



# Clasificación basada en la estimación de Parzen en espacios generalizados de disimilitudes

Carlos Alberto Trujillo Pulgarín

Universidad Nacional de Colombia  
Facultad de Ciencias Exactas y Naturales  
Departamento de Matemáticas y Estadística  
Manizales, Colombia  
2012

# Clasificación basada en la estimación de Parzen en espacios generalizados de disimilitudes

Carlos Alberto Trujillo Pulgarín

Tesis de grado presentada como requisito parcial para optar al título de:  
**Magíster en ciencias Matemática Aplicada**

Director:

Dr. Ing. Mauricio Orozco Alzate

Línea de Investigación:

Clasificación y Reconocimiento de Patrones

Grupo de Investigación:

Cálculo Científico y Modelamiento Matemático

Universidad Nacional de Colombia

Facultad de Ciencias Exactas y Naturales

Departamento de Matemáticas y Estadística

Manizales, Colombia

2012

# Classification Based on the Parzen estimation in generalized dissimilarity Spaces

Carlos Alberto Trujillo Pulgarín

Thesis for the degree of  
**Master in Science - Applied Mathematics**

Supervisor:  
Dr. Eng. Mauricio Orozco-Alzate

Line of Research:  
Classification and Pattern Recognition  
Research Group:  
Cálculo Científico y Modelamiento Matemático

Universidad Nacional de Colombia  
Faculty of Natural and Exact Sciences  
Department of Mathematics and Statistics  
Manizales, Colombia  
2012

A mi familia...

# Agradecimientos

Quiero agradecer a mi director de trabajo de grado, Mauricio Orozco Alzate, por la orientación y apoyo en la realización de este trabajo, por cada una de las horas dedicadas, y la guía en todo lo necesario para la buena realización del mismo.

A mis compañeros en la sala de Maestría en Ciencias Matemática Aplicada por cada momento de risas y alegría porque sin esos momentos la monotonía y aburrimiento hubiesen abundado en este camino recorrido. Al grupo de cálculo científico y modelamiento matemático, por el apoyo y espacio de trabajo brindado, importante para consecución de este proyecto, en particular quiero agradecer al director del grupo, profesor Carlos Daniel Acosta. Finalmente a cada uno de los profesores por el tiempo y ayuda en la concepción de todo este gran proyecto.



## Resumen

Generalmente, en reconocimiento automático de patrones, un objeto está representado por sus características a través de un vector de  $m$  componentes, es decir perteneciente al espacio  $m$ -dimensional. El estudio sobre el espacio de características es la forma usual de trabajo en clasificación. Los objetivos de esta tesis incluyen examinar y estudiar una representación alternativa de los objetos, basada en medidas de disimilitud que, en este caso, no serán determinadas de la forma usual, la cual consiste en hallar la distancia entre los objetos en el espacio de disimilitudes, sino en hallar la distancia de objetos a líneas de características [21]. Siendo el concepto de línea de características el que permite realizar una generalización del espacio de disimilitudes, obteniendo de esta manera el denominado espacio generalizado de disimilitudes. Se pretende mostrar qué propiedades tiene el espacio generalizado de disimilitudes y como objetivo principalmente, si en él es posible estimar densidades de probabilidad y hacer uso de ellas para la clasificación de patrones. Tal estimación se basa en la estimación de densidades de probabilidad de Parzen, la cual se encuentra definida sobre el espacio de características; por consiguiente, modificando la estructura de este método, en particular sobre el algoritmo de la estimación de Parzen y adaptándolo para el espacio generalizado de disimilitudes, se pretende implementar el clasificador correspondientes a las modificaciones y adaptaciones del estimador de Parzen. Con el algoritmo modificado, la implementación del clasificador permitirá mostrar la eficiencia o ineficiencia de este método de clasificación sobre el espacio generalizado de disimilitudes, para lo cual se requiere de la adquisición de bases de datos -artificiales, obtenidas haciendo uso de sistemas de sensores o que ya estén dadas-, la medición del desempeño obtenido por este nuevo clasificador y la comparación con los resultados de clasificadores ya establecidos mediante el uso de diferentes métodos de validación. Es importante notar la importancia de la noción de línea de característica y tener presente el algoritmo de estimación de probabilidad de Parzen para espacios de características el cual ya se encuentra establecido.

**Palabras Claves:** Espacio generalizado de disimilitudes, espacio de características, espacio de disimilitudes, línea de características, representación de disimilitudes, representación generalizada de disimilitudes, estimación de densidades de probabilidad de Parzen, estimación del parámetro de suavizado, inmersiones, reconocimiento de patrones.

# Abstract

Usually in automatic pattern recognition, an object is represented by its features, through a  $d$  components vector, i.e. belonging to the  $d$ -dimensional space. The study on the feature space is the usual form to work in classification. Part of the objectives is to examine and explore an alternative objects representation, based on dissimilarity measures, in this case these measures are determined, not in the usual way, which is to find the distance between objects in the dissimilarity space, in this case these measures are determined in a generalized dissimilarity space, this generalization is performed using features lines. We try to show which properties have this new space and the principal objective, estimate probability densities and used for classification, this estimate is based on the Parzen's estimated probability densities which in principle is defined on the feature space, therefore modifying the structure of the estimation algorithm and adapting it for Parzen generalized dissimilarity space, is to implement the classifier corresponding to the modifications and adaptations of the Parzen estimator. With the modified algorithm, the creation of the classifier will show the efficiency or inefficiency of this alternative method of classification, which is required for the acquisition of datasets, whether artificial, obtained using sensor systems or already given, to observe the performance obtained by this new classification and compare the results of classifications established by using different methods of validation. It is important to note the importance of the notion of feature line and Parzen probability estimation in the feature space.

**Keywords:** Generalized dissimilarity space, feature space, dissimilarity space, feature line, dissimilarity representation, generalized dissimilarity representation, Parzen probability estimation, smoothing parameter estimation, embedding, pattern recognition.



# Contenido

Agradecimientos	v
Resumen	vii
Lista de símbolos	xi
<b>1. Introducción</b>	<b>1</b>
<b>2. Marco teórico</b>	<b>3</b>
2.1. Espacio de características . . . . .	5
2.2. Espacio de disimilitudes . . . . .	6
2.2.1. Representación de disimilitudes . . . . .	9
2.3. Clasificadores . . . . .	12
2.4. Estimador de Parzen . . . . .	13
2.4.1. Algoritmo de clasificación . . . . .	16
<b>3. Estado del arte</b>	<b>18</b>
3.1. Espacio generalizado de disimilitudes . . . . .	18
3.1.1. Línea de características . . . . .	19
3.1.2. Representación generalizada de disimilitudes . . . . .	21
<b>4. Estimación y clasificador de Parzen en espacios generalizados de disimilitudes</b>	<b>23</b>
4.1. Estimador de Parzen . . . . .	23
4.2. Estimación del parámetro de suavizado . . . . .	25
4.3. Propiedades del espacio generalizado de disimilitudes . . . . .	26
4.3.1. Inmersiones (embedding) . . . . .	30
4.4. Propiedades de la matriz generalizada de disimilitudes . . . . .	33
<b>5. Marco experimental</b>	<b>35</b>
5.1. Bases de datos . . . . .	35
5.1.1. Clases difíciles distribuidas normalmente . . . . .	36
5.1.2. Clases de Highleyman . . . . .	39
5.1.3. Conjunto de datos Laryngeal 1 . . . . .	40
5.1.4. Conjunto de datos Laryngeal 2 . . . . .	42

5.1.5. Conjunto de Laryngeal 3 . . . . .	43
5.1.6. Conjunto de datos Iris . . . . .	45
5.1.7. Conjunto de datos banana . . . . .	47
5.1.8. Evaluación y comparación de la clasificación . . . . .	51
5.2. Metodología de evaluación . . . . .	52
<b>6. Resultados y discusiones</b>	<b>55</b>
6.1. Resultados . . . . .	55
6.1.1. Clases difíciles distribuidas normalmente . . . . .	57
6.1.2. Clases de Highleyman . . . . .	69
6.1.3. Conjunto banana . . . . .	75
6.1.4. Conjunto Círculos concéntricos . . . . .	78
6.1.5. Conjunto clases lituanas . . . . .	81
6.1.6. Problema de clasificación simple de dos clases gaussianas . . . . .	84
6.1.7. Conjunto problema de truncamiento . . . . .	87
6.1.8. Conjunto Laryngeal 1 . . . . .	90
6.1.9. Conjunto Laryngeal 2 . . . . .	96
6.1.10. Conjunto Laryngeal 3 . . . . .	105
6.1.11. Conjunto iris . . . . .	113
6.1.12. Conjunto wine . . . . .	121
6.1.13. Conjunto Malaysia . . . . .	124
6.1.14. Conjunto diabetes . . . . .	127
6.1.15. Conjunto biomed . . . . .	130
6.1.16. Conjunto heart . . . . .	133
6.1.17. Conjunto ionosphere . . . . .	141
6.1.18. Conjunto IMOX . . . . .	144
6.1.19. Conjunto glass . . . . .	147
6.1.20. Conjunto de imágenes ORL . . . . .	150
6.2. Discusiones . . . . .	162
<b>7. Conclusiones</b>	<b>173</b>
<b>A. Anexo: PRTools</b>	<b>176</b>
<b>B. Valores del parametro de suavizado <math>\sigma_h</math></b>	<b>177</b>
<b>Bibliografía</b>	<b>181</b>

# Lista de símbolos

## Símbolos con letras latinas

Símbolo	Definición
$A, B, \dots, Z$	Matrices, conjuntos, espacios vectoriales
$C$	Número de clases
$S$	Matriz de covarianzas
$a, b, \dots, z$	Escalares, vectores, objetos
$d$	Medida de disimilitud
$D$	Matriz de disimilitud
$p$	Función de probabilidad
$N_s$	Número de elementos en un conjunto de entrenamiento
$Z_n$	Vector de medida (o Características)
$T_S$	Conjunto de entrenamiento
$p(\cdot \cdot)$	Probabilidad condicional
$K$	$ \Omega $
$k = 1, 2, \dots, K$	Número de vectores de medida
$n = 1, 2, \dots, N_s$	Número de vectores en el conjunto de entrenamiento
$N_k$	Número de elementos con clase $\omega_k$
$(z_{k,n})$	Vector de medida y su clase verdadera
$D(T, R)$	Representación de disimilitudes
$D(T_L, R_L)$	Representación generalizada de disimilitudes

## Símbolos con letras griegas

Símbolo	Definición
$\theta_b$	Clase verdadera del $n$ -esimo objeto
$\Omega$	Conjunto de clases
$(z_n, \theta_n)$	Vector de medida y su clase verdadera
$\omega_k$	Muestra con clase k
$\delta$	Métrica

# 1. Introducción

Usualmente, en reconocimiento automático de patrones, los estudios de clasificación se realizan sobre el espacio de características. En él, los clasificadores se sintonizan haciendo uso de objetos de muestra para entrenamiento. Sin embargo, el espacio de características, no es el único conjunto de trabajo y estudio en el cual la tarea de clasificación se puede desarrollar, teniendo en cuenta que, en algunos casos la definición de funciones no resulta ser sencilla de comprender. Por esto se plantea una alternativa: las disimilitudes entre nuevos objetos y objetos prototipo (objetos de muestra) que pueden ser utilizadas directamente en el procesamiento de medidas de entrada. La regla del vecino más cercano (*1-NN*) se aplica convencionalmente para clasificar un nuevo objeto sobre la base de sus disimilitudes al conjunto de objetos de prueba. Este procedimiento resulta ser generalmente, desde el punto de vista computacional, costoso o pierde precisión cuando se selecciona sólo un número pequeño de objetos prototipo. Por esto, en estudios recientes se ha examinado la generalización del espacio de disimilitudes mediante la definición de líneas de características [21], modelos ocultos de Markov [2] o agrupamiento de prototipos [14]. En la generalización mediante líneas de características, se generaliza cada par de objetos prototipo de la misma clase mediante una función lineal, que se denomina, línea de características. Un nuevo objeto se clasifica sobre la base de sus disimilitudes al conjunto de prueba, en este caso no a los objetos de prueba usuales, sino al conjunto de líneas de características generadas por objetos prototipo previamente seleccionados.

En el momento de la clasificación de un nuevo objeto, se propone en esta tesis la utilización de la estimación de densidades de probabilidad de Parzen, que permite el desarrollo de un clasificador basado en tal estimación. En esta, no es necesario tener conocimiento de la distribución de densidad de probabilidad del conjunto de datos, caso contrario de los clasificadores usuales (bayesianos) que asumen que la distribución de los datos es normal, lo cual podría llegar a condicionar los resultados. En este trabajo se propone, mediante el uso de un conjunto de representación [24], hacer uso de las líneas de características generadas a través de este conjunto y del estimador de Parzen modificado para el espacio generalizado de disimilitudes, realizar la clasificación de objetos y mostrar el beneficio que trae el hecho de calcular la representación, no sobre los objetos prototipo en un conjunto de representación, sino con las líneas de características generadas por los mismos, además de mostrar que el conjunto de entrenamiento necesario para realizar la clasificación, no debe ser grande para que los resultados obtenidos sean óptimos. Los experimentos muestran los beneficios obtenidos

al realizar el trabajo de clasificación mediante líneas de características, a través de la comparación con otros clasificadores previamente establecidos, así como el costo computacional que implica este método y el tipo de conjuntos sobre el cual el desempeño del mismo es mejor.

## 2. Marco teórico

Desde temprana edad el ser humano adquiere la capacidad de reconocer diversidad de caracteres y hechos físicos que se desarrollan en su entorno. El tamaño de los caracteres, si son escritos a mano o a computador, la forma, la textura, el color, son fácilmente reconocidos desde la juventud. Esta habilidad y capacidad de reconocimiento se adquiere mediante la experiencia y aprendizaje continuo, a través de la enseñanza y práctica guiada, de manera empírica o científica.

De manera similar, al igual que para los seres humanos, las máquinas siguen este mismo comportamiento, siendo en este caso nosotros quienes les enseñamos a tomar esta habilidad de reconocimiento. Por lo tanto diremos que *el reconocimiento de patrones es el estudio de cómo las máquinas pueden observar el medio ambiente, aprender a distinguir los patrones de interés y sus antecedentes y tomar decisiones acertadas y razonables sobre la categoría de los patrones* [7].

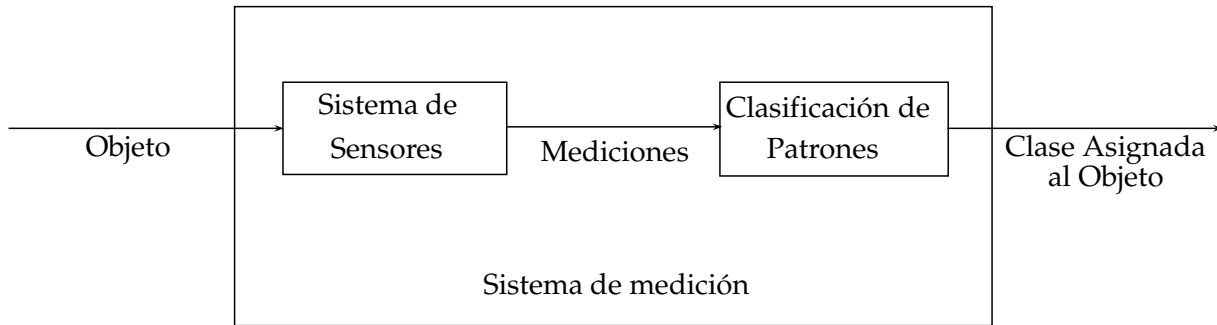
El reconocimiento automático, la descripción, la clasificación y el agrupamiento de patrones son problemas de relevante interés en diversos campos del conocimiento, como lo son las ingenierías y las ciencias, entendiendo como patrón *"lo opuesto del caos, una entidad, vagamente definida, que se le podría dar un nombre"* [29].

Dado un patrón, su reconocimiento o clasificación se ve abarcado en los siguientes tipos de tareas [29],

1. Clasificación supervisada, en la cual los patrones de entrada son identificados como miembros de una clase predefinida.
2. Clasificación no supervisada, en la cual el patrón es asignado a una clase desconocida hasta ese momento.

Luego el problema de reconocimiento se transforma en una tarea de clasificación o categorización, la cual se desarrolla como se muestra en la figura 2-1, las clases son una a una definidas por sistemas diseñados o son adquiridas sobre la base de las disimilitudes de los patrones, dependiendo si es clasificación supervisada o no supervisada respectivamente.

Resulta fundamental para el reconocimiento automático de patrones el uso de herramientas computacionales y tecnológicas, que gracias a su crecimiento y desarrollo, permiten que el procesamiento de grandes conjuntos de datos sea realizado, facilitando el uso de elaborados y diversos métodos para su análisis, contribuyendo a que la tarea de clasificación sea llevada a cabo. Al mismo tiempo en que el poder computacional crece, son necesarios los sistemas



**Figura 2-1.:** Clasificación de Patrones

de reconocimiento automático de patrones, dado que el tamaño de los conjuntos de datos también aumenta gracias a la posibilidad que ofrecen los avances tecnológicos. El diseño de estos sistemas de reconocimiento consiste básicamente en tres aspectos [7]:

1. Adquisición y procesamiento de los datos.
2. Representación de los datos.
3. Toma decisiones.

Parte del problema de clasificación se basa en la selección de sensores, técnicas de pre-procesamiento, esquemas de representación y realización de modelos de decisión.

Aprender de un conjunto de muestras (conjunto de entrenamiento) es una de las características más importantes y necesarias en los sistemas de reconocimiento de patrones. Los cuatro enfoques más conocidos y desarrollados en este campo son [7, 4]:

1. Geométrico (Clustering), Los patrones deben poderse graficar. En éste enfoque se emplea el cálculo de distancias, geometría de formas, vectores numéricos, puntos de atracción, etc.
2. Ajuste de plantilla.
3. Ajuste sintáctico o estructural, se basa en encontrar las relaciones estructurales que guardan los objetos de estudio, utilizando la teoría de lenguajes formales, teoría de autómatas, etc. El objetivo es construir una gramática que describa la estructura del universo de objetos.
4. Redes neuronales.
5. Clasificación estadística, Se basa en la teoría de la probabilidad y la estadística, utiliza análisis de varianzas, covarianzas, dispersión, distribución, etc.
6. Lógico-combinatorio: se basa en la idea de que el modelado del problema debe ser lo más cercano posible a la realidad del mismo, sin hacer suposiciones que no estén



fundamentadas. Se utiliza para conjuntos difusos y utiliza lógica simbólica, circuitos combinatoriales y secuenciales, etc.

## 7. Computación bioinspirada.

Siendo el enfoque de clasificación estadística el de mayor estudio y que ha presentado grandes avances, gracias a sus posibilidades de estudio en diferentes campos. Resulta importante tomar en cuenta el conocimiento previo que se debe tener acerca del manejo de densidades de probabilidad condicional, probabilidad a priori y posterior, y el conocimiento de clasificadores bayesianos que permitirá un mejor entendimiento de los temas aquí presentes [4, 7, 12].

