normado** sobre un cuerpo \mathbb{K} (\mathbb{R} ó \mathbb{C}) si existe una norma $\| \cdot \|: \mathcal{F} \rightarrow \mathbb{R}$ donde para cualesquiera dos elementos $f, g \in \mathcal{F}$ se cumple:

1. $\| f \| \geq 0$ y $\| f \| = 0 \Leftrightarrow f = 0$,
2. $\forall \lambda \in \mathbb{K}, \| \lambda f \| = |\lambda| \| f \|$,
3. $\| f + g \| \leq \| f \| + \| g \|$.

Todo espacio normado se puede convertir en un espacio métrico con la distancia inducida por la norma $d(f, g) = \| f - g \|$. Además, si el espacio es completo (toda sucesión de Cauchy converge) se dice que es un espacio de Banach. Los espacios de Banach pueden ser un buen punto de partida, pero generalmente se pide un poco más, esto es, conocer las orientaciones, posiciones relativas o ángulos.

Definición 1.5. Sea \mathcal{F} un espacio funcional, \mathcal{F} es un **espacio funcional con producto interno** sobre un cuerpo \mathbb{K} si existe un producto escalar $\langle \cdot, \cdot \rangle: \mathcal{F} \times \mathcal{F} \rightarrow \mathbb{K}$ donde para cualesquiera elementos $f, g, h \in \mathcal{F}$ se cumple:

1. $\langle f, f \rangle \geq 0$ y $\langle f, f \rangle = 0 \Leftrightarrow f = 0$,
2. $\forall a, b \in \mathbb{K}, \langle af + bg, h \rangle = a\langle f, h \rangle + b\langle g, h \rangle$, análogamente existe linealidad por la derecha,
3. $\langle f, g \rangle = \overline{\langle g, f \rangle}$.

Todo espacio funcional con producto interno se convierte en un espacio normado con la norma inducida por el producto escalar, $\|f\| = \sqrt{\langle f, f \rangle}$. Si el espacio además es completo se dice que es un espacio de Hilbert. Generalmente se intenta

trabajar en espacios de Hilbert por la comodidad, aunque no siempre es posible. Existen los L -espacios definidos como:

$$L_p(S, \mu) = \left\{ f : S \rightarrow \mathbb{K} \text{ medible, } \|f\|_p = \left(\int |f|^p d\mu \right)^{1/p} < \infty \right\}, \quad (1.1)$$

donde (S, μ) es el espacio de medida. Estos espacios, en el caso más usual $L_p[a, b]$ con $\mathbb{K} = \mathbb{R}$, son de Hilbert para $p = 2$ y de Banach para $p \neq 2$. Además de estas consideraciones sobre Análisis Funcional serán necesarias las herramientas básicas de Estadística como son el cálculo de medias, varianzas y correlaciones.

En el caso finito, típicamente \mathbb{R}^d , hay una equivalencia entre todas las normas en el sentido de que definen la misma métrica, y por tanto, la misma topología. Por esto, la elección de una u otra puede tener menos relevancia, aunque repercutirá en aspectos computacionales y puede tener alguna implicación más importante, incluso decisiva, dependiendo del tipo de problema al que nos enfrentemos (por ejemplo, la norma euclídea usual tiene asociado un producto escalar con el que se puede medir ángulos). Sin embargo, en el caso funcional no se da esta equivalencia y los problemas serán atacados de forma distinta en función de la métrica que elijamos, lo que convierte la elección de dicha métrica en una decisión importante. En este sentido, puede que incluso la elección de cierto espacio métrico sea insuficiente y sea necesario el uso de espacios semi-métricos.

Definición 1.6. Sea \mathcal{F} un espacio funcional, \mathcal{F} es un **espacio semi-métrico** si existe una semi-métrica $d : \mathcal{F} \times \mathcal{F} \rightarrow \mathbb{R}$ que cumple:

1. $\forall f \in \mathcal{F}, d(f, f) = 0,$
2. $\forall f, g \in \mathcal{F}, d(f, g) = d(g, f),$
3. $\forall f, g, h \in \mathcal{F}, d(f, g) \leq d(f, h) + d(h, g).$

Notemos que una semi-métrica es una métrica, excepto en que $d(f, g) = 0$ no implica que $f = g$. Análogamente se puede definir una semi-norma y un espacio semi-normado con las mismas propiedades que la norma, haciendo que elementos no nulos puedan tener norma cero, $\|f\| = 0 \not\Rightarrow f = 0$.

1.2. Singularidades y dificultades

Como hemos podido percibir, utilizar datos funcionales conlleva ciertas particularidades que hacen que los métodos tradicionales no sirvan o no sean los más adecuados. Esto es debido principalmente a tres de sus características: la dimensión, la correlación y el trabajar en espacios funcionales. El primer problema lo encontramos en la propia captura de datos, puesto que en general, no es posible capturar la curva integra. En este sentido los avances tecnológicos palían esta pérdida de información con rejillas cada vez más finas. Una vez capturado el dato, además del procesado propio de cualquier problema (imperfecciones, outliers, entre otros), la alta dimensión en la práctica, e infinita en la teoría, hacen que haya que elegir una representación adecuada para trabajar. El problema de la representación es muy importante en FDA y hay mucho material al respecto. Por otra parte, cuando se trabaja con datos funcionales, los distintos puntos de la curva están altamente correlacionados. Esta correlación no indica más que cercanía, pero hace que se introduzca redundancia al modelo, sobreajustando, provocando matrices cuasi-singulares, donde los algoritmos clásicos fallan. Por lo tanto, es necesario buscar nuevos métodos o cambios sustanciales en los algoritmos clásicos.

Finalmente, la propia naturaleza de los datos plantea otra dificultad. Dependiendo del espacio en el que vivan las funciones puede que no tengamos ni siquiera una métrica, pero aún teniéndola no es trivial definir el concepto de cercanía o de

similaridad entre dos funciones o imágenes, incluso puede que, dependiendo del problema, nos interese más un criterio u otro. Además, en esta misma línea, es complicado establecer cuál es el elemento central de un conjunto de funciones, o qué observaciones están en el extremo. Para ello, se han propuesto nuevas medidas de profundidad, Cuevas *et al.* (2007); Lopez-Pintado y Romo (2009); Sguera *et al.* (2014) que pueden ser más útiles para este tipo de datos que los estadísticos clásicos.

1.3. Representación

La dimensión infinita en los datos funcionales dificulta cualquier análisis que se desee realizar. Por lo tanto, es necesaria una representación en dimensión finita, en lo posible reducida. Éste es un problema muy importante, ya que una mala conversión puede hacer que se pierda información relevante. Los métodos existentes de representación se pueden agrupar en dos grandes familias: la discretización y la elección de una base de funciones reducida. Los sistemas que se engloban bajo estos epígrafes son ampliamente utilizados y no se puede decir que una filosofía sea mejor que otra; en general los autores las eligen según el problema al que se enfrentan.

- **Discretización:** Los datos funcionales se toman en función de un continuo. Podemos suponer, sin pérdida de generalidad, que ese continuo es el tiempo t , y que tenemos un dato funcional $f(t)$ que toma valores en el intervalo $[t_{\text{mín}}, t_{\text{máx}}]$. La discretización más sencilla consiste en tomar una partición del tiempo $\{t_i\}_{i=0}^m$ tal que $t_{\text{mín}} \leq t_0 < t_1 < \dots < t_m \leq t_{\text{máx}}$ y tomar como atributos los valores $\{f(t_i)\}_{i=0}^m$. Sin embargo, aún en el caso más sencillo tenemos que tener en cuenta múltiples factores, tales como, elegir el número de elementos de la partición m ,

definir la separación entre los elementos de la partición, seleccionar los valores para los elementos de la partición, entre otros.

• **Representación en una base:** La otra opción más utilizada para representar funciones es dar las coordenadas de su proyección en algún sub-espacio funcional de dimensión finita. Esto se hace generalmente considerando su desarrollo en una cierta base prefijada y truncando este desarrollo para quedarnos sólo con un número finito de términos. De esta manera, los datos siguen siendo funciones y por tanto conservan sus propiedades, aunque al ser aproximaciones los resultados se verán influidos por el tipo de base elegida. De esta manera, la observación funcional puede aproximarse como:

$$X(t) \approx \sum_{k=1}^K c_k \varphi_k(t), \quad (1.2)$$

donde los c_k son los coeficientes y las φ_k son las funciones de la base. Así, tendremos observaciones en una base finita. El problema en juego ahora es la elección de K , que se comporta como otro parámetro del modelo midiendo el grado de suavizado de la función. La elección de K es a menudo un punto de discusión y no hay un método claro para hacerlo, se suele elegir con algún mecanismo de validación. Por otra parte, las bases más utilizadas en la literatura de FDA son: Fourier, B-Splines y Wavelets. Para más detalles ver capítulo 3 del libro de [Ramsay y Silverman \(2005\)](#).

1.4. Problemas importantes en FDA

Los problemas a los que se debe enfrentar la Estadística con datos funcionales responde a las mismas necesidades que la Estadística clásica. Estos se podrían resumir de la siguiente manera:

- Explorar y describir el conjunto de datos funcionales resaltando sus características más importantes.
- Explicar y modelar la relación entre las variables dependientes y las independientes.
- Aplicar métodos de clasificación a conjuntos de datos respecto a alguna característica.
- Contraste, validación y predicción.

Para alcanzar estos objetivos surge la necesidad de crear metodologías en el ámbito funcional, tales como: Análisis de componentes principales , Regresión, Modelos lineales generalizados y Clasificación.

- **Análisis de Componentes Principales Funcionales (FPCA):** El objetivo principal de FPCA, es al igual que en el caso multivariado, dar una representación de los datos mediante el criterio de conservación de la máxima varianza en una dimensión menor de tal manera que se observen características latentes de los datos. Esta representación suele utilizarse para la visualización o estudio preliminar de los datos, pero a menudo es una herramienta para otros procesos posteriores como la clasificación o la detección de outliers. Por estos motivos, el Análisis de componentes principales fue uno de los primeros métodos adaptados en FDA, el cual cuenta con un gran número de trabajos, donde se estudian sus propiedades y se extienden sus aplicaciones. Por ejemplo, [Benko *et al.* \(2009\)](#); [Hall *et al.* \(2006\)](#) pueden ser dos citas importantes en el tema, o más recientemente podemos ver el capítulo 3 del libro de [Horváth y Kokoszka \(2012\)](#), los cuales hacen un buen resumen de la técnica en general.

- **Regresión Funcional:** El problema de la regresión ha suscitado siempre

mucho interés. Se pueden distinguir distintos tipos de regresión en función de las características de las variables regresoras y de la variable respuesta, pero el problema siempre es el mismo, se trata de predecir o modelar la respuesta a partir de las variables regresoras. En el ámbito funcional, existen diversas referencias importantes, que dedican capítulos enteros a distintas variantes del problema desde un enfoque práctico. Además existen otras vías de investigación entre las que están, por ejemplo, las ideas funcionales para trabajar con respuestas binarias, es decir, regresión logística funcional. Como referencia podemos ver [Morris \(2015\)](#).

- **Modelos Lineales Funcionales Generalizados (GFLM):** La idea principal es ampliar los GLM, pero ahora para datos funcionales. Aquí podemos citar a [Gareth \(2002\)](#); [Müller y Stadtmüller \(2005\)](#), donde el trabajo realizado consiste en aplicar Modelos lineales generalizados a situaciones donde la variable respuesta es escalar y las variables de predicción son observaciones funcionales. También podríamos hablar de Análisis de correlación canónica funcional (FCCA) en donde se puede citar el capítulo 4 del libro de [Horváth y Kokoszka \(2012\)](#).

- **Clasificación Funcional:** Independientemente de las divisiones entre métodos paramétricos y no-paramétricos, la clasificación engloba dos grandes ramas conocidas como clasificación supervisada y clasificación no supervisada.

El objetivo principal de la **Clasificación no supervisada**, también llamada Clustering, es dividir un conjunto de datos en un cierto número definido de grupos k , de forma que los miembros de cada grupo o cluster tengan algún tipo de similitud. Esta similitud vendrá definida por el algoritmo y la métrica utilizada. Clustering se suele utilizar cuando se sospecha de una división en subgrupos de un conjunto de datos y está encaminado a revelar dicha estructura y facilitar la comprensión del problema. Uno de los principales retos es la elección del número de grupos, y aunque algunos algoritmos aportan sugerencias para esta elección, el

problema principal será elegir un método, como lo pueden ser, K -medias o métodos Jerárquicos. Ambos enfoques han tenido su adaptación al entorno funcional. Aquí podemos citar a [Cuesta-Albertos y Fraiman \(2007\)](#); [Fraiman *et al.* \(2008\)](#).

Por otra parte, la **Clasificación supervisada** se aplica a problemas en los que k poblaciones están dadas. En este marco, se tiene disponible un conjunto de datos de entrenamiento, para los cuales sabemos a que clase o grupo pertenecen. Por lo tanto, la estructura de grupo se conoce a priori y el objetivo es conseguir una regla de decisión para clasificar nuevos datos de manera eficiente. En la siguiente sección se profundizará sobre el tema y se comentarán algunos de los métodos más utilizados en la actualidad.

1.5. Clasificación supervisada de datos funcionales

Entre los enfoques existentes en la literatura para Clasificación supervisada de datos funcionales, tenemos por ejemplo, [James y Hastie \(2001\)](#), en donde modelan los coeficientes con la distribución Gaussiana con matriz de covarianza común para todos los grupos. En [Hall y Presnell \(2001\)](#) maximizan la verosimilitud, y aunque proponen una estimación de la densidad totalmente no-paramétrica, en la práctica se consideran densidades Gaussianas multivariadas, que conduce al Análisis de discriminante cuadrático. [Biau *et al.* \(2003\)](#), aplican k -vecinos más cercanos a los coeficientes. [Ferraty y Vieu \(2003\)](#), clasifican las nuevas curvas en el grupo con mayor probabilidad posterior dada por el kernel estimado. Por otro lado, [Lopez-Pintado y Romo \(2006\)](#) tienen en cuenta la característica de continuidad de los datos y proponen dos métodos de clasificación basados en la idea de “profun-

didad” de las curvas. En [Alonso *et al.* \(2012\)](#), proponen un enfoque de distancias ponderadas, donde se proyectan los datos en un espacio finito-dimensional, esta técnica es llamada “filtrado”. [Cuesta-Albertos *et al.* \(2011\)](#), proporcionan algunos resultados teóricos sobre el clasificador funcional de k -vecinos para unos procesos Gaussianos en particular, y sugieren, como una respuesta parcial, que este método podría jugar el mismo papel central para datos funcionales como el método de Fisher para el caso finito dimensional. Por último, podemos mencionar el trabajo de [Sguera *et al.* \(2014\)](#), donde complementan ideas de profundidad espacial, dadas por [Chakraborty y Chaudhuri \(2014\)](#).

Abordaremos a continuación los fundamentos del problema de Clasificación supervisada en dimensión infinita. Hay muchos puntos de coincidencia con el Análisis de discriminante multivariado, pero hay importantes diferencias que hacen necesaria la adaptación de los métodos, tanto formalmente como en los aspectos computacionales.

Sea $D_n = \{(x_i, y_i), 1 \leq i \leq n\}$ el conjunto de entrenamiento con observaciones independientes, que siguen la distribución de la variable aleatoria $\mathcal{X} \in \mathcal{F}$, y $y_i \in Y$ la etiqueta de la i -ésima observación. El problema matemático es encontrar un **Clasificador** g ,

$$g : \mathcal{X} \longrightarrow Y, \tag{1.3}$$

que minimice el error de clasificación, $P(g_n(x_i) \neq y_i)$.

Supongamos que π_1 y π_2 son las probabilidades previas y las probabilidades condicionales de cada grupo como

$$f_i(x) = P(\mathcal{X} = x | Y = y_i), \quad i = 1, 2. \tag{1.4}$$

Aplicando el Teorema de Bayes, podemos encontrar las probabilidades posteriores

$$P(Y = 1|\mathcal{X} = x) = \frac{f_1(x)\pi_1}{f_1(x)\pi_1 + f_2(x)\pi_2}, \quad P(Y = 2|\mathcal{X} = x) = \frac{f_2(x)\pi_1}{f_1(x)\pi_1 + f_2(x)\pi_2}. \quad (1.5)$$

Para un valor de x dado, una forma razonable de clasificación es, asignar x a la clase con mayor probabilidad posterior. Esta estrategia es conocida como la **Regla de Bayes**. En otras palabras, asignamos x al grupo 1 si

$$\frac{P(Y = 1|\mathcal{X} = x)}{P(Y = 2|\mathcal{X} = x)} > 1, \quad (1.6)$$

y al grupo 2 en otro caso. Notemos que en la ecuación (1.5) los denominadores son iguales, entonces la regla de Bayes asigna x al grupo 1 si

$$\frac{f_1(x)}{f_2(x)} > \frac{\pi_2}{\pi_1}, \quad (1.7)$$

y al grupo 2 en otro caso. Por otra parte, si suponemos que las probabilidades previas son iguales, esto es suponer igual costo de clasificación, la regla de Bayes se reduce a asignar x al grupo 1 si

$$f_1(x) > f_2(x), \quad (1.8)$$

y al grupo 2 en otro caso. Para este caso, podemos ver un resultado importante en el Teorema 2.1, página 10 en [Devroye *et al.* \(1996\)](#), donde demuestran que la regla óptima de clasificación esta dada por

$$g^*(x) = \mathbb{I}_{\eta(x) > 1/2} \quad (1.9)$$

donde $\eta(x) = P(Y = 1|\mathcal{X} = x) = E(Y|\mathcal{X} = x)$. Dado que se desconoce la expresión exacta de η , la aplicación g es usualmente desconocida. Entonces el objetivo de todos los clasificadores que se construyan será acercarse lo más posible al clasificador óptimo de Bayes.

Los clasificadores $g_n(X)$ se construyen en la práctica a partir de la información del conjunto de entrenamiento. El error condicional de estos clasificadores es $L_n = P(g_n(X) \neq Y|D_n)$ y debe tender al óptimo $L^* = P(g^*(X) \neq Y)$. Así, el problema de clasificación puede ser visto como un problema de predicción, ya que el objetivo sería estimar la esperanza condicional.

Hay múltiples formas de crear estas aplicaciones, dos de las más comunes son las siguientes:

1. Los llamados métodos de plug-in sustituyen η y aplican el clasificador de Bayes con esta nueva η . De nuevo, al desconocer la distribución conjunta de $(X; Y)$ no se puede calcular el error del clasificador, teniendo que recurrir a un estimador llamado riesgo empírico, que se define así:

$$\tilde{L}_n(g_n) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{\{g_n(x_i) \neq y_i\}}. \quad (1.10)$$

2. A partir de esta expresión se pueden definir los clasificadores basados en riesgo. La idea es de una familia de clasificadores, encontrar el clasificador que minimice el riesgo empírico:

$$g_n^*(x) = \arg \min_{g_n \in \mathcal{C}} \tilde{L}_n(g_n). \quad (1.11)$$

1.5.1. Métodos destacados

Discriminante Lineal de Fisher

Supongamos que tenemos dos muestras tal que, $(\mu_0 \neq \mu_1, \Sigma_0 = \Sigma_1)$. En **Análisis Multivariado** clásico, el método se basa en una transformación lineal β y se asigna al individuo x a la clase cuya media proyectada sea la más cercana a su propia proyección $\beta^T x$. La proyección trata de maximizar la distancia entre las

medias y minimizar la varianza dentro de cada clase, sujeto a $\beta^T \Sigma \beta = 1$, donde Σ es la varianza total. De esta manera, el β óptimo se encuentra con la expresión:

$$\beta^* = \arg \max_{\beta \in \mathbb{R}^d} \frac{(\beta^T \mu_0 - \beta^T \mu_1)^2}{\beta^T \Sigma \beta}. \quad (1.12)$$

Para más detalle podemos consultar el capítulo 13 del libro de [Peña \(2002\)](#). En este caso, la solución es equivalente a minimizar la distancia de Mahalanobis y se puede concluir que x se asigna a la población 1 si: $(x - \mu_0)^T \Sigma^{-1} (x - \mu_0) < (x - \mu_1)^T \Sigma^{-1} (x - \mu_1)$. Notemos que la solución óptima de β , depende de la inversa de la matriz Σ , la cual es cada vez más difícil de calcular cuando d aumenta.

En **Análisis de Datos Funcionales**, el clasificador se obtiene proyectando x en \mathbb{R} y comparándolo con las proyecciones de las medias, donde $\mu_i = E(X|Y = i)$ y sus proyecciones $\langle \beta, \mu_i \rangle$. El objetivo vuelve a ser maximizar las distancias entre las medias, pero ahora la restricción es $\int_a^b \beta(t) \langle \beta, c(t, \cdot) \rangle ds = 1$, con $c(t, s) = \text{cov}(x(t), x(s))$. Siguiendo la misma idea, el β óptimo se encuentra de la siguiente forma:

$$\beta^* = \arg \max_{\beta \in \mathcal{F}} \frac{\text{var}(E(\langle \beta, x \rangle | Y))}{E(\text{var}(\langle \beta, x \rangle | Y))}. \quad (1.13)$$

El problema en dimensión infinita, es que no hay solución, ya que $c(s, t)$ no es invertible, así la aplicabilidad de la metodología lineal de Fisher para datos funcionales no es trivial. Algunas adecuaciones a este método para el caso funcional se pueden encontrar en los trabajos hechos por: [Alonso *et al.* \(2012\)](#); [James y Hastie \(2001\)](#); [Li y Yu \(2008\)](#).

KNN

El método de k -vecinos más cercanos es válido tanto para datos funcionales como en multivariado. Sea \mathcal{F} un espacio métrico, se clasifica x en función de los k

elementos de entrenamiento más cercanos según la métrica. El parámetro k juega en este método el papel de parámetro de suavizado y es hallado comúnmente por medio de validación cruzada, minimizando la función de pérdida. La única diferencia entre funcional y multivariado es la elección de una distancia de forma adecuada. KNN es un método ampliamente utilizado por su buen desempeño y sus resultados de consistencia ya existentes. Algunos trabajos alrededor de KNN son: [Baíllo y Cuevas \(2008\)](#); [Baíllo *et al.* \(2011\)](#); [Cuesta-Albertos *et al.* \(2011\)](#); [Cérou y Guyader \(2006\)](#).

Kernel

En los métodos tipo Kernel se fija una distancia h y “votan” todos los elementos que no disten más de ésta. En general, si definimos una función Kernel $K : [0, \infty) \rightarrow [0, \infty)$, el elemento x pertenecerá a la población 1 si:

$$\sum_{i=1}^n \mathbb{I}_{\{y_i=0\}} K \left[\frac{d(x, x_i)}{h} \right] > \sum_{i=1}^n \mathbb{I}_{\{y_i=1\}} K \left[\frac{d(x, x_i)}{h} \right]. \quad (1.14)$$

Las dificultades de este método radican en la elección de la distancia, similar a KNN y en la selección de h . Para ajustar este parámetro se puede realizar una validación cruzada que disminuya el error como en el caso anterior, sin embargo, la complicación es mayor al ser h un parámetro continuo.

Para el caso de dimensión finita se han conseguido resultados de consistencia fuerte universal. Sin embargo, el estudio de las condiciones suficientes de consistencia en el caso funcional es un problema reciente de estudio. [Abraham *et al.* \(2006\)](#) han ofrecido algunos resultados en cuanto a la consistencia débil. Otros trabajos realizados con este método se pueden ver en: [Charlie y Ventura \(2012\)](#) y en el capítulo 8 del libro de [Ferraty y Vieu \(2006\)](#).

Medidas de profundidad

La noción de profundidad fue desarrollada en primer lugar en el contexto multivariado para generar estadísticos de orden y listas ordenadas en altas dimensiones. Desde los primeros estudios de Mahalanobis en 1936, varios autores han propuesto y analizado distintas medidas de profundidad para datos multivariados. Se podría pensar en la generalización de estas medidas a los datos funcionales, sin embargo, esto puede conducir a medidas computacionalmente intratables o ineficientes. Así, en los últimos años han surgido formulaciones de medidas de profundidad específicas para datos funcionales. A continuación enumeramos las más destacadas en la literatura.

1. Fraiman y Muñiz (FMD), [Fraiman y Muñiz \(2001\)](#).
2. Moda- h (HMD), [Cuevas *et al.* \(2006\)](#).
3. Proyecciones Aleatorias (RPD), [Cuesta-Albertos y Nieto-Reyes \(2008\)](#).
4. Profundidad de Banda Generalizada (GBD), [Lopez-Pintado y Romo \(2009\)](#).
5. Profundidad Espacial, [Sguera *et al.* \(2014\)](#).

Después de calcular una medida de profundidad, los datos podrían ser ordenados. Con esto, se puede pensar en clasificar un nuevo dato en la clase entre cuyos representantes el nuevo elemento sea más profundo. Es decir, se puede clasificar definiendo,

$$g_n(x) = \mathbb{I}_{\{D_{n,0}(x) > D_{n,1}(x)\}}, \quad (1.15)$$

siendo $D_{n,i}(x)$ la estimación empírica de la medida de profundidad obtenida del conjunto de entrenamiento.

Capítulo 2

Boosting para datos finito-dimensionales

Los Algoritmos Boosting son métodos iterativos que combinan modelos sobre un mismo conjunto de datos de forma automática con el fin de obtener un modelo predictivo con un mejor rendimiento que los individuales. Los mismos han sido ampliamente estudiados en el campo de datos finito-dimensionales y han dado lugar a trabajos tanto experimentales como teóricos.

Boosting está basado en el cuestionamiento planteado por Kearns y Valiant en 1988, el cual fue: ¿puede un conjunto de clasificadores débiles crear un clasificador robusto?. Un clasificador débil está definido como un clasificador el cuál está solo débilmente correlacionado con la clasificación correcta (simplemente algo mejor que el azar). En contraste, un clasificador robusto es un clasificador que tiene un mejor desempeño que un clasificador débil, ya que sus clasificaciones se aproximan más a las verdaderas clases.

En 1990 Robert Schapire responde afirmativamente a la pregunta de Kearns y Valiant en el artículo que llevó como título *The Strength of Weak Learnability*, [Schapire \(1990\)](#). Dicha respuesta tuvo repercusiones significativas en el aprendi-

zaje automático y la Estadística, esta potente influencia llevó al desarrollo del Boosting. Más adelante, [Freund y Schapire \(1996\)](#), demostraron que se podía aumentar el rendimiento de cualquier clasificador, siempre que este sea mejor que el azar. Además, algo importante es que el error de prueba sigue disminuyendo, incluso después de que el error de entrenamiento llegue a cero. De este modo Boosting es muy resistente al sobreajuste.

En vista del éxito empírico, los estadísticos comenzaron a interesarse en estos métodos, por ejemplo, [Breiman \(1998\)](#) mostró que Boosting puede ser interpretado como una forma de descenso de gradiente en el espacio funcional. Este punto de vista se amplió en [Friedman *et al.* \(2000\)](#), que mostraron cómo Boosting podría extenderse para manejar una variedad de funciones de pérdida, incluyendo regresión, regresión robusta y regresión de Poisson. En trabajos más recientes como [Buhlmann y Hothorn \(2007\)](#); [Freund y Schapire \(2012\)](#) muestran cómo se generan los diferentes algoritmos Boosting.

Boosting es un algoritmo útil para ajustar modelos aditivos. La función de predicción tiene la forma $f(\mathbf{x}) = \sum_{m=1}^M \alpha_m \phi_m(\mathbf{x})$, donde las α_m 's son las ponderaciones para cada modelo, calculadas por el algoritmo. Su efecto es dar mayor peso a los modelos más precisos. Por otra parte, en cada iteración los datos se van ponderando, proporcionando mayor peso a los datos que fueron mal clasificados en la iteración anterior. Estos modelos de predicción pueden ser clasificadores o modelos de regresión. Para nuestro caso, de aquí en adelante los mencionaremos como clasificadores.

Una característica singular de Boosting es su resistencia al sobreajuste, siendo capaz de seguir añadiendo clasificadores débiles y mejorando su desempeño sobre el conjunto de prueba, incluso cuando el error de entrenamiento se ha hecho nulo. Este comportamiento ha sido analizado por varios autores; por ejemplo

Buhlmann y Hothorn (2007) estudian a detalle el comportamiento del parámetro M_{stop} (número de iteraciones Boosting), concluyendo que es un parámetro de regularización, el cual se puede estimar por algún método de validación.

Ahora bien, el objetivo de Boosting es resolver el problema de optimización,

$$\min_f \sum_{i=1}^n L(y_i, f(x_i)) \quad (2.1)$$

donde $L(y, \cdot)$ es alguna función de pérdida y f el modelo aditivo. La función objetivo de la expresión (2.1) puede ser considerada como una estimación de la pérdida esperada,

$$\mathbb{E}_{(x,y) \sim D} [L(yf(x))]. \quad (2.2)$$

Sea $p(x) = \mathbb{P}[y = 1|x]$ la probabilidad condicional, supondremos que $p(x)$ existe y que la función f es arbitraria. Por lo tanto, la esperanza de la expresión (2.2) puede reescribirse condicionando con respecto a x ,

$$\mathbb{E}_x [\mathbb{E}_y [L(yf(x))|x]] = \mathbb{E}_x [p(x)L(f(x)) + (1 - p(x))L(-f(x))]. \quad (2.3)$$

Así, elegir la función óptima f es buscar una función que minimice,

$$p(x)L(f(x)) + (1 - p(x))L(-f(x)). \quad (2.4)$$

Es decir,

$$p(x)L'(f(x)) - (1 - p(x))L'(-f(x)) = 0. \quad (2.5)$$

Dependiendo de la elección de la función de pérdida podremos encontrar una forma cerrada para la función f óptima (f^*). En general, si resolvemos la ecuación anterior con respecto a $p(x)$ obtenemos:

$$p(x) = \frac{1}{1 + \frac{L'(f(x))}{L'(-f(x))}}. \quad (2.6)$$

Para más detalles podemos ver [Friedman *et al.* \(2000\)](#). Observemos que la función f^* depende de la probabilidad condicional, la cual no es posible calcularla exactamente. Por lo tanto, es indispensable utilizar algún método de optimización.

2.1. Métodos de optimización

Retomando el problema planteado en la expresión (2.1), y reemplazando f como un modelo aditivo, podemos ver que el problema se convierte en:

$$\min_{\alpha, \phi} \sum_{i=1}^n L \left(y_i, \sum_{m=1}^M \alpha \phi(x_i) \right). \quad (2.7)$$

En la literatura observamos que los métodos más utilizados para optimizar este tipo de funciones son:

- *Forward stagewise additive modeling* (FSAM).
- *Functional gradient descent* (FGD).
- *Newton-like updates* (NU).

2.1.1. *Forward stagewise additive modeling* (FSAM)

El algoritmo FSAM aproxima la solución de la expresión (2.7) secuencialmente adicionando nuevas funciones a la expansión sin ajustar los parámetros y coeficientes que ya se han añadido. En cada iteración m , encontramos la función óptima ϕ_m y su correspondiente coeficiente α_m que minimice cierta función de pérdida. Esta actualización sumada con el proceso realizado en la etapa anterior $f_{m-1}(x)$, producen el nuevo $f_m(x)$. El proceso se puede resumir como se muestra en el [Algoritmo 1](#). Para más detalles ver sección 10.2 del libro de [Hastie *et al.* \(2008\)](#).

-
1. Inicie $f_0(x) = 0$.
 2. Para $m = 1, \dots, M$:

a) Calcular

$$(\alpha_m, \phi_m) = \arg \min_{\alpha, \phi} \sum_{i=1}^n L(y_i, f_{m-1}(x_i) + \alpha \phi(x_i)).$$

b) Actualizar $f_m(x) = f_{m-1}(x) + \alpha_m \phi_m(x)$.

3. Salida, $f_M(x)$.
-

Algoritmo 1: *Forward stagewise additive modeling*

2.1.2. *Functional gradient descent (FGD)*

El método anterior es bastante útil y general, pero puede volverse un problema difícil a la hora de encontrar la función y su respectivo coeficiente para algunas funciones de pérdida. Un segundo método para optimizar es FGD, en donde calcular el gradiente de la función lo hace un proceso más sencillo y útil para diferentes funciones de pérdida. Este proceso consiste en encontrar el gradiente de la función de pérdida evaluado en $f = f_{m-1}$, para luego ajustar ϕ_m por medio de un método de regresión. En el [Algoritmo 2](#) describimos su proceso. Para más detalle podemos ver la sección 16.4.5 del libro de [Murphy \(2012\)](#).

Este procedimiento es bastante sencillo, aunque el problema principal ahora radica en ajustar el tamaño del parámetro α , el cual es complicado, ya que si usamos una tasa demasiado pequeña, la convergencia será muy lenta; pero si lo hacemos demasiado grande, puede fallar la convergencia del algoritmo.

-
1. Inicie $f_0(x) = 0$.
 2. Para $m = 1, \dots, M$:

- a) Calcular el gradiente

$$\nabla L(y, f(x)) = \left. \frac{\partial L(y, f(x))}{\partial f(x)} \right|_{f(x)=f_{m-1}(x)}.$$

- b) Encontrar

$$\phi_m = \arg \min_{\phi} \sum_{i=1}^n (\nabla L(y, f_{m-1}(x_i)) - \phi_m(x))^2.$$

- c) Elegir $\alpha_m > 0$ adecuadamente.

- d) Actualizar $f_m(x) = f_{m-1}(x) + \alpha_m \phi_m(x)$

3. Salida, $f_M(x)$.
-

Algoritmo 2: *Functional gradient descent*

2.1.3. *Newton-like updates (NU)*

Otro de los enfoques de optimización se basan en la curvatura del espacio. Estos métodos de optimización son llamados de segundo orden. El representante de este tipo de métodos es el algoritmo de Newton, el cual es un algoritmo iterativo que consiste en cambios de la forma:

$$f_{m+1}(x) = f_m(x) - \alpha_m H_m^{-1}(x) S_m(x), \quad (2.8)$$

donde $S_m(x)$ es la primera derivada de la función objetivo y $H_m(x)$ es la segunda derivada, ambas evaluadas en $f(x) = 0$. Este método requiere que $H_m(x)$ sea

definida positiva, lo que se consigue si la función objetivo es estrictamente convexa. En el [Algoritmo 3](#) podemos observar su mecanismo. Para más detalles podemos ver la sección 8.3.3 del libro de [Murphy \(2012\)](#)

-
1. Inicie $f_0(x) = 0$.
 2. Para $m = 1, \dots, M$:
 - a) Evaluar $S_m = \nabla L(y, f(x))$.
 - b) Hallar $H_m = \nabla^2 L(y, f(x))$.
 - c) Actualizar $f_m(x) = f_{m-1}(x) - \frac{S_m}{H_m}$.
 3. Salida, $f_M(x)$.
-

Algoritmo 3: *Newton-like updates*

Este método puede no ser adecuado si $H_m(x) \rightarrow 0$, pues la convergencia podría ser muy lenta.

2.2. Funciones de pérdida

La regla para discriminar es $F(x) = \text{signo}[f(x)]$. Esto implica que las observaciones con margen positivo ($yf(x) \geq 0$) son clasificadas correctamente, mientras que las observaciones con margen negativo ($yf(x) < 0$) estarán mal clasificadas. La finalidad es producir la mayor cantidad de márgenes positivos, y cualquier criterio de pérdida utilizado para la clasificación debería penalizar en mayor medida a los márgenes negativos, como medida de corrección. En la [tabla 2.1](#) podemos observar algunas funciones de pérdida y en la [figura 2.1](#) sus respectivos comportamientos.

Nombre	Función de pérdida	Rango
Exponencial	$e^{-yf(x)}$	$y \in \{-1, 1\}$
Logística	$\log(1 + e^{-yf(x)})$	$y \in \{-1, 1\}$
Devianza Binomial	$\log(1 + e^{-2yf(x)})$	$y \in \{-1, 1\}$
Error cuadrático	$(y - f(x))^2$	$y \in \{-1, 1\}$
Vector soporte	$(1 - yf(x))_+$	$y \in \{-1, 1\}$
Log-verosimilitud Binomial	$2yf(x) - \log(1 + e^{-2f(x)})$	$y \in \{0, 1\}$

Tabla 2.1: Algunas funciones de pérdida destacadas en la literatura de Boosting.

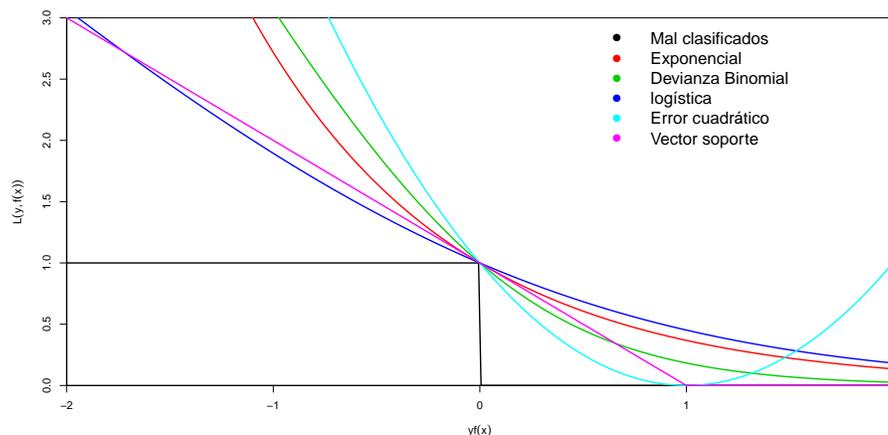


Figura 2.1: Comportamiento de las funciones de pérdida para un problema de clasificación binario.

2.3. Algoritmos Boosting

Como ya mencionamos, la idea principal de Boosting es combinar los resultados de varios clasificadores débiles para obtener un clasificador robusto. Cuando se añaden estos clasificadores débiles, se hace de modo que estos tengan diferente

peso en función de la exactitud de sus predicciones. Luego de que se añade un clasificador débil, los datos cambian su estructura de peso, de forma que los casos que son mal clasificados aumenten su peso y los que son clasificados correctamente disminuyan su peso. Así, el nuevo clasificador débil que se agregue, se centrará en los casos que fueron mal clasificados por el clasificador débil anterior.

En la literatura existen diferentes algoritmos Boosting, los cuales fueron apareciendo dependiendo tanto de la función de pérdida como del método de optimización que se utilice. A continuación expondremos brevemente los algoritmos más reconocidos en la literatura.

2.3.1. AdaBoost

Freund y Schapire desarrollaron el algoritmo AdaBoost. Éste es considerado uno de los más importantes algoritmos en Boosting, el cual ganó el prestigioso Premio Gödel en el año 2003. AdaBoost fue la primera formulación de un algoritmo que pudo aprender a partir de clasificadores débiles. Este algoritmo surge de tomar la función de pérdida exponencial y aplicar el método de optimización FSAM. Por lo tanto, con este método de optimización, en la iteración m del algoritmo debemos encontrar:

$$(\alpha_m, \phi_m) = \arg \min_{\alpha, \phi} \sum_{i=1}^n \exp[-y_i(f_{m-1}(x_i) + \alpha\phi(x_i))], \quad (2.9)$$

tanto α_m como ϕ_m son agregados a $f_{m-1}(x)$ para actualizar $f_m(x)$. Notemos que podemos reescribir esta expresión así:

$$(\alpha_m, \phi_m) = \arg \min_{\alpha, \phi} \sum_{i=1}^n w_m(x_i) \cdot e^{-y_i\alpha\phi(x_i)}, \quad (2.10)$$

donde $w_m(x_i) = e^{-y_i f_{m-1}(x_i)}$, la cual denominamos la ponderación de cada iteración m . Observemos que:

$$\begin{aligned} \sum_{i=1}^n w_m(x_i) \cdot e^{-y_i \alpha \phi(x_i)} &= e^{\alpha} \sum_{y_i \neq \phi(x_i)} w_m(x_i) + e^{-\alpha} \sum_{y_i = \phi(x_i)} w_m(x_i) \\ &= (e^{\alpha} - e^{-\alpha}) \sum_{i=1}^n w_m(x_i) \cdot \mathbb{I}(y_i \neq \phi(x_i)) + e^{-\alpha} \sum_{i=1}^n w_m(x_i). \end{aligned} \quad (2.11)$$

Por lo tanto, para cualquier $\alpha > 0$ la función ϕ óptima es:

$$\phi_m = \arg \min_{\phi} [w_m(x_i) \cdot \mathbb{I}(y_i \neq \phi(x_i))]. \quad (2.12)$$

Sustituyendo ϕ_m en la expresión (2.11) y resolviendo para α tenemos:

$$\alpha_m = \frac{1}{2} \log \left(\frac{1 - err_m}{err_m} \right), \quad (2.13)$$

donde,

$$err_m = \sum_{i=1}^n w_m(x_i) \cdot \mathbb{I}(y_i \neq \phi(x_i)). \quad (2.14)$$

Finalmente, los pesos se actualizarán, $w_{m+1}(x) = w_m(x) \cdot e^{-y_i \alpha_m \phi_m(x_i)}$. Estos se normalizarán en cada iteración. Por lo tanto,

$$w_{m+1}(x) = \frac{w_m(x) \cdot e^{-y_i \alpha_m \phi_m(x_i)}}{\sum_{i=1}^n w_m(x)}. \quad (2.15)$$

En el [Algoritmo 4](#) describimos su proceso.

El problema con la pérdida exponencial es que el peso puede aumentar muy rápido para un dato difícil de clasificar. Esto hace que el método sea muy sensible a valores atípicos.

-
1. Inicie $w_1(x_i) = 1/n$.
 2. Para $m = 1, \dots, M$:
 - a) Ajuste un clasificador ϕ_m a la muestra de entrenamiento ponderada.
 - b) Calcule $err_m = \sum_{i=1}^n w_m(x_i) \cdot \mathbb{I}_{\{y_i \neq \phi_m(x_i)\}}$.
 - c) Si $err_m < 0.5$, calcule $\alpha_m = \frac{1}{2} \log \left(\frac{1-err_m}{err_m} \right)$.
 - d) Actualice $w_{m+1}(x_i) = \frac{w_m(x_i) \cdot e^{-y_i \alpha_m \phi_m(x_i)}}{\sum_{i=1}^n w_m(x_i)}$.
 3. $F(x) = \text{signo} \left[\sum_{m=1}^M \alpha_m \phi_m(x) \right]$.
-

Algoritmo 4: AdaBoost

2.3.2. LogitBoost

En la teoría del aprendizaje computacional, LogitBoost es un algoritmo formulado por Friedman, Hastie y Tibshirani. En el documento original de [Friedman *et al.* \(2000\)](#), podemos observar como se deriva el algoritmo LogitBoost. Este algoritmo Boosting surge de tomar la función de pérdida log-verosimilitud Binomial y el método de optimización NU. Sabemos que la log-verosimilitud Bernoulli se puede escribir como:

$$l(y, p(x)) = y \log(p(x)) + (1 - y) \log(1 - p(x)), \quad y \in \{0, 1\}. \quad (2.16)$$

Parametrizando la probabilidad: $p(x) = \mathbb{P}[y = 1|x] = \frac{1}{1 + e^{-2f(x)}}$, podemos

reescribir la ecuación anterior,

$$\begin{aligned}
l(y, p(x)) &= y \log \left[\frac{1}{1 + e^{-2f(x)}} \right] + (1 - y) \log \left[1 - \frac{1}{1 + e^{-2f(x)}} \right] \\
&= -y \log(1 + e^{-2f(x)}) + (1 - y) \log \left[\frac{e^{-2f(x)}}{1 + e^{-2f(x)}} \right] \\
&= 2yf(x) - \log(1 + e^{2f(x)}). \tag{2.17}
\end{aligned}$$

donde $f(x)$ es la función de predicción. Ahora, si calculamos el valor esperado de la log-verosimilitud y consideramos la actualización $f(x) + \phi(x)$ tenemos:

$$\mathbb{E}[l(y, f(x) + \phi(x))] = \mathbb{E}[2y(f(x) + \phi(x)) - \log(1 + e^{2(f(x) + \phi(x))})]. \tag{2.18}$$

Condicionando con respecto a x , podemos calcular la primera y segunda derivada evaluada en $\phi(x) = 0$. Términos necesarios para aplicar el método de optimización NU,

$$\begin{aligned}
S(x) &= \left. \frac{\partial \mathbb{E}[l(y, f(x) + \phi(x))]}{\partial \phi(x)} \right|_{\phi(x)=0} \\
&= 2\mathbb{E}[y - p(x)|x] \tag{2.19}
\end{aligned}$$

$$\begin{aligned}
H(x) &= \left. \frac{\partial^2 \mathbb{E}[l(f(x) + \phi(x))]}{\partial \phi(x)^2} \right|_{\phi(x)=0} \\
&= -4\mathbb{E}[p(x)(1 - p(x))|x] \tag{2.20}
\end{aligned}$$

Por lo tanto, la actualización para la m -ésima iteración es:

$$\begin{aligned}
f_m(x) &= f_{m-1}(x) - \frac{S_m}{H_m} \\
&= f_{m-1}(x) + \frac{1}{2} \frac{\mathbb{E}[y - p(x)|x]}{\mathbb{E}[p(x)(1 - p(x))|x]} \\
&= f_{m-1}(x) + \frac{1}{2} \mathbb{E}_w \left[\frac{y - p(x)}{p(x)(1 - p(x))} \middle| x \right], \tag{2.21}
\end{aligned}$$

donde $w(x) = p(x)(1 - p(x))$. Equivalentemente, la actualización de $\phi_m(x)$ por el método de Newton resuelve la aproximación de mínimos cuadrados ponderados sobre $f_{m-1}(x)$ de la log-verosimilitud. Por lo tanto,

$$\phi_m(x) = \arg \min_{\phi} \mathbb{E}_w \left[f_{m-1}(x) + \frac{1}{2} \frac{(y - p(x))}{(p(x)(1 - p(x)))} - (f_{m-1}(x) + \phi(x)) \right]^2. \quad (2.22)$$

Finalmente, actualizamos:

$$f_{m+1}(x_i) = f_m(x_i) + \frac{1}{2} \phi_m(x_i), \quad (2.23)$$

y las probabilidades condicionales

$$p_{m+1}(x_i) = \frac{1}{1 + e^{-2f_{m+1}(x_i)}} \quad (2.24)$$

El proceso descrito se puede ver en el [algoritmo 5](#).

En algunas ocasiones $w_m(x)$ puede obtener valores muy pequeños, cuando $p(x)$ está cerca de 0 o 1. Esto puede causar problemas numéricos en la construcción de z (ver [algoritmo 5](#)), y dar lugar a la siguiente implementación.