## 2.1. Espacio de características

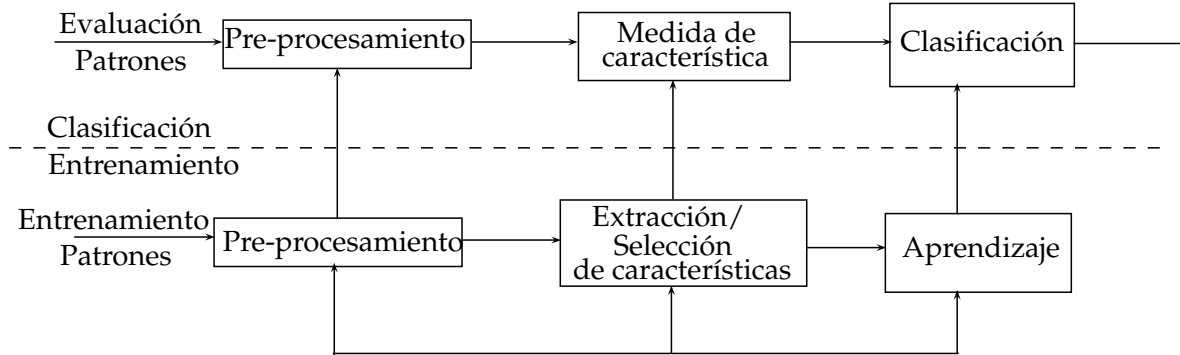
En clasificación estadística o reconocimiento estadístico de patrones, cada patrón es representado en términos de  $m$  características o medidas y es visto como un punto en el espacio  $m$ -dimensional, es decir como un vector con  $m$ -componentes. El objetivo es elegir las características que permitan a los vectores patrón, pertenecer a diferentes clases que ocupan regiones compactas y disjuntas en el espacio  $m$ -dimensional de características.

Diremos entonces que, en reconocimiento estadístico de patrones, los estudios de clasificación se realizan sobre el espacio de características. La "efectividad" del espacio de representación (conjunto de características) es determinada por cómo los patrones, que denominaremos objetos, de diferentes clases, pueden ser separados.

Para un conjunto de entrenamiento dado de cada clase, el objetivo es establecer límites de decisión en el espacio de características, que permitan separar los objetos que pertenecen a clases diferentes. Estos límites de decisión son establecidos por la distribución de probabilidad de los objetos pertenecientes a cada clase. En reconocimiento estadístico de patrones los sistemas de reconocimiento son operados de dos maneras como se muestra en la figura 2-2:

1. Entrenamiento (Aprendizaje) y
2. Clasificación (Prueba)

El papel de la etapa de pre-procesamiento es segmentar el objeto de interés desde el fondo, eliminar el ruido, normalizar el patrón, y cualquier otra operación que contribuya en la definición de una representación compacta del objeto. En la etapa de entrenamiento, el paso de extracción/selección de características encuentra las características adecuadas para representar los objetos de entrada y el clasificador se entrena para la partición del espacio de características. El camino de re-alimentación permite al diseñador optimizar el procesamiento previo y las estrategias de extracción/selección de características. En la etapa de clasificación, el clasificador entrenado asigna al objeto de entrada, una de las clases de objeto en consideración, con base en las medidas de características.



**Figura 2-2.:** Reconocimiento estadístico de patrones

La toma de decisiones en reconocimiento estadístico de patrones se pueden resumir de la siguiente manera: Dado un objeto, se le asignará una de las clases,  $\omega_1, \omega_2, \dots, \omega_c$ , con base en el vector de  $m$  valores de características  $x = (x_1, x_2, \dots, x_d)$ . Se asume que las características tienen una función condicionada de densidad de probabilidad o de masa (dependiendo de si las funciones son continuas o discretas), sobre las clases de objetos. Por lo tanto, un vector de objetos  $x$  pertenece a la clase  $\omega_i$  si es visto como una observación tomada aleatoriamente de la función de clases de probabilidad condicional  $p(x|w_i)$ . Una serie de reglas de decisión conocidas, incluida la regla de decisión de Bayes, la regla de máxima verosimilitud (que puede ser vista como un caso particular de la regla de Bayes), y la regla de Neyman-Pearson, son tomadas en cuenta para poder definir los límites de decisión [7].

Al momento de realizar la clasificación es necesario, para un conjunto de datos dado, determinar los conjuntos de entrenamiento y prueba, con los cuales el clasificador tomado sea entrenado y luego realice la clasificación de los objetos presentes en el conjunto de evaluación. El modo de escoger esto dos conjuntos, entrenamiento y prueba, usualmente depende del problema en el cual se esté trabajando, aunque es común seleccionarlos de manera aleatoria. En casi todos los casos se asume que las muestras (objetos de muestra o prototipo) están independiente e idénticamente distribuidas. Esto quiere decir que las muestras son seleccionadas de la misma población de objetos y, en el caso más simple, con igual probabilidad. Además, la probabilidad de que un miembro de la población sea seleccionado no depende de la selección de otros miembros de la población [12].

## 2.2. Espacio de disimilitudes

En ocasiones la definición de funciones en el espacio de características no resulta ser sencilla de comprender, se requiere de conocimiento extra acerca del problema que se esté tratando, para determinar un modelo que permita realizar la clasificación de un conjunto de datos, bien sea porque no se pueden resaltar de manera clara las clases presentes en él, porque re-

sultan ser difíciles de obtener, o porque no resulta fácil poder determinar las características apropiadas para solucionar el problema; en otras palabras tienen poco poder discriminatorio [24, 25].

En el enfoque tradicional de reconocimiento automático de patrones, los objetos se representan como vectores en el espacio  $m$ -dimensional, en los cuales cada una de las  $m$  componentes representa una característica del objeto, este espacio  $m$  dimensional es el que se denomina espacio de características, sin embargo no es el único espacio donde los problemas de clasificación pueden ser discutidos [27]; por esto, y a raíz de los diversos inconvenientes que en ocasiones puede generar el trabajo con características, se plantea una alternativa: un razonamiento y método diferente de estudiar el problema de la clasificación de objetos, las similitudes entre nuevos objetos y objetos prototipo, que pueden ser utilizadas directamente en el procesamiento de medidas de entrada.

Desde hace unos años el trabajo con este método ha tomado gran importancia debido a la posibilidad que brinda el manejo de las diferencias o similitudes entre los objetos de un conjunto, sin que sus características tengan gran relevancia al momento de tomar decisiones. Resulta importante aclarar, que van a existir problemas en los cuales el uso de las características de los objetos va a facilitar su solución, pero la posibilidad de tener una diversidad de métodos ayuda a facilitar la tarea de clasificación.

La noción de similitud juega un papel importante en la formación de clases, por esto se dice que debe observarse como un puente entre las observaciones de objetos y las propiedades de los mismos. En esencia, objetos similares pueden agruparse de manera que formen una clase, y consecuentemente una clase es un conjunto de objetos similares, aunque esta noción es relativa, dado que en ocasiones dos objetos de una misma clase pueden, de acuerdo a sus similitudes, ser muy diferentes. Supongamos, por ejemplo, que queremos clasificar a un grupo de personas de acuerdo a la familia que pertenezcan, se diría que cada uno de los miembros de una familia pertenece a una clase, si nos basamos en las similitudes entre las personas podría darse el caso en que dos de ellas sean parecidas, pero pertenezcan a familias diferentes, lo que al final sería un error de clasificación, así que el concepto de formar clases a partir de las similitudes (recordando que similitud y disimilitud son conceptos inversos) suele ser un concepto relativo de acuerdo al contexto de desarrollo o procedimiento.

Las disimilitudes se representan de forma general, como una matriz de medidas de disimilitudes, en la cual se relacionan los objetos del conjunto de entrenamiento, de la siguiente manera:

Sea  $X$ , conjunto de datos, para un subconjunto  $T = \{x_1, x_2, \dots, x_N\}$  de  $X$ , que denota el conjunto de entrenamiento, una representación de disimilitudes es una matriz de valores de disimilitud de tamaño  $N \times N$ ,  $D(T, T)$ :

$$D(T, T) = \begin{matrix} & \begin{matrix} x_1 & x_2 & \cdots & x_n & \cdots & x_N \end{matrix} \\ \begin{matrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_N \end{matrix} & \begin{pmatrix} d_{11} & d_{12} & \cdots & d_{1n} & \cdots & d_{1N} \\ d_{21} & d_{22} & \cdots & d_{2n} & \cdots & d_{2N} \\ d_{31} & d_{32} & \cdots & d_{3n} & \cdots & d_{3N} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ d_{N1} & d_{N2} & \cdots & d_{Nn} & \cdots & d_{NN} \end{pmatrix} \end{matrix} \quad (2-1)$$

Un objeto no etiquetado  $x$ , que será clasificado, se representa como,

$$d_x = (d_1 \ d_2 \ \cdots \ d_n \ \cdots \ d_N) \quad (2-2)$$

Es decir, la clasificación de objetos, en el espacio de disimilitudes, se determina teniendo en cuenta que nuevo objeto se representa por su distancia al conjunto de entrenamiento, en otras palabras a los objetos en el conjunto de entrenamiento [27]. Estas distancias se calculan de acuerdo sea el caso, generalmente se utiliza la distancia euclidiana, aunque existen diversas definiciones de distancias que se pueden emplear. Las medidas de disimilitudes difieren de acuerdo a varios conjuntos de datos o aplicaciones. Se asume que las distancias son no negativas y que los objetos tienen distancia cero si y solo si son idénticos, es decir las medidas de disimilitud (distancias) en el espacio de disimilitudes son métricas, lo que implica que las siguientes afirmaciones son ciertas:

**Definición 1** (Métrica). *Una métrica sobre un conjunto  $X$  es una función*

$$\delta : X \times X \rightarrow \mathbb{R} \quad (2-3)$$

*que cumple las siguientes propiedades:*

1.  $\delta(x, y) \geq 0$ , para todo  $x, y \in X$
2.  $\delta(x, y) = 0$  si y solamente si  $x = y$  Reflexividad
3.  $\delta(x, y) = \delta(y, x)$ , para todo  $x, y \in X$  Simetría
4.  $\delta(x, y) + \delta(y, z) \geq \delta(x, z)$  para todo  $x, y, z \in X$  Desigualdad triangular

A la pareja  $(X, \delta)$ , con  $\delta$  métrica, se le denomina **espacio métrico** [19].

Garantizar que las distancias empleadas, para determinar las disimilitudes entre los objetos, son métricas en el espacio de disimilitudes garantizará la definición de clasificadores sobre representaciones de disimilitudes.

Para la tarea de clasificación se plantean dos métodos:

- Basado en características (interpretando los datos de las distancias como en el espacio de características),
- Basado en reglas de decisión.

Una forma sencilla de afrontar el problema de clasificación se basa en la relación-distancia entre los objetos, lo que naturalmente conduce a métodos de reglas de decisión, por ejemplo la regla del vecino más cercano ( $1\text{-}NN$ ) [11][Pág. 155-162], [12], el clasificador de Parzen [12][Pág. 150-155], máquina de vectores de soporte ( $SVC$ ) [12][Pág. 168-163], [27].

El uso de disimilitudes entre los objetos puede ser tratado como una descripción específica de un objeto en el espacio de características y no cambia de manera esencial el enfoque clásico basado en características, aunque en este caso el número de muestras es igual a su dimensión (por lo que se consideran matrices cuadradas) y cada valor expresa la disimilitud entre dos objetos. Por lo tanto, en general, cualquier clasificador sobre el espacio de características puede ser empleado sobre una representación en el espacio de disimilitudes.

Se plantean dos enfoques para usar la matriz de representación de disimilitudes con el propósito de clasificación:

- El enfoque del vecino más cercano,
- El enfoque de la representación de disimilitudes,

en donde el primero es el que conocemos como la regla del vecino más cercano ( $1\text{-}NN$ ), y el segundo resulta ser de gran relevancia al momento de plantear la generalización del espacio de disimilitudes.

### 2.2.1. Representación de disimilitudes

En reconocimiento de patrones, un conjunto de datos es representado mediante una matriz, en la cual cada una de sus filas representa un objeto y cada una de las columnas una característica de tal objeto, esto basado en el espacio de características. Al momento de hacer uso de las disimilitudes entre los objetos, consideramos cada posición en la matriz como el valor de la distancia entre dos objetos del conjunto, es decir cada columna pasa ahora a representar un objeto, obteniendo lo que denominamos la *matriz de disimilitudes*.

Para la tarea de clasificación, la manera de proceder, tanto en el espacio de disimilitudes como en el espacio de características, es generar conjuntos de entrenamiento y prueba a partir del conjunto de datos original, teniendo en cuenta que en el uso de disimilitudes surge el problema de la dimensión con respecto a la representación matricial, dado que se consideran las distancias entre todos los objetos del conjunto. Esto inevitablemente lleva al aumento del tamaño de la matriz, que finalmente produce altos costos computacionales y problemas en

la tarea de los expertos al momento de definir modelos de clasificación. Por esto se plantea la idea de sólo seleccionar un número representativo de objetos del conjunto, para generar el conjunto de entrenamiento, que denominaremos objetos prototipo, de manera que la clasificación no se vea afectada y se reduzcan considerablemente los altos costos que trae consigo el uso de conjuntos de alta dimensión.

Anteriormente se planteó definir a una clase como un conjunto de objetos similares, esto implica que el concepto de "similitud" resulta ser más básico que el de "característica" o el de "clase". Utilizando la noción de proximidad (en lugar de características) como un concepto primario, se renueva el área de aprendizaje automático en una de sus bases, es decir, la representación de objetos. En las medidas de proximidad se capturan tanto la información estadística como la estructural de los objetos y, por tanto, se forma un puente natural entre estos enfoques del reconocimiento de patrones: el estructural y el estadístico. Existen dos tipos principales de las representaciones que se consideran: las que son aprendidas y las que están optimizadas (o fijas). Las representaciones de proximidad se pueden dividir en relativas y conceptuales. En una representación relativa, un par de objetos están relacionados entre sí para medir el valor de la proximidad entre ellas. La representación de proximidad puede ser extendida para describir una relación de una entidad a una serie de ellas, o una relación de un modelo para todo el concepto. Estas representaciones se llaman conceptuales. Ahora, respecto a representaciones relativas, encontramos la representación de disimilitudes.

**Definición 2** (Representación de disimilitudes). *Sea  $R := \{p_1, p_2, \dots, p_n\}$  una colección de objetos llamado conjunto de representación o conjunto de prototipos y  $d$  una medida de disimilitud, la disimilitud  $d$  es computada o derivada de los objetos directamente, sus representaciones de sensores, representaciones en cadena u otra representación intermedia,  $d(p_i, p_j)$  es usada para referirnos a  $d(f(p_i), f(p_j))$  donde  $f(p_i)$  corresponde a alguna representación intermedia de  $p_i$ . Una representación de disimilitud de un objeto  $x$  es un conjunto de disimilitudes entre  $x$  y los objetos de  $R$  expresados como un vector  $D(x, R) = [d(x, p_1), d(x, p_2), \dots, d(x, p_n)]$ . Consecuentemente, para una colección de objetos  $T$  se extiende a la matriz de disimilitudes  $D(T, R)$ . La ventaja de un conjunto de representación  $R$  es que éste es un conjunto relativamente pequeño de objetos representativo para un dominio considerado. La representación más simple de disimilitud es cuando  $D(R, R)$ , siendo una matriz cuadrada con ceros en su diagonal. En general  $R$  será un subconjunto de  $T$ , aunque podría ser completamente distinto[24][Pág. 168].  $T$  es el conjunto de entrenamiento y  $R \subseteq T$ , recordando que un clasificador se debe primero entrenar y luego evaluar, en este caso la matriz de disimilitudes no necesariamente será una matriz cuadrada.*

La intención que se presenta con esta definición es que en algunos casos, la descripción posible basada en las características de los objetos podría ser difícil de obtener o ineficiente para fines de aprendizaje; por ejemplo, cuando los expertos no pueden definir características de una manera directa, cuando los datos tienen una alta dimensión, o cuando las características

consisten de ambas variables, tanto continuas como categóricas. En el uso de disimilitudes entre los objetos se presenta el problema en que conjuntos de entrenamiento pequeños pueden complicar la clasificación.

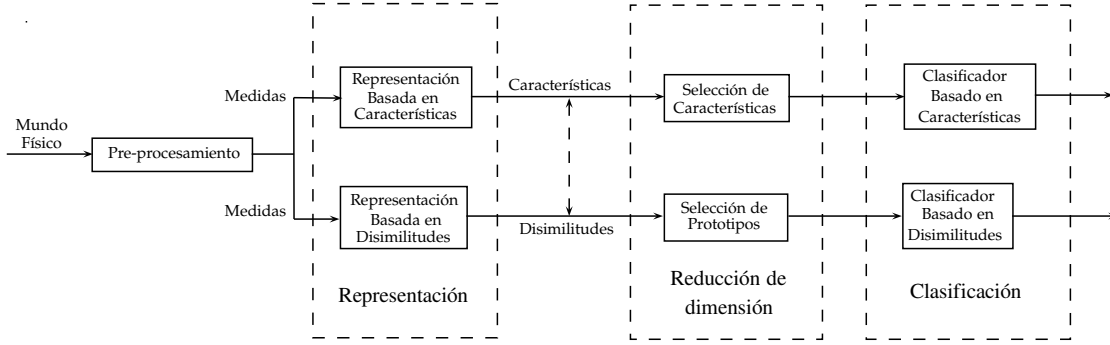
De la definición 2, bajo las condiciones más generales, las disimilitudes presentes en  $R$  pueden derivarse de los objetos directamente, sus medidas, o alguna representación intermedia, es decir, las características de los objetos no se ven involucradas [20][Pág. 66].

Matricialmente, tenemos que la representación de disimilitudes se nota de la siguiente manera:

$$D(T, R) = \begin{matrix} & \begin{matrix} p_1 & p_2 & p_3 & \cdots & p_n \end{matrix} \\ \begin{matrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_N \end{matrix} & \begin{pmatrix} d_{11} & d_{12} & d_{13} & \cdots & d_{1n} \\ d_{21} & d_{22} & d_{23} & \cdots & d_{2n} \\ d_{31} & d_{32} & d_{33} & \cdots & d_{3n} \\ \vdots & \vdots & & \vdots & \ddots \\ d_{N1} & d_{N2} & d_{N3} & \cdots & d_{Nn} \end{pmatrix} \end{matrix} \quad (2-4)$$

$D$  es una matriz de tamaño  $N \times n$ , en el caso en que  $D$  fuera de tamaño  $N \times N$ , se estaría refiriendo a la matriz de disimilitudes  $D(T, T)$ , ver matriz 2-1.