Si $y = 1$, entonces $z = 1/p$. Dado que este número puede ser muy grande, si $p(x)$ es pequeño, el umbral de esta relación es z máx. El valor particular elegido para z máx no es crucial; [Friedman et al. \(2000\)](#) encontraron empíricamente que z máx $\in [2, 4]$ funciona bien. Del mismo modo, si $y = 0$, $z = -1/(1 - p(x))$ con un umbral más bajo que $-z$ máx. Por lo tanto, hay que forzar un umbral si en algún momento los pesos son muy pequeños, $w = \text{máx}(w, u_0)$, con u_0 por ejemplo de 10^{-6} .

2.3.3. Binomial Boosting

Este algoritmo es una variante de LogitBoost, surge de tomar de igual manera la función de pérdida log-verosimilitud Binomial pero ahora con el método de opti-

-
1. Inicie $p_1(x_i) = 1/2$ y $f_1(x_i) = 0$.
 2. Para $m = 1, \dots, M$:
 - a) Calcule la variable respuesta $z_m(x_i) = \frac{y_i - p_m(x_i)}{p_m(x_i)(1 - p_m(x_i))}$.
 - b) Calcule la ponderación $w_m(x_i) = p_m(x_i)(1 - p_m(x_i))$.
 - c) Encuentre $\phi_m(x_i) = \arg \min_{\phi} \sum_{i=1}^n w_m(x_i) (z_m(x_i) - \phi(x_i))^2$.
 - d) Actualice $f_{m+1}(x_i) = f_m(x_i) + \frac{1}{2}\phi_m(x_i)$.
 - e) Actualice $p_{m+1}(x_i) = \frac{1}{1 + e^{-2f_{m+1}(x_i)}}$.
 3. $F(x) = \text{signo} \left[\sum_{m=1}^M f_m(x) \right]$.
-

Algoritmo 5: LogitBoost

mización FGD. En el artículo de [Buhlmann y Hothorn \(2007\)](#) explican su estructura. Recordamos de la expresión (2.17), la función de pérdida log-verosimilitud Binomial y calculando nuevamente la esperanza obtenemos:

$$\mathbb{E} [l(y, f(x))] = \mathbb{E} [2yf(x) - \log(1 + e^{2f(x)})]. \quad (2.25)$$

Condicionando con respecto a x , calculamos la primera derivada evaluada en $f(x) = f_{m-1}(x)$,

$$\left. \frac{\partial \mathbb{E} [l(y, f(x))]}{\partial f(x)} \right|_{f(x)=f_{m-1}(x)} = 2\mathbb{E}_p [y - p(x) | x]. \quad (2.26)$$

Por lo tanto, el enfoque del Gradiente actualiza $\phi_m(x)$ en la m -ésima iteración, equivalentemente resolviendo la aproximación de mínimos cuadrados ponderados

sobre $f_{m-1}(x)$ de la log-verosimilitud.

$$\phi_m(x_i) = \arg \min_{\phi} \mathbb{E}_p [f_{m-1}(x) + 2(y - p(x)) - (f_{m-1}(x) + \phi(x))]^2. \quad (2.27)$$

Finalmente, actualizamos

$$f_m(x) = f_{m-1}(x) + \phi_m(x). \quad (2.28)$$

y las probabilidades condicionales

$$p_{m+1}(x_i) = \frac{1}{1 + e^{-2f_{m+1}(x_i)}} \quad (2.29)$$

En el [algoritmo 6](#) podemos observar su proceso.

-
1. Inicie $p_1(x_i) = 1/2$ y $f_1(x_i) = 0$.
 2. Para $m = 1, \dots, M$:
 - a) Calcule la variable respuesta $z_m(x_i) = y_i - p_m(x_i)$.
 - b) Encuentre $\phi_m(x_i) = \arg \min_{\phi} \sum_{i=1}^n p_m(x_i) (z_m(x_i) - \phi(x_i))^2$.
 - c) Actualice $f_{m+1}(x_i) = f_m(x_i) + \phi_m(x_i)$.
 - d) Actualice $p_{m+1}(x_i) = \frac{1}{1 + e^{-2f_{m+1}(x_i)}}$.
 3. $F(x) = \text{signo} \left[\sum_{m=1}^M f_m(x) \right]$.
-

Algoritmo 6: Binomial Boosting

2.3.4. Ponderación Alternativa

Este nuevo algoritmo propuesto en esta tesis, no está diseñado con la misma idea que las anteriores, ya que no depende de una función de pérdida o del método

de optimización. Este algoritmo fue diseñado a partir de las ideas presentadas en [Charlie y Ventura \(2012\)](#), donde proponen un algoritmo llamado **Boosted KNN**. La idea principal de nuestra propuesta es incorporar mejoras a las deficiencias que presentaba este algoritmo, mencionadas por los mismos autores. Además de generalizar su funcionamiento para poder aplicar cualquier otro clasificador distinto a KNN, (Clasificador con el cual ésta implementado el algoritmo original).

El algoritmo Ponderación Alternativa sigue la misma estructura de AdaBoost, pero tiene un cambio sustancial en la estructura de las ponderaciones, donde juega un papel importante la noción de cercanía, la cual en el espacio finito dimensional será la distancia Euclidiana, pero en el espacio infinito dimensional dependerá de la métrica utilizada. A continuación explicaremos su procedimiento.

Dado un elemento x_j , calculamos una función de distancia, que denominamos función de disimilitud, a cada las observación x_i con la función:

$$D(x_j, x_i) = \frac{1}{[1 + e^{-w(x_i)}] d(x_j, x_i)}, \quad (2.30)$$

donde $w(x_i)$ es la ponderación de la observación x_i y $d(x_j, x_i)$ es la distancia entre las dos observaciones.

Calculamos el máximo $D(x_j, \cdot)$, que denominamos $Dmax_j$, el cual actuará en nuestro algoritmo como la ponderación para cada observación. Luego, ajustamos un clasificador ϕ a la muestra ponderada por el vector $Dmax$, encontrando una primera discriminación.

De acuerdo a este resultado, buscamos el conjunto S_q , que ésta formado por los elementos mal clasificados, $S_q = \{x_q : \phi(x_q) \neq y_q\}$. Los elementos de este conjunto serán las observaciones a las cuales se les modifique su peso en cada

iteración. Primero, construimos vectores de pesos para cada elemento x_q :

$$w_{\{m+1, x_q\}} = \begin{cases} w_m - \frac{\lambda}{d(x_q, x_i)}, & \text{si } y_q \neq y_i \\ w_m + \frac{\lambda}{d(x_q, x_i)}, & \text{si } y_q = y_i, \end{cases}$$

para finalmente actualizar los pesos así:

$$w_{m+1}(x_q) = \max_{x_q} w_{\{m+1, x_q\}}. \quad (2.31)$$

En el [Algoritmo 7](#) podemos ver su estructura.

1. Inicie $w_1(x_i) = 0$.
2. Para $m = 1, \dots, M$:
 - a) Calcule el vector $Dmax$.
 - b) Ajuste un clasificador ϕ_m a la muestra de entrenamiento ponderada por $Dmax$.
 - c) Encuentre $S_q = \{x_q : \phi_m(x_q) \neq y_q\}$.
 - d) Para los elementos de S_q , calcule

$$w_{\{m+1, x_q\}} = \begin{cases} w_m - \frac{\lambda}{d(x_q, x_i)}, & \text{si } \phi_m(x_i) \neq y_i \\ w_m + \frac{\lambda}{d(x_q, x_i)}, & \text{si } \phi_m(x_i) = y_i \end{cases}$$

- e) Actualice $w_{m+1}(x_q) = \max_{x_q} w_{\{m+1, x_q\}}$.

3. $F(x) = \text{signo} \left[\sum_{m=1}^M \phi_m(x) \right]$.

Algoritmo 7: Ponderación Alternativa

El parámetro λ juega un papel que se podría interpretar como un ancho de banda. Los autores del artículo destacan la falta de estudio de este parámetro y lo proponen como un trabajo futuro. Nosotros analizamos su comportamiento y conseguimos estimarlo por medio de validación cruzada obteniendo resultados favorables.

En el siguiente Capítulo mostraremos a detalle el trabajo realizado sobre Boosting para datos funcionales, el cual consiste en la extensión de estos métodos para aplicarlos a datos infinito dimensionales.

Capítulo 3

Boosting para datos infinito-dimensionales

Consideramos un problema relevante de investigación trabajar en el mejoramiento del desempeño de clasificadores para datos funcionales. Un ejemplo de motivación a este problema, lo presentamos en [Figura 3.1](#). Son trayectorias de un estudio de neuroimagen de pacientes con Esclerosis Múltiple (EM); estas trayectorias son representadas por medio de imágenes de tensores de difusión. Para más detalles ver [Sørensen *et al.* \(2013\)](#). Observamos en este conjunto de datos reales la complejidad presente para clasificar. Como parte del análisis de estos datos, hemos aplicado diferentes métodos de clasificación funcional, obteniendo muy malos resultados, lo que hace imposible obtener una predicción aceptable.

Una forma de atacar el problema fue considerando la combinación de diferentes clasificadores de modo que en forma grupal el error de clasificación sea menor, o un solo clasificador que tenga algún método de aprendizaje de tal forma que al aplicarlo iterativamente mejore su desempeño. Esto nos llevo a pensar en cómo adecuar los algoritmos Boosting para obtener su aplicabilidad en datos funcionales. Este problema lo analizamos desde dos puntos de vista. Primero, adaptamos algunos

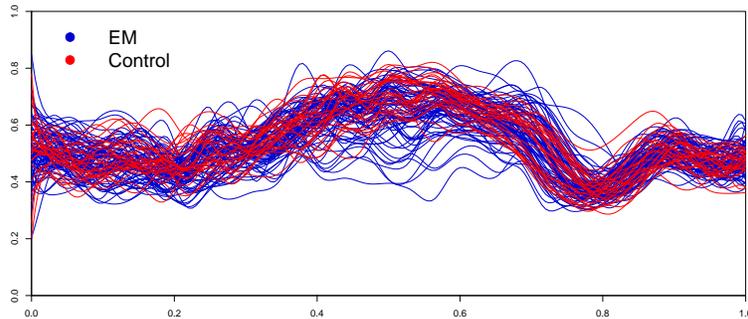


Figura 3.1: 74 pacientes con EM en color azul y 26 pacientes control en control rojo.

clasificadores funcionales logrando incorporarlos en ciertos algoritmos Boosting. Segundo, modificamos algunos algoritmos Boosting extendiendo su aplicabilidad a datos funcionales. A continuación discutiremos los resultados obtenidos de estas dos investigaciones.

Es importante resaltar que la literatura de Boosting para datos funcionales es bastante reducida; las únicas obras conocidas encontradas es [Krämer \(2006\)](#) en problemas de clasificación, donde expone la idea de discretizar los datos funcionales, para utilizar su estructura de forma multivariada y de esta manera poder aplicar los algoritmos Boosting. Este trabajo tan solo es un reporte técnico y no se ha podido determinar si en algún momento fue publicado. Además de éste hay dos artículos de [Fernandez *et al.* \(2005\)](#); [Ferraty y Vieu \(2009\)](#) en aplicaciones de regresión funcional.

3.1. Boosting con Clasificadores funcionales

Consideramos tres clasificadores funcionales de la literatura de FDA. Estos clasificadores se tomaron por su alto reconocimiento en la literatura, y principalmente

por tener mecanismos naturales en su desarrollo, los cuales nos permitieron adaptarlos para nuestro objetivo. Para ello, fue necesario implementar una forma útil de ponderar los datos de entrenamiento, y así lograr incorporarlos en dos algoritmos Boosting, (AdaBoost y Ponderación alternativa) estudiando su desempeño. Finalmente, nuestras conclusiones para este primer trabajo están apoyadas por un estudio de simulación, analizado en dos escenarios diferentes y se ilustran además con dos conjuntos de datos Benchmark.

3.1.1. Clasificadores funcionales ponderados

Los tres métodos de clasificación que utilizamos fueron:

- Distancias ponderadas, [Alonso *et al.* \(2012\)](#).
- KNN, capítulo 8 del libro de [Ferraty y Vieu \(2006\)](#).
- Profundidad de Banda Generalizada, [Lopez-Pintado y Romo \(2009\)](#).

Veamos a continuación como implementamos la ponderación en cada clasificador, así como su funcionamiento. Es importante resaltar que utilizamos la misma nomenclatura que usaron los autores en cada clasificador, y por ello en el escrito se redefinen variables de diferentes formas.

Clasificador 1. Distancias ponderadas:

Sean f_1, \dots, f_{n_1} y g_1, \dots, g_{n_2} muestras funcionales de dos poblaciones.

1. Transformamos los datos funcionales a multivariados: Para cada f_j se construye el siguiente vector:

$$\mathbf{x}_j^{(f)} = \left(x_{j1}^{(f)}, x_{j2}^{(f)}, x_{j3}^{(f)} \right)^T, \quad (3.1)$$

donde,

$$x_{ji}^{(f)} = d\left(f_j^{(i-1)}, \bar{f}^{(i-1)}\right) - d\left(f_j^{(i-1)}, \bar{g}^{(i-1)}\right), \quad (3.2)$$

para $i = 1, 2, 3$ y $\bar{f}^{(i-1)} = \frac{1}{n_1} \sum_{j=1}^{n_1} f_j^{(i-1)}$ y $\bar{g}^{(i-1)} = \frac{1}{n_2} \sum_{j=1}^{n_2} g_j^{(i-1)}$. Es decir, cada $x_{ji}^{(f)}$ es la diferencia entre las distancias de $f_j^{(i-1)}$ a la $(i-1)$ -ésima derivada de las medias. Estos vectores constituyen la muestra multivariada,

$$\left(\mathbf{x}_1^{(f)}, \dots, \mathbf{x}_{n_1}^{(f)}\right) = \begin{pmatrix} x_{11}^{(f)} & \cdots & x_{n_1 1}^{(f)} \\ x_{12}^{(f)} & \cdots & x_{n_1 2}^{(f)} \\ x_{13}^{(f)} & \cdots & x_{n_1 3}^{(f)} \end{pmatrix}. \quad (3.3)$$

La muestra multivariada $\left(\mathbf{x}_1^{(g)}, \dots, \mathbf{x}_{n_2}^{(g)}\right)$ se define de manera análoga.

2. La función discriminante ponderada: Estas muestras se utilizan como entrada en el problema de optimización para obtener la función discriminante:

$$\hat{y}(\mathbf{x}) = \hat{\mathbf{a}}^T \mathbf{x}. \quad (3.4)$$

El discriminante $\hat{\mathbf{a}}$ se obtiene al resolver:

$$\hat{\mathbf{a}} = \arg \max_a \left\{ \frac{a^T \mathbf{B} a}{a^T \mathbf{W} a} \right\}, \quad \text{con } a^T \mathbf{W} a = 1. \quad (3.5)$$

Dicho valor esta dado por $\hat{\mathbf{a}}^T = (\bar{\mathbf{x}}^{(f)} - \bar{\mathbf{x}}^{(g)})^T \mathbf{W}^{*-1}$. La matriz de variabilidad entre las clases \mathbf{B} es:

$$\mathbf{B} = n_1 (\bar{\mathbf{x}}^{(f)} - \bar{\mathbf{x}}) (\bar{\mathbf{x}}^{(f)} - \bar{\mathbf{x}})^T + n_2 (\bar{\mathbf{x}}^{(g)} - \bar{\mathbf{x}}) (\bar{\mathbf{x}}^{(g)} - \bar{\mathbf{x}})^T. \quad (3.6)$$

La Matriz \mathbf{W}^* es la matriz de covarianza modificada. Es aquí donde implementamos una forma útil de ponderar este clasificador, la cual obtenemos con la expresión:

$$\begin{aligned} \mathbf{W}^* = & \sum_{j=1}^{n_1} \frac{1}{w_m(\mathbf{x}_j^{(f)})} \left(\mathbf{x}_j^{(f)} - \bar{\mathbf{x}}^{(f)}\right) \left(\mathbf{x}_j^{(f)} - \bar{\mathbf{x}}^{(f)}\right)^T \\ & + \sum_{k=1}^{n_2} \frac{1}{w_m(\mathbf{x}_k^{(g)})} \left(\mathbf{x}_k^{(g)} - \bar{\mathbf{x}}^{(g)}\right) \left(\mathbf{x}_k^{(g)} - \bar{\mathbf{x}}^{(g)}\right)^T, \end{aligned} \quad (3.7)$$

donde $w_m(\mathbf{x}_j)$ es la ponderación para cada \mathbf{x}_j .

3. Clasificación de un nuevo dato: Para clasificar una nueva curva h , se construye su vector multivariado,

$$\mathbf{x}^{(h)} = \left(x_1^{(h)}, x_2^{(h)}, x_3^{(h)} \right)^T, \quad (3.8)$$

con $x_i^{(h)} = d(h^{(i-1)}, \bar{f}^{(i-1)}) - d(h^{(i-1)}, \bar{g}^{(i-1)})$. Finalmente, el valor $\hat{y}(\mathbf{x}^{(h)})$ se utiliza para asignar la curva h a una de las dos poblaciones, donde la regla de clasificación es:

$$\begin{cases} k = 1 & \text{si } \hat{y}(\mathbf{x}^{(h)}) > \frac{1}{2} (\bar{\mathbf{x}}^{(f)} - \bar{\mathbf{x}}^{(g)})^t \mathbf{W}^{-1} (\bar{\mathbf{x}}^{(f)} + \bar{\mathbf{x}}^{(g)}), \\ k = 2 & \text{otro caso.} \end{cases} \quad (3.9)$$

Clasificador 2. KNN:

Sean $x_1^{(1)}, \dots, x_{n_1}^{(1)}$ y $x_1^{(2)}, \dots, x_{n_2}^{(2)}$ muestras funcionales de dos poblaciones.

1. Probabilidad posterior: Estimamos la probabilidad posterior de cada grupo, $\hat{P}_{\{y=-1\}}(\mathbf{x})$ y $\hat{P}_{\{y=1\}}(\mathbf{x})$. Esta regla de clasificación es conocida como la **Regla de Bayes**, donde:

$$P_y(x) = P(Y = y | X = x), \quad y \in \{-1, 1\}. \quad (3.10)$$

Notemos que:

$$P_y(x) = E(\mathbb{I}_{Y=y} | X = x). \quad (3.11)$$

Por lo tanto, se puede utilizar un estimador tipo Kernel para la predicción a través de la esperanza condicional.

2. La función discriminante ponderada: En este paso implementamos una for-

ma útil de ponderar este clasificador,

$$\hat{P}_y(x) = \frac{\sum_{\{i:y_i=y\}} w_m(x_i) K \left[\frac{d(x_i, x)}{h_k} \right]}{\sum_{i=1}^N w_m(x_i) K \left[\frac{d(x_i, x)}{h_k} \right]}, \quad (3.12)$$

donde $w_m(x_i)$ es la ponderación para cada x_i y h_k es el ancho de banda tal que:

$$k = \#\{i : d(x, x_i) < h_k\}. \quad (3.13)$$

Para elegir el parámetro k , usamos un proceso de validación cruzada,

$$k_{vc}^* = \arg \min_k L_n(k). \quad (3.14)$$

3. Clasificación de un nuevo dato: Con ayuda de las probabilidades posteriores ponderadas estimadas, asignamos una nueva curva z a una de las dos poblaciones sí:

$$\begin{cases} k = 1 & \text{si } \hat{P}_{\{y=-1\}}(z) > \hat{P}_{\{y=1\}}(z), \\ k = 2 & \text{otro caso.} \end{cases} \quad (3.15)$$

Clasificador 3. Profundidad de Banda Generalizada:

Sean $x_1^{(1)}, \dots, x_{n_1}^{(1)}$ y $x_1^{(2)}, \dots, x_{n_2}^{(2)}$ dos muestras funcionales de dos poblaciones.

1. Banda generalizada: Se define la banda generalizada entre dos curvas como:

$$GB_y(x; x_{i_1}, x_{i_2}) \equiv \left\{ t \in [0, 1] : \min_{r=\{i_1, i_2\}} x_r(t) \leq x(t) \leq \max_{r=\{i_1, i_2\}} x_r(t) \right\}. \quad (3.16)$$

2. Profundidad de banda generalizada: A partir de GB, calculamos la profundidad de cada curva en su respectivo grupo:

$$GBD_y(x_i) = \binom{n}{2}^{-1} \sum_{1 \leq i_1 \leq i_2 \leq n} \lambda(GB_y(x_i; x_{i_1}, x_{i_2})), \quad (3.17)$$

donde λ es la medida de Lebesgue en el intervalo $[0, 1]$. La “proporción de tiempo” que la curva x está dentro de la banda es $\lambda(GB(x))$.

De esta manera, obtenemos los datos ordenados de cada grupo. En la [Figura 3.2](#) podemos ver un ejemplo de una banda generada por dos curvas y la proporción de tiempo que una curva x está dentro de ella.

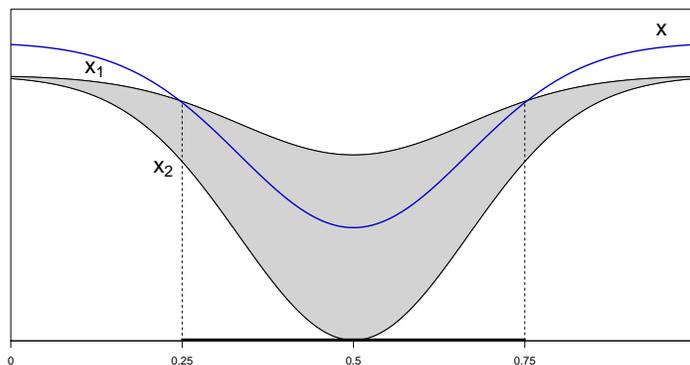


Figura 3.2: Ejemplo de un curva x contenida parcialmente en la banda generalizada $GB(x_1, x_2)$

3. La función discriminante ponderada: Nuevamente, en este paso implementamos una forma útil de ponderar este clasificador. Se calcula la distancia media truncada ponderada de cada grupo ($WTAD$),

$$WTAD_y(x) = \frac{\sum_{i=1}^M \frac{1}{w_m(x_{(i)})} (d(x, x_{(i)}) \cdot GBD_y(x_{(i)}))}{\sum_{i=1}^M \frac{1}{w_m(x_{(i)})} (GBD_y(x_{(i)}))}, \quad (3.18)$$

donde $w_m(x_{(i)})$ es la ponderación para cada elemento ordenado $x_{(i)}$ y M es el parámetro de truncamiento, con $M < \min(n_1, n_2)$.

4. Clasificación de un nuevo dato: Con ayuda de la distancia media truncada ponderada estimada de cada grupo, asignamos una nueva curva x a una de las

dos poblaciones sí:

$$\begin{cases} k = 1 & \text{i } WTAD_{\{y=-1\}}(x) \leq WTAD_{\{y=1\}}(x), \\ k = 2 & \text{otro caso.} \end{cases} \quad (3.19)$$

Luego de implementar la ponderación en cada clasificador, los incorporamos en cada algoritmo Boosting. El parámetro de regularización $Mstop$ lo estimamos en cada réplica, logrando parar las iteraciones cuando la diferencia entre estimaciones fuera mayor a 10^{-6} .

Analizamos el mejoramiento de su desempeño respecto al mismo clasificador, que denominamos, clasificador estándar. Para adecuar las necesidades de las ponderaciones propuestas, fue necesario construir todos los códigos utilizando el software estadístico **R**.

3.1.2. Resultados y discusión

Para ilustrar el desempeño de los métodos propuestos utilizamos cuatro conjuntos de datos: un conjunto de datos simulado analizado desde dos escenarios distintos y dos conjuntos de datos Benchmark. En cada escenario realizamos 1000 repeticiones para estimar la media y la desviación estándar de la tasa de clasificación correcta.

Datos simulados:

Los modelos presentados para simular los datos fueron obtenidos de una idea presentada en sección 5.2 del libro de [Horváth y Kokoszka \(2012\)](#).

Consideramos dos muestras x_1, \dots, x_{n_1} y z_1, \dots, z_{n_2} , que satisfacen:

$$\begin{aligned} x_i(t) &= \mu_1(t) + \varepsilon_i(t), & 1 \leq i \leq n_1, \\ y_j(t) &= \mu_2(t) + \delta_j(t), & 1 \leq j \leq n_2, \end{aligned} \quad (3.20)$$

donde las medias de cada modelo son: $\mu_1(t) = 0$ y $\mu_2(t) = 0.8t(1 - t)$, los errores $\varepsilon_i(t)$ y $\delta_j(t)$ son puentes Brownianos.

Para esta simulación, trabajamos dos escenarios, el primero con muestras de igual tamaño ($n_1 = n_2 = 100$) y el segundo con muestras de diferente tamaño ($n_1 = 100$ y $n_2 = 50$). Convertimos las trayectorias en objetos funcionales por medio de una base de Fourier con 49 funciones, como lo hacen en la sección 5.2 de Horváth y Kokoszka (2012). En la Figura 3.3 podemos observar ejemplos de trayectorias de cada muestra diferenciadas por color.

El conjunto compuesto por las dos muestras simuladas se dividió en cuatro partes: las primeras tres fueron tomadas como el conjunto de entrenamiento y la cuarta como el conjunto de prueba.

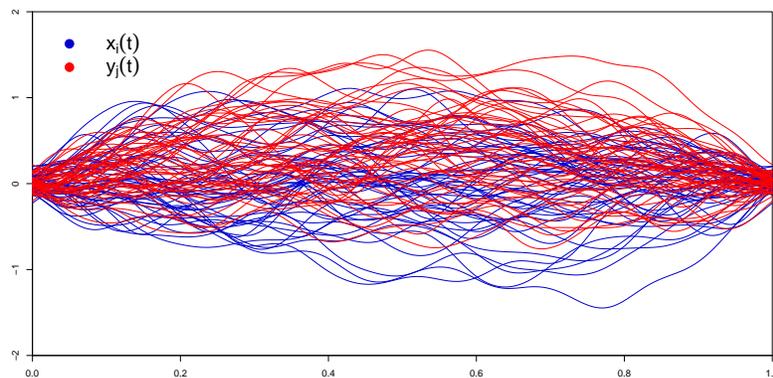
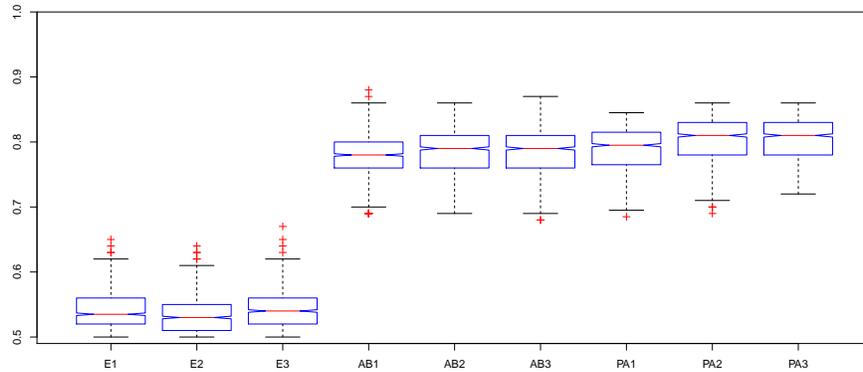


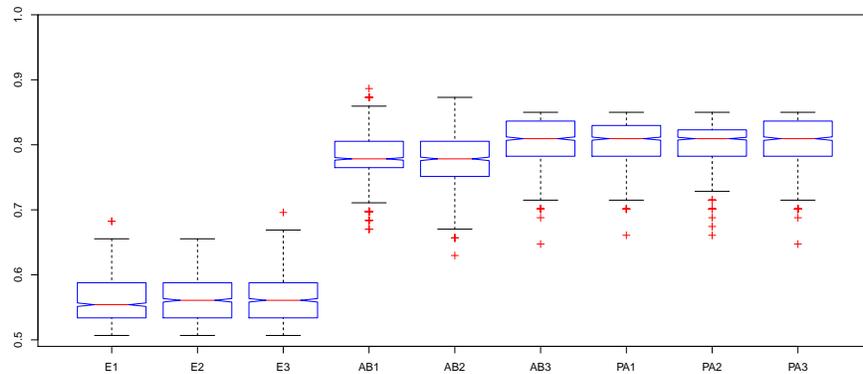
Figura 3.3: Trayectorias simuladas de cada muestra diferenciadas por color.

En la Figura 3.4 se muestran los *boxplots* el comportamiento de la tasa de clasificación correcta y en la Tabla 3.1 la estimación de la media y la desviación estándar para cada escenario. Simbolizamos con las letras **E**: clasificador estándar, significa que solo actúa el clasificador sin ningún tipo de ponderación, es decir, el clasificador sin implementar ningún algoritmo Boosting; **AB** implementando

AdaBoost y PA implementando Ponderación alternativa”.



(a) Tamaños de muestras iguales.



(b) Tamaños de muestras diferentes.

Figura 3.4: Comportamiento de la tasa de clasificación correcta para cada escenario.

Observando el comportamiento de la tasa de clasificación correcta, podemos resaltar una mejora significativa en los seis escenarios con respecto a los clasificadores estándar. Además, no se percibe diferencia en los resultados de la estimación de las medias o de las desviaciones estándar cuando se tienen muestras de igual o

	E1	E2	E3	AB1	AB2	AB3	PA1	PA2	PA3
Media	0.5379	0.5374	0.5389	0.7779	0.7855	0.7867	0.7888	0.8041	0.8034
Sd	0.0279	0.0282	0.0282	0.0330	0.0328	0.0335	0.0369	0.0351	0.0354

(a) Tamaños de muestras iguales.

	E1	E2	E3	AB1	AB2	AB3	PA1	PA2	PA3
Media	0.5564	0.5574	0.5587	0.7818	0.7802	0.8044	0.8038	0.8024	0.8044
Sd	0.0347	0.0341	0.0347	0.0377	0.0403	0.0336	0.0338	0.0347	0.0336

(b) Tamaños de muestras diferentes.

Tabla 3.1: Media y desviación estándar de la tasa de clasificación correcta para cada escenario.

diferente tamaño.

Datos reales

Trabajamos con dos conjuntos de datos Benchmark, los cuales describimos a continuación:

Datos de Espectrometría (Tecator): Nuestro primer conjunto de datos Benchmark, son los datos de espectrometría, nombrado en la literatura como **Tecator**. Está compuesto por 215 muestras de carne molida a las cuales se les mide su nivel de grasa por medio de su espectro de 100 canales de absorbencia. La longitud de onda está medida entre 850-1050 nm. de forma equiespaciada. Los datos pueden encontrarse en lib.stat.cmu.edu/datasets/tecator. Para la clasificación tomamos su índice de grasa, así, el conjunto lo dividimos en dos grupos, alto contenido de grasa (HF), más del 20%, y bajo contenido de grasa (LF), menor o igual al 20%. Los grupos establecidos poseen 77 espectros de HF y 138 espectros de LF. En la [Figura 3.5](#) podemos observar los espectros de cada grupo diferencia-

dos por color. Para representarlos como objetos funcionales se utilizó una base de B-Splines con 15 funciones, como lo hacen en el paquete `fda.usc` de **R**, [Febrero *et al.* \(2014\)](#).

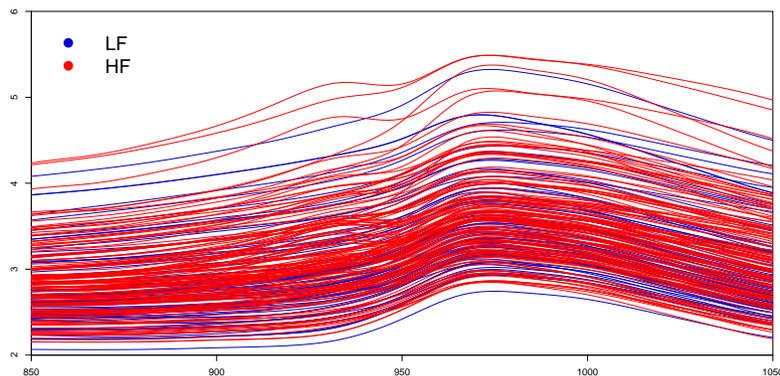


Figura 3.5: Conjunto de datos **Tecator** de cada grupo diferenciados por color.

Para evaluar el desempeño de nuestra propuesta, tomamos 120 espectros como el conjunto de entrenamiento y los restantes 95 espectros como el conjunto de prueba, como en [Alonso *et al.* \(2012\)](#). En la [Figura 3.7 \(a\)](#) se muestran los resultados de clasificación y en la [Tabla 3.2 \(a\)](#) la estimación de la media y la desviación estándar de la tasa de clasificación correcta para cada escenario.

Con el animo de comparar los resultados obtenidos, en [Alonso *et al.* \(2012\)](#) en la sección 4.1, reportan la tasa de clasificación correcta, para los cuales nuestros resultados tienen rangos similares con cualquiera de nuestros seis escenarios reportados en la [Tabla 3.2 \(a\)](#) de igual manera que lo hacemos con respecto a los clasificadores estándar reportados en la misma tabla.

Datos de Altura (Growth): El segundo conjunto de datos Benchmark, conocido en la literatura como **Growth**, está compuesto por 93 curvas de alturas medidas en centímetros, de las cuales 39 son de niños y 54 de niñas. Se tomaron

31 mediciones de 1 a 18 años, las cuales no están equiespaciadas. En la [Figura 3.6](#) podemos observar las trayectorias de los dos grupos diferenciadas por color. Los datos pueden encontrarse en el paquete **fda** de **R**. Para representarlos como objetos funcionales se utilizó una base de B-Splines con 12 funciones, como lo hacen en la sección 5.1 en el libro de [Ramsay *et al.* \(2009\)](#).

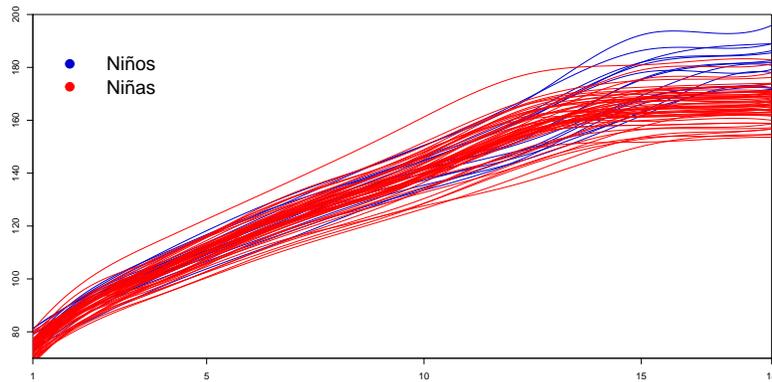
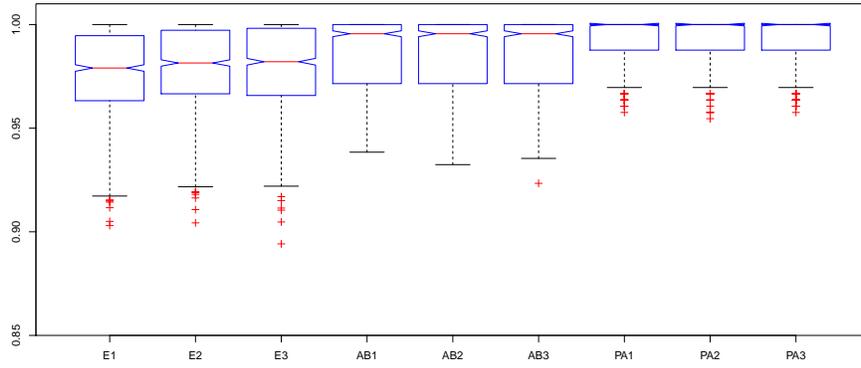


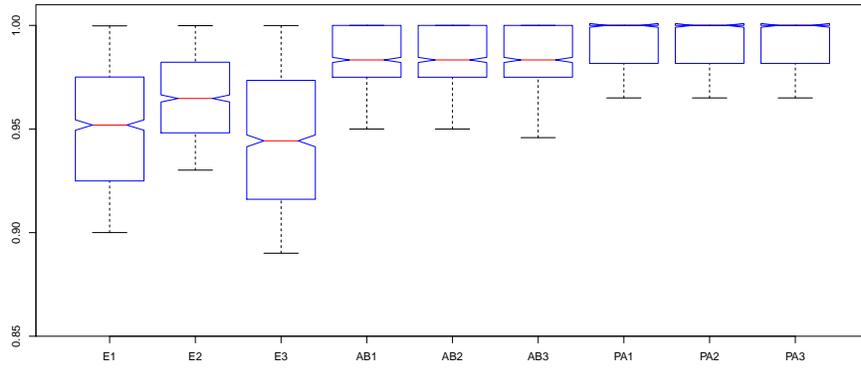
Figura 3.6: Conjunto de datos **Growth** de cada grupo diferenciados por color.

Para evaluar el desempeño de nuestra propuesta, dividimos el conjunto de datos en 60 curvas como el conjunto de entrenamiento y las restantes 23 como el conjunto de prueba, como en [Alonso *et al.* \(2012\)](#). En la [Figura 3.7 \(b\)](#) se muestra el comportamiento de la tasa de clasificación correcta y en la [Tabla 3.2 \(b\)](#) la estimación de la media y la desviación estándar en cada escenario.

Con el animo de comparar los resultados obtenidos, en [Alonso *et al.* \(2012\)](#) en la sección 4.2, reportan la tasa de clasificación correcta, para los cuales nuestros resultados tienen rangos similares con cualquiera de nuestros seis escenarios reportados en la [Tabla 3.2 \(a\)](#) de igual manera que lo hacemos con respecto a los clasificadores estándar reportados en la misma tabla.



(a) Datos Tecator



(b) Datos Growth

Figura 3.7: Comportamiento de la tasa de clasificación correcta para cada conjunto de datos.

	E1	E2	E3	AB1	AB2	AB3	PA1	PA2	PA3
Media	0.9765	0.9788	0.9788	0.9862	0.9861	0.9860	0.9937	0.9936	0.9937
Sd	0.0202	0.0190	0.0198	0.0169	0.0170	0.0169	0.0097	0.0097	0.0098

(a) Datos Tecator

	E1	E2	E3	AB1	AB2	AB3	PA1	PA2	PA3
Media	0.9507	0.9652	0.9450	0.9839	0.9838	0.9837	0.9917	0.9914	0.9912
Sd	0.0286	0.0203	0.0324	0.0127	0.0128	0.0129	0.0121	0.0122	0.0121

(b) Datos Growth

Tabla 3.2: Media y desviación estándar de la tasa de clasificación correcta para cada conjunto de datos.

3.1.3. Conclusiones

En esta primera investigación, denominada Boosting con clasificadores funcionales, hemos propuesto un nuevo enfoque para el mejoramiento del desempeño de clasificadores funcionales. Este enfoque se basa en la adaptación de algunos clasificadores funcionales, logrando incorporarlos en ciertos algoritmos Boosting. Nuestros resultados en los diferentes estudios muestran que el método funciona muy bien. Esto se evidencia en dos diferentes escenarios, que son, cuando tenemos datos difíciles de clasificar, como son los datos simulados, obtenemos mejoras significativas comparadas con los clasificadores estándar. Además, es importante resaltar que los algoritmos no presentan diferencias en sus resultados cuando se utilizan tamaños de muestras iguales o diferentes entre grupos. Ahora, cuando tenemos datos con bajo error de clasificación, como son los dos conjuntos de datos reales, las ganancias aunque son marginales, siempre logramos mejoras, tanto con

los clasificadores estándar como con los resultados reportados en las referencias consultadas.

Por otra parte, resaltamos los buenos resultados obtenidos con el algoritmo propuesto, Ponderación alternativa, sin embargo no logramos concluir cual algoritmo tiene mejor rendimiento con respecto a Adaboost funcional, aunque la evidencia experimental es prometedora. Este algoritmo es aún parte de estudio, tanto en la parte computacional como en sus fundamentos teóricos, ya que aun cuando comparte la misma filosofía que el algoritmo AdaBoost, no es clara su equivalencia y, por lo tanto, los resultados existentes sobre cotas para el error no son necesariamente aplicables.

3.2. Boosting con estructura funcional

Para este segundo trabajo consideramos dos algoritmos Boosting, LogitBoost y Binomial Boosting. La idea principal fue desarrollar la misma estructura existente de cada algoritmo, pero realizando los cambios necesarios para incorporar datos funcionales. Para ello fue necesario cambiar procesos puntuales de los algoritmos que se hacen para datos multivariados por procesos para datos funcionales y estudiamos su desempeño. Nuestras conclusiones para este trabajo están apoyadas por cuatro estudios de simulación y se ilustran además con tres conjuntos de datos Benchmark.

3.2.1. Modificaciones realizadas

Un paso importante tanto para el algoritmo LogitBoost como en Binomial Boosting, es actualizar la función discriminante $\phi_m(x)$ en cada iteración m por medio

de mínimos cuadrados ponderados. Ahora, para diseñar algoritmos con estructura funcional, incorporamos en ese paso un proceso funcional, es decir, realizamos mínimos cuadrados ponderados funcionales. A continuación mostramos brevemente la modificación realizada.

Recordemos que la función objetivo en dicho paso era:

$$\phi_m(x) = \arg \min_{\phi} \mathbb{E}_w (z_m(x) - \phi(x))^2 = \mathbb{E}_w (z_m(x)|x) \quad (3.21)$$

retomando la ecuación (2.22), para LogitBoost,

$$w_m(x) = p_m(x)(1 - p_m(x)), \quad z_m(x) = \frac{y - p_m(x)}{p_m(x)(1 - p_m(x))}, \quad (3.22)$$

y remonando la ecuación (2.27), para Binomial Boosting,

$$w_m(x) = p_m(x), \quad z_m(x) = y - p_m(x). \quad (3.23)$$

En la literatura de FDA, podemos encontrar el método regresión por medio de mínimos cuadrados ponderados funcionales, y como referencia podemos ver el capítulo 4 del libro de [Ramsay y Silverman \(2005\)](#). Aquí, la idea es la misma y solo se hace hincapié al cambio de estructura funcional. Este modelo se conoce como modelo de **regresión con respuesta escalar y covariable funcional**. En este modelo se tiene que:

$$\mathbb{E}_w (z|x) = \int \beta(t)x(t) dt, \quad (3.24)$$

donde la representación de x y β la podemos hacer de varias maneras. Nosotros trabajamos con la representación en una base conocida. Por lo tanto,

$$x(t) \approx \sum_{k=1}^{K_x} c_k \varphi_k(t) \quad \text{y} \quad \beta(t) \approx \sum_{k=1}^{K_\beta} b_k \varphi_k(t), \quad (3.25)$$

donde los c_k , b_k son los coeficientes y las φ_k son las funciones de la base. Esta base la elegimos de igual manera como lo han hecho las referencias consultadas para cada conjunto de datos utilizado. De este modo, el estimador por mínimos cuadrados ponderados los coeficientes b_k resulta:

$$\hat{b}_k = (\varphi_k^T(t)W\varphi_k(t))^{-1} \varphi_k^T(t)Wx, \quad (3.26)$$

donde W es una matriz diagonal con valores w_i .

Por lo tanto, nuestro trabajo fue aplicar los algoritmos vistos LogitBoost, [Sección 2.3.2](#) y Binomial Boosting, [Sección 2.3.3](#), con los cambios anteriormente mencionados, y así poder tener como entrada, datos funcionales.

3.2.2. Resultados y Discusión

Para ilustrar el desempeño de los métodos propuestos para este trabajo utilizamos siete conjuntos de datos: cuatro conjuntos de datos simulados y tres conjuntos de datos Benchmark. En cada escenario realizamos 1000 repeticiones para estimar la media y la desviación estándar de la tasa de clasificación correcta.

Datos simulados

Analizamos cuatro conjuntos de datos simulados, algunos de estos ejemplos fueron tomados de [Alonso *et al.* \(2012\)](#); [Horváth y Kokoszka \(2012\)](#). Generamos 200 funciones para cada población, donde el conjunto de entrenamiento está formado por las primeras 150 funciones de cada muestra y las restantes 50 observaciones de cada muestra son el conjunto de prueba.

Simulación 1:

1. **Muestra 1a:** $f_i(t) = t + u_i$, donde u_i es una variable aleatoria uniforme en el intervalo $(0, 1)$.
2. **Muestra 1b:** $g_i(t) = t + v_i$, donde v_i es una variable aleatoria uniforme en el intervalo $(1/2, 3/2)$.

Las funciones han sido evaluadas en 30 puntos entre 0 y 1. En la [Figura 3.8](#) observamos los datos simulados de las dos muestras. Notemos que la muestra 1a y 1b difieren en nivel cuando u_i toma valores en $(0, 1/2)$ y v_i en $(1, 3/2)$ pero coinciden cuando u_i y v_i toman valores en $(1/2, 1)$. Esta intersección causa una tasa de error de clasificación teórica igual al 25 %, que logramos observar en los resultados.

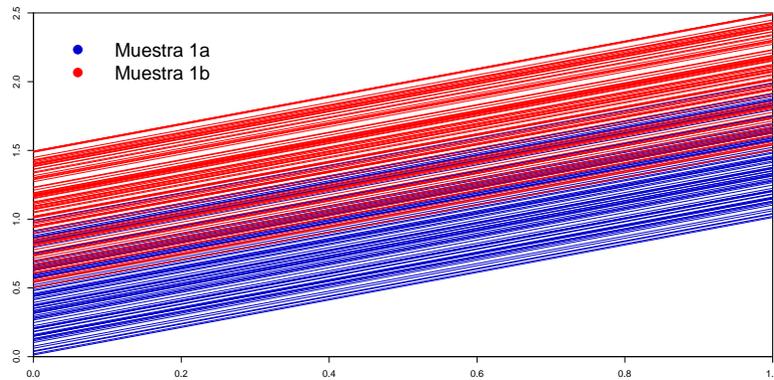


Figura 3.8: Conjunto de datos de la simulación 1 de cada muestra diferenciadas por color.

Simulación 2:

1. **Muestra 2a:** $f_i(t) = (t + u_i)^2$, donde u_i es una variable aleatoria uniforme en el intervalo $(0, 1)$.
2. **Muestra 2b:** $g_i(t) = t^2 + v_i$, donde v_i es una variable aleatoria uniforme en el intervalo $(0, 1)$.

Las funciones han sido evaluadas en 30 puntos entre 0 y 1. En la [Figura 3.9](#) observamos el comportamiento de estas dos muestras. Notemos que la muestra 2a y la muestra 2b generan observaciones funcionales que se cruzan, pero los clasificadores logran obtener la diferencia en su forma.