La pregunta que surge ahora es ¿cómo se escogen los elementos de  $R$ ? La selección de los objetos prototipo de  $R$  para la construcción del clasificador en el espacio de disimilitudes, es un objetivo similar a la selección de prototipos usando la regla 1-NN: *La reducción al mínimo del conjunto de disimilitudes que deben medirse para la clasificación de nuevos objetos entrantes*. La selección de objetos similares producirá contribuciones similares a la representación, por esto vale la pena evitar este tipo de selección, es decir, por cada clase en el conjunto de entrenamiento se seleccionan objetos de manera que sean lo más diferentes posible [26][Pág. 192]. La selección aleatoria de los elementos en el conjunto de representación resulta ser un método básico y adecuado para la formación de dicho conjunto sin que los resultados en la clasificación se vean alterados o sean erróneos [26], el esquema de clasificación en el espacio de disimilitudes se muestra en la figura 2-3.



**Figura 2-3.:** Clasificación en el espacio de disimilitudes

## 2.3. Clasificadores

Para clasificadores definidos sobre el espacio de características, se buscan técnicas simples de reconocimiento de patrones que los generalicen mejor, ya que menos parámetros deben ser determinados sobre la base de las muestras dadas. Por lo tanto, son de interés los clasificadores lineales o cuadráticos. Dentro del grupo clasificadores en el espacio de características o basados en reglas de decisión, los siguientes fueron estudiados y seleccionados:

1. Clasificador normal lineal (*NLC*), asume que todas las clases se caracterizan por múltiples distribuciones normales con la misma matriz de covarianza  $S$ . Para un problema de dos clases, el clasificador *NLC* está dado por:

$$f(x) = \left[ x - \frac{\bar{x}_{(1)} + \bar{x}_{(2)}}{2} \right]^t S^{-1} (\bar{x}_{(1)} - \bar{x}_{(2)}) + 2 \log \frac{P_{(1)}}{P_{(2)}} \quad (2-5)$$

Donde  $P_{(i)}$ ,  $i = 1, 2$  son las probabilidades a priori. Dado que el rango de la estimación de la matriz de covarianza  $S$  es no más de  $n - 1$  para los datos de nuestra disimilitud, resulta imposible determinar su inversa. Por lo tanto, se usan una versión regularizada  $S_r$  y un ejemplo de regla de decisión llamada clasificador normal lineal regularizado (*RNLC*). La regularización hace que la operación inversa sea posible, haciendo énfasis en (por ejemplo, aumentar) los valores de la diagonal (varianzas) de la matriz  $S$  con referencia a los términos fuera de la diagonal principal (covarianzas) [12][Pág. 29][27].

2. Clasificador normal cuadrático (*NQC*), asume que las clases tienen múltiples distribuciones normales, cada una caracterizada por una matriz de covarianzas diferente. Para un problema de dos clases, el clasificador *NQC* con las matrices de covarianzas,  $S_{(1)}$  y  $S_{(2)}$ , viene dado por:

$$f(x) = \sum_{i=1}^2 (-1)^i (x - \bar{x}_{(i)})^t S_{(i)}^{-1} (x - \bar{x}_{(i)}) + 2 \log \frac{|P_{(1)}|}{|P_{(2)}|} + \log \frac{|S_{(1)}|}{|S_{(2)}|} \quad (2-6)$$



Donde  $P_{(i)}$ ,  $i = 1, 2$  son probabilidades a priori. Cuando las matrices de covarianzas estimadas son singulares, las versiones de regularización se utilizan para los datos de disimilitud y el clasificador es llamado clasificador normal cuadrático regularizado (*RNQC*) [12][Pág. 27][27].

3. Clasificador basado en Maquinas de vectores de soporte (*SVM* por sus siglas en inglés), se busca un hiperplano que separe de forma óptima a los puntos de una clase de la de otra, que eventualmente han podido ser previamente proyectados a un espacio de dimensión superior.

En el concepto de "separación óptima" reside la característica fundamental de las *SVM*: este tipo de algoritmos buscan el hiperplano que tenga la máxima distancia (margen) con los puntos que estén más cerca de él mismo. Por eso también a veces se les conoce a las *SVM* como clasificadores de margen máximo. De esta forma, los puntos del vector que se etiquetaron con una clase estarán a un lado del hiperplano y los casos que se encuentren en la otra clase estarán al otro lado.

Las maquinas de vectores de soporte pertenecen a la familia de los clasificadores lineales. Al vector formado por los puntos más cercanos al hiperplano se le llama vector de soporte. Los modelos basados en *SVMs* están estrechamente relacionados con las redes neuronales [4, 24, 12].

Otros clasificadores, que suelen ser utilizados son: el clasificador del vecino más cercano (*knn*) ( $1 - NN$ ) [12][Pág. 155-162][27], el clasificador de Parzen [12] el cual está basado en la estimación de densidades de probabilidad de Parzen.

## 2.4. Estimador de Parzen

Un método para el desarrollo de un clasificador o un estimador es el llamado enfoque basado en modelos. En este caso, la disponibilidad requerida de la densidad de probabilidad condicional y las probabilidades a priori se obtienen por medio del conocimiento general del proceso físico y sensorial del sistema en términos de modelos matemáticos. En el aprendizaje supervisado, dos enfoques se discuten usualmente, el aprendizaje paramétrico y el aprendizaje no paramétrico.

En el aprendizaje paramétrico, la suposición base es que los únicos factores desconocidos son los parámetros de las densidades de probabilidad involucradas. Por lo tanto, el aprendizaje a partir de muestras se reduce hasta encontrar los valores adecuados de estos parámetros.

En el aprendizaje no paramétrico, el conocimiento previo acerca de la forma funcional de las distribuciones de probabilidad condicional no está disponible o no se utiliza de forma explícita. El nombre puede sugerir que no hay parámetros implicados. Sin embargo, y de hecho, estos métodos a menudo requieren más parámetros que los métodos paramétricos. La diferencia es que, en los métodos no paramétricos, los parámetros no son aquellos de la

distribución condicional. A partir del concepto básico del aprendizaje no paramétrico surge la posibilidad de definir el clasificador de Parzen basado en la estimación de densidades de probabilidad de Parzen.

El objetivo de la estimación de Parzen es obtener estimaciones de densidades de probabilidad condicional  $p(z|\omega_k)$ , las cuales se hacen sin tener mucho conocimiento a priori de las mismas. Un camino sencillo para llegar al objetivo planteado es particionar el espacio de medida en un número finito de regiones disyuntas  $R_i$ , llamadas cajas, y se cuentan las muestras que caen en cada una de estas cajas. La estimación de la densidad de probabilidad dentro de la caja es proporcional a tal número.

$N_{k,i}$  denota el número de muestras con clase  $\omega_k$  que caen dentro de la  $i$ -ésima caja, entonces la densidad de probabilidad dentro de la  $i$ -ésima caja se estima como:

$$\hat{p}(z|\omega_k) = \frac{N_{k,i}}{\text{volumen}(R_i) \cdot N_k}, \quad \text{con } z \in R_i \quad (2-7)$$

En donde  $N_k$  es el número de muestras con clase  $\omega_k$ . Para cada clase, el número  $N_{x,i}$  tiene una distribución multinomial con parámetros

$$P_{k,i} = \int_{z \in R_i} p(z|\omega_k) \quad \text{con } i = 1, \dots, N_{bin}, \quad N_{bin} \equiv \text{Número de cajas.} \quad (2-8)$$

El método anterior se denomina **histograma** [12][Pág. 150][25] (ver figura 2-4), este método se desempeña bien si el número de muestras dentro de cada caja es suficientemente grande. El método de Parzen es considerado como un refinamiento del histograma (ver figura 2-5). El primer paso en el desarrollo de la estimación es considerar solamente una muestra del conjunto de entrenamiento, supongamos que  $z_j \in T_k$ , para  $z_j$  se tiene claramente que si  $z_j \in T_k$ , en esta posición en el espacio de características la densidad de probabilidad es diferente de cero, es decir,  $p(z_j|\omega_k) \neq 0$ , bajo el supuesto de que  $p(z|\omega_k)$  es una función continua sobre todo el espacio de características se tiene que, en una pequeña vecindad de  $z_j$  la densidad de probabilidad es también cero. Sin embargo, mientras más lejos de esta de  $z_j$ , menos se puede decir acerca de la densidad de probabilidad  $p(z|\omega_k)$ . La idea básica de la estimación Parzen es que el conocimiento obtenido por la observación de  $z_j$  se representa por una función de posicionada en  $z_j$  y con una influencia restringida a una pequeña vecindad de  $z_j$ . Tal función se llama el **kernel** del estimador, este representa la contribución de  $z_j$  a la estimación. En resumen, junto a los aportes de todos los vectores en el conjunto de entrenamiento, se obtiene la estimación final.

Sea  $\rho(z, z_j)$  una medida de distancia definida en el espacio de características. El conocimiento ganado por la observación  $z_j \in T_k$  se representa por el kernel,  $h(\rho(z, z_j))$ , donde  $h(\cdot)$  es una

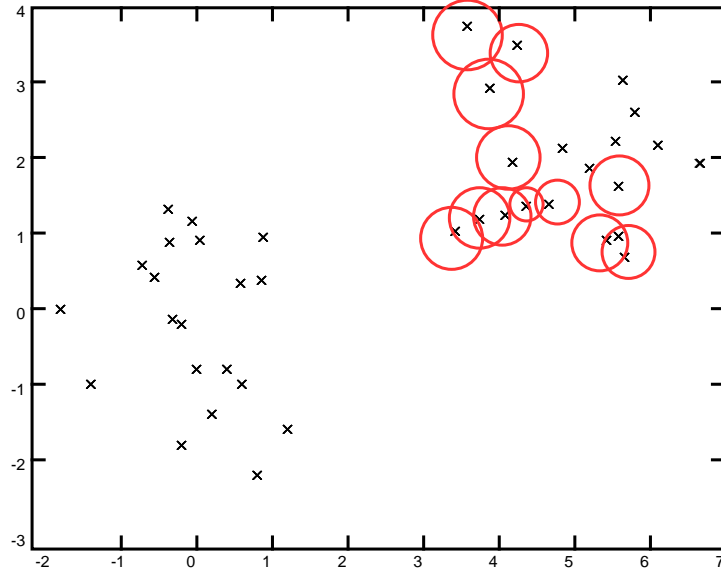


Figura 2-4.: Histograma

función  $\mathbb{R}^+ \rightarrow \mathbb{R}^+$ ,  $h(\rho(z, z_j))$ , la cual tiene máximo en  $z = z_j$ , es decir cuando  $\rho(z, z_j) = 0$ .  $\rho(\cdot, \cdot)$ , es monótona creciente y normalizada, es decir:

$$\int h(\rho(z, z_j)) dz = 1 \quad (2-9)$$

La ecuación 2-9 es la integral sobre el espacio de características, la contribución de todas las observaciones se obtiene de la siguiente manera:

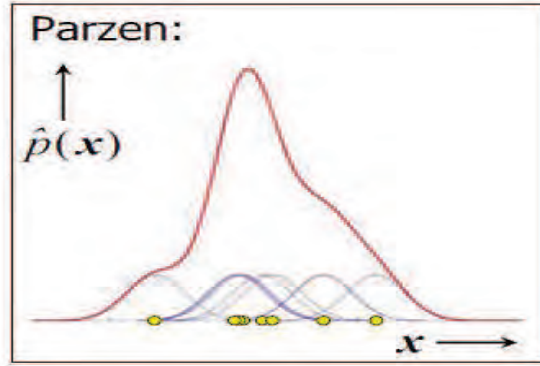
$$\hat{p}(z|\omega_k) = \frac{1}{N_k} \sum h(\rho(z, z_j)) \quad (2-10)$$

Así el kernel puede observarse como una función de interpolación.

En el caso de dimensión  $N$  son útiles varias funciones de interpolación, una popular es la función gaussiana,

$$\rho(z, z_j) = \sqrt{(z - z_j)^T S^{-1} (z - z_j)} \quad (2-11)$$

$$h(\rho) = \frac{1}{\sigma_h^N \sqrt{(2\pi)^N |S|}} e^{\left(-\frac{\rho^2}{2\sigma_h^2}\right)} \quad (2-12)$$



**Figura 2-5.:** Estimacion de densidades de probabilidad de Parzen

La constante  $\sigma_h$  controla el tamaño de la zona de influencia de  $h(\cdot)$ . Puede ser elegido tan pequeño como el número de muestras en el conjunto de entrenamiento se incrementa. La matriz  $S$  debe ser simétrica y definida positiva. Para realizar la estimación de probabilidad mediante el método de Parzen se plantea el siguiente algoritmo [12][Pág. 154]:

### 2.4.1. Algoritmo de clasificación basado en la estimación de densidades de probabilidad de Parzen

Se ingresan un conjunto de entrenamiento etiquetado  $T_s$ , y un conjunto de prueba sin etiquetar  $T$ .

#### 1. Determinar $\sigma_h$

Maximizar el logaritmo de verosimilitud del conjunto de entrenamiento  $T_s$  variando  $\sigma_h$ , usando la estimación de *leave – one – out*. En otras palabras seleccionar  $\sigma_h$  tal que:

$$\sum_{k=1}^K \sum_{j=1}^{N_k} \ln(\hat{p}(z_{k,j}|\omega_k)), \quad (2-13)$$

sea máxima [12].  $z_{k,j}$  es la muestra  $j$  de la clase  $k$ ,  $\sigma_h$  se denomina usualmente **parámetro de suavizado**.

#### 2. Estimación de la densidad

Para cada  $z$  en el conjunto de evaluación

$$\hat{p}(z|\omega_k) = \frac{1}{N_k} \sum_{z_j \in T_k} \frac{1}{\sigma_h^N \sqrt{(2\pi)^N}} e^{\left( \frac{-\|z - z_j\|^2}{2\sigma_h^2} \right)} \quad [12, 23] \quad (2-14)$$

### 3. Clasificación

$$\hat{\omega} = \omega_k \text{ con } k = \operatorname{argmax}_{i=1,\dots,K} \left\{ \hat{p}(z|\omega_i) \hat{P}(\omega_i) \right\} \quad (2-15)$$

La implementación computacional de este algoritmo determina el clasificador que denominaremos clasificador de Parzen. Dicha implementación se hace mediante el software Matlab, y en particular, a través del toolbox PRTools (ver [anexo A](#)).

## 3. Estado del arte

El uso de disimilitudes entre objetos de un conjunto de datos, permite realizar la tarea de clasificación directamente con los objetos, dejando a un lado el uso de características. Planteando el razonamiento para el uso de medidas de disimilitud y la representación de disimilitudes definida en el capítulo anterior, surgen diversos métodos, alternativos, de manera que se genera un cambio en la forma de plantear el uso de dichas medidas de disimilitud, permitiendo al igual determinar una forma alterna de definir el conjunto de representación de disimilitudes, reemplazando los objetos prototipo por conjuntos de objetos prototipo, por modelos basados en tales conjuntos o por prototipos contruidos artificialmente. El uso de estos métodos se conoce como la **generalización del espacio de disimilitudes, o generalización de las representaciones de disimilitudes**.

### 3.1. Espacio generalizado de disimilitudes

A partir de la noción de disimilitud, y de la formación del espacio de disimilitudes, se plantea un método alternativo a éste, mediante la definición de **líneas de características**, de tal manera que los conjuntos de representación de disimilitudes van a consistir en distancias del conjunto de objetos prototipo seleccionados, bien sea siguiendo un modelo de criterios para la selección o de manera aleatoria que resulta ser más sencilla, a un conjunto conformado por líneas de características. La noción de línea de característica generaliza cada par de objetos prototipo de la misma clase mediante una función lineal que se denomina línea de características [21, 22]. Un nuevo objeto se clasifica sobre la base de sus disimilitudes al conjunto de objetos prototipo (conjunto de representación), que en este caso está constituido por líneas de características generadas a partir de los objetos prototipo seleccionados. Para fines prácticos a este conjunto se le denomina conjunto generalizado de representación de disimilitudes. Mediante el uso de un conjunto de representación [24][Pág. 192], se genera el conjunto generalizado de representación para el espacio generalizado de disimilitudes, formado por los nuevos objetos prototipo (líneas de características), y a este conjunto generalizado de representación de disimilitudes se le asocia una matriz de representación que se denomina **matriz generalizada de disimilitudes**.

### 3.1.1. Línea de características

El espacio de disimilitudes se puede generalizar de diversas maneras, es decir, incluyendo conceptos que modifican la idea principal del manejo de las distancias entre objetos, en este caso se hace esta generalización mediante el uso de líneas de características.

**Definición 3** (Línea de características). Sean  $x_{ci}, x_{cj}$ , un par de objetos prototipo pertenecientes a la misma clase, mediante una función lineal  $L_{ij}^c$ , se generalizan este par de puntos. Esta función se denomina **línea de características**.

Para  $x$  un nuevo punto dado, este punto se proyecta sobre  $L_{ij}^c$  como un punto  $p_{ij}^c$  como se muestra en la figura 3-1. Esta proyección se obtiene como  $p_{ij}^c = x_{ci} + \tau(x_{cj} - x_{ci})$ , donde  $\tau = (x - x_{ci}) \cdot (x_{cj} - x_{ci}) / \|x_{cj} - x_{ci}\|^2 \in \mathbb{R}$ ;  $\tau$  se denomina **parámetro de posición**. Entonces la regla de la  $k$  líneas de características más cercanas [18][Pág. 439-443] permite clasificar los objetos de la siguiente manera:

- La clasificación de  $x$  se hace mediante la asignación de la etiqueta de clase  $\hat{c}$  asociada a la línea de características más cercana, esto es:

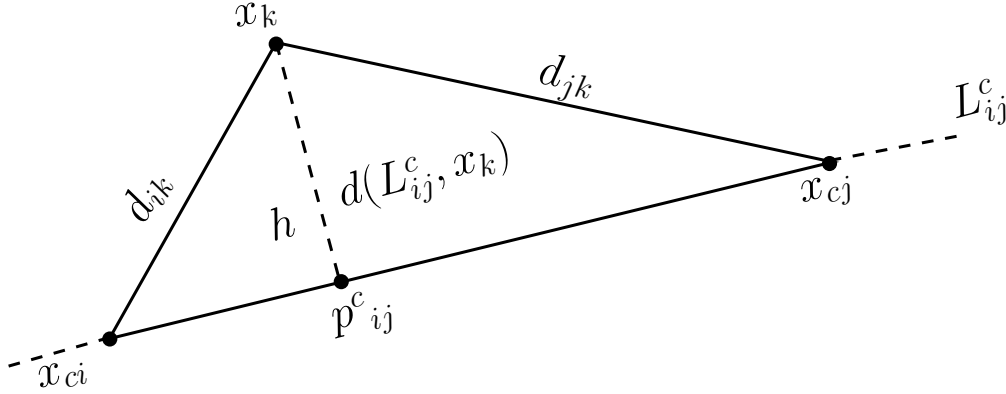
$$d(x, L_{ij}^{\hat{c}}) = \min_{\substack{1 \leq c \leq C, 1 \leq i, j \leq n_c \\ i \neq j}} d(x, L_{ij}^c) \quad (3-1)$$

donde,  $d(x, L_{ij}^c) = \|x - p_{ij}^c\|$

El número de distancias calculadas es:

$$n_L = \sum_{c=1}^C n_c(n_c - 1)/2 \quad (3-2)$$

Donde  $C$  es el número de clases y  $n_c$  es el número de objetos en una clase [20][Pág. 66].