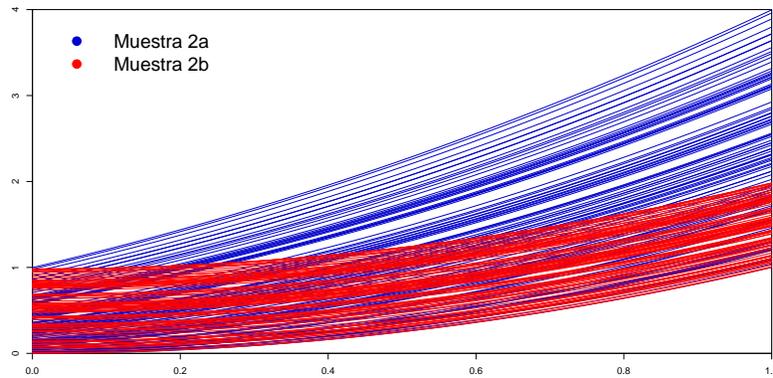


Figura 3.9: Conjunto de datos de la simulación 2 de cada muestra diferenciadas por color.

Simulación 3:

1. **Muestra 3a:** $f_i(t) = (t + u_i)^2 + 1/4$, donde u_i es una variable aleatoria uniforme en el intervalo $(0, 1)$.
2. **Muestra 3b:** $g_i(t) = (t + v_i)^2$, donde v_i es una variable aleatoria uniforme en el intervalo $(1/2, 3/2)$.

Las funciones han sido evaluadas en 30 puntos entre 0 y 1. En la [Figura 3.10](#) observamos las muestras simuladas. Notemos que la muestra 3a y la muestra 3b, generan observaciones funcionales que se cruzan, (el término $+1/4$ se añade a f para maximizar el cruce). Se observa un problema de discriminación con mayor dificultad que los anteriores, pero al aplicar los algoritmos podemos observar como mejoran la discriminación.

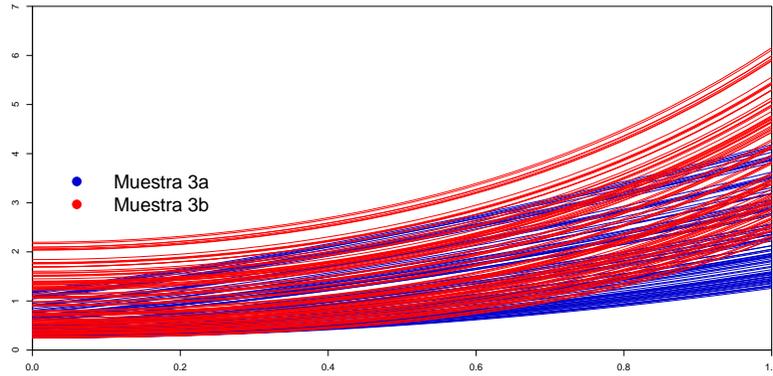


Figura 3.10: Conjunto de datos de la simulación 3 de cada muestra diferenciadas por color.

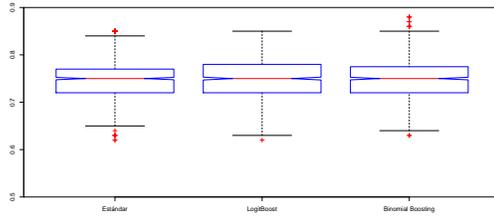
Simulación 4:

1. **Muestra 4a:** $f_i(t) = \varepsilon_i(t)$,
2. **Muestra 4b:** $g_i(t) = .8t(1 - t) + \varepsilon_i(t)$, donde ε_i son puentes Brownianos.

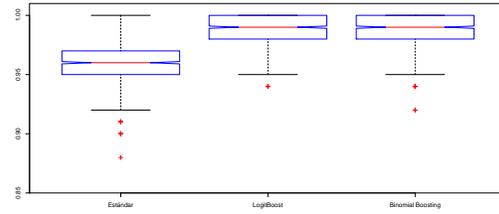
Este conjunto de datos simulados es el mismo que utilizamos en el trabajo anterior, el cual describimos en la [Sección 3.1.2](#). Además, sus trayectorias fueron ilustradas en la [Figura 3.3](#). Estas muestras simuladas son muy similares, ocasionando un problema complejo de discriminación, pero nuevamente los algoritmos logran mejorar sustancialmente la discriminación.

En la [Figura 3.11](#) se muestran los *boxplots* de la estimación de la tasa de clasificación correcta y en la [Tabla 3.3](#) la estimación de la media y la desviación estándar para cada conjunto simulado. Simbolizamos con las letras **E**: clasificador estándar, **LB** implementando LogitBoost y **BB** implementando Binomial Boosting”.

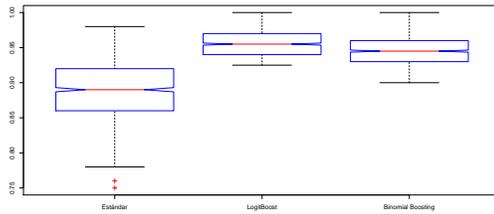
Observando el comportamiento de la tasa de clasificación correcta, podemos resaltar una mejora significativa en el desempeño de los algoritmos con respecto a los clasificadores estándar. Los resultados experimentales son bastante prometedores, aunque no logramos diferenciar cual de los dos algoritmos lo hace mejor.



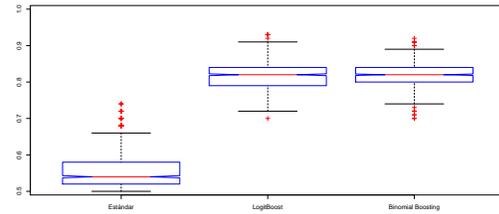
(a) Simulación 1



(b) Simulación 2



(c) Simulación 3



(d) Simulación 4

Figura 3.11: Comportamiento de la tasa de clasificación correcta para cada simulación.

	E	LB	BB
Media	0.7470	0.7510	0.7483
Sd	0.0424	0.0438	0.0421

(a) Simulación 1

	E	LB	BB
Media	0.9600	0.9895	0.9897
Sd	0.0197	0.0115	0.0123

(b) Simulación 2

	E	LB	BB
Media	0.8850	0.9580	0.9439
Sd	0.0409	0.0233	0.0203

(c) Simulación 3

	E	LB	BB
Media	0.5542	0.8191	0.8187
Sd	0.0490	0.0359	0.0347

(d) Simulación 4

Tabla 3.3: Media y desviación estándar de la tasa de clasificación correcta para cada simulación.

Datos reales

En esta sección ilustramos la propuesta con tres conjuntos de datos Benchmark.

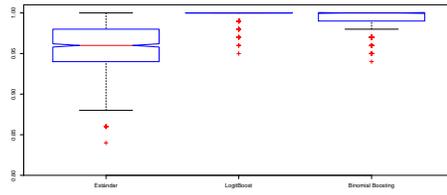
Datos de espectrometría (Tecator): Nuestro primer conjunto de datos Benchmark fue explicado en la [Sección 3.1.2](#), donde en la [Figura 3.5](#) podemos ver sus trayectorias.

Datos de altura (Growth): El segundo conjunto de datos Benchmark fue explicado en la [sección 3.1.2](#), de igual manera, en la [Figura 3.6](#) muestra las trayectorias de las dos muestras

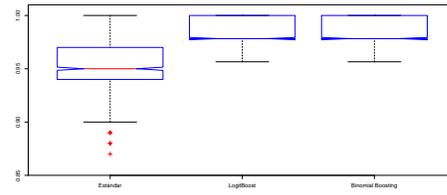
Datos de Esclerosis Múltiple (DTI): Nuestro tercer conjunto de datos viene de un estudio de neuroimagen de Esclerosis Múltiple (EM). Esta es una enfermedad del sistema nervioso que afecta al cerebro y la médula espinal. Para cuantificar de forma no invasiva la micro-estructura de la materia blanca utilizan los tensores de difusión (DTI). Los datos pueden encontrarse en el paquete **refund** de **R**. Este conjunto de datos consta de 382 perfiles de las vías de Anisotropía fraccional del tracto corticoespinal derecho, de las cuales tomamos 100 de estas, 74 pacientes con EM y 26 sujetos de control. Al inicio de este capítulo, como ejemplo de motivación, en la [Figura 3.1](#) mostramos sus trayectorias de los dos grupos diferencias por color.

Tomamos las primeras 76 curvas como el conjunto de entrenamiento y las restantes 24 como el conjunto de prueba. Para representarlos como objetos funcionales se utilizó una base de B-Splines penalizados con 35 funciones y un parámetro de penalización $\lambda = e^{-12}$ como lo hacen en la [sección 2.3](#) de [Sørensen *et al.* \(2013\)](#).

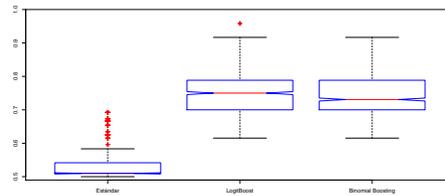
En la [Figura 3.12](#) se muestran los *boxplots* de la estimación de la tasa de



(a) Datos Tecator



(b) Datos Growth



(c) Datos DTI

Figura 3.12: Comportamiento de la tasa clasificación correcta para cada conjunto de datos Benchmark.

clasificación correcta y en la [Tabla 3.4](#) podemos observar la media y la desviación estándar para cada conjunto de datos.

Con el animo de comparar los resultados obtenidos, [Alonso *et al.* \(2012\)](#) en la sección 4.1 y 4.2 [Lopez-Pintado y Romo \(2006\)](#) en la sección 4 reportan los valores de la tasa de clasificación correcta tanto para los datos Tecator como para los datos Growth por sus métodos propuestos, para los cuales nuestros resultados tienen rangos similares en los dos escenarios reportados en la [Tabla 3.4 \(a\) y \(b\)](#) respectivamente para cada conjunto de datos, de igual manera que lo hacemos con respecto a los clasificadores estándar reportados en la misma tabla.

	E	LB	BB
Media	0.9522	0.9968	0.9921
Sd	0.0280	0.0068	0.0110

(a) Datos Tecator

	E	LB	BB
Media	0.9501	0.9830	0.9817
Sd	0.0220	0.0163	0.0163

(b) Datos Growth

	E	LB	BB
Media	0.5419	0.7402	0.7372
Sd	0.0578	0.0619	0.0626

(c) Datos DTI

Tabla 3.4: Media y desviación estándar de la tasa de clasificación correcta para cada conjunto de datos Benchmark.

3.2.3. Conclusiones

En esta segunda investigación, denominada Boosting con estructura funcional, hemos propuesto un segundo enfoque para el mejoramiento del desempeño de clasificadores funcionales. En este enfoque extendemos la utilidad de los algoritmos Boosting, proponiendo algoritmos con la misma estructura que los actuales, pero incorporando procesos para datos funcionales. Nuestros resultados en los diferentes escenarios simulados muestran que el método funciona muy bien, en algunos con ganancias marginales y en otras con una ganancia significativa. De igual forma se puede evidenciar la misma conclusión en los conjuntos de datos reales utilizados, aunque las ganancias sean marginales, pero siempre se logra mejorar, tanto con los clasificadores estándar como con los resultados reportados en las referencias consultadas.

No logramos concluir cual de los dos algoritmos Boosting tiene mejor rendimiento, pues los valores obtenidos tanto en la estimación de la media y la desviación estándar es muy similar, y en algunos casos uno es mejor que otra y viceversa. Por otra parte, los algoritmos propuestos necesitan de un mayor estudio tanto en la parte computacional como en sus fundamentos teóricos. Además, los resultados obtenidos solo son resultados empíricos, aunque la evidencia experimental es prometedora.

Capítulo 4

Análisis de imágenes digitales de mamografías

El Cáncer de mama es el cáncer más común en las mujeres a nivel mundial, según las estadísticas publicadas por la Organización Mundial de la Salud (www.who.int/es/), existen 1.7 millones de nuevos casos y 522,000 muertes por año. En la actualidad, no existen mecanismos eficaces para prevenirlo, debido a que la causa de origen es desconocida. Sin embargo, un diagnóstico eficaz en su etapa inicial brinda una gran probabilidad de ser tratado, [Global Cancer Observatory \(2017\)](#). Por lo tanto, la detección temprana del Cáncer de mama juega un papel importante para la reducción de la mortalidad y el único método que ha demostrado ser efectivo para este proceso es la Mamografía.

La Mamografía es una imagen plana de la mama obtenida a través de rayos X. Este examen puede ser utilizado en mujeres sanas para detectar tempranamente signos de cáncer o para diagnóstico después de haberse encontrado un abultamiento u otro signo o síntoma. En la Mamografía podemos obtener información acerca de la localización, el número y las características de una lesión palpable. Estas imágenes se pueden obtener al proyectar la mama de diferentes ángulos, las

dos proyecciones más comunes son: Cráneo Caudal (CC) y Medio Lateral Oblicua (MLO). En la [figura 4.1](#) podemos ver un ejemplo de cada una. La vista CC es una proyección vertical tomada de la parte superior de la mama, en la que se puede observar con mayor precisión el tejido medial. La vista MLO es una proyección a 45° con respecto del eje de simetría del cuerpo de la paciente. La ventaja de esta proyección es la visibilidad de la mama por completo, y por ello generalmente es la proyección más utilizada. Estas imágenes son el caso A-0072-1 de la base de datos *Digital Database for Screening Mammography* (DDSM), [Heath et al. \(2001\)](#).

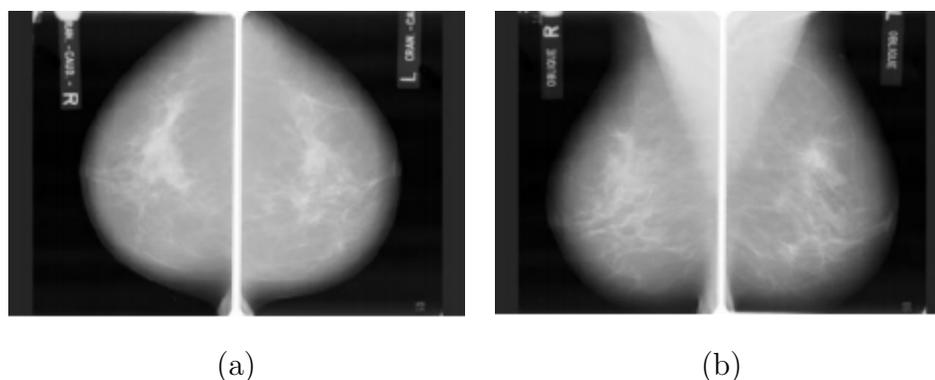


Figura 4.1: Caso A-0072-1 de DDSM. (a) Vista cráneo caudal (CC), (b) Vista medio lateral oblicua (MLO).

El Análisis de estas imágenes es una tarea muy importante tanto para la detección temprana como para el diagnóstico de alguna anomalía presente. Sin embargo, la interpretación de una Mamografía es un trabajo para el radiólogo de gran complejidad, debido a diversos factores, como: alta heterogeneidad presente entre mamas (tamaño, forma, textura, color), variabilidad en la apariencia de las anomalías, características del tejido mamario, y la calidad de las imágenes. En algunas situaciones, las anomalías no pueden ser visualizadas debido a que la mama está constituida por tejidos muy similares entre sí, y porque las lesiones

son muy pequeñas y se confunden con el tejido mamario normal. Por otra parte, la interpretación humana de una Mamografía es subjetiva, y requiere mucho tiempo del radiólogo. Por lo que la interpretación correcta de estos estudios, depende en gran medida de la experiencia del radiólogo, y aún entre radiólogos experimentados puede existir una alta variabilidad. Estudios retrospectivos han demostrado que entre un 10 % y un 25 % de los tumores pasan inadvertidos a los ojos del radiólogo, [Cheng et al. \(2006\)](#). Otros estudios confirman estos mismos resultados al estimar que la sensibilidad alcanzada en la detección de tumores mediante la inspección visual de la Mamografía por parte de un radiólogo no supera el 75 %, [Ball y Bruce \(2007\)](#).

Con el fin de mejorar el desempeño en el Análisis de mamografías, se han propuesto en la literatura diferentes métodos computacionales, para la detección o en el diagnóstico de lesiones. Estos sistemas ayudan a localizar lesiones, caracterizarlas y a determinar su grado de malignidad o benignidad. Las etapas que se deben realizar según lo que se ha establecido en la literatura (ver [Hela et al. \(2013\)](#)), son:

1. Pre-procesamiento: Se implementa para reducir el ruido y mejorar la calidad de la imagen. En general, en esta etapa es donde se elimina el fondo de la imagen y el músculo pectoral.
2. Segmentación: Se localizan y se busca aislar regiones sospechosas que podrían contener alguna anormalidad.
3. Extracción de características: Se obtienen características de las regiones sospechosas obtenidas en la etapa anterior.
4. Clasificación: Una vez seleccionadas las características, estas deben ser cla-

sificadas para determinar el grado de malignidad.

Es importante resaltar, que los resultados que se tienen aún no son lo suficientemente buenos y conlleva a ser un tema de investigación actual. Es por ello, que iniciamos un trabajo de investigación donde se estudie y se proponga un método eficaz de clasificación en el análisis mamografías abordando las herramientas de FDA, tema principal de mi tesis. Este trabajo se inició en una estancia académica realizada en la Universidad Carlos III de Madrid, España, como parte de mi trabajo doctoral en conjunto con los investigadores Juan Romo y Rosa E. Lillo. Veamos a continuación un breve resumen de los conocimientos básicos que debemos tener sobre este tipo de imágenes y más adelante expondremos el trabajo realizado hasta el momento y los avances obtenidos.

4.1. Anatomía de la mama

La mama se encuentra ubicada sobre el músculo pectoral que lo separa del resto del cuerpo. Se ha podido discriminar sobre el tipo de severidad de la anomalía, dependiendo del tipo de tejido y el tipo de anomalía. A continuación explicaremos brevemente cada una de ellas.

4.1.1. Tipo de severidad

La clasificación final que se desea hacer en una Mamografía que presenta alguna anomalía es discriminar en anomalía Benigna o Maligna.

Denominamos anomalía **Benigna** a cualquier abultamiento que no genere cáncer. Estos no se propagan fuera de la mama hacia otros órganos. Si el abultamiento Benigno es grande, puede cambiar el tamaño y la forma de la mama.

Llamamos anomalía **Maligna** a los abultamientos que indiquen la presencia de cáncer. Estos abultamientos se componen de las células que invaden y dañan los tejidos y órganos cercanos, así mismo, estas células malignas se pueden desprender del abultamiento y entrar en el sistema circulatorio o sistema linfático.

4.1.2. Tipo de tejido

El tipo de tejido se puede clasificar en tres grupos: Graso, Fibroso o Denso. Para más detalle ver [He et al. \(2015\)](#).

Se denomina **Graso** cuando el seno está compuesto casi por completo por tejido grasoso, aunque puede tener una pequeña presencia de tejido fibroso, ($\leq 25\%$). Este tipo de tejido, también denominado el menos denso, se presenta en la Mamografía de color oscuro, ya que son tejidos blandos y por tanto permiten el paso de la luz (radiolúcido). El tejido Graso ayuda a darle al seno su tamaño y forma. En la figura [figura 4.2 \(a\)](#) se muestra un ejemplo de este tipo de tejido.

Se llama **Fibroso** cuando la mayoría del tejido es graso, pero se observan áreas un poco más amplias de tejido fibro-glandular ($> 25\%$). El tejido Fibroso incluye lóbulos que producen leche durante la lactancia y conductos que llevan la leche desde los lóbulos hasta el pezón durante el amamantamiento. Este tipo de tejido también llamado conectivo, sirve de soporte y mantiene el tejido glandular en su lugar, ver [figura 4.2 \(b\)](#) como ejemplo.

El ultimo tipo de tejido es denominado **Denso**, y hace referencia a cuando el seno está casi todo compuesto de tejido glandular, fibroso y poco tejido graso. En una Mamografía, este tipo de tejido aparece de color blanco (radiopacos). Esencialmente, cuanto más “blanco” se ve más denso es el seno. En la [figura 4.2 \(c\)](#) podemos observar un ejemplo. Las siguientes imágenes presentadas en

el escrito provienen de la base de datos *Mammographic Image Analysis Society* (MIAS), [Sukling et al. \(1994\)](#). Citaremos en cada una de ellas el número exacto como aparecen en su base de datos.

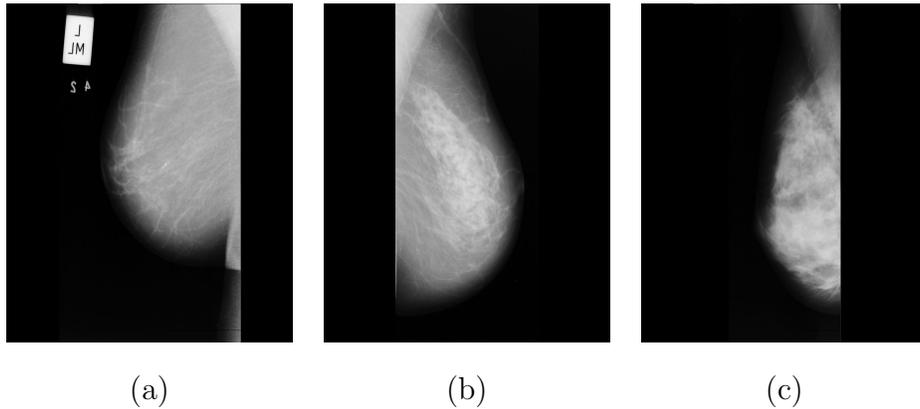


Figura 4.2: Tipos de tejido en la mama: (a) Graso, mdb011 de mini-MIAS. (b) Fibroso, mdb022 de mini-MIAS. (c) Denso, mdb033 de mini-MIAS.

No es claro porqué ciertas mujeres tienen senos densos y otras no. En general, las mujeres más jóvenes tienden a tener senos más densos, y algunas pueden perder densidad debido a cambios hormonales que se producen durante la menopausia. Sin embargo, esto no es una regla exacta. El tipo de tejido está determinado probablemente por factores genéticos, pero la dieta, la nutrición, la ganancia o pérdida de peso, y factores hormonales también pueden afectar.

Tener senos densos no es una condición anormal, de hecho, aproximadamente la mitad de todas las mujeres de más de 40 años tienen senos densos. La relación exacta entre el tipo de tejido y el cáncer de mama todavía no se ha podido establecer y aún es un tema relevante de investigación. Es más fácil detectar un cáncer subyacente en mujeres con senos grasos y más difícil detectarlo en mujeres con senos densos. Esto se debe a que el tejido denso normal podría ocultar

anomalías que también, por lo general, aparecen de color blanco en la mamografía. Lo que se ha podido concluir, de forma empírica, es que el tener senos con tejido denso esta asociado al riesgo a desarrollar cáncer de mama. [Radiological Society of North America \(2017\)](#).

4.1.3. Tipo de anomalía

Ahora bien, existe un gran número de lesiones o anomalías que pueden estar presentes en la mama. El estudio de estas anomalías es vital, ya que algunas de estas pueden ser indicadores de malignidad y son aquellas las que un radiólogo busca al analizar una Mamografía. Las anomalías que se pueden presentar son: Asimetría, Distorsión de la Arquitectura, Calcificaciones y Masas, [Bozek *et al.* \(2009\)](#).

La **Asimetría** son áreas del tejido fibro-glandular que son más extensas en una mama que en la otra, [Sickles \(2007\)](#). Esta puede ser el indicador de la presencia de una masa o de una variación en la distribución del tejido normal, ver [figura 4.3](#) como ejemplo.

La **Distorsión de la arquitectura** se refiere a un cambio anormal de la mama formándose lesiones finas que no están asociadas a la presencia de una masa. Se puede visualizar como espículas que irradian desde un punto común, formando un patrón similar al de una estrella, [Shantanu *et al.* \(2013\)](#). Esta puede ser un indicador de malignidad cuando es complementada con la identificación de Masas, Calcificaciones o Asimetrías. En la [figura 4.4 \(a\)](#) podemos observar un ejemplo de Distorsión de la Arquitectura.