**Figura 3-1.:** Línea de características

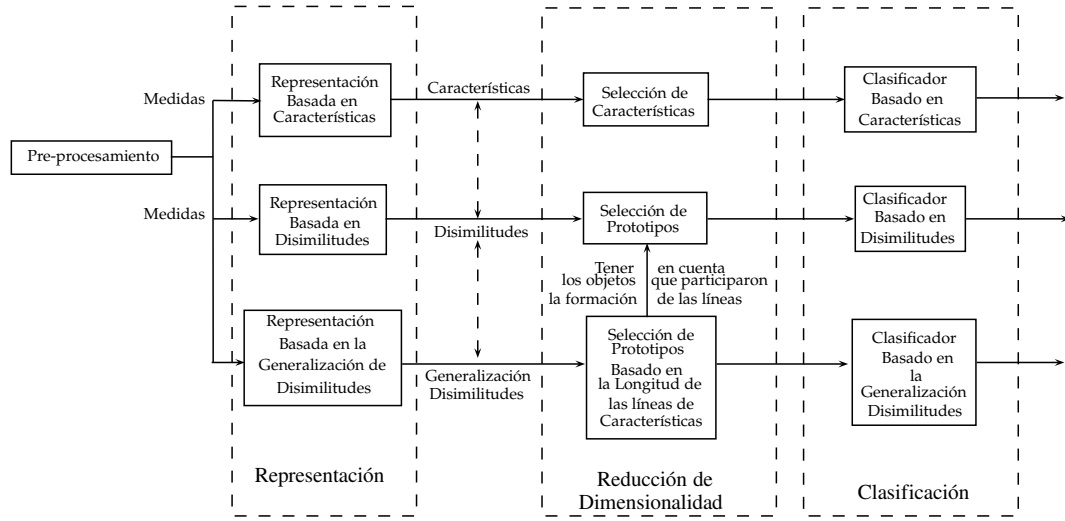
La generalización del espacio de disimilitudes, a partir de la definición de línea de características, consiste en crear la representación  $D_L(T, R_L)$ , matriz de disimilitudes de tamaño  $N \times n_L$ , donde  $L$  denota que el conjunto de representación está compuesto por líneas de características. Esta matriz está dada como:

$$D(T, R_L) = \begin{matrix} & L_1 & L_2 & L_3 & \cdots & L_{n_L} \\ \begin{matrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_N \end{matrix} & \begin{pmatrix} d_{11} & d_{12} & d_{13} & \cdots & d_{1n_L} \\ d_{21} & d_{22} & d_{23} & \cdots & d_{2n_L} \\ d_{31} & d_{32} & d_{33} & \cdots & d_{3n_L} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ d_{N1} & d_{N2} & d_{N3} & \cdots & d_{Nn_L} \end{pmatrix} \end{matrix} \quad (3-3)$$

Donde  $d_{jk} = D_L(x_j, L_k)$  [21][Pág. 242-254]. Entonces como resultado, la representación generalizada del objeto a clasificar  $x$  corresponde al vector

$$D(x, R) = [d(x, L_1), d(x, L_2), \cdots, d(x, L_{n_L})] \quad (3-4)$$





**Figura 3-2.:** Clasificación en el espacio generalizado de disimilitudes

### 3.1.2. Representación generalizada de disimilitudes

Al igual que para el espacio de disimilitudes, se plantea la noción de conjunto de representación de disimilitudes de manera análoga:

**Definición 4** (Representación Generalizada de disimilitudes). *Sean una colección de objetos  $R := \{p_1, p_2, \dots, p_n\}$ , llamado conjunto de representación o conjunto de prototipos y  $d$  una medida de disimilitud. Para esta medida de disimilitud se genera el conjunto  $R_L = \{L_1, L_2, \dots, L_{n_L}\}$ , en donde  $L_i$  es una línea de característica generada a partir de un par de objetos de la misma clase pertenecientes a  $R$ . Una representación generalizada de disimilitudes de un objetos  $x$  es un conjunto de disimilitudes entre  $x$  y los objetos de  $R_L$  expresados como un vector  $D(x, R_L) = [d(x, L_1), d(x, L_2), \dots, d(x, L_{n_L})]$ . Consecuentemente, para una colección de objetos  $T$  se extiende a la matriz de disimilitudes  $D(T, R_L)$ , ver matriz 3-3.*

La ventaja del conjunto de representación  $R_L$ , para la tarea de clasificación, es que éste es un conjunto relativamente pequeño, ya que en él están asociados pares de objetos del conjunto  $R$  mediante una función lineal. El conjunto  $R$  no debe ser muy grande debido a que el número de líneas de características en  $R_L$  crece de manera exponencial según el número de objetos de representación en  $R$ , como se puede notar al ver ecuación 3-2. Una vez que se ha definido la generalización del espacio de disimilitudes mediante línea de características, la tarea de clasificación se plantea en la figura 3-2.

Tenemos entonces que, planteado el razonamiento para el uso de disimilitudes y la representación de disimilitudes, surgen métodos alternativos de generalización del espacio de disimilitudes de manera que, reemplazando los objetos prototipo del espacio de disimilitudes por conjuntos de prototipos, por modelos basados en tales conjuntos o por prototipos construidos artificialmente, se generaliza este espacio. Se mostró la generalización mediante líneas

de características [21], y se puede hacer una extensión y discutir también sobre planos de características, en el caso que la generalización no se haga mediante el uso de una función lineal sino de una función en el espacio [21, 20].

Entre los métodos alternativos que han sido estudiados en los últimos años, se encuentra la generalización del espacio de disimilitudes mediante modelos ocultos de Markov [1, 2, 3] y un método propuesto para superar el problema de tamaño de muestra pequeña en reconocimiento facial [14, 15]. Se referencian estos métodos, mas no se explican a fondo dado que no serán utilizados en las pruebas de desempeño de los clasificadores.

## 4. Estimación y clasificador de Parzen en espacios generalizados de disimilitudes

A partir del conocimiento de la estimación de Parzen en el espacio de características, se plantea la posibilidad de hacer uso de este método en el espacio generalizado de disimilitudes, dado que este ofrece ventajas tales como que la distribución de los datos no es de gran relevancia. A diferencia de la mayoría de métodos de estimación en los cuales se asume que la distribución de los datos es normal, en el método de Parzen este hecho no toma gran importancia dado que se ajusta a la distribución real que tienen los datos. Ahora, si verificamos en el algoritmo para la estimación de Parzen, en el paso de la estimación de densidades de probabilidad, la ecuación 4-2 muestra cómo la potencia en la función exponencial representa el cálculo de las distancias entre los objetos del conjunto de datos. Se propone entonces no realizar este cálculo, de lo contrario, a partir de la determinación de la representación generalizada de disimilitudes, con su correspondiente representación matricial, tomar los valores presentes en esta matriz y reemplazarlos en la formulación de la estimación de Parzen, en la ecuación 4-2.

### 4.1. Estimador de Parzen

Recordemos que el objetivo de la estimación de Parzen es obtener estimaciones de densidades de probabilidad condicional, sin tener mucho conocimiento a priori de estas densidades, se requiere estimar  $p(z|\omega_k)$ . A partir del concepto de histograma [12][Pág. 150][25], y teniendo en cuenta que este método trabaja bien si el número de muestras dentro de cada caja es suficientemente grande, se considera al método de estimación de Parzen, descrito en la sección 2.4, que como un refinamiento del método de histograma. Para la estimación de Parzen se planteó el siguiente algoritmo:

**Algoritmo.** Se ingresan un conjunto de entrenamiento etiquetado  $T_s$ , y un conjunto de prueba sin etiquetar  $T$ .

1. **Determinar  $\sigma_h$**

Maximizar el logaritmo de verosimilitud del conjunto de entrenamiento  $T_s$  variando  $\sigma_h$ , usando la estimación de leave – one – out. En otras palabras seleccionar  $\sigma_h$  tal que:

$$\sum_{k=1}^K \sum_{j=1}^{N_k} \ln(\hat{p}(z_{k,j}|\omega_k)), \quad (4-1)$$

sea máxima [12].  $z_{k,j}$  es la muestra  $j$  de la clase  $k$ ,  $\sigma_h$  se denomina usualmente **parámetro de suavizado**.

2. **Estimación de la densidad**

Para cada  $z$  en el conjunto de evaluación

$$\hat{p}(z|\omega_k) = \frac{1}{N_k} \sum_{z_j \in T_k} \frac{1}{\sigma_h^N \sqrt{(2\pi)^N}} e^{\left( \frac{-\|z - z_j\|^2}{2\sigma_h^2} \right)} \quad [12, 23] \quad (4-2)$$

3. **Clasificación**

$$\hat{\omega} = \omega_k \text{ con } k = \underset{i=1, \dots, K}{\operatorname{argmax}} \left\{ \hat{p}(z|\omega_i) \hat{P}(\omega_i) \right\} \quad (4-3)$$

$\hat{P}(\omega_i)$  es la probabilidad a priori.

La modificación fundamental sobre este algoritmo está presente en la ecuación 4-2, dado que las distancias presentes en:

$$e^{\left( \frac{-\|z - z_j\|^2}{2\sigma_h^2} \right)} \quad (4-4)$$

ya han sido calculadas, y están almacenadas en la matriz  $D_L(T, R_L)$  las cuales pertenecen al espacio generalizado de disimilitudes. En consecuencia, dado que estas distancias están representadas mediante una matriz generalizada de disimilitudes, no es necesario calcular las distancias en el exponente de la ecuación 4-2, bastará con encontrar el valor en las posiciones que se requieran, directamente en la matriz generalizada de disimilitudes.

## 4.2. Estimación del parámetro de suavizado

Ahora teniendo en cuenta los costos computacionales que implica realizar la estimación del parámetro de suavizado  $\sigma_h$  mediante *Leave-One-Out*, como se plantea en el algoritmo de la estimación de densidades de probabilidad de Parzen, y además, con el fin de hacer competitivo el método propuesto, en cuanto al tiempo de cómputo, se buscan alternativas para la estimación del parámetro  $\sigma_h$  que permitan ahorrar tiempo sin perder precisión en la estimación del parámetro. En general el método de la estimación de Parzen [6, 23] está dado por:

$$\frac{1}{n} \sum_{i=1}^n h(n)^{-m} K e \left( \frac{-\|x - x_i\|}{h(n)} \right)$$

en donde  $K$  es el kernel, en este caso gaussiano,  $n$  es el número de muestras,  $h(n)$  es una sucesión de números positivos. Koontz y Fukunaga [9], usan un método en el cual  $h$  está en función de  $n$  y de la estimación de la matriz de covarianzas  $S$ :

$$h(n) = \left( \frac{1}{m} \text{tr}(S) \right)^{\frac{1}{2}} n^{\frac{-\alpha}{m}}$$

en donde  $\alpha$  es una constante que satisface la condición  $0 < \alpha < 0.5$ [9]. Para estimar el valor óptimo de  $h$  se toma el siguiente método, debidamente expuesto y demostrado en [6]. A partir del método de Koontz y Fukunaga, se toma el producto de las densidades estimadas en los puntos de muestra y se trata de encontrar el valor de  $h$  de manera que optimice este producto, es decir:

$$\max_h L'(h) = \prod_{j=1}^n \frac{1}{n} \sum_{i=1}^n \frac{1}{(h\sqrt{2\pi})^m} e \left( \frac{-\|x_j - x_i\|}{2h^2} \right)$$

incluso, se logra establecer que siempre que los objetos sean diferentes, el método converge más rápidamente [9], que para el caso del espacio generalizado de disimilitudes, será para aquellos objetos que participaron en la construcción de cada una de las líneas de características que conocemos si una posición de la matriz generalizada es cero. Entonces se tiene que el valor de  $h$  óptimo es aquél que hace máximo el valor de los productos estimados siempre y cuando los objetos sean diferentes, es decir:

$$\max_h L'(h) = \prod_{j=1}^n \frac{1}{n} \sum_{i \neq j}^n \frac{1}{(h\sqrt{2\pi})^m} e \left( \frac{-\|x_j - x_i\|}{2h^2} \right)$$

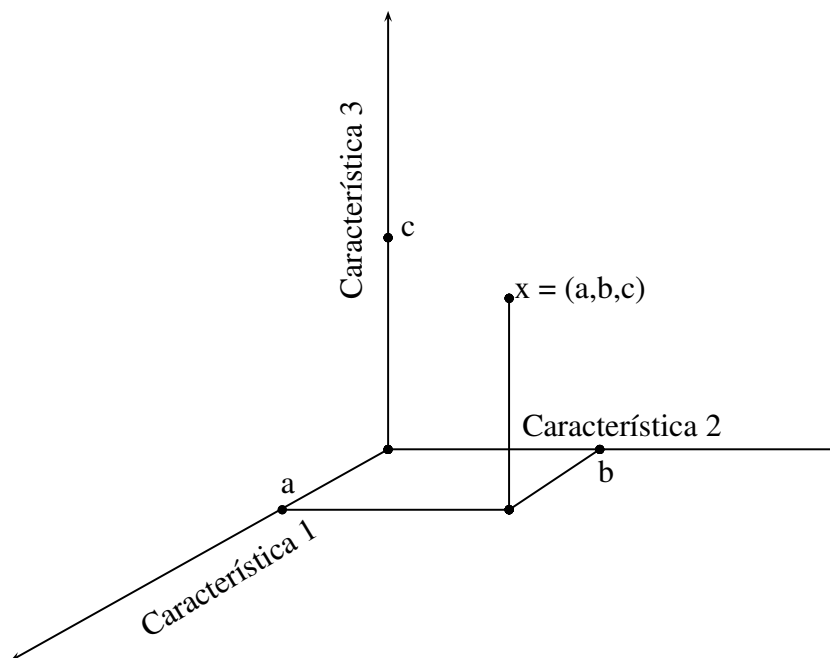
De esta manera entonces se procederá a determinar el valor de  $h$  optimizando el tiempo de cómputo y haciendo, de paso, competitivo el método presentado.

Hechas las modificaciones necesarias, la estimación de probabilidades se podrá realizar en el espacio generalizado de disimilitudes y, por consiguiente, se utilizarán éstas para la clasificación logrando determinar un nuevo clasificador sobre el espacio generalizado de disimilitudes basado en la estimación de Parzen que es el objetivo principal del presente documento. Con el algoritmo modificado, la implementación computacional de éste permitirá realizar dicha tarea, provocando entonces que el objetivo fundamental sea probar el desempeño del clasificador determinado, comparándolo con clasificadores que previamente han sido seleccionados y probados sobre el espacio generalizado de disimilitudes [21].

### 4.3. Propiedades del espacio generalizado de disimilitudes

Con el fin de tener una idea clara del procedimiento a seguir, se iniciara con un ejemplo particular que permita vislumbrar el camino a seguir en el intento de lograr una caracterización del espacio generalizado de disimilitudes. Lo que se plantea es un primer paso en este objetivo, teniendo en cuenta que lograrlo podría necesitar de un tiempo moderado y adecuado para su consecución.

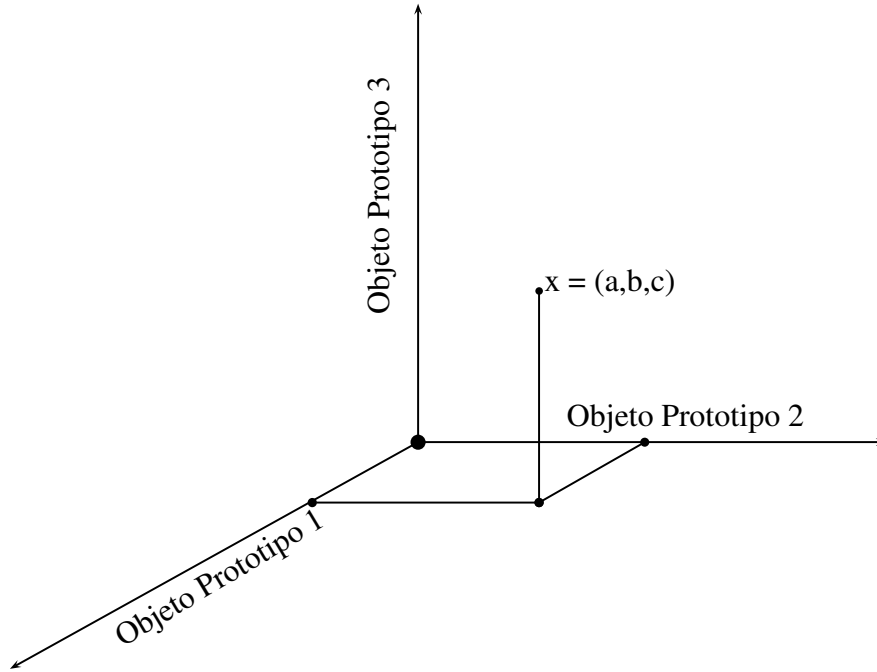
Partiremos del siguiente supuesto: los objetos de estudio tienen tres características, así, un objeto  $x$  en el espacio de características se representa como un vector representado como:  $x = (a, b, c)$ , en donde cada una de sus componentes es una característica del objeto. Para un conjunto de  $n$  objetos, todos tienen en común como referencia las mismas tres características, así, gráficamente, para este caso, el espacio de características se puede notar como el espacio tridimensional, en donde los ejes denotan cada una de las características, como se muestra en la figura 4-1. De manera clara se podría definir una métrica sobre el espacio de características lo que implica que, este conjunto es espacio métrico.



**Figura 4-1.:** Espacio de características

Ahora, supongamos que se escogen, aleatoriamente, tres objetos en el espacio de características, los cuales se denominan objetos prototipo u objetos de representación, tenemos que todos los objetos en el espacio de características se pueden describir de acuerdo a la similitud o disimilitud con estos tres objetos de representación, en otras palabras, tomando la distancia de todos los objetos del espacio de características a los objetos de representación. Cada uno de los objetos se describe en termino de estos, es decir ahora los ejes en el espacio tridimensional denotan a cada uno de los objetos de representación como se muestra en la figura 4-2, y se obtiene de esta manera un nuevo espacio de trabajo denominado, espacio de disimilitudes.

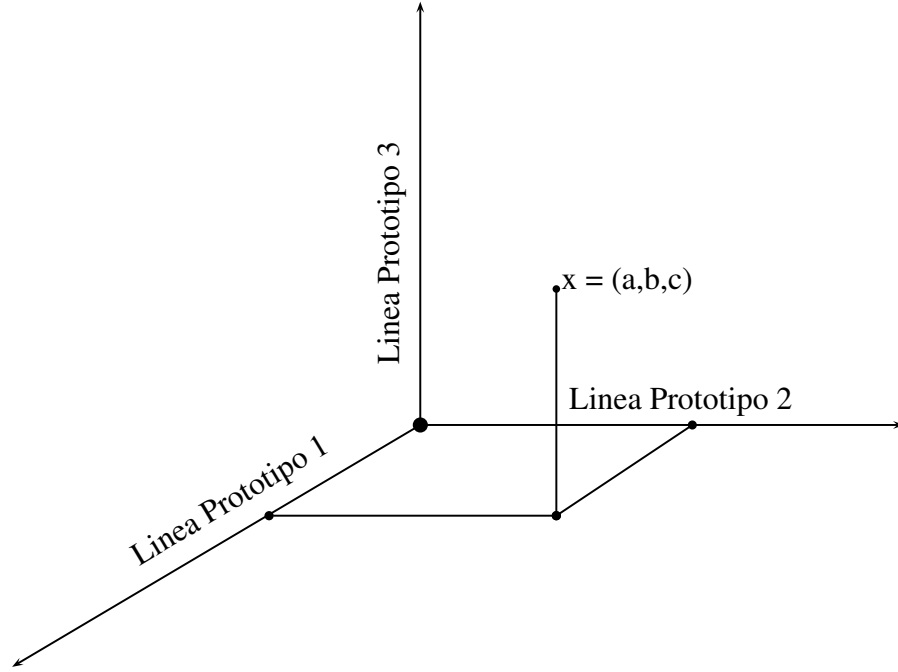
Un objeto  $x$ , en el espacio de disimilitudes, se representa como un vector, partiendo del supuesto que tenemos tres objetos prototipo,  $x = (a, b, c)$ , en donde cada una de sus componentes representa la distancia del objeto a cada uno de los objetos de representación. al igual que para el espacio de características, para una métrica  $d$ , el espacio de disimilitudes es un espacio métrico.



**Figura 4-2.:** Espacio de disimilitudes

Ahora supongamos que tomamos tres objetos del espacio de disimilitudes, que denominaremos objetos prototipo. A partir de la noción de línea de característica, para cada par de objetos obtenemos una línea, es decir una por cada par de puntos. Estas tres líneas de características son las que nos permitirán describir cada uno de los objetos en el que denominaremos espacio generalizado de disimilitudes, de tal manera que cada uno se describe de acuerdo a la distancia a cada una de las líneas de características, es decir, un objetos  $x$  se representa como un vector, partiendo del supuesto que tenemos tres líneas de características,  $x = (a, b, c)$ , en donde cada una de las componentes representa la distancia a cada una de las líneas de características, es decir ahora los ejes del espacio tridimensional son cada una de las líneas de características, como se muestra en la figura 4-3.





**Figura 4-3.:** Espacio generalizado de disimilitudes

Entonces para todo objeto en el espacio generalizado de disimilitudes, definimos por ejemplo:

$$d : X \times X \rightarrow \mathbb{R}$$

$$(x, y) \mapsto d(x, y) = \|x - y\|$$

en donde  $X$  denota el espacio generalizado de disimilitudes,  $d$  es métrica, luego tenemos que el espacio generalizado de disimilitudes es espacio métrico.

Definimos de esta manera para un objeto  $x$ , la distancia de este punto a una línea de características como:

$$\tilde{d}(x, L_{ij}^c) = \|x - p_{ij}^c\|$$

en donde  $p_{ij}^c$  es la proyección de  $x$  sobre la línea  $L_{ij}^c$ .

Ahora, para un conjunto de  $n_c$  líneas de características, supongamos que pertenecen a la clase  $c$ , definimos la distancia de  $x$  a este conjunto de líneas como:

$$\tilde{d}(x, \mathbb{L}_{ij}^c) = \min_{1 \leq i, j \leq n_c, 1 \leq c \leq k, i \neq j} \tilde{d}(x, L_{ij}^c)$$

Es importante notar que cada uno de los tres espacios, características, disimilitudes, generalizado de disimilitudes, se pueden relacionar de manera natural con  $\mathbb{R}^3$ , la diferencia entre cada uno de ellos es la manera de interpretar los ejes coordenados.

En un primer intento de caracterizar el espacio generalizado de disimilitudes mostramos los siguientes resultados:

### 4.3.1. Inmersiones (embedding)

Las Inmersiones son una herramienta útil en problemas prácticos donde las representaciones de disimilitudes finitas son consideradas, esto es, espacios métricos finitos  $(X, d)$  definidos por la correspondiente matriz de disimilitudes  $D$ . Si se logra establecer una equivalencia entre tales espacios y otros espacios conocidos, los últimos, si poseen propiedades favorables, pueden ser usados para establecer reglas de aprendizaje.

Consideremos los siguientes espacios métricos:

1.  $\ell_p^m = (\mathbb{R}, d_p)$  donde  $d_p(x, y) = \left( \sum_{i=1}^m |x_i - y_i|^p \right)^{\frac{1}{p}}$  y  $p > 0$
2.  $\ell_\infty^m = (\mathbb{R}, d_\infty)$  donde  $d_\infty(x, y) = d_{\max}(x, y) = \max_i |x_i - y_i|$

El método a seguir será establecer una relación entre el espacio de disimilitudes y los espacio métricos considerados de manera tal que las propiedades de estos sean validos también sobre el espacio generalizado de disimilitudes. Un posible camino a seguir es hacer uso de las isometrías, que en este caso denominaremos Inmersiones isométricas.

**Definición 5** (Inmersión isométrica). *Sean  $(X, d_x)$  y  $(Y, d_y)$  espacios métricos.  $(X, d_x)$  es isométricamente inmersible en  $(Y, d_y)$  si existe una isométrica  $\phi : X \rightarrow Y$  es decir, una asignación  $\phi$  talque*

$$d_X(x_1, x_2) = d_Y(\phi(x_1), \phi(x_2))$$

para todo  $x_1, x_2 \in X$

A partir de la definición de Inmersión isométrica, se presenta la definición de  $\ell_p$ -Inmersión, la cual permitirá dar un primer paso en el objetivo de encontrar una relación, y en este caso una aplicación entre el espacio generalizado de disimilitudes y el espacio métrico  $\ell_p$ .

**Definición 6** ( $\ell_p$ -Inmersión). *Un espacio métrico  $(X, d)$  es  $\ell_p$ -incorporable si  $(X, d)$  es isométricamente inmersible en el espacio  $\ell_p^m$  para algún entero  $m \geq 1$ . El entero más pequeño es llamado  $\ell_p$ -dimensión*

Las isometrías son asignaciones inyectivas según la definición. Diremos que dos espacio son isométricamente isomorfos si existe una isométrica biyectiva entre ellos. En este caso, los dos espacio son esencialmente idénticos. Todo espacio métrico es isométricamente isomorfo a un subconjunto de algún espacio vectorial normado. Todo espacio métrico completo es isométricamente isomorfo a algún subconjunto cerrado de algún espacio de Banach.

**Teorema 1.** *En un espacio métrico  $(X, d)$ , la desigualdad triangular hacia atrás,  $|d(x_i, x_k) - d(x_j, x_k)| \leq d(x_i, x_j)$ , se tiene para todo  $x, y, z$  en  $X$*

*Demostración.* De la definición 1, la desigualdad triangular establece que

$$d(x, y) + d(y, z) \geq d(x, z) \quad (4-5)$$

para todo  $x, y, z \in X$ , es decir

$$d(x, z) - d(y, z) \leq d(x, y) \quad (4-6)$$

y

$$d(y, z) - d(x, z) \leq d(y, x) \quad (4-7)$$

Dado que  $d(x, y) = d(y, x)$ , la ultima desigualdad es igual a

$$-(d(x, z) - d(y, z)) \leq d(x, y) \quad (4-8)$$

Por lo tanto la desigualdad se tiene.  $\square$

**Teorema 2.** *El espacio euclidiano  $\mathbb{R}^m$  puede ser incorporado en un espacio de Hilbert. Todo subconjunto finito de  $m$  elementos en un espacio de Hilbert puede ser incorporable en  $\mathbb{R}^{m-1}$ .*

En otras palabras todo espacio de Hilbert finito es isomorfo a  $\mathbb{R}^m$ .

**Lema 1** (Sobre la incorporabilidad de un espacio métrico en espacios con la norma del máximo). *Cualquier espacio métrico finito  $(X, d)$  es  $\ell_\infty$ -incorporable.*

*Demostración.* Supongamos que una colección de  $n$  objetos es dada como  $X = x_1, x_2, \dots, x_n$ . Entonces para una métrica  $d$  el espacio métrico  $(x, d)$  es incorporable en  $\ell_\infty^n$ . Defínase la aplicación,

$$\begin{aligned} \phi : X &\longrightarrow \mathbb{R}^n \\ x &\longmapsto \phi(x) = [d(x, x_1), d(x, x_2), \dots, d(x, x_n)]^T \end{aligned} \quad (4-9)$$

denotamos  $z_i = \phi(x_i)$  para  $i = 1, 2, \dots, n$ .

Entonces tenemos que

$$d_\infty(\phi(x_i), \phi(x_j)) = d_\infty(z_i, z_j) \quad (4-10)$$

$$= \max_{1 \leq k \leq n} |z_{ik} - z_{jk}| \quad (4-11)$$

$$= \max_{1 \leq k \leq n} |d(x_i, x_k) - d(x_j, x_k)| \quad (4-12)$$

Gracias a la desigualdad triangular

$$|d(x_i, x_k) - d(x_j, x_k)| \leq d(x_i, x_j) \quad (4-13)$$

se tiene para cualquier  $k = 1, 2, \dots, n$  y la igualdad de  $d(x_i, x_j)$  se tiene para  $k = j$ . Por lo tanto,

$$\max_{1 \leq k \leq n} |d(x_i, x_k) - d(x_j, x_k)| = d(x_i, x_j) \quad (4-14)$$

y  $d_\infty(\phi(x_i), \phi(x_j)) = d(x_i, x_j)$ . Luego,  $\phi$  juega el rol de una Inmersión isométrica de  $(X, d)$  en  $\ell_\infty^n$   $\square$

Tenemos así que, en el espacio generalizado de disimilitudes, para una métrica  $d$ , como  $X$  es un conjunto finito, entonces  $(X, d)$  es un espacio métrico finito, lo que nos permite establecer que:

- El espacio generalizado de disimilitudes es  $\ell_\infty$ -incorporable.
- En particular, el espacio generalizado de disimilitudes es incorporable en un espacio de Hilbert, en particular a  $\mathbb{R}^m$

Respecto a la matriz generalizada de disimilitudes debido a las desventajas que esta presenta en comparación con la matriz de disimilitudes, la cual se deja caracterizar de una manera clara, no es evidente el poder caracterizarla de manera adecuada de acuerdo a ciertas propiedades como, simetría, matriz inversa asociada; aclarando que se intenta seguir con el esquema utilizado para la matriz de disimilitudes. En general, la gran desventaja de la matriz de disimilitudes parte del hecho de no ser cuadrada, lo que lleva a desechar muchas de las propiedades el álgebra de matrices, esto no quiere decir que no se pueda realizar una caracterización de la misma solo resulta ser menos sencillo llegar al propósito establecido.

## 4.4. Propiedades de la matriz generalizada de disimilitudes

Tomando en cuenta las propiedades de la matriz de disimilitudes, al igual que para el espacio generalizado de disimilitudes, se pretende hacer un trabajo análogo de manera que se pueda presentar una buena caracterización de la matriz generalizada de disimilitudes.

En principio, para  $X = \{x_1, x_2, \dots, x_N\}$ , la matriz de disimilitudes  $D(T, T)$  es una matriz de orden  $N \times N$ ,

$$D(T, T) = \begin{matrix} & x_1 & x_2 & \cdots & x_n & \cdots & x_N \\ \begin{matrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_N \end{matrix} & \begin{pmatrix} d_{11} & d_{12} & \cdots & d_{1n} & \cdots & d_{1N} \\ d_{21} & d_{22} & \cdots & d_{2n} & \cdots & d_{2N} \\ d_{31} & d_{32} & \cdots & d_{3n} & \cdots & d_{3N} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ d_{N1} & d_{N2} & \cdots & d_{Nn} & \cdots & d_{NN} \end{pmatrix} \end{matrix} \quad (4-15)$$

que puede variar en su tamaño si se toma en cuenta el conjunto de representación  $R$ , obteniendo una matriz de orden  $N \times n$ ,

$$D(T, R) = \begin{matrix} & p_1 & p_2 & p_3 & \cdots & p_n \\ \begin{matrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_N \end{matrix} & \begin{pmatrix} d_{11} & d_{12} & d_{13} & \cdots & d_{1n} \\ d_{21} & d_{22} & d_{23} & \cdots & d_{2n} \\ d_{31} & d_{32} & d_{33} & \cdots & d_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ d_{N1} & d_{N2} & d_{N3} & \cdots & d_{Nn} \end{pmatrix} \end{matrix} \quad (4-16)$$

En cuanto a la matriz generalizada de disimilitudes la primera gran diferencia es que esta matriz

$$D(T, R_L) = \begin{matrix} & L_1 & L_2 & L_3 & \cdots & L_{n_L} \\ \begin{matrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_N \end{matrix} & \begin{pmatrix} d_{11} & d_{12} & d_{13} & \cdots & d_{1n_L} \\ d_{21} & d_{22} & d_{23} & \cdots & d_{2n_L} \\ d_{31} & d_{32} & d_{33} & \cdots & d_{3n_L} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ d_{N1} & d_{N2} & d_{N3} & \cdots & d_{Nn_L} \end{pmatrix} \end{matrix} \quad (4-17)$$

no necesariamente es cuadrada, incluso se podría decir que extrañamente lo será, dado que el número de líneas de características en el conjunto de representación depende de la cantidad de objetos prototipos seleccionados, debido a que el número de líneas crece de manera exponencial. Luego, desde este punto de vista la matriz generalizada de disimilitudes no es simétrica, además que las distancias son entre elementos y líneas de características lo que imposibilita esta opción. En esta matriz, si un elemento es cero, implica que el objeto correspondiente contribuyó a la formación de la respectiva línea de características o esta sobre ella, así a diferencia de la matriz de disimilitudes, en ésta muchos elementos pueden ser cero. También en esta matriz los elementos son mayores o iguales a cero dado que cada valor proviene de una medida que es métrica.

## 5. Marco experimental

Establecido el algoritmo para la estimación de densidades de probabilidad de Parzen y la manera como se hallará el parámetro de suavizado para esta estimación, se hace la implementación computacional de dicho algoritmo y se presentan las condiciones bajo las cuales se realizarán los experimentos que permitan mostrar la conveniencia del método planteado. En particular, resulta importante describir las características de cada una de las bases de datos utilizadas (artificiales y del mundo real), la metodología de evaluación y la comparación del desempeño de los clasificadores, en los cuales se incluye el clasificador de Parzen para el espacio generalizado de disimilitudes, el cual es importante resaltar, dado que es de mayor interés para el método y razonamiento expuestos en esta tesis.

### 5.1. Bases de datos

El uso de bases de datos artificiales, generadas a partir de funciones en Matlab, específicamente del toolbox PRTools, permitirá mostrar la ventaja del método, a través del clasificador implementado gracias al algoritmo de la estimación de densidades de probabilidad de Parzen. También bases de datos reales obtenidas de diferentes recursos virtuales [8, 16] y que ayudarán a no quedar en supuesto, sino en mostrar las ventajas de este método con este tipo de bases de datos. La descripción de cada una de las bases de datos se muestra en las tablas 5-2 a 5-4.

Para algunas bases de datos de alta dimensión, se hace necesario hacer una reducción de la misma, de manera que se pueda proyectar el conjunto de datos en 2D y observar la distribución de los objetos en cada clase. Los métodos utilizados para hacer la reducción de dimensión y obtener una gráfica de dispersión de los datos en 2D son el análisis de componentes principales y el escalamiento multidimensional.

En análisis de componentes principales (*PCA*), la reducción se hace tomando como criterio la variabilidad de los valores propios obtenidos a partir de la matriz de covarianzas correspondiente al conjunto de datos, además, las componentes principales se hallan graficando la variabilidad de los valores propios y el lugar donde se encuentre el codo de la gráfica, indicara el número y las componentes principales del conjunto de datos, logrando de esta

manera realizar la proyección del conjunto de datos a partir de las componentes principales. Se hace también escalamiento multidimensional (*MDS*) teniendo en cuenta que, en *PCA* se puede descartar información importante para lograr una mejor proyección de los datos y además, en escalamiento multidimensional la idea principal es lograr una proyección en  $2D$  ó  $3D$  de las bases de alta dimensión, preservando la distancia entre los objetos lo mejor posible mediante combinaciones no lineales de las características originales. Para el escalamiento multidimensional se hace uso de la proyección o asignación de *Sammon* [17, 24], la cual se propuso en reconocimiento de patrones por John W. Sammon Jr. [Sammon Jr., 1969] como una herramienta para una proyección no lineal de un espacio euclidiano de alta dimensión a un espacio de menor dimensión.

A continuación se presenta una descripción de algunas de las bases de datos utilizadas para las pruebas:

### 5.1.1. Clases difíciles distribuidas normalmente

El conjunto de clases difíciles distribuidas normalmente [5], corresponde a un conjunto de  $N$  objetos  $k$ -dimensional con dos clases, este conjunto tiene clases de varianza muy diferentes para las primeras dos dimensiones. La separación es por lo tanto, para muestras pequeñas, "difícil", es decir, establecer los límites de decisión para muestras de tamaño pequeño resulta ser complicado (ver gráficos 5-1 y 5-4).