Las **Calcificaciones** son pequeños acumulaciones cristalinas de minerales (calcio) que se encuentran en la mama en regiones de alta intensidad, y se pue-

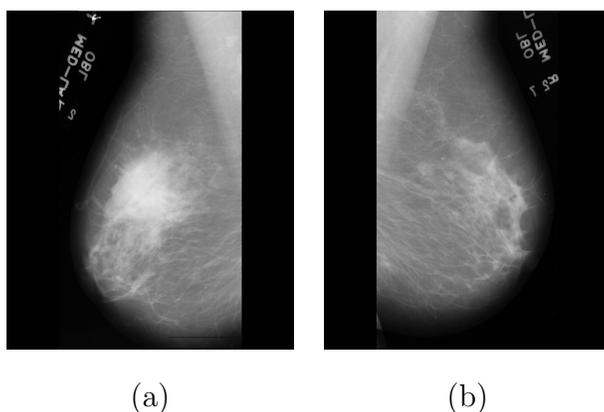


Figura 4.3: Ejemplo de asimetría apreciada en la vista MLO de la mama izquierda, mdb081 y mdb082 de mini-MIAS.

den visualizar como manchas en la Mamografía. En la [figura 4.4 \(b\)](#) podemos ver un ejemplo de calcificaciones. Existen dos tipos de calcificaciones: macro-calcificaciones y micro-calcificaciones. Las macro-calcificaciones son acumulaciones de forma dispersa y gruesa. Por estudios anteriores se sabe que estas tienen una forma oval o redonda y tamaño uniforme (entre calcificaciones), las cuales están asociadas a procesos benignos. En cambio, las micro-calcificaciones pueden encontrarse de forma aislada o incrustadas en una masa. El diámetro de las micro-calcificaciones varía desde 0.1 mm. hasta 0.2 mm. y su patrón de agrupación y morfología son indicadores de malignidad. Las micro-calcificaciones asociadas a procesos malignos son irregulares, polimórficas y con variaciones en su tamaño. Alrededor del 30% al 50% de los casos no palpables se detectan inicialmente debido a la presencia de grupos de micro-calcificaciones, [He et al. \(2104\)](#).

Una **Masa** es una estructura tridimensional y es visible en al menos dos proyecciones. Con frecuencia el cáncer de mama se presenta como una masa con o sin la presencia de calcificaciones. Ésta representa un quiste que se forma a partir

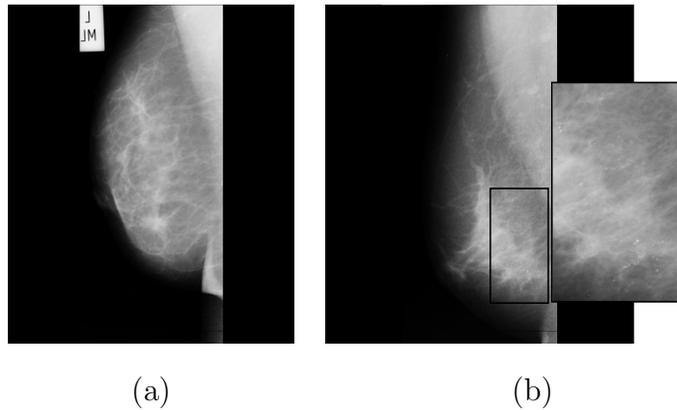


Figura 4.4: (a) Distorsión de la arquitectura, mdb181 de mini-MIAS. (b) Agrupaciones de calcificaciones, mdb322 de mini-MIAS.

de una colección de líquidos. La dificultad para decidir si es una masa, se debe a la similitud en los niveles de gris con las del tejido normal. Regularmente, se caracteriza por su densidad, sus bordes y su forma. Los bordes de una masa deben ser cuidadosamente clasificados debido a que es uno de los criterios más importantes en la determinación de la malignidad o benignidad, [Malagelada \(2007\)](#). Los bordes de una masa se clasifican en: Circunscritos, Espiculados y Mal definidos.

Los **Circunscritos** son masas con bordes bien definidos y demarcados con una abrupta transición entre la masa y el tejido que lo rodea. Los **Espiculados** se caracterizan por tener líneas que irradian desde el centro de la masa y los **Mal definidos** serían las demás masas o aquellas que tengan un borde pobremente definido y disperso. En la [figura 4.5](#) podemos ver un ejemplo de cada uno de ellos.

Las diferentes formas que se pueden presentar en una masa son: Redonda, Ovalada, Lobulada e Irregular. En la [figura 4.6](#) podemos ver como deberían ser las diferentes formas de una masa.

Como parte del trabajo, hemos implementado un método automático para el

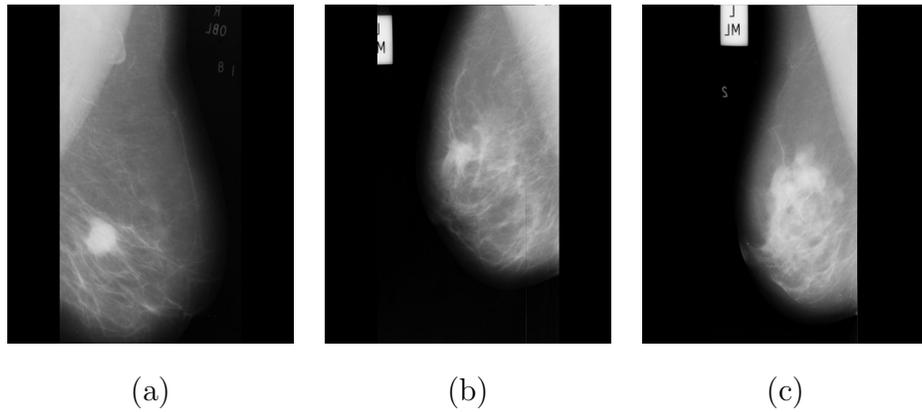


Figura 4.5: Tipos de bordes de una masa: (a) Circunscrito, mdb028 de mini-MIAS. (b) Espiculado, mdb117 de mini-MIAS. (c) Mal definido, mdb265 de mini-MIAS.

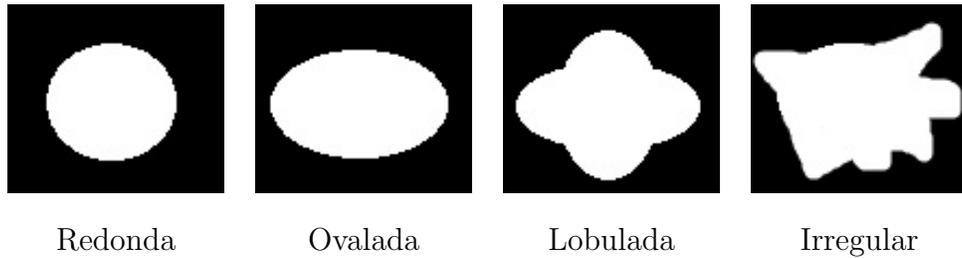


Figura 4.6: Ejemplos de la forma de la masa.

pre-procesamiento con resultados óptimos y estamos trabajando en el desarrollo de ideas para realizar el proceso de segmentación. En el siguiente capítulo explicaremos a detalle el trabajo realizado, así como el trabajo futuro a realizar.

Capítulo 5

Clasificación automática de anomalías presentes en una mamografía

Con el animo de apoyar a los especialistas médicos en el diagnóstico de anomalías presentes en las mamografías, queremos proponer un método automático de clasificación fundamentado en técnicas emergentes de la literatura de Clasificación de Datos Funcionales.

Como mencionamos en el capítulo anterior, se ha establecido en la literatura ciertos pasos a seguir para desarrollar un mecanismo de clasificación, los cuales nosotros adoptamos. Por lo tanto, se requiere como primer paso, desarrollar un pre-procesamiento de la imagen. Esto debido a ruidos en el proceso de recolección de los datos. Este problema es particularmente evidente en la vista MLO, donde las etiquetas, cuñas, marcadores, arañazos, cinta adhesiva y el músculo pectoral no permiten enfocarse en el análisis del tejido mamario. El pre-procesamiento es una tarea exigente y desafiante, ya que las mamografías son muy heterogéneas en tamaño, intensidad, forma y textura, posición del paciente durante la toma

de la imagen y en muchas ocasiones el tejido mamario puede ocultar el borde del músculo pectoral, lo cual dificulta la eliminación del músculo. Es importante resaltar que las fallas en este proceso pueden ocasionar importantes sesgos en el resultado, por lo que el pre-proceso es una tarea bastante relevante.

5.1. Pre-procesamiento

En la literatura existente sobre esta etapa se han propuesto diversos métodos, entre los más destacados aparecen [Ferrari *et al.* \(2004\)](#), los cuales adoptaron una técnica de multiresolución usando Wavelets de Gabor y transformaciones de Hough. [Kwok *et al.* \(2004\)](#), propusieron un método basado en “Cliff detection”. [Ma *et al.* \(2007\)](#), segmentaron el músculo pectoral por medio de “Adaptive Pyramids” (AP) y “Minimum Spanning Trees” (MST). [Wang *et al.* \(2010\)](#), presentaron un método basado en Cadenas de Markov a tiempo discreto (DTMC). [Camilus *et al.* \(2010\)](#), presentan un enfoque usando un método de fusión basado en Cortes gráficos, que luego suavizan utilizando las curvas de Bezier. [Chen *et al.* \(2012\)](#), encuentran el borde del músculo pectoral combinando el umbral iterativo de Otsu y procesamientos morfológicos. [Li *et al.* \(2013\)](#), segmentan el borde combinando características de homogeneidad y desviación de intensidad, el cual refinan a través del filtro de Kalman.

Los métodos mencionados tienen una gran validez en la literatura, pero aún tienen un margen alto de error. Por lo tanto, como un primer trabajo de nuestra investigación, desarrollamos un método automático de baja complejidad computacional que elimine tanto los ruidos del fondo de la imagen como el músculo pectoral presentes en la Mamografía, buscando márgenes de error menores a los que ya existen. El método propuesto tiene dos pasos: primero, removemos el ruido

presente en el fondo de la imagen (etiquetas, cuñas, marcadores, arañazos, cinta adhesiva, entre otros), usando una técnica de umbral. Segundo, aprovechando algunas características anatómicas de la mama ya estudiadas anteriormente por otros autores, aplicamos una función de transformación de intensidad con ayuda de la distribución Beta, proporcionando en este paso un nuevo mecanismo en la transformación de intensidad. Esta idea se propone debido a las ventajas que proporciona las diferentes formas que toma la distribución Beta debido a sus parámetros. De esta manera, podemos proponer diferentes transformaciones que se ajusten a las características de cada Mamografía. Con ayuda de esta transformación identificamos el contorno inicial del borde del músculo, luego, este borde pasa por un proceso de selección que nos proporciona un borde rugoso, el cual se suaviza por medio de un B-Spline cúbico. Finalmente, ya teniendo el borde identificado, eliminamos la región del músculo pectoral.

5.1.1. Metodología

Es importante identificar en una Mamografía, dos zonas importantes, la región del músculo pectoral y la región de la mama. En las vistas MLO, el músculo se ubica en la parte superior y está contenido dentro de un “triángulo recto” (colocamos entre comillas porque en la realidad no necesariamente el borde del músculo es una línea recta). La parte inferior de la Mamografía, donde se encuentra el tejido mamario puede ser aproximado como un óvalo (ver [figura 5.1](#)).

Por lo tanto, teniendo en cuenta estas consideraciones, orientamos todas las mamografías hacia la izquierda, esto es, todas las imágenes donde la mama aparece en la parte derecha las reflejaron verticalmente. Así, cada imagen tendrá el músculo pectoral en la parte superior izquierda. El origen del sistema de coorde-

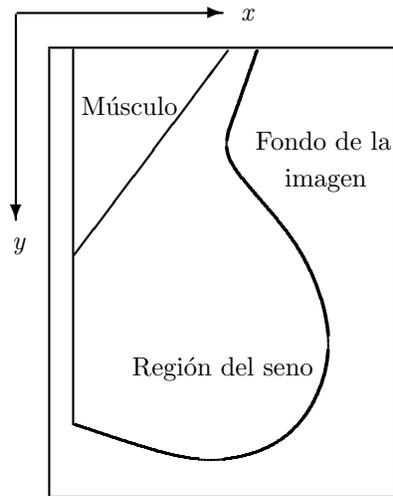


Figura 5.1: Regiones importantes en cada Mamografía.

nadas está en la esquina superior izquierda de la imagen, donde x se representa el eje horizontal y y el eje vertical. La intensidad de la escala de gris en la imagen se denota por $I(x, y)$, donde $x = 1, \dots, n_x$, y $y = 1, \dots, n_y$ (ver [figura 5.1](#)). A continuación, exponemos a detalle los dos pasos principales para desarrollar el método propuesto; remover los elementos de ruido del fondo de la imagen y eliminar el músculo pectoral.

Remover los elementos de ruido del fondo de la imagen

Como mencionamos, cada imagen contiene muchos objetos que deben ser removidos del fondo de la imagen. En la [figura 5.2](#) podemos observar un ejemplo. Para remover estos objetos usamos una **Técnica de umbral basado en el histograma de frecuencias** de los niveles de gris de cada pixel, ver [Ratna y Neelima \(2012\)](#).

Este método es usado para separar regiones en una imagen, separando de tal forma que puedan ser reconocidas como objetos diferentes. La idea es crear

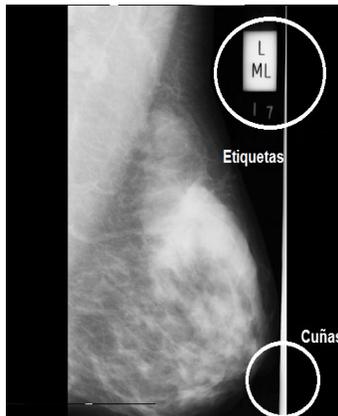


Figura 5.2: Imagen con algunos objetos en el fondo, mdb115 de mini-MIAS.

una imagen binaria por medio de un umbral, para conseguir dicha separación. El parámetro clave en este proceso es la elección del valor umbral. Existen diferentes métodos para su elección, de tal forma que se puede elegir manualmente o se puede calcular un valor automático. Para nuestro proceso elegimos un parámetro automático que dependa de los niveles de gris de cada Mamografía. Este valor se consigue al analizar un cambio significativo existente en los niveles de gris entre el fondo de la imagen y la región del seno, para más detalles ver [Moumena y Ali \(2013\)](#). El proceso puede resumirse en los siguientes pasos:

1. Construimos la densidad empírica de la escala de gris en la imagen, en donde ubicamos el valor umbral, que denominamos c , donde aparece el primer mínimo local de izquierda a derecha. En la [figura 5.3](#) podemos ver un ejemplo de un histograma y la ubicación del valor umbral c .
2. Con ayuda del valor umbral c , transformamos la Mamografía en una imagen

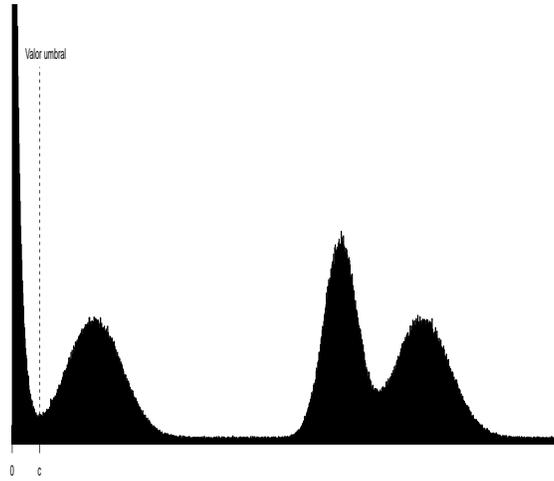
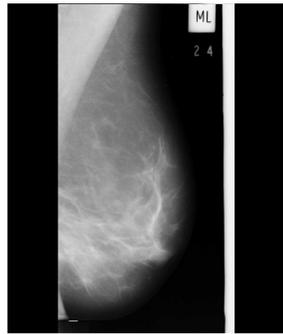


Figura 5.3: Ejemplo de una densidad empírica, donde localizamos c

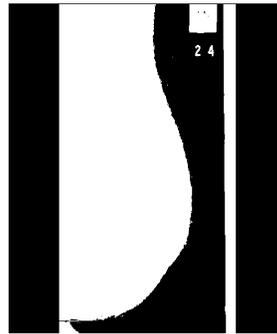
binaria con la siguiente función, (ver [Figura 5.4 \(b\)](#)).

$$D(x, y) = \begin{cases} 1 & \text{si } I(x, y) > c, \\ 0 & \text{si } I(x, y) \leq c. \end{cases} \quad (5.1)$$

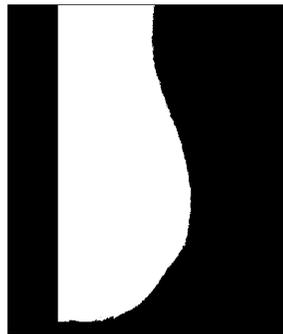
3. La matriz obtenida de la imagen binaria, la utilizamos como elemento de entrada en la función “bwlable” del paquete EImage en R. Esta función conecta los conjuntos de pixeles diferentes de cero y los etiqueta separándolos por grupos, localizando así todos los objetos que pueda tener la imagen. Luego, elegimos el objeto encontrado de mayor tamaño, el cual será la región del seno, y eliminamos todos los demás objetos, los cuales serán los elementos de ruido existentes en el fondo de la imagen, (ver [Figura 5.4 \(c\)](#)).
4. Ya con el fondo de la imagen limpia, retornamos los valores de los pixeles originales, (ver [Figura 5.4 \(d\)](#)).



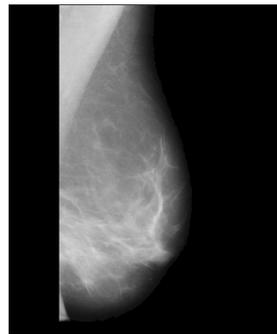
(a)



(b)



(c)



(d)

Figura 5.4: Imágenes obtenidas en el proceso para remover los elementos de ruido del fondo de la imagen, mdb051 de mini-MIAS. (a) Imagen original. (b) Imagen binaria. (c) Región del seno. (d) Imagen con el fondo limpio.

Eliminación del músculo pectoral

En la literatura encontramos algunos estudios donde se han analizado características anatómicas importantes del músculo pectoral, tales como:

- [Ferrari et al. \(2004\)](#), comentan que el músculo pectoral tiene niveles de gris casi homogéneos.
- [Wang et al. \(2010\)](#), mencionan acerca de las diferencias existentes de la intensidad de gris entre la región del músculo pectoral y el tejido mamario.
- [Ma et al. \(2007\)](#), discuten el hecho de que el músculo pectoral tiene una forma aproximadamente triangular.
- [Kwok et al. \(2004\)](#), muestran cómo la forma del músculo pectoral disminuye gradualmente su ancho de arriba hacia abajo.

Entonces, tomando ventaja de estos estudios, proponemos un nuevo enfoque para eliminar el músculo pectoral. Este método consiste en aplicar una **Transformación de intensidad** (ver [Gonzalez y Woods \(2008\)](#)), para resaltar la región del músculo pectoral y así identificarlo correctamente. Esta técnica se basa en operaciones espaciales, que se realizan directamente sobre cada pixel. La función de transformación T , es de la forma:

$$g(x, y) = T [I(x, y)], \quad (5.2)$$

donde $I(x, y)$ es el nivel de intensidad de entrada, $g(x, y)$ es la salida (transformada).

Hay tres formas básicas de funciones de transformación, que frecuentemente se usan para realizar características en una imagen, las cuales son: lineal, logarítmica y potencia (para más detalles ver [Singh et al. \(2010\)](#)). Tratamos con estas

transformaciones, pero no obtuvimos resultados favorables para nuestro objetivo, que es identificar el borde del músculo pectoral en las diferentes formas como se pueden apreciar las mamografías. Consideramos que se debe a la heterogeneidad existente (tamaño, forma, anomalías presentes, contraste, intensidad, entre otros). Para solucionar este inconveniente, introducimos una nueva forma de realizar este proceso, que fue utilizando como función de transformación la distribución Beta.

$$\begin{aligned}
 g(x, y) &= BD [I(x, y); \alpha, \beta] \\
 &= \frac{1}{B(\alpha, \beta)} \int_0^{I(x, y)} t^{\alpha-1} (1-t)^{\beta-1} dt.
 \end{aligned} \tag{5.3}$$

Esta nueva idea nace de las ventajas que tienen las diferentes formas que toma la función debido a sus parámetros. De esta manera, podemos ajustar los parámetros tanto como sea necesario para resaltar los píxeles donde se encuentra el músculo pectoral dependiendo de la escala de gris de cada Mamografía.

Ahora, el problema se centra en ajustar de manera adecuada los parámetros de la distribución Beta. En esta función tanto α como β son parámetros de estiramiento que nos ayudan a conseguir la transformación adecuada. Estas transformaciones se usan para comprimir o expandir los niveles de intensidad de cada píxel. Por lo tanto, basados principalmente en la forma que toma la función, realizamos un análisis exploratorio a diferentes mamografías, decidimos trabajar con el parámetro fijo $\alpha = 5$ y el parámetro de ajuste $\beta \in [2, 6]$. En la [Figura 5.5](#) podemos observar como varía la forma de la distribución Beta con estos parámetros.

Además, en las [Figuras 5.6](#) (a), (b) y (c) podemos ver un ejemplo de una Mamografía transformada con diferentes valores de β .

A continuación, resumimos el proceso realizado en los siguientes pasos:

1. Retomamos la imagen final del proceso anterior.

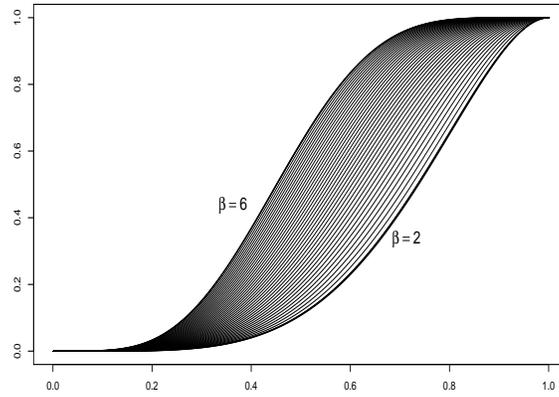


Figura 5.5: Forma que toma la distribución Beta con los parámetros $\alpha = 5$ y $\beta \in [2, 6]$.

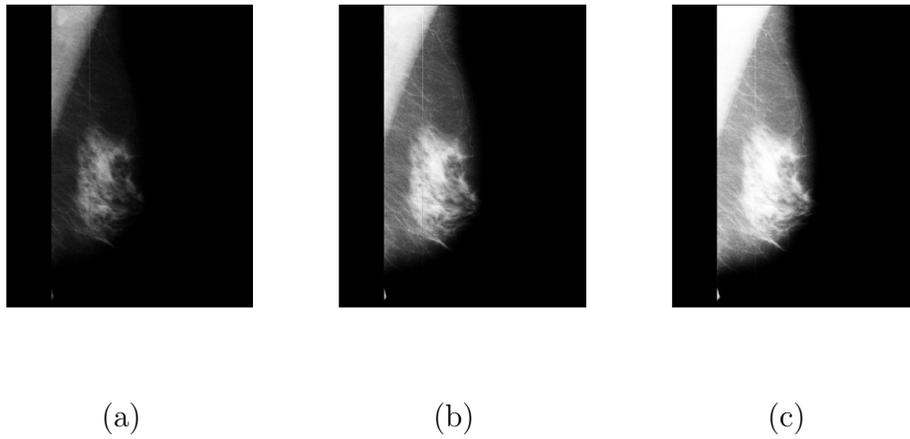


Figura 5.6: Imagen transformada con los valores de $\beta = \{2, 4, 6\}$ respectivamente, mdb042 de mini-MIAS.