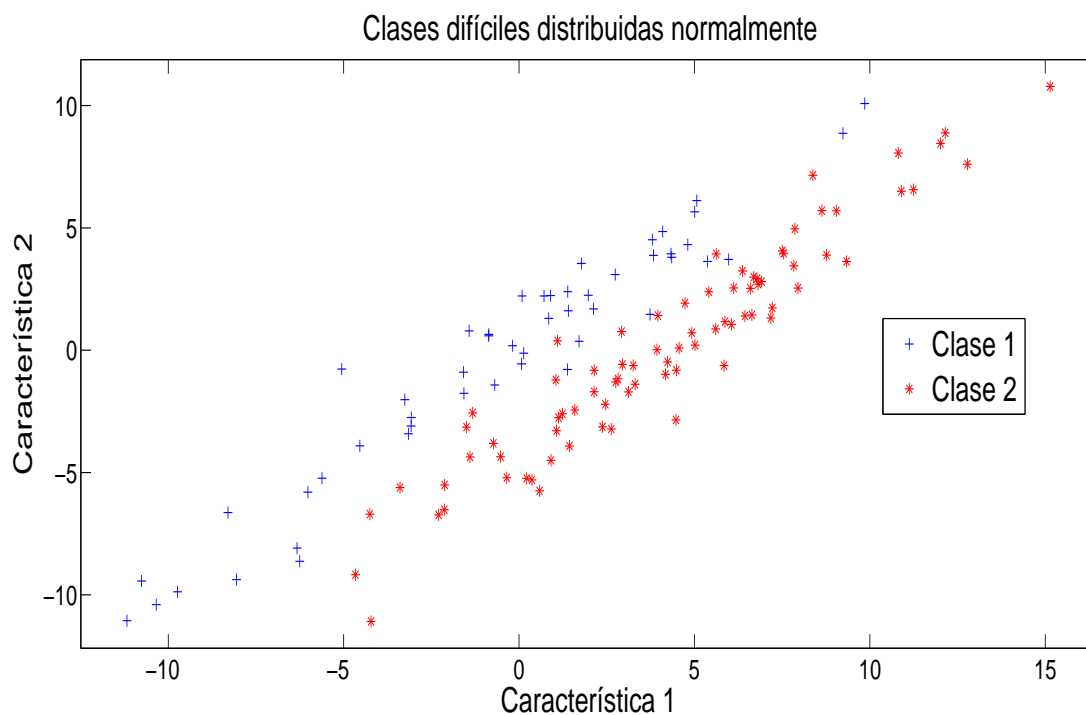
$D1$  es la diferencia entre los promedios para la primera característica,  $D2$  es la diferencia entre los promedios para la segunda característica. En todas las demás direcciones los promedios son iguales. Las dos matrices de covarianzas son iguales, con una varianza de 1 en todas las direcciones, a excepción de la segunda característica, que tiene una varianza de 40. Las dos primeras se rotan más de 45 grados para la construcción de una fuerte correlación. Las probabilidades a priori de las clases son  $P(1) = P(2) = 0,5$ . Si  $N$  es un vector de medida, es decir un vector que indica el número de objetos por clase en el conjunto de datos, exactamente  $N(I)$  objetos se generan para la clase  $I$ ,  $I = 1, 2$ , de manera simplificada se puede observar el resumen de este conjunto de datos en la tabla 5-2, la cantidad de objetos por clase varía según como lo desee el investigador y no es necesario que la cantidad de objetos por clase sea balanceada.

Gráficamente las distribuciones de los datos para el conjunto de clases difíciles distribuidas normalmente se muestran en la figuras 5-1 a 5-4, las cuales tienen 150, 150 y 200 objetos para la clase 1 respectivamente; 225, 225 y 125 objetos para la clase dos respectivamente. Las figura 5-2 y 5-3 corresponden al mismo conjunto de datos, al cual se hizo una reducción de dimensión, dado que se genero con una dimensión igual a 20. La reducción de dimensión se hizo mediante el análisis de componentes principales y escalamiento multidimensional. La razón por la cual en la gráfica 5-3 los ejes no tienen etiquetas es dado a que estos corres-

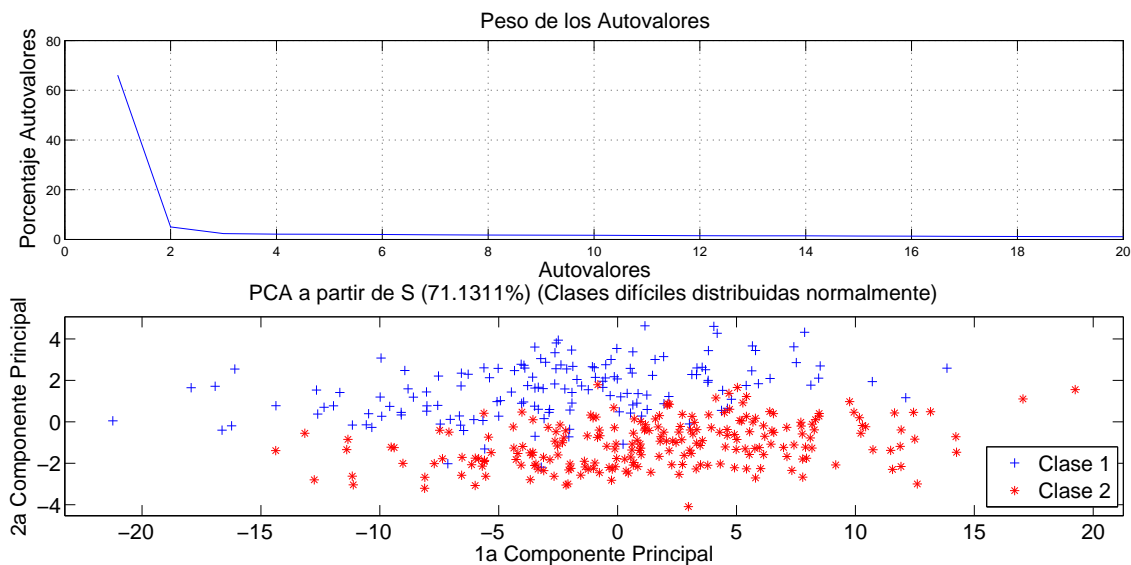


ponden a combinaciones no lineales de las características originales. Por ello, no tienen una interpretación directa, recordando que *MDS* intenta preservar las distancias entre los puntos tan fielmente como sea posible, con el fin de obtener una visualización en  $2D$  ó  $3D$  de los datos de alta dimensión.

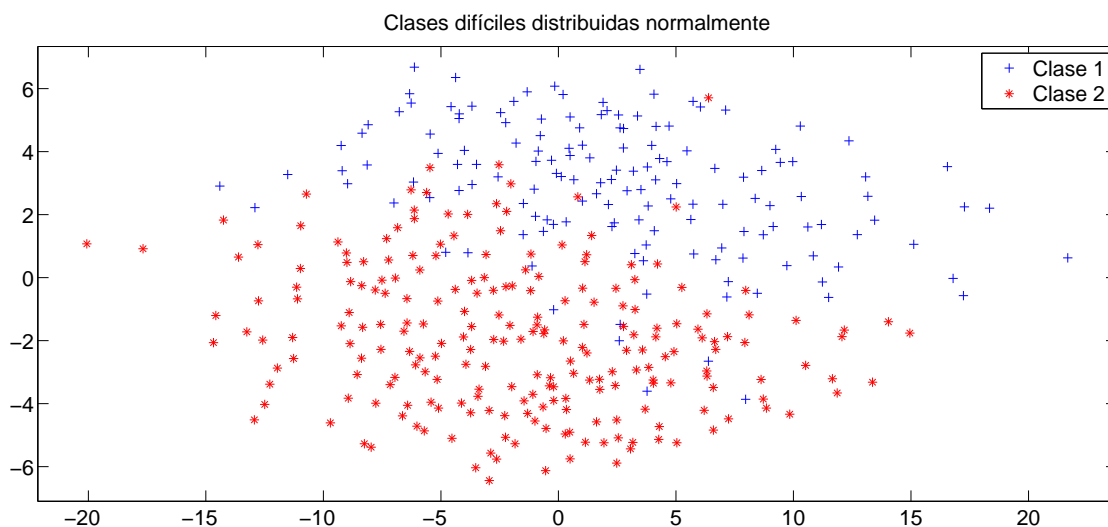
(La generación del conjunto de clases difíciles distribuidas normalmente se hace mediante la función **gendatd** [5], del toolbox de matlab PRTools).



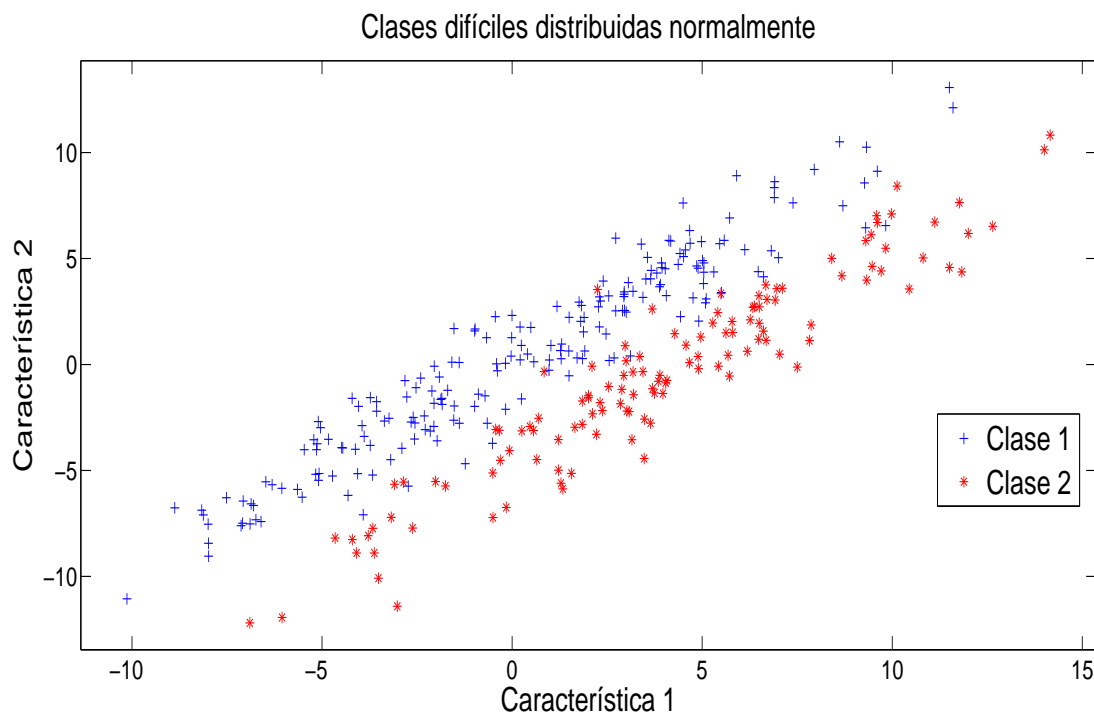
**Figura 5-1.:** Clases difíciles distribuidas normalmente con 50 objetos en la clase 1 y 80 objetos en la clase 2 (conjunto de datos generado mediante la función **gendatd** del toolbox PRTools [5]).



**Figura 5-2.:** Clases difíciles distribuidas normalmente con 150 objetos en la clase 1 y 225 objetos en la clase 2, dimensión 20, (*PCA*). (Conjunto de datos generado mediante la función **gendatd** del toolbox **PRTools** [5]).



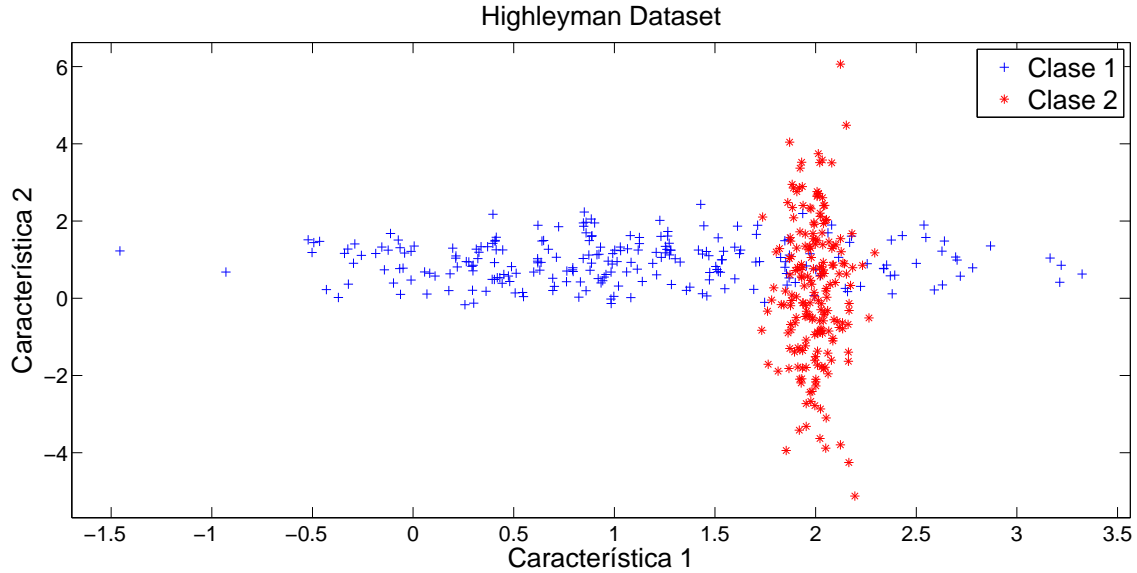
**Figura 5-3.:** Clases difíciles distribuidas normalmente con 200 objetos en la clase 1 y 125 objetos en la clase 2, dimensión 20, (*MDS*). (Conjunto de datos generado mediante la función **gendatd** del toolbox **PRTools** [5]).



**Figura 5-4.:** Clases difíciles distribuidas normalmente con 200 objetos en la clase 1 y 125 objetos en la clase 2. (Conjunto de datos generado mediante la función **gendatd** del toolbox PRTools [5]).

### 5.1.2. Clases de Highleyman

El conjunto de clases de Highleyman, corresponde a un conjunto bi-dimensional con dos clases, basado en la distribución de Highleyman [13]. De manera simplificada se puede observar el resumen de este conjunto de datos en la tabla 5-2. Gráficamente las distribución de los datos para el conjunto de clases de Highleyman se muestran en la figuras 5-5, la cual tiene 240 objetos para la clase 1 y 200 objetos para la clase 2. (La generación de este conjunto de datos se hace mediante la función **gendath** [5], del toolbox de matlab PRTools).

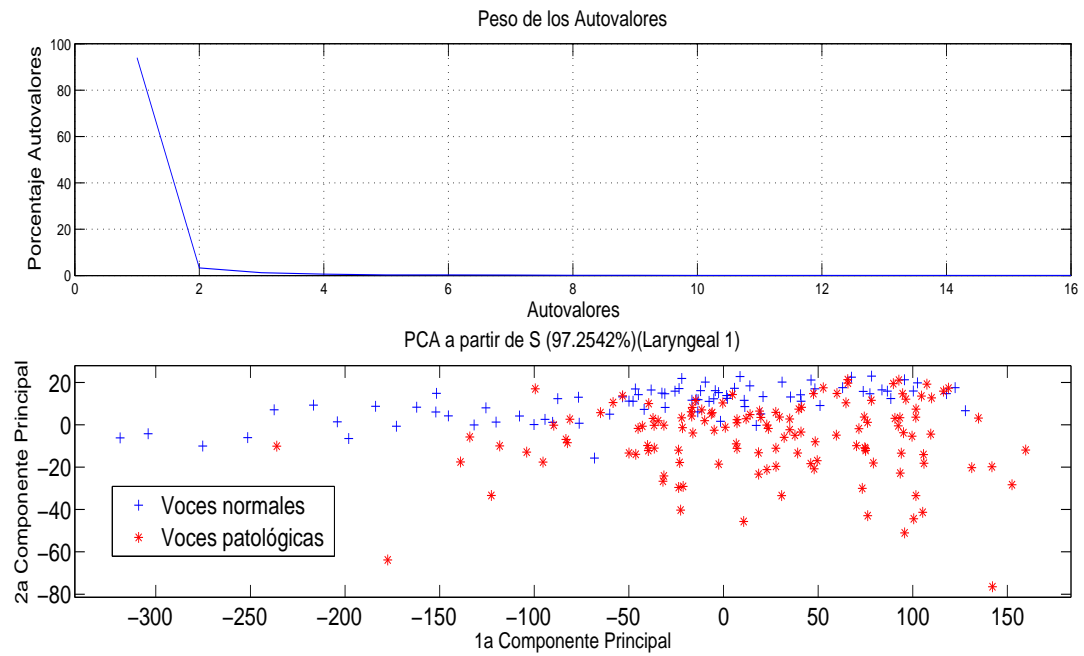


**Figura 5-5.:** Clases de Highleyman con 240 objetos en la clase 1 y 200 objetos en la clase 2 (conjunto de datos generado mediante la función **gendath** del toolbox PRTools [5]).

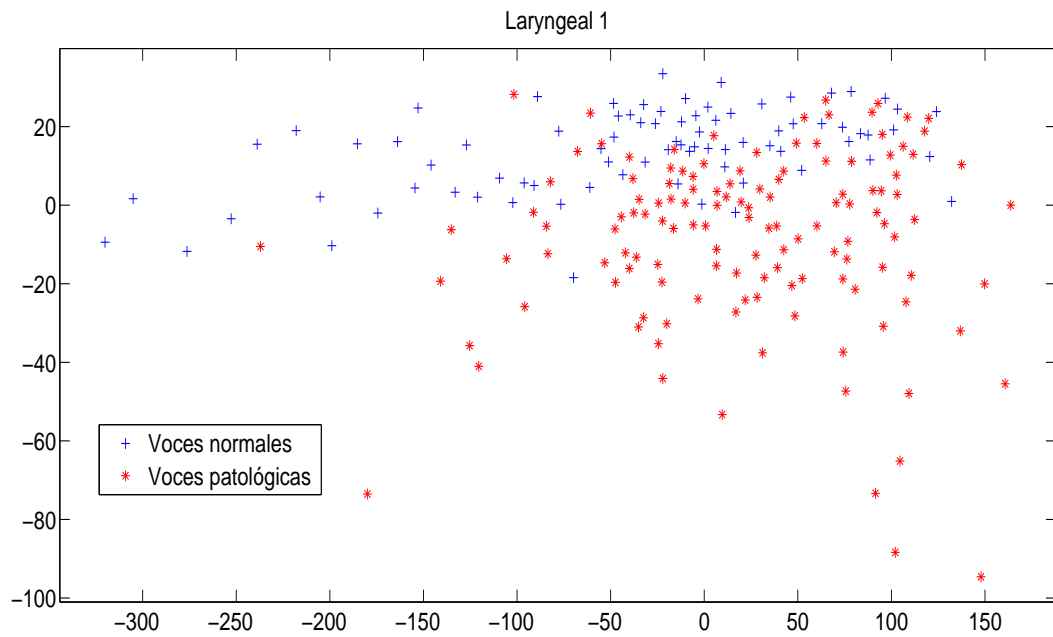
### 5.1.3. Conjunto de datos Laryngeal 1

El conjunto de datos Laryngeal-1 proviene de la Academia Búlgara de Ciencias y está disponible en [16]. El conjunto fue utilizado originalmente para un sistema informático de soporte de decisiones, con el fin de ayudar al diagnóstico de patología de la laringe y en especial en la detección de sus primeras etapas. Voces normales y patológicas se describen en 16 parámetros en el tiempo, los dominios espectral y cepstral.

El número de muestras en este conjunto es de 213, las cuales están separadas de la siguiente manera: 81 Voces Normales y 132 Voces patológicas. De manera simplificada se puede observar el resumen de este conjunto de datos en la tabla 5-2. Las figura 5-6 y 5-7 corresponden a la proyección del conjunto de datos en  $2D$ , al cual se hizo una reducción de dimensión mediante el análisis de componentes principales y el escalamiento multidimensional. La razón por la cual en la gráfica 5-7 los ejes no tienen etiquetas es dado a que estos corresponden a combinaciones no lineales de las características originales. Por ello, no tienen una interpretación directa, recordando que *MDS* intenta preservar las distancias entre los puntos tan fielmente como sea posible, con el fin de obtener una visualización en  $2D$  ó  $3D$  de los datos de alta dimensión.



**Figura 5-6.:** Base de datos Laryngeal 1, el cual se a proyectado en 2D mediante *PCA*.

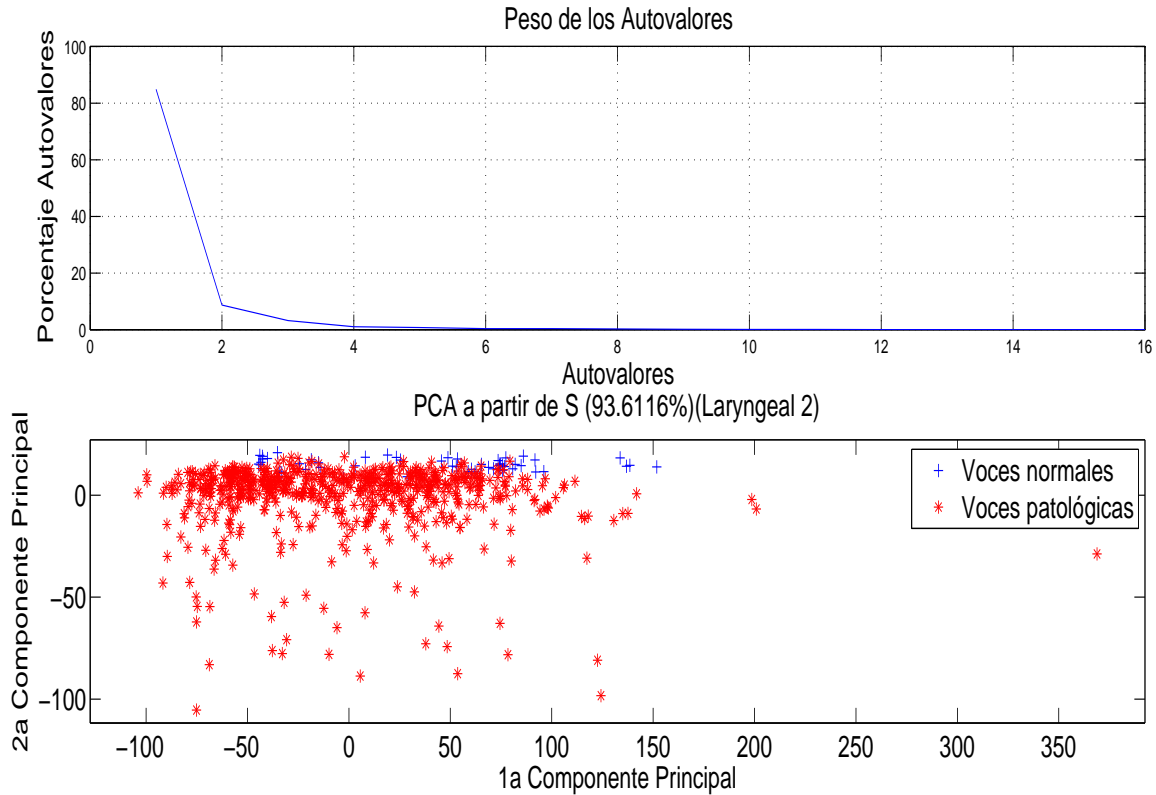


**Figura 5-7.:** Base de datos Laryngeal 1, el cual se a proyectado en 2D mediante *MDS*.

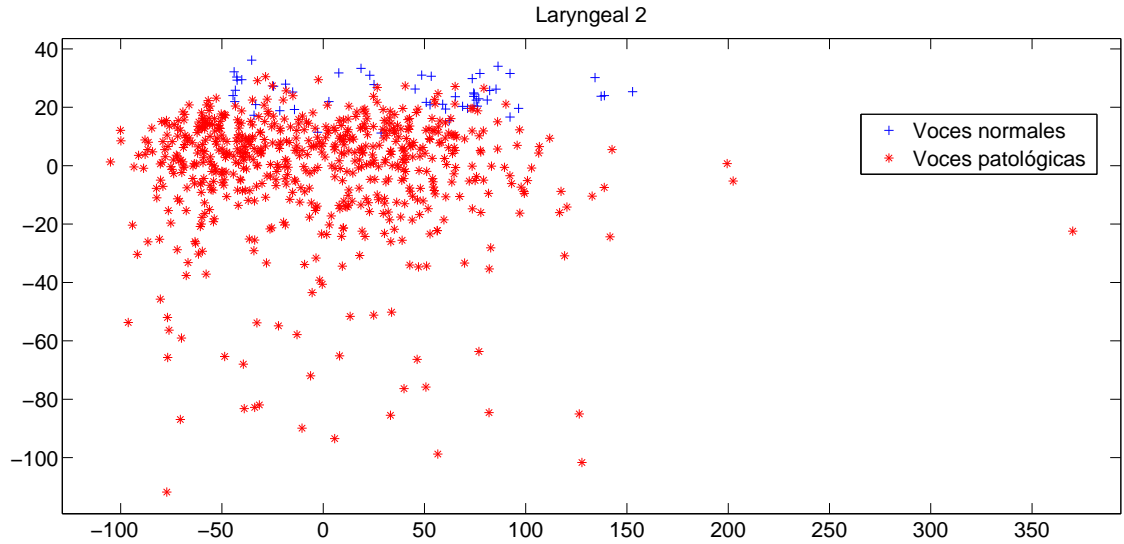
#### 5.1.4. Conjunto de datos Laryngeal 2

El conjunto de datos Laryngeal-2 proviene de la Academia Búlgara de Ciencias y está disponible en [16]. El conjunto fue utilizado originalmente para un sistema informático de soporte de decisiones, con el fin de ayudar al diagnóstico de patologías de la laringe y en especial en la detección de sus primeras etapas. Voces normales y patológicas se describen en 16 parámetros en el tiempo, los dominios espectral y cepstral.

El número de muestras en este conjunto es de 692, las cuales están separadas de la siguiente manera: 53 Voces Normales y 639 Voces patológicas. De manera simplificada se puede observar el resumen de este conjunto de datos en la tabla 5-2. Las figura 5-8 y 5-9 corresponden a la proyección del conjunto de datos en 2D, al cual se hizo una reducción de dimensión mediante el análisis de componentes principales y escalamiento multidimensional. La razón por la cual en la gráfica 5-9 los ejes no tienen etiquetas es dado a que estos corresponden a combinaciones no lineales de las características originales. Por ello, no tienen una interpretación directa, recordando que *MDS* intenta preservar las distancias entre los puntos tan fielmente como sea posible, con el fin de obtener una visualización en 2D ó 3D de los datos de alta dimensión.



**Figura 5-8.:** Base de datos Laryngeal 2, el cual se a proyectado en 2D mediante *PCA*.

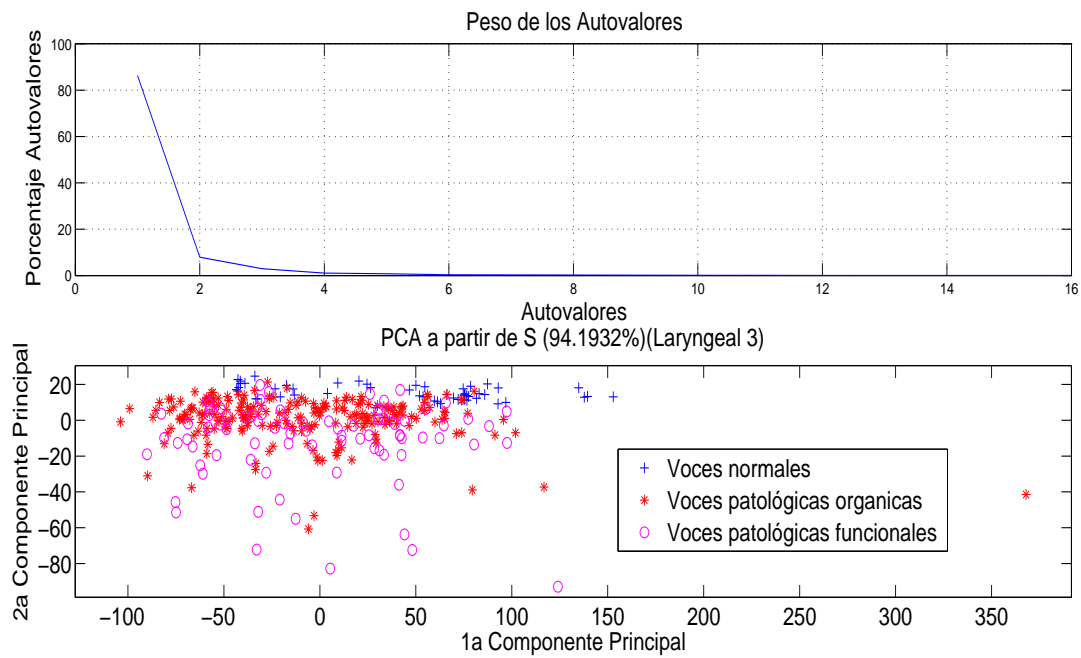


**Figura 5-9.:** Base de datos Laryngeal 2, el cual se a proyectado en 2D mediante *MDS*.

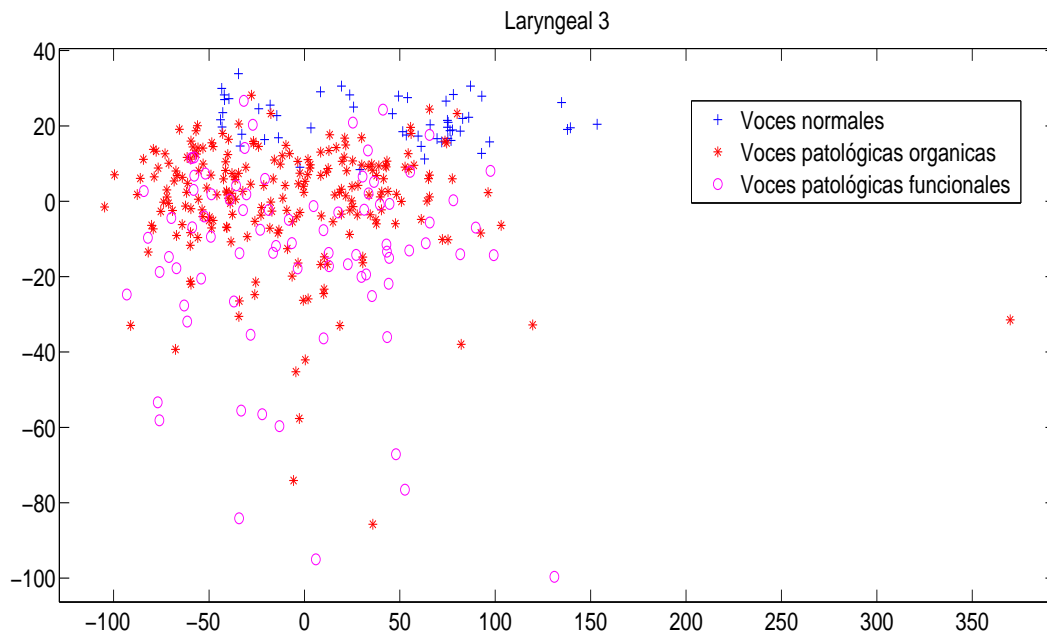
### 5.1.5. Conjunto de Laryngeal 3

El conjunto de datos Laryngeal-3 proviene de la Academia Búlgara de Ciencias y está disponible en [16]. El conjunto fue utilizado originalmente para un sistema informático de soporte de decisiones, con el fin de ayudar al diagnóstico de patología de la laringe y en especial en la detección de sus primeras etapas. Voces normales y patológicas se describen en 16 parámetros en el tiempo, los dominios espectral y cepstral.

El número de muestras en este conjunto es de 353, las cuales están separadas de la siguiente manera: 53 voces normales, 218 voces patológicas orgánicas y 82 voces patológicas funcionales (neurológico). De manera simplificada se puede observar el resumen de este conjunto de datos en la tabla 5-2. Las figura 5-10 y 5-11 corresponden a la proyección del conjunto de datos en 2D, al cual se hizo una reducción de dimensión mediante el análisis de componentes principales (*PCA*) y el escalamiento multidimensional (*MDS*). La razón por la cual en la gráfica 5-11 los ejes no tienen etiquetas es dado a que estos corresponden a combinaciones no lineales de las características originales. Por ello, no tienen una interpretación directa, recordando que *MDS* intenta preservar las distancias entre los puntos tan fielmente como sea posible, con el fin de obtener una visualización en 2D ó 3D de los datos de alta dimensión.



**Figura 5-10.:** Base de datos Laryngeal 3, el cual se a proyectado en 2D mediante *PCA*.



**Figura 5-11.:** Base de datos Laryngeal 3, el cual se a proyectado en 2D mediante *MDS*.



### 5.1.6. Conjunto de datos Iris

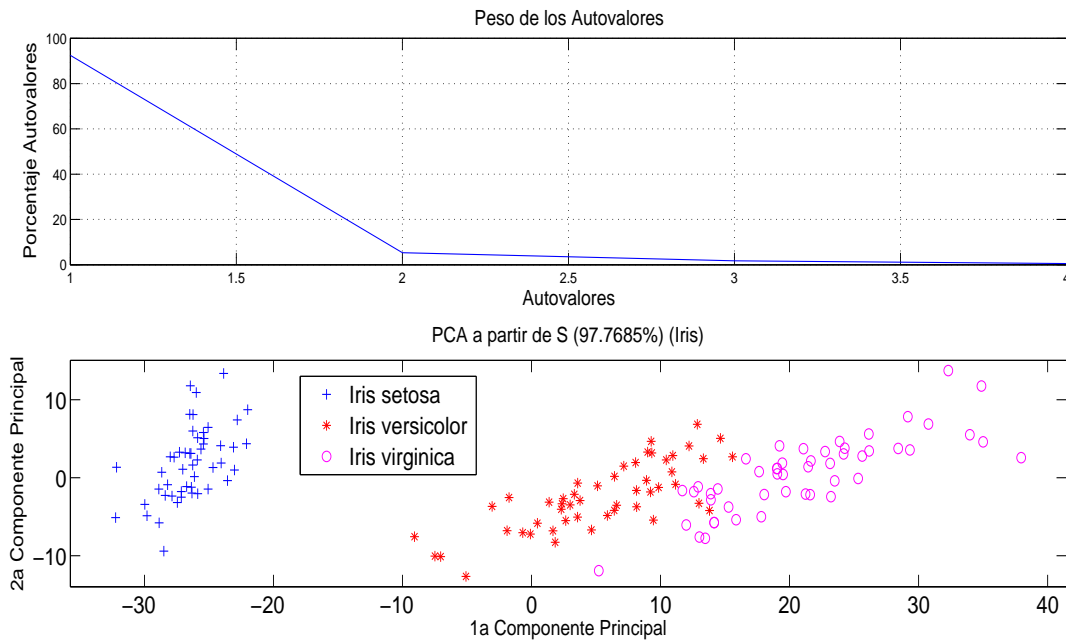
El conjunto de datos iris consta de 150 muestras, las cuales están divididas en 3 clases, cada una con 50 muestras. Cada clase se refiere a un tipo de planta, Iris setosa, Iris versicolor e Iris virginica. La primera clase es linealmente separable de las otras dos, mientras que las últimas no lo son una de otra. El número de características de las muestras son 4: longitud de los sépalos en centímetros, ancho de los sépalos en centímetros, longitud de los pétalos en centímetros y ancho de los pétalos en centímetros [8]. El resumen estadístico de este conjunto de datos se presenta en la tabla 5-2

Las figura 5-12 y 5-13 corresponden a la proyección del conjunto de datos en 2D, al cual se

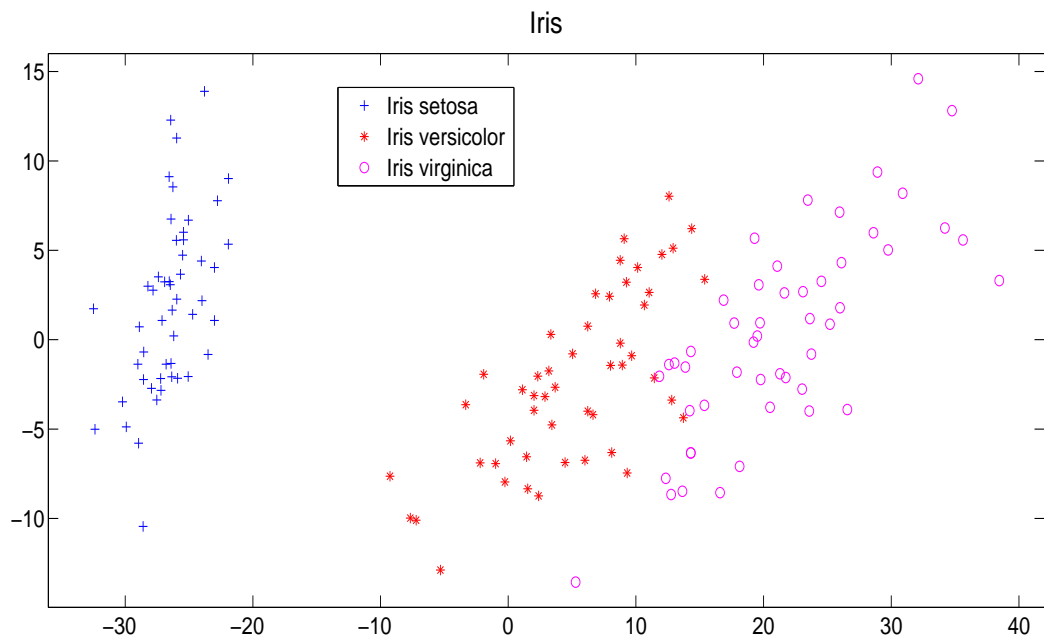
**Tabla 5-1.:** Resumen estadístico base de datos iris

	min	max	Promedio	Desviación estándar	Correlación entre las clases
Longitud del sépalo	4.3	7.9	5.84	0.83	0.7826
Ancho del sépalo	2.0	4.4	3.05	0.43	-0.4194
Longitud del pétalo	1.0	6.9	3.76	1.76	0.9490 (Alta!)
Ancho del pétalo	0.1	2.5	1.20	0.76	0.9565 (Alta!)

hizo una reducción de dimensión mediante el análisis de componentes principales (*PCA*) y escalamiento multidimensional (*MDS*). Los ejes en la gráfica 5-13 no tienen etiquetas dado a que estos corresponden a combinaciones no lineales de las características originales. Por ello, no tienen una interpretación directa, recordando que *MDS* intenta preservar las distancias entre los puntos tan fielmente como sea posible, con el fin de obtener una visualización en 2D ó 3D de los datos de alta dimensión.