2. Basados en las dos primeras características anatómicas mencionadas anteriormente, las cuales discuten sobre la homogeneidad de los niveles de gris en el músculo y las diferencias existentes de la intensidad de gris entre la región del músculo y el tejido mamario, consideramos que el músculo pectoral puede separarse del tejido mamario aplicando una transformación de intensidad con ayuda de la distribución Beta, $T[I(x, y)] = BD[I(x, y); 5, \beta]$, donde el parámetro β varía en la partición $\mathcal{P} = \{2 + 0.1i, i = 0, \dots, 40\}$. En el paso 7 realizaremos el ajuste de este parámetro.

3. Para cada imagen k , calculamos la media de todos los píxeles distintos de cero,

$$\mu_{0,k} = \frac{1}{m} \sum_{x=1}^{n_x} \sum_{y=1}^{n_y} I_k(x, y), \quad (5.4)$$

donde m es el número de píxeles distintos de cero de la imagen k .

4. Para cada imagen k y cada β de la partición, seleccionamos en cada fila de izquierda a derecha la primera coordenada (x, y) que sea menor a $\mu_{0,k}$ y diferente de cero.

5. Teniendo en cuenta ahora las dos últimas características anatómicas, que discuten sobre la forma aproximadamente triangular del músculo pectoral y la disminución de su ancho gradualmente de arriba a abajo, consideramos realizar un proceso de selección con las coordenadas obtenidas en la salida del paso anterior. Este proceso de selección consiste en aceptar cada coordenada, si es menor o igual que el anterior. Estos puntos seleccionados forman un borde rugoso inicial del músculo para cada valor de β y cada imagen k .

6. Suavizamos por medio de un B-Spline cúbico cada borde y así obtenemos una estimación del borde del músculo pectoral. Las coordenadas de cada

estimación del borde serán: $(x_{\beta,1}^k, 1), (x_{\beta,2}^k, 2), \dots, (x_{\beta,n_{\beta,k}}^k, n_{\beta,k})$, donde $n_{\beta,k}$ será la longitud del borde para cada β en cada imagen k .

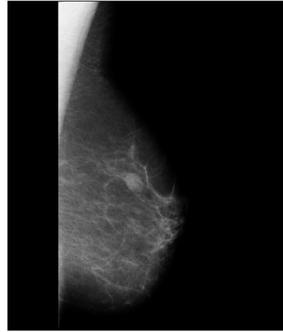
7. En este paso, elegimos un valor de β de forma automática para cada imagen. Dicho valor se obtiene de modo que la estimación del borde con este valor de β sea el borde que más se aproxime al borde real de acuerdo a la siguiente expresión:

$$\hat{\beta}_k = \max_{\beta \in \mathcal{P}} \sum_{y=1}^{n_{\beta,k}} [g(x_{\beta,y}^k - 2, y) - g(x_{\beta,y}^k + 2, y)]. \quad (5.5)$$

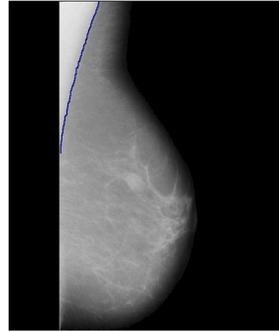
8. Finalmente, ya teniendo ubicado la estimación de borde del músculo, eliminamos la región del músculo pectoral y así obtenemos una imagen donde solo aparece el tejido mamario.

5.1.2. Resultados y discusión

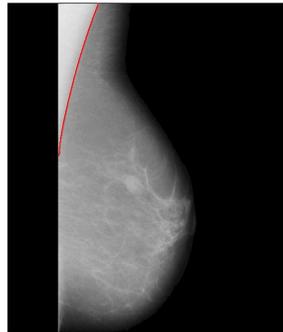
Consideramos que utilizar la distribución beta como una transformación de intensidad, proporciona un método eficaz para identificar el borde del músculo pectoral. El parámetro β juega un papel importante, ya que podemos ajustarlo tanto como sea necesario en cada Mamografía hasta encontrar la estimación del borde que más se aproxime al borde exacto del músculo. Realizamos varios estudios para analizar el comportamiento de este parámetro, y consideramos que la manera como lo ajustamos fue bastante adecuada, ya que logra capturar en general la heterogeneidad existente en las mamografías. Un resultado empírico que logramos obtener es que cuando el músculo pectoral tiene altos niveles de intensidad, el valor de β es pequeño y viceversa. También encontramos que cuando el tejido mamario está muy cerca o sobrepuesto al borde del músculo, el ajuste del parámetro es bastante sensible.



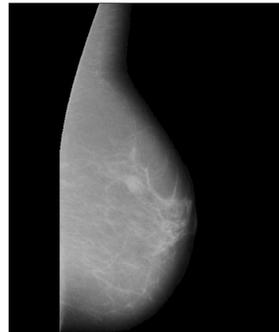
(a)



(b)



(c)



(d)

Figura 5.7: Ejemplo del proceso para eliminar el músculo pectoral, mdb012 de mini-MIAS. (a) Imagen transformada con la estimación de $\beta = 3.5$ encontrado en el paso 7. (b) Borde rugoso inicial en azul. (c) Estimación de borde en rojo. (d) Imagen del seno sin músculo pectoral.

El método propuesto fue probado con la base de datos mini-MIAS. Ésta es un conjunto de imágenes digitales reunida por “Mammographic Image Analysis Society (MIAS)”, <http://peipa.essex.ac.uk/info/mias.html>. Esta base de datos contiene 322 imágenes digitales de mamografías de $200 \mu\text{m}/\text{pixel}$, 8 bits/pixel y tamaño de 1024×1024 píxeles. La base de datos incluye imágenes con vista MLO del seno izquierdo y derecho de la misma paciente, la clasificación del tipo de tejido, las coordenadas de las anormalidades (si las hay) y el tipo de severidad. Estos resultados fueron desarrollados por expertos del tema por medio de estudios clínicos, para más detalles ver [Sukling *et al.* \(1994\)](#).

Métodos de evaluación

El rendimiento de nuestra propuesta se evaluó por medio de dos métodos conocidos en la literatura para este tipo de imágenes, que son: **Tasa de aceptación** y el **Error de área normalizado**, los cuales explicaremos a detalle a continuación. Comparamos los resultados obtenidos con trabajos previos que utilizaron la misma mecánica de evaluación y la misma base de datos. Este trabajo fue implementado en R.

Tasa de aceptación: Cada borde obtenido se organiza en las siguientes tres categorías:

- Exacta: El borde estimado se ajusta exactamente en el margen pectoral. Cualquier desviación del margen es imperceptible o insignificante.
- Adecuada: La línea encontrada identifica el margen del músculo inexactamente, pero con una precisión suficiente para localizar el margen pectoral ($> 70\%$).

- Inadecuada: El borde sugerido se ajusta de manera inexacta al margen pectoral o no logra encontrarlo.

La tasa de aceptación la conforman las dos primeras categorías, Exacta y Adecuada. Para este método, evaluamos las 322 imágenes, y la clasificación obtenida la podemos apreciar en la [Tabla 5.1](#). La tasa de aceptación conseguida fue del 95.34 %, y es importante resaltar que el porcentaje de resultados exactos es bastante alto, con 91.61 %.

Categoría	Cantidad	Porcentaje	Tasa de aceptación
Exacta	295	91.61 %	95.34 %
Adecuada	12	3.73 %	
Inadecuada	15	4.66 %	

Tabla 5.1: Clasificación de la tasa de aceptación del método propuesto.

Luego, en la [Tabla 5.2](#) observamos la comparación entre el método propuesto y estudios anteriores, donde evalúan las 322 imágenes. Podemos apreciar un mejor desempeño de nuestro método.

Referencias	Tasa de aceptación
Kwok <i>et al.</i> (2004)	83.9 %
Li <i>et al.</i> (2013)	90.06 %
Método propuesto	95.34 %

Tabla 5.2: Comparación entre el método propuesto y estudios anteriores.

El conocimiento previo de las características anatómicas permitió adaptar el algoritmo propuesto a mamografías donde el músculo pectoral podría tener dife-

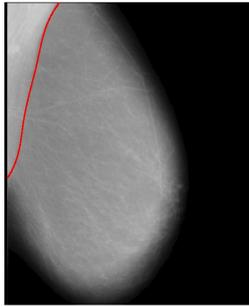
rente tamaño, forma e intensidad en los niveles de gris. Por ejemplo, el método mostró resultados exactos cuando el músculo pectoral era pequeño (Figura 5.8 (c)) o grande (Figura 5.8 (d)), incluso cuando la diferencia de los niveles de gris entre el músculo pectoral y sus alrededores era baja, Como en la Figura 5.8 (a). Además, en algunos casos, la compresión de la mama cuando se toma la Mamografía, puede introducir ruidos que generan bordes del músculo pectoral falsamente reconocidos. Sin embargo, en estos casos, el método también reconoció exactamente los margenes como se muestra en la Figura 5.8 (b) y (e).

Los casos adecuados ocurrieron principalmente cuando el tejido mamario se superponía al músculo pectoral y por lo tanto, el método no podía hacerlo exactamente, como en la Figura 5.8 (f) y Figura 5.9 (g) y (h).

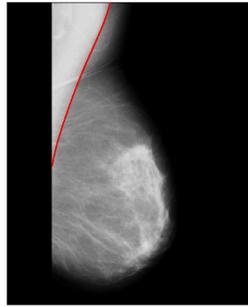
Existen algunas imágenes en las cuales la escala de gris dentro del músculo pectoral es muy heterogénea, como se aprecia en la Figura 5.8 (i) y (j). Para dichas imágenes nuestro método encontró bordes inadecuados. En otras imágenes, la intensidad de los niveles de gris es muy alta y la diferencia entre la región pectoral y el tejido mamario es mínima, lo cual no permitió encontrar el borde muscular, como se muestra en la Figura 5.9 (k) y (l).

El uso de la tasa de aceptación como método de evaluación puede conducir a discrepancias entre los diferentes autores, ya que lo que constituye un resultado “aceptable” puede diferir significativamente entre especialistas, y se basa en la opinión visual subjetiva con poco respaldo cuantitativo. Es por ello que consideramos aplicar un segundo método de evaluación, el cual presentamos a continuación.

Error de área normalizado: Este método de evaluación se basa en un conjunto particular de datos, utilizado para validar este tipo de estudios. El conjunto está compuesto por las coordenadas del margen del músculo pectoral de 84 imágenes



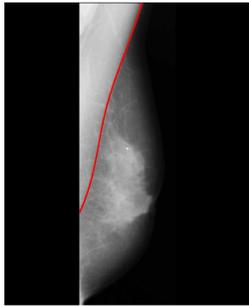
(a)



(b)



(c)



(d)



(e)



(f)

Figura 5.8: Ejemplos de resultados exactos (a) mdb136 de mini-MIAS, (b) mdb095 de mini-MIAS, (c) mdb070 de mini-MIAS, (d) mdb214 de mini-MIAS y (e) mdb161 de mini-MIAS. Ejemplos de resultados adecuados (f) mdb226 de mini-MIAS.



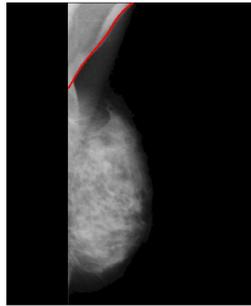
(g)



(h)



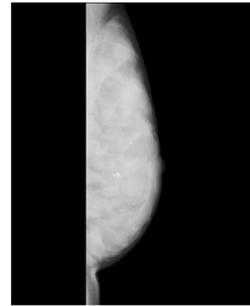
(i)



(j)



(k)



(l)

Figura 5.9: Ejemplos de resultados adecuados (g) mdb242 de mini-MIAS y (h) mdb285 de mini-MIAS. Ejemplos de resultados inadecuados o no encontrados. (i) mdb040 de mini-MIAS, (j) mdb066 de mini-MIAS, (k) mdb061 de mini-MIAS y (l) mdb318 de mini-MIAS.

de la base de datos mini-MIAS. Estas coordenadas fueron marcadas por especialistas y se toman como los bordes reales, ver (Ferrari *et al.*, 2004). Dichos datos fueron amablemente proporcionados por el Dr. Ricardo Ferrari.

El error de área normalizado se construye a partir de porcentajes de pixeles de falsos positivos (FP) y falsos negativos (FN). Tal como se define en Ferrari *et al.* (2004), un FP es un pixel que está en la región propuesta, pero no en la región de referencia. Similarmente, un FN es un pixel en la región de referencia, pero no en la región propuesta. La proporción de pixeles de falsos positivos (FP_I) y la proporción de pixeles de falsos negativos (FN_I) para una imagen I se calcula por medio de las expresiones:

$$\begin{aligned} FP_I &= \frac{1}{A(I)} \sum_{y=1}^p \text{máx} \{0, E_{pro}(y) - E_{ref}(y)\}, \\ FN_I &= \frac{1}{A(I)} \sum_{y=1}^p \text{máx} \{0, E_{ref}(y) - E_{pro}(y)\}, \end{aligned} \quad (5.6)$$

donde $A(I)$ es el área del músculo pectoral en la Mamografía I marcado por los especialistas, p es el número de filas donde el músculo pectoral aparece, $E_{ref}(y)$ es la coordenada del borde en la fila y determinada por los especialistas y $E_{pro}(y)$ es la coordenada del borde en la fila y determinada por el método propuesto. El error medio de FP y FN para el conjunto de imágenes $I = \{1, \dots, N\}$ se obtiene por medio de las expresiones:

$$\begin{aligned} FP_m &= \frac{1}{N} \sum_{I=1}^N FP_I, \\ FN_m &= \frac{1}{N} \sum_{I=1}^N FN_I. \end{aligned} \quad (5.7)$$

La comparación del rendimiento a través de este método de evaluación, entre el método propuesto y estudios anteriores la podemos observar en la [Tabla 5.3](#).

Esta tabla se compone de los resultados obtenidos al calcular el error medio tanto de FP como de FN . Además, contiene ciertos conteos específicos que involucran los porcentajes de FP_I y FN_I .

A continuación presentamos un breve resumen de los resultados obtenidos por los otros autores. En la transformación de Hough, [Ferrari et al. \(2004\)](#), el error medio de falsos positivos es aceptable, pero el error medio de falsos negativos es en gran medida mayor que todos los otros métodos. Además, aproximadamente el 79% de los resultados se encuentran en el rango de mayor error. En el método que utiliza “Gabor wavelets”, [Ferrari et al. \(2004\)](#), el error medio de falsos positivos es el mejor de todos los métodos comparados y el error medio de falsos negativos también es bueno. Sin embargo, más del 26% de los resultados tienen un error entre 0.05 y 0.10, y más del 20% de los resultados se encuentran en el rango de mayor error. En el algoritmo AP, [Ma et al. \(2007\)](#), el error medio de falsos positivos es el mayor de todos los métodos y el error medio de falsos negativos es también alto, (aunque en un rango comparable con los otros métodos). Además, aproximadamente el 6% de los resultados están en el rango de mayor error. Sin embargo, este método tiene aproximadamente el 59% de resultados con menor error. En el método de “Graph cut”, [Camilus et al. \(2010\)](#), el error medio de falsos positivos es muy bueno (muy cerca del mejor resultado obtenido en esta evaluación), y el error medio de falsos negativos tiene un rango comparable con los otros métodos. La distribución de las cantidades del error es aceptable, y además no obtiene resultados en el rango de mayor error. En el método I.S, [Li et al. \(2013\)](#), el error medio de falsos positivos es bueno, el error medio de falsos negativos tiene un resultado comparable con los otros métodos, la distribución de las cantidades del error es aceptable y no obtiene ningún resultado en el rango de mayor error.

	Hough Ferrari et al. (2004)	Gabor Ferrari et al. (2004)	A.P. Ma et al. (2007)	Graph cut Camilus et al. (2010)	I.S. Li et al. (2013)	Método propuesto
FP_m	0.0198	0.0058	0.0371	0.0064	0.0145	0.0196
FN_m	0.2519	0.0577	0.0595	0.0558	0.0552	0.0486
$FP_I < 0.05$ y $FN_I < 0.05$	10	45	50	43	48	52
$\min(FP_I, FN_I) < 0.05$ y $0.05 < \max(FP_I, FN_I) < 0.10$	0	0	18	19	28	18
$\min(FP_I, FN_I) < 0.05$ y $\max(FP_I, FN_I) > 0.10$	0	0	11	33	7	14
$0.05 < FP_I < 0.10$ y $0.05 < FN_I < 0.10$	8	22	0	0	0	0
$0.05 < \min(FP_I, FN_I) < 0.10$ y $\max(FP_I, FN_I) > 0.10$	0	0	0	0	1	0
$FP_I > 0.10$ y $FN_I > 0.10$	66	17	5	0	0	0

Tabla 5.3: Comparación del rendimiento por medio del error de área normalizado.

En lo que sigue, presentamos un resumen del rendimiento obtenido por el método propuesto. El error medio de falsos positivos es bueno y el error medio de falsos negativos es el mejor comparado con los demás métodos. Además, conseguimos la mayor cantidad de resultados con menor error, aproximadamente el 62% y no obtuvimos ningún resultado en el rango de mayor error.

5.1.3. Conclusiones

La etapa de pre-procesamiento es una tarea importante y compleja. Como mencionamos al principio, un mal pre-procesamiento o inclusive no realizar este paso, conlleva a sesgos significativos en los procedimientos de diagnóstico, ya que en muchos casos las mamografías presentan la misma intensidad de grises tanto en el músculo pectoral como en las anormalidades. Para superar esta limitación, en este trabajo se desarrolló un enfoque automático para eliminar tanto el ruido en el fondo de la imagen como la región del músculo pectoral.

El método propuesto, el cual aprovecha algunas características anatómicas, es bastante sencillo de implementar computacionalmente. El uso de la distribución Beta como función de transformación de intensidades, proporcionó una visión novedosa con resultados bastante favorables.

El algoritmo propuesto fue evaluado con 322 imágenes digitales de mamografías de la base de datos mini-MIAS, consiguiendo una tasa de aceptación del 95.34%. Además se determinó el error de área normalizado con un conjunto de 84 imágenes digitales de mamografías, las cuales tienen las coordenadas del borde del músculo pectoral marcadas por especialistas. Mediante dicho método conseguimos resultados favorables, tales como el menor error medio de falsos negativos y la mayor cantidad de resultados con menor error, comparado con los demás métodos.

Además y no se obtuvo ningún resultado en el rango de mayor error. Nuestros resultados demuestran claramente que el método propuesto puede ser utilizado satisfactoriamente como un paso de pre-procesamiento para trabajos adicionales en el Análisis de mamografías.

5.2. Segmentación

La fase de segmentación es la que aísla las regiones sospechosas del resto de la imagen. Este proceso determinará la sensibilidad del sistema, esto es, su capacidad para detectar correctamente el tejido canceroso. La segmentación debería aislar el mayor número posible de anomalías, aunque aún en esta etapa se permite encontrar regiones correspondientes a tejido sano, las cuales deberán ser descartadas en una fase posterior. Por lo tanto, cuando se trabaja con imágenes digitales de mamografías en la detección de lesiones, lo habitual es tratar de diferenciar entre dos clases: tejido con lesión y tejido normal. Aunque discriminar únicamente entre dos clases puede parecer una reducción de la complejidad del problema, realmente no lo es, especialmente porque el tejido normal tiene apariencias muy variables, y discriminar este tipo de tejido suele ser una tarea bastante compleja. En la [Figura 5.10](#) podemos ver un ejemplo del trabajo que se desea realizar.

La mayoría de los métodos de segmentación pueden ser vistos como problemas de optimización, donde la solución deseada es la que minimiza alguna función de costo, definida para una aplicación en particular. La ventaja de ver la segmentación como un problema de optimización es que define de manera precisa los aspectos deseables.

En este paso, tenemos un trabajo parcial realizado, donde incorporamos las herramientas de FDA, las cuales esperamos nos brinden apoyo para proponer un

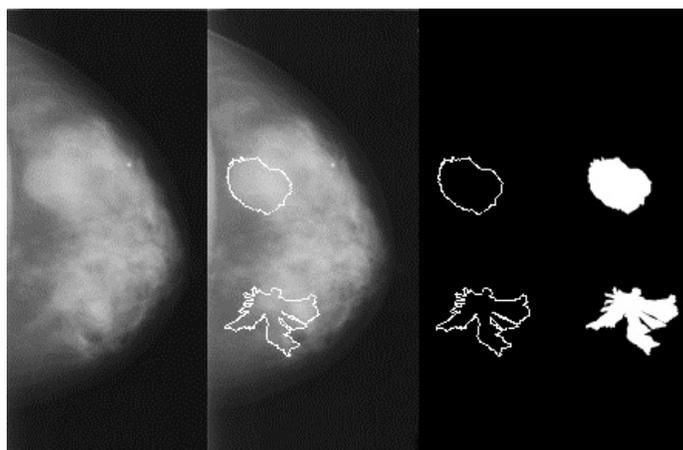


Figura 5.10: Ejemplo de un proceso de segmentación.

mecanismo novedoso y óptimo en esta etapa. Discutiremos brevemente a continuación el trabajo realizado hasta el momento, así como el desempeño encontrado, el cual es bastante prometedor.

5.2.1. Ideas trabajadas

Es importante resaltar que este es solo un trabajo parcial, el cual está en desarrollo. Para cada imagen pre-procesada, definiremos D a la matriz de información y creamos un arreglo de matrices de vectores de incrementos. La idea es encontrar los incrementos de cada pixel con respecto a sus 24 vecinos más cercanos. Dichos vecinos son buscados de manera particular, es decir, creamos ventanas de tamaño 5×5 , y calculamos los incrementos respecto al pixel central en cada ventana. Este proceso es realizado a cada pixel.

$$\begin{pmatrix} d_{j-2,i-2} & d_{j-2,i-1} & d_{j-2,i} & d_{j-2,i+1} & d_{j-2,i+2} \\ d_{j-1,i-2} & d_{j-1,i-1} & d_{j-1,i} & d_{j-1,i+1} & d_{j-1,i+2} \\ d_{j,i-2} & d_{j,i-1} & d_{j,i} & d_{j,i+1} & d_{j,i+2} \\ d_{j+1,i-2} & d_{j+1,i-1} & d_{j+1,i} & d_{j+1,i+1} & d_{j+1,i+2} \\ d_{j+2,i-2} & d_{j+2,i-1} & d_{j+2,i} & d_{j+2,i+1} & d_{j+2,i+2} \end{pmatrix}$$

Para cada fila j de la matriz D , creamos en una matriz que denominamos A_j , donde se guardan los vectores de incrementos.

$$\nabla d_{j,i} = a_{i,l} = (a_{i,1}, a_{i,2}, \dots, a_{i,24}). \quad (5.8)$$

De esta manera, en cada imagen k , tendremos un arreglo de matrices $A_{i,j}^k$.

Estos vectores de incrementos son los elementos que deseamos tratar como datos funcionales. Calculamos la norma L_2 de cada vector y lo guardamos en una matriz N , definida como:

$$n_{i,j} = \|\nabla d_{j,i}\|_2. \quad (5.9)$$

Notemos que un valor $n_{i,j}$ grande es indicio de la presencia del borde de una anomalía, mientras que valores pequeños indican uniformidad en la intensidad de una región. Por lo tanto, consideramos que nuestro interés radica en los vectores con norma grande, es decir, aquellos vectores donde hay un cambio significativo de color.

En un primer análisis, hicimos un filtro y nos quedamos solo con los vectores que tienen norma mayor a 0.5. Con los vectores seleccionados, podemos localizar los pixeles asociados a ellos, y así buscar información relevante que nos pueda servir en el proceso de segmentación. En la [Figura 5.11](#) presenta algunos ejemplos, donde se observa en la parte superior, la imagen de la mamografía con los pixeles resaltados mediante el proceso antes descrito. Además, en la parte inferior de

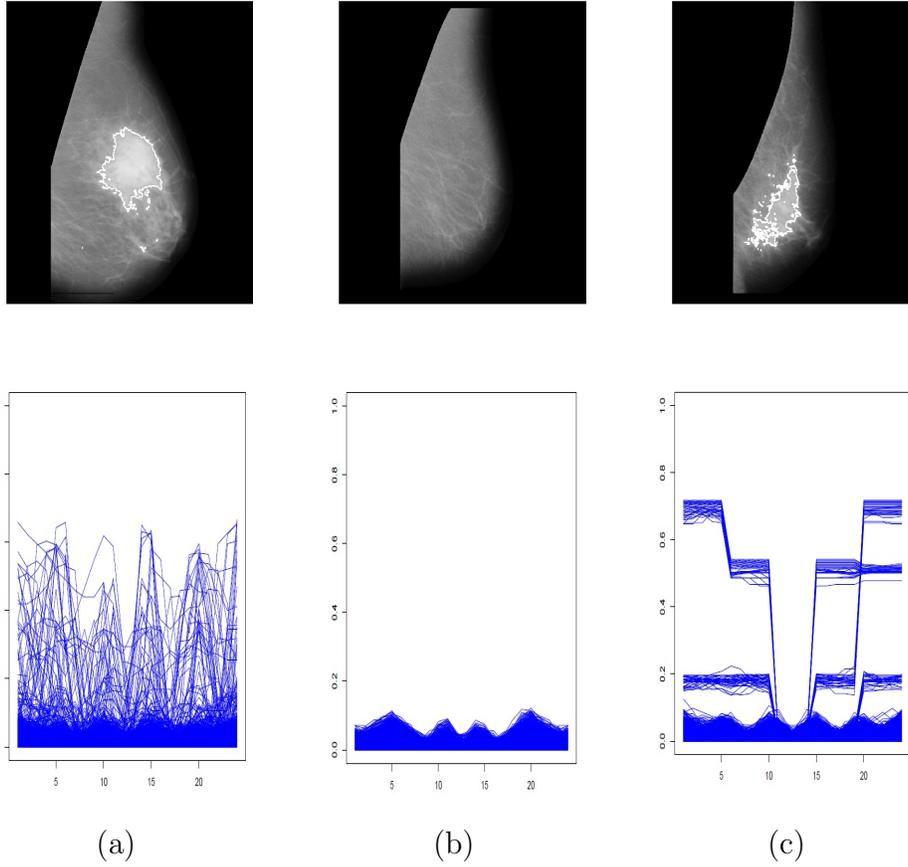


Figura 5.11: Algunos resultados obtenidos. (a) mdb081 de mini-MIAS , (b) mdb272 de mini-MIAS y (c) mdb102 de mini-MIAS.

cada mamografía, está la imagen del comportamiento de los vectores de derivadas direccionales asociados a ella.

En los resultados obtenidos con esta primera idea, observamos patrones interesantes, ya que se logra apreciar que los pixeles seleccionados están asociados a alguna anomalía y a su vez, existe una relación importante entre la anomalía presente y la forma que toman sus vectores direccionales. El proceso a seguir es

agregar herramientas de la literatura de FDA, para aprovechar la mayor cantidad de información que nos puedan brindar estos vectores, y así obtener una propuesta sólida para la etapa de segmentación.

Capítulo 6

Conclusiones y trabajos futuros

6.1. Conclusiones generales

El Análisis de datos funcionales es una rama de la estadística extensa y compleja que aborda gran variedad de problemas y encierra distintos enfoques dentro de ella. Aunque un estudio detallado en un trabajo de esta índole es difícil de lograr, se ha realizado un esfuerzo de síntesis para dar una visión general y actual de las características, metodologías y problemáticas del área, colocando la mayor relevancia en clasificación supervisada, tema donde se desarrolla esta tesis. De la revisión bibliográfica se deduce que FDA es un campo de gran interés y en expansión, como muestra el creciente número de publicaciones y aplicaciones. La importancia del Análisis de datos funcionales radica en gran medida en el valor y alcance de los indicadores o magnitudes con naturaleza funcional. Este tipo de datos se encuentran cada vez más en áreas como la Medicina, Biomecánica, Química, Econometría, Genética, Geofísica, Meteorología, Geología, entre otras. Además, FDA es un campo de investigación relativamente reciente debido a los avances tecnológicos recientes, los cuales permiten capturar cantidades cada vez más grandes y con mayor precisión. Por lo tanto, todavía existe trabajo pendiente, tanto

en aspectos teóricos y computacionales, como en nuevas líneas de investigación.

En esta tesis se presentaron dos propuestas para el mejoramiento del desempeño de clasificadores funcionales, las cuales son: Boosting con clasificadores funcionales y Boosting estructura funcional.

En el primer trabajo, adaptamos algunos clasificadores funcionales, logrando incorporarlos en ciertos algoritmos Boosting. Estas adaptaciones son ideas novedosas propuestas, las cuales fueron estudiadas analíticamente y desarrolladas computacionalmente. Para esto, tuvimos que estudiar el lugar exacto para realizar la ponderación, logrando obtener el mismo efecto que proponen los algoritmos Boosting. En la segunda investigación, modificamos algunos algoritmos Boosting extendiendo su aplicabilidad a datos funcionales. Estas modificaciones fueron estudiadas, buscando obtener la misma estructura del algoritmo para datos multivariados, pero pudiendo incorporar datos funcionales. Esto nos llevó a trabajar con procesos de datos funcionales conocidos en la literatura, pero nunca antes relacionados con algoritmos Boosting.

Los resultados obtenidos con los dos trabajos de investigación muestran un camino novedoso y prometedor. Observamos que cuando nos enfrentamos a datos difíciles de discriminar, nuestras propuestas siempre lograron mejoras significativas. Además, cuando se trabajó con datos reales, conseguimos siempre mejores resultados que los reportados en las referencias consultadas. Por lo tanto, si bien se trata de trabajos donde aún hay bastante por concretar tanto en la parte teórica como en la computacional, se puede concluir que esta línea de investigación es perfectamente aplicable a datos funcionales y abre una puerta a un mayor estudio en el ámbito funcional.

Por otra parte, el estudio que se inició sobre Análisis de imágenes digitales de mamografías, forma parte del problema en particular que trabajamos, tomando

como dato funcional las imágenes. Pudimos desarrollar un método automático y sencillo de implementar computacionalmente, para la etapa de pre-procesamiento, con resultados satisfactorios. Además, se encuentra en desarrollo una propuesta para las demás etapas, con el fin de obtener un método con un desempeño favorable en la clasificación automática de anomalías presentes en una mamografía. Consideramos que el Análisis de imágenes mediante las técnicas de FDA, abre un línea de investigación importante para la utilidad de dichas técnicas.

6.2. Trabajos futuros

En este trabajo nos hemos enfocado principalmente en clasificación supervisada de datos funcionales. Concretamente, hemos trabajado en el mejoramiento del desempeño de clasificadores funcionales, lo que abre varias líneas de estudio a tener en cuenta.

El trabajo realizado con los algoritmos Boosting se ha aplicado directamente a ciertos algoritmos y los resultados son únicamente empíricos. Por lo tanto, uno de los primeros esfuerzos de investigación es precisar desde un marco formal, si la teoría existente para datos finito dimensionales se puede extender para datos funcionales. Además, si es posible, buscar nuevos resultados teóricos para este tipo de datos.

Por otra parte, es necesario precisar las características de cada método, buscando ampliar su aplicabilidad con otros algoritmos. Para concluir, aunque sólo se ha tocado tangencialmente en este trabajo, otro tema interesante es el estudio para formalizar la definición de clasificador base para datos funcionales. De los experimentos realizados no se puede extraer dicha definición, ya que no estaban encaminados a tal efecto. Sin embargo se puede, tratar de obtener una definición

formal de clasificador base, buscando estructurar una teoría de Boosting para datos funcionales.

Finalmente, en el trabajo realizado sobre Análisis de imágenes, es necesario concretar las ideas desarrolladas en la etapa de segmentación, para continuar con la extracción de características, pero desde un punto de vista funcional, igualmente en la etapa de clasificación. Esperamos que la inclusión de las técnicas de FDA en este problema, conlleve a desarrollar nuevos mecanismos que abran paso a nuevas líneas de investigación.

Bibliografía

- Abraham, C., Biau, G. y Cadre, B. (2006). On the kernel rule for function classification. *Ann. Inst. Stat. Math.* 58, 619-633.
- Alonso, A., Casado, D. y Romo, J. (2012). Supervised classification for functional data: A weighted distance approach. *Computational Statistics and Data Analysis.* 56, 2334-2346.
- Baíllo, A. y Cuevas, A. (2008). Supervised functional classification: A theoretical remark and some comparisons. *Arxiv preprint arXiv* 0806.2831.
- Baíllo, A., Cuevas, A. y Fraiman, R. (2011). Classification methods for functional data. In *The Oxford Handbook of Functional Data Analysis*, Ferraty and Romain, Eds. Oxford University Press, USA, 259-297.
- Baíllo, A., Cuesta-Albertos, J.A. y Cuevas, A. (2011). Supervised classification for a family of Gaussian functional models. *Scandinavian Journal of Statistics.* 38, 480-498.
- Ball, J. y Bruce, L. (2007). Digital mammogram spiculated mass detection and spicule segmentation using level sets. *29th Annual International Conference of the IEEE.* 4979-4984.

- Benko, M., Härdle, H. y Kneip, A. Common functional principal components. *The Annals of Statistics*. 37 (1), 1-34.
- Biau, G., Bunea, F. y Wegkamp, M.H. (2003). Functional classification in Hilbert spaces. *IEEE Transactions on Information Theory*. 1(11), 1-8.
- Boor, C. (2001). *A Practical Guide to Splines*. Springer-Verlag, New York.
- Bozek, J., Mustra, M., Delac, K. y Grgic, M. (2009). A survey of image processing algorithms in digital mammography. *Recent Advances in Multimedia Signal Processing and Communications*. 231, 631-657.
- Buhlmann, P. y Hothorn, T. (2007). Boosting algorithms: Regularization, prediction and model fitting. *Statistical Science* 22(4), 477-505.
- Breiman, L. (1998). Arcing classifiers. *Annals of Statistics*. 26, 801-849.
- Cérou, F. y Guyader, A. (2006). Nearest neighbor classification in infinite dimension. *ESAIM Probab. Stat.* 10, 340-355.
- Camilus, K., Govindan, V. y Sathidevi, P. (2010). Computer-aided identification of the pectoral muscle in digitized mammograms. *Journal of Digital Imaging*. 25(5) 562-580.
- Chakraborty, A. y Chaudhuri, P. (2014). On data depth in infinite dimensional spaces. *Annals of the Institute of Statistical Mathematics*. 66(2), 303-324.
- Charlie, T. y Ventura, D. (2012). A direct boosting algorithm for the k -nearest neighbor classifier via local warping of the distance metric. *Pattern Recognition Letters*. 33, 92-102.

- Cheng, H., Shi, X., Min, R., Hu, L., Cai, X. y Du, H. (2006). Approaches for automated detection and classification of masses in mammograms. *Pattern Recognition*. 39(4), 646-668.
- Chen-Chung, L., Chung-Yen, T., Jui, L., Chun-Yuan, Y., y Shyr-Shen, Y. (2012). A pectoral muscle segmentation algorithm for digital mammograms using otsu thresholding and multiple regression analysis. *Computers & Mathematics with Applications*. 64(5) 1100-1107.
- Cuesta-Albertos, J.A. y Fraiman, R. (2007). Impartial trimmed k -means for functional data. *Computational Statistics and Data Analysis*. 51, 4864-4877.
- Cuesta-Albertos, J.A., Fraiman, R. y Ransford, T. (2006). Random projections and goodness-of-fit tests in infinite-dimensional spaces. *Bull Braz Math Soc*. 37(4), 477-501.
- Cuesta-Albertos, J. A. y Nieto-Reyes, A. (2008). The random Tukey depth. *Computational Statistics and Data Analysis* 52 4979-4988.
- Cuevas, A., Febrero, M. y Fraiman, R. (2004). An anova test for functional data. *Computational Statistics and Data Analysis*. 47, 111-122.
- Cuevas, A., Febrero, M. y Fraiman, R. (2006). On the use of the bootstrap for estimating functions with functional data. *Computational Statistics and Data Analysis*. 51, 1063-1074.
- Cuevas, A., Febrero, M. y Fraiman, R. (2007). Robust estimation and classification for functional data via projection-based depth notions. *Computational Statistics and Data Analysis*. 22,481-496.

- Cuevas, A. (2013). A partial overview of the theory of statistics with functional data. *Journal of Statistical Planning and Inference*. 147, 1-23.
- Cuevas, A. y Fraiman, R. (2009). On depth measures and dual statistics: A methodology for dealing with general data. *Journal of Multivariate Analysis*. 100 753-766.
- Dauxois, J., Pousse, A. y Romain, Y. (1982). Asymptotic theory for the principal component analysis of a vector random function: Some applications to statistical inference. *J. Multivariate Anal.* 12, 136-154.
- Delaigle, A., Hall, P., y Apanasovich, T. (2009). Weighted least squares methods for prediction in the functional data linear model. *Electronic Journal of Statistics*. 3, 865-885.
- Delicado, P. (2007). Functional k -sample problem when data are density functions. *Computational Statistics*. 22 (3), 391-410.
- Deville, J. (1974). Méthodes statistiques et numériques de l'analyse harmonique. *Annales de l'INSEE*. 15, 7-97.
- Devroye, L, Györfi, L. y Lugosi, G. (1996). *A Probabilistic Theory of Pattern Reconigition*. Springer-Verlag, New York.
- Escabias, M., Aguilera, A. y Valderrama, M. (2004). Principal components estimation of functional logistic regression: Discussion of two different approaches. *Journal of Nonparametric Statistics*. 16(3-4), 365-384.
- Febrero-Bande, M. y Oviedo M (2012). Statistical Computing in Functional Data Analysis: The R Package *fda.usc*. *Journal of Statistical Software*. 51(4), 1-28.

- Fernandez, B., Guillas, S., y Gonzalez, W. (2005). Functional samples and bootstrap for predicting sulfur dioxide levels. *Technometrics*. 47, 212-22.
- Ferrari, R., Rangayyan, R., Desautels, J., Borges J. y Frère, A. (2004). Automatic identification of the pectoral muscle in mammograms. *Medical Imaging*. 23(2) 232-245.
- Ferraty F. y Vieu P.(2003). Curves discrimination: A nonparametric functional approach. *Computational Statistics and Data Analysis*. 44, 161-173.
- Ferraty, F. y Vieu, P. (2006). *Nonparametric functional data Analysis. Theory and Practice*. Springer-Verlag, New York.
- Ferraty, F. y Vieu, P. (2009). Additive prediction and boosting for functional data. *Computational Statistics and Data Analysis*. 53, 1400-1413.
- Ferraty, F. y Romain, Y. (2011). *The Oxford Handbook of Functional Data Analysis*. Oxford University Press.
- Fraiman, R., Justel, A. y Svarc, M. (2008). Selection of variables for cluster analysis and classification rules. *Journal of the American Statistical Association*. 103, 1294-1303.
- Fraiman, R, y Muñoz, G. (2001). Trimmed means for functional data. *Test*. 10: 419-440.
- Freund, Y. y Schapire, R.E. (1996). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*. 55, 119-139.

- Freund, Y. y Schapire, R.E. (2012). *Boosting: Foundations and Algorithms*. MIT Press, Cambridge, MA.
- Friedman, J., Hastie, T. y Tibshirani, R. (2000). Additive logistic regression: a statistical view of boosting. *Annals of Statistics* 28(2), 337-374.
- Gareth, M. (2002). Generalized linear models with functional predictors. *Journal of the Royal Statistical Society*. 64 (3), 411-432.
- James, G. y Hastie, T. (2001) Functional linear discriminant analysis for irregularly sampled curves. *Journal of the Royal Statistical Society*. 63(3), 533-550.
- Global Cancer Observatory (2017). *International Agency for Research on Cancer*, <http://gco.iarc.fr>.
- Gonzalez, R. y Woods, R. (2008). *Digital Image Processing, 3rd Edition*. Prentice Hall, New Jersey.
- Hastie, T., Tibshirani, T. y Friedman, J. (2008). *The Elements of Statistical Learning*. Springer-Verlag, New York.
- Hall, P. y Presnell, B. (2001). A functional data-analytic approach to signal discrimination. *Technometrics*. 43(1), 1-9.
- Hall, P., Müller, H. y Wang, J. (2006). Properties of principal component methods for functional and longitudinal data analysis. *The Annals of Statistics*. 34(3), 1493-1517.
- He, W., Denton, E. y Zwiggelaar, R. (2014) A novel breast image preprocessing for full field digital mammographic segmentation and risk classification. *Med. Image Underst. Anal.* 40-47.

- He, W., Juette, A., Denton, E., Oliver, A., Martí, R. y Zwiggelaar, R. (2015) A review on automatic mammographic density and parenchymal segmentation. *Int. J. Breast Cancer*. ID 276217, 31.
- Hela, B., Hela, M., Kamel, H., Sana, B. y Najla, M. (2013). Breast cancer detection: a review on mammograms analysis techniques. *10th International Multi-Conference on Systems, Signals and Devices (SSD)*. 1-6.
- Heath, M., Bowyer, K., Kopans, D., Moore R, y Kegelmeyer F. (2001). Digital Database for Screening Mammography. *Medical Physics Publishing*. ISBN 1-930524-00-5.
- Horváth, L. y Kokoszka, P. (2012). *Inference for Functional Data with Applications*. Springer-Verlag, New York.
- James, G. y Hastie, T. (2001). Functional linear discriminant analysis for irregularly sampled curves. *Journal of the Royal Statistical Society*. 63(3), 533-550.
- Krämer, N. (2006). Boosting functional data. In COMPSTAT2006. *Proceedings in Computational Statistics*. 1121-1128.
- Kwok, S., Chandrasekhar, R., Attikiouzel, Y. y Rickard, M. (2004). Automatic pectoral muscle segmentation on mediolateral oblique view mammograms. *IEEE Transactions on Medical Imaging*. 23(9) 1129-1140.
- Li, Y., Chen, H., Yang, Y. y Yang, N. (2013). Pectoral muscle segmentation in mammograms based on homogenous texture and intensity deviation. *Pattern Recognition*. 46(3) 681-691.

- Li, Y. y Hsing, T. (2010). Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data. *The Annals of Statistics*. 38(6), 3321-3351.
- Li, B. y Yu, Q. (2008). Classification of functional data: a segmentation approach. *Computational Statistics and Data Analysis* 52, 4790-4800.
- Lopez-Pintado, S. y Romo, J. (2006). Depth-based classification for functional data. *DIMACS Series in Discrete Mathematics and Theoretical Computer Science. American Mathematical Society*. 72, 103-120.
- Lopez-Pintado, S. y Romo, J. (2009). On the concept of depth for functional data. *Journal of the American Statistical Association*. 104:486, 718-734.
- Ma, F., Bajger, M., Slavotinek, J. y Bottema, M. (2007). Two graph theory based methods for identifying the pectoral muscle in mammograms. *Pattern Recognition*. 40(9) 2592-2602.
- Malagelada, A. (2007). Automatic mass segmentation in mammographic images. *PhD thesis, Universidad de Girona, España*.
- Morris, F. (2015). Functional regression. *Annual Review of Statistics and Its Application*. 2, 1-385.
- Moumena, A. y Ali, E.(2013) Mammogram images thresholding for breast cancer detection using different thresholding methods. *Advances in Breast Cancer Research* 2 72-77.
- Müller, H. y Stadtmüller, U. (2005). Generalized functional linear models. *The Annals of Statistics*. 33 (2), 774-805.

- Murphy, K. (2012). *Machine Learning. A Probabilistic Perspective*. MIT Press, Cambridge, MA.
- Peña D. (2002). *Analisis de datos multivariantes*. McGraw-Hill, España.
- Quin, J. y Choi, T. (2011). *Gaussian Process Regression Analysis for Functional Data*. Taylor & Francis Group, New York.
- Ramsay, J.O. y Dalzell, C.J. (1991). Some tools for functional data analysis. *Journal of the Royal Statistical Society*. 53, 539-572.
- Ramsay, J.O. y Silverman, B.W. (2002). *Applied Functional Data Analysis: Methods and Case Studies*. Springer-Verlag, New York.
- Ramsay, J.O. y Silverman, B.W. (2005). *Functional Data Analysis, Second Edition*. Springer-Verlag, New York.
- Ramsay, J.O., Hooker, G. y Graves, S. (2009). *Functional Data Analysis with R and MATLAB*. Springer-Verlag, New York.
- Ratna, D. y Neelima, G. (2012). Image segmentation by using histogram thresholding. *IJCSET*. 776-779.
- RSNA (2017). *Radiological Society of North America*. <http://www.rsna.org/>.
- Shantanu, G., Vandana, D., Priscilla, J. y Ronald, L. (2013). Architectural distortion of the breast. *American Journal of Roentgenology*. 201(5), 662-670.
- Schapire, R.E. (1990). The Strength of weak learnability. *Machine Learning*. 5, 197-227.

- Schapire, R.E., Freund, Y. y Lee, W.S. (1998). Boosting the margin: a new explanation for the effectiveness of voting methods. *The Annals of Statistics*. 26(5) 1651-1686.
- Sguera, C., Galeano, P. y Lillo, R. (2014). Spatial depth-based classification for functional data. *Test*. 23 (4), 725-750.
- Sickles, E. (2007). El espectro de las asimetrías en la mama: características radiológicas, estudio diagnóstico y tratamiento. *Clínicas Radiológicas de Norteamérica*. 45, 765-771.
- Singh, T., Singh, O., Singh, K., Sinamand, T. y Singh, T.R. (2010). Image enhancement by adaptive power-law transformations. *Bahria University Journal of Information and Communication Technology*. 3 (1) 29-37.
- Sørensen, H., Goldsmith, J. y Sangalli, L. (2013). An introduction with medical applications to functional data analysis. *Statistics in Medicine*. 32, 5222-5240.
- Suckling, J., Parker, J., Dance, D., Astley, S., Hutt, I., Boggis, C., Ricketts, I., Stamatakis, E., Cerneaz, N., Kok, S., Taylor, P., Betal, D. y Savage, J. (1994). The mammographic image analysis society digital mammogram database, Excerpta Medica. *International Congress Series*. 1069 375-378.
- Wang, L., Zhu, M., Deng, L. y Xin, Y. (2010). Automatic pectoral muscle boundary detection in mammograms based on markov chain and active contour model. *Journal of Zhejiang University SCIENCE C*. 11(2) 111-118.