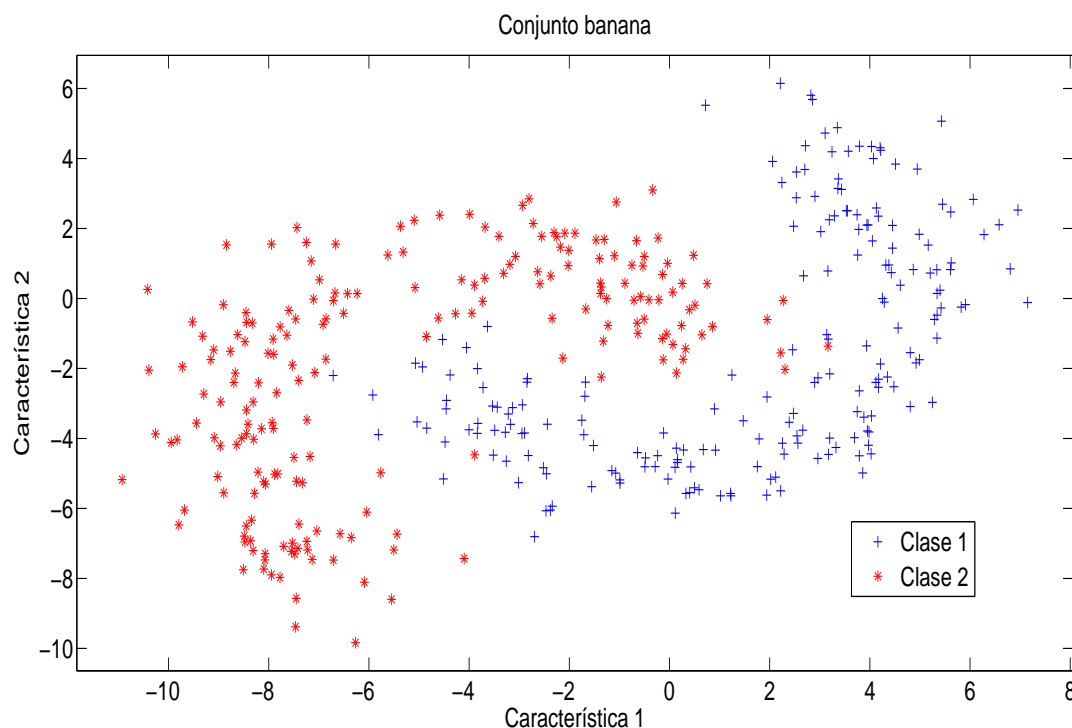
**Figura 5-12.:** Base de datos Iris, la cual se a proyectado en 2D mediante *PCA*.



**Figura 5-13.:** Base de datos Iris, la cual se a proyectado en 2D mediante *MDS*.

### 5.1.7. Conjunto de datos banana

El conjunto banana es un conjunto de  $m$  objetos bi-dimensional con dos clases. Los objetos se distribuye uniformemente a lo largo de las bananas y se superpone con una distribución normal con una desviación estándar  $S$  en todas las dos direcciones. La probabilidad de las clases es  $P(1) = P(2) = 0.5$ . (La generación de este conjunto de datos se hace mediante la función **gendatb** [5], del toolbox de matlab PRTools).



**Figura 5-14.:** Conjunto banana con 200 objetos en la clase 1 y 250 objetos en la clase 2 (conjunto de datos generado mediante la función **gendatb** del toolbox PRTools [5]).

Además de las bases de datos expuestas anteriormente, se hará para las pruebas de desempeño de los clasificadores de las bases de datos referenciadas en las tablas 5-2 a 5-4 . Cada una de estas bases de datos es tomada del toolbox PRTools [5] y de Machine Learning Repository [8].

**Tabla 5-2.:** Conjuntos de datos seleccionados

Conjunto de datos	Clases	Descripción	Dimensión
Clases Difíciles	2	Base de datos artificial	$2 \leq k \leq n, k \in \mathbb{N}$
Clases de Highleyman	2	Base de datos artificial	2
Conjunto banana	2	Base de datos artificial	2
Círculos concéntricos	2	Base de datos artificial, con clases distribuidas de manera concéntricas, [5][gendatc]	$2 \leq k \leq n, k \in \mathbb{N}$
Clases lituanas	2	Base de datos artificial. Clases lituanas [5][gendatl]	2
Clases distribuidas de manera gaussiana	2	Base de datos artificial. Conjunto de dos clases con distribución gaussiana y matrices de covarianzas idénticas [5][gendats]	$2 \leq k \leq n, k \in \mathbb{N}$
Truncamiento de 2 clases gaussianas	2	Base de datos artificial. Problema de clasificación de truncamiento de dos clases gaussianas	$2 \leq k \leq n, k \in \mathbb{N}$
Laryngeal-1	2	Características de 213 voces, normales o patológicas	16
Laryngeal-2	2	Características de 692 voces, normales o patológicas	16

Tabla 5-3.: Conjuntos de datos seleccionados

Conjunto de datos	Clases	Descripción	Dimensión
<b>Laryngeal-3</b>	3	Características de 353 voces, normales, patológicas funcionales o patológicas orgánicas	16
<b>Iris</b>	3	Características de plantas	3
<b>Wine</b>	3	Estos datos son los resultados de un análisis químico de los vinos cultivados en la misma región en Italia, pero derivada de tres cultivares diferentes. El análisis determina las cantidades de 13 componentes que se encuentran en cada uno de los tres tipos de vinos.	13
<b>Malaysia</b>	20	Conjunto de datos que se refiere a medidas simples en los segmentos de los símbolos de servicios públicos de malaysia	8
<b>Diabetes</b>	2	Características de 768 pacientes, con diabetes o sin diabetes	8
<b>Imox</b>	4	Las características son las distancias desde un cuadro de límite a los píxeles de un conjunto de caracteres escritos a mano 'i', 'M', 'O', y 'X', medidos desde las esquinas a lo largo de las diagonales y desde los puntos medios del borde a lo largo de la horizontal y vertical de los ejes centrales	8
<b>ORL</b>	40	40 sujetos, representados por 10 fotos cada uno. Cada imagen tiene un tamaño de $112 \times 92$ píxeles.	10304

Tabla 5-4.: Conjuntos de datos seleccionados

Conjunto de datos	Clases	Descripción	Dimensión
<b>Biomed</b>	2	El conjunto de datos de biomedicina se tiene para desarrollar métodos de detección para identificar a los portadores de un trastorno genético poco común. Debido a que la enfermedad es rara, sólo hay unos pocos transportistas de la enfermedad de la que se dispone de datos.	5
<b>Heart</b>	2	Esta base de datos contiene 76 atributos, pero todos los experimentos publicados se refieren al uso de un subconjunto de 14 de ellos. El "objetivo" es determinar la presencia de la enfermedad cardíaca en el paciente.	13
<b>Glass</b>	4	Conjunto de datos que describe los componentes químicos en varios tipos de gafas	9
<b>Ionosphere</b>	2	En este conjunto de datos se describe por 2 atributos por número de impulsos, que corresponden a los valores complejos devueltos por la función resultante de la señal electromagnética compleja	34

### 5.1.8. Evaluación y comparación de la clasificación

Dado que para la estimación de densidades de probabilidad de Parzen, es necesaria la estimación del parámetro de suavizado  $\sigma_h$ , en cada uno de los experimentos, el valor de este parámetro, se obtiene de acuerdo al procedimiento presentado en la sección 4.2, a partir de la generación del conjunto de entrenamiento y prueba. Para obtener estos conjuntos, el método a seguir es generar la representación del conjunto de datos para el espacio de disimilitudes, luego determinar el conjunto de representación mediante la selección de objetos prototipo y finalmente con estos objetos prototipo obtener el conjunto de representación conformado por líneas de características.

Se consideran cuatro clasificadores implementados sobre el espacio de características, los cuales también son válidos para el espacio de disimilitudes. Estos clasificadores son:

- Clasificador normal lineal [5][Observe la función *ldc*].
- Clasificador normal cuadrático [5][Observe la función *qdc*].
- Clasificador de Parzen [5][Observe la función *parzenc*].
- Máquina de soporte de vectores[5][Observe la función *svc*].

Y se considera también el clasificador basado en la estimación de Parzen para el espacio generalizado de disimilitudes, que se ha denominado *parzenc\_dis*.

La comparación de los clasificadores se hace a partir del desempeño de cada uno respecto a las bases de datos, desempeño en cuanto al error de clasificación y al tiempo de cómputo (medido en segundos) de cada uno de ellos. Además, con el fin de obtener un punto adecuado de comparación entre los clasificadores, con una confianza del 99 % calculada con una distribución t-student y 99 grados de libertad, se encuentra un intervalo de confianza para el error promedio cometido por cada uno de ellos, de acuerdo al número de objetos de representación seleccionados para la generación de las líneas de características.

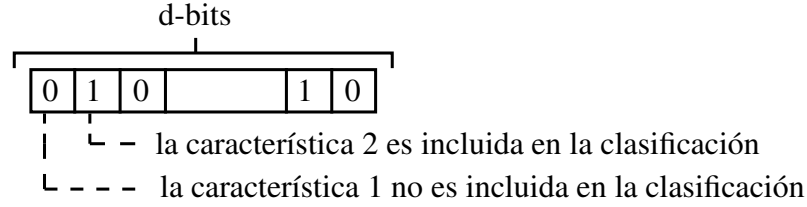
## 5.2. Metodología de evaluación

Para evaluar el desempeño de cada uno de los clasificadores, una vez determinados los conjuntos de entrenamiento y prueba, se debe hacer una reducción de dimensión, dado que se podrían generar problemas debido a la alta dimensión que pueden tener las representaciones de disimilitudes de los conjunto de datos. Existen diversas maneras de hacer esta reducción de dimensión, tales como análisis de componentes principales (*PCA*) [28, 12], análisis de agrupamiento (cluster)[28], análisis factorial [28, 12], y en el caso de las disimilitudes se ha mostrado que la selección aleatoria de objetos resulta ser adecuada y eficiente [26]. Para el número de objetos prototipo seleccionados de manera aleatoria, el método de *Kcentros* [26, 5], funciona bien para ejercer un control de este número, este método consiste encontrar  $K$  objetos-centro de una matriz de disimilitud  $D$ , los objetos-centro son escogidos del conjunto de datos de manera tal que el máximo de todas las distancias de los objetos a el más cercano de los objetos-centro, es minimizada [26, 5], siendo el número  $K$  el número de objetos a seleccionar por clase.

Como la selección aleatoria de objetos prototipo es un método adecuado y eficiente para la reducción de dimensión en el espacio de disimilitudes, para el espacio generalizado de disimilitudes se realiza también la selección aleatoria de objetos prototipo, de acuerdo a la longitud de cada una de las líneas de características, representadas mediante las columnas de la matriz generalizada de disimilitudes. Se seleccionan las líneas de mayor longitud, longitud media y las de menor longitud, para generar el conjunto de entrenamiento, reduciendo la dimensión del mismo y minimizando las posibles dificultades que trae consigo el manejo de representaciones de alta dimensión. Para cada uno de los tres casos, de acuerdo a la longitud, se realiza una comparación en cuanto al desempeño, con el fin de estudiar y verificar el comportamiento de la longitud de las líneas y su incidencia en el desempeño del clasificador. Es importante hacer la siguiente aclaración: en la sección de resultados, se compararan aquellos resultados que muestran una diferencia significativa en cuanto al desempeño, es decir se comparan los métodos de selección de las líneas de características de acuerdo a su longitud. El número de líneas a tomar está a juicio del investigador, aunque como se plantea una reducción de la dimensión, resulta conveniente tomar un número pequeño, que por el momento, a falta de un razonamiento matemático, depende de la experiencia y del tipo de datos. Se debe tener también en cuenta que cada línea de características en el conjunto de representación ha sido generada por objetos del conjunto de representación en el espacio de disimilitudes.

Por otra parte, para ilustrar un método de selección diferente a la longitud de las líneas, se presenta un método basado en algoritmo genético (*GA*)[10], este método consiste en encontrar un vector binario óptimo, donde cada bit está asociado con una característica, como se ilustra en la figura 5-15. Si el  $i$ -ésimo bit de este vector es igual a 1, entonces la





**Figura 5-15.:** Selección de características con el algoritmo genético simple

$i$ -ésima característica (columna de la matriz generalizada de disimilitudes) participa en la clasificación; si el bit es 0, entonces la característica correspondiente no es tomada en cuenta en la clasificación.

Respecto a los resultados para  $GA$  se muestran unos pocos debido a diversos inconvenientes en cuanto al tiempo de cómputo para este método, éste es excesivo teniendo en cuenta que se hace de una manera sencilla, como lo es repitiendo el experimento una única vez, una generación y 3 sujetos por generación. Este método resulta ser inapropiado para la estimación de densidades de probabilidad de Parzen en el espacio generalizado de disimilitudes debido al alto costo computacional.

Existen diferentes criterios para comprobar el desempeño de un clasificador. La probabilidad del error de clasificación, es decir, la tasa de error, es el más utilizado. La expresión analítica para la tasa de error está dada como:

$$Error_{\min} = \int_z e_{\min}(z)p(z)dz \quad (5-1)$$

en donde  $z$  es un vector de medida con dimensión  $m$ , la integral sobre  $z$  significa que la probabilidad de error promediado sobre todos los objetos, se puede encontrar por un promedio de  $e(z)$  sobre todos los vectores de medición posibles y  $e(z) = 1 - P(\hat{\omega}(z)|z)$ . La expresión para la tasa de error no es muy útil porque, en la práctica, las densidades de probabilidad condicionales son desconocidas. Sin embargo, se puede obtener fácilmente una estimación de la tasa de error sometiendo el clasificador a un conjunto de validación. El porcentaje de error estimado, es la fracción de muestras mal clasificadas con respecto al tamaño del conjunto de validación. Cuando la clasificación usa la opción de rechazo, se necesita considerar dos medidas: la tasa de error y la tasa de rechazo. Para determinar el desempeño de los clasificadores se particiona entonces el conjunto de datos, en un conjunto de entrenamiento y otro conjunto de prueba. De esta manera, suponga que el conjunto de validación o conjunto de prueba consiste de  $N_{prueba}$  muestras etiquetadas.  $n_{error}$  es el número de muestras mal clasificadas luego de entrenar el clasificador con el conjunto de entrenamiento. Así la tasa de error del clasificador denotada como  $E$ , es:

$$\hat{E} = \frac{n_{error}}{N_{prueba}} \quad (5-2)$$

De esta manera se realizara la comparación de desempeños de los clasificadores (La función *testc* [5], permite hallar este desempeño de los clasificadores).

Ahora, al momento de realizar la reducción de dimensión del conjunto representación de disimilitudes conformado por líneas de características, se tomarán en cuenta cada uno de los objetos que participaron en la conformación de las líneas seleccionadas, de manera que estos mismos objetos serán seleccionados en la matriz de disimilitudes. Teniendo en cuenta que, un objeto participo en la formación de una línea de características, si la distancia del objeto a la línea es cero, es decir, si en la posición correspondiente en la matriz generalizada de disimilitudes, el valor es cero. El correspondiente objeto en el espacio de disimilitudes será seleccionado de la matriz de disimilitudes, de esta manera se garantiza que la comparación entre los desempeños de los clasificadores en el espacio de disimilitudes frente al clasificador en el espacio generalizado de disimilitudes, basado en la estimación de Parzen, sea la adecuada y correcta para mostrar los resultados esperados de desempeño.

Cada uno de los experimentos se repite 100 veces y los errores de clasificación son promediados; 100 veces porque para 100 la distribución t-student aproxima la normal y t-student para los intervalos porque no se conoce la desviación exacta. Para la formación de las líneas de características a partir de los objetos prototipos en el espacio de disimilitudes, se varía el número de objetos seleccionados para dicha formación, entre 4 y 20 objetos, teniendo en cuenta que el número de líneas crece de manera exponencial de acuerdo al número de objetos prototipo tomados (ver la ecuación 3-2). Finalmente, para cada una de las bases de datos, se presentarán gráficos en los cuales se compara el desempeño de los clasificadores, de acuerdo al error cometido por cada uno de ellos, según el número de objetos de representación tomados para la generación de las líneas de características. Además, para cada una de las bases de datos, se presentan tablas con los intervalos de confianza para los errores, la desviación estándar de cada uno de los errores, el promedio de computo (medido en segundos) de cada uno de los clasificadores luego de las 100 repeticiones, y además para el caso del clasificador *Parzenc\_dis* se muestra, en el anexo B, el promedio del valor del parámetro de suavizado obtenido de acuerdo al número de objetos de representación tomados.

## 6. Resultados y discusiones

### 6.1. Resultados

Tomando como referencia los clasificadores, *ldc*, *qdc*, *svc* y *parzenc*, se evalúan sus desempeños, de acuerdo a el error y tiempo de cómputo, y se realiza la comparación con el clasificador *parzenc\_dis*, basado en la estimación de Parzen sobre el espacio generalizado de disimilitudes. A través del uso del toolbox de Matlab, PRTools [5], y de la función *testc* (ver capítulo 5, sección 5.2), se obtienen las estimaciones del error de los cuatro primeros clasificadores, ya que estos mismos están implementados en dicho toolbox, mientras que para *parzenc\_dis*, se hará la estimación del error a partir del conocimiento de los objetos y las clases en el conjunto de prueba, contando el número de objetos mal clasificados y luego calcular el cociente del número de objetos mal clasificados sobre el número de objetos en el conjunto de prueba. Los conjuntos de entrenamiento y prueba para los cuatro clasificadores tomados dependen de la selección y formación del conjunto de entrenamiento para *parzenc\_dis*, dado que, para realizar la correcta comparación de los desempeños de los clasificadores se necesitan obtener los respectivos objetos, en el espacio de disimilitudes, que participaron en la construcción de las líneas de características seleccionadas para conformar el conjunto de entrenamiento en el espacio generalizado de disimilitudes.

Los resultados de estos desempeños se evaluarán de acuerdo al número de objetos tomados para generar el conjunto de representación conformado por líneas de características, de la formación de cada uno de los conjuntos de entrenamiento y prueba, y de los errores obtenidos para cada uno de los números de objetos de representación considerados, recordando que varían entre 4 y 20, y serán mostrados mediante gráficas del número de objetos en el conjunto de representación para determinar las líneas de características, contra el error del clasificador para el conjunto de prueba conformado por líneas.

Para los conjuntos de datos artificiales (ver tablas 5-2 y 5-3), y de manera aleatoria se seleccionaron los números de objetos por clase de con el fin de responder a dudas tales como: si el número de objetos por clase debe ser balanceado, o muy grande o muy pequeño. Para las demás bases de datos se mostrarán gráficas con el desempeño de los clasificadores. Además, para cada una de las pruebas representada por las gráficas, se presentan una tabla con los intervalos de confianza para los errores, la desviación estándar de cada uno de los errores, el promedio de cómputo (medido en segundos) de cada uno de los clasificadores luego de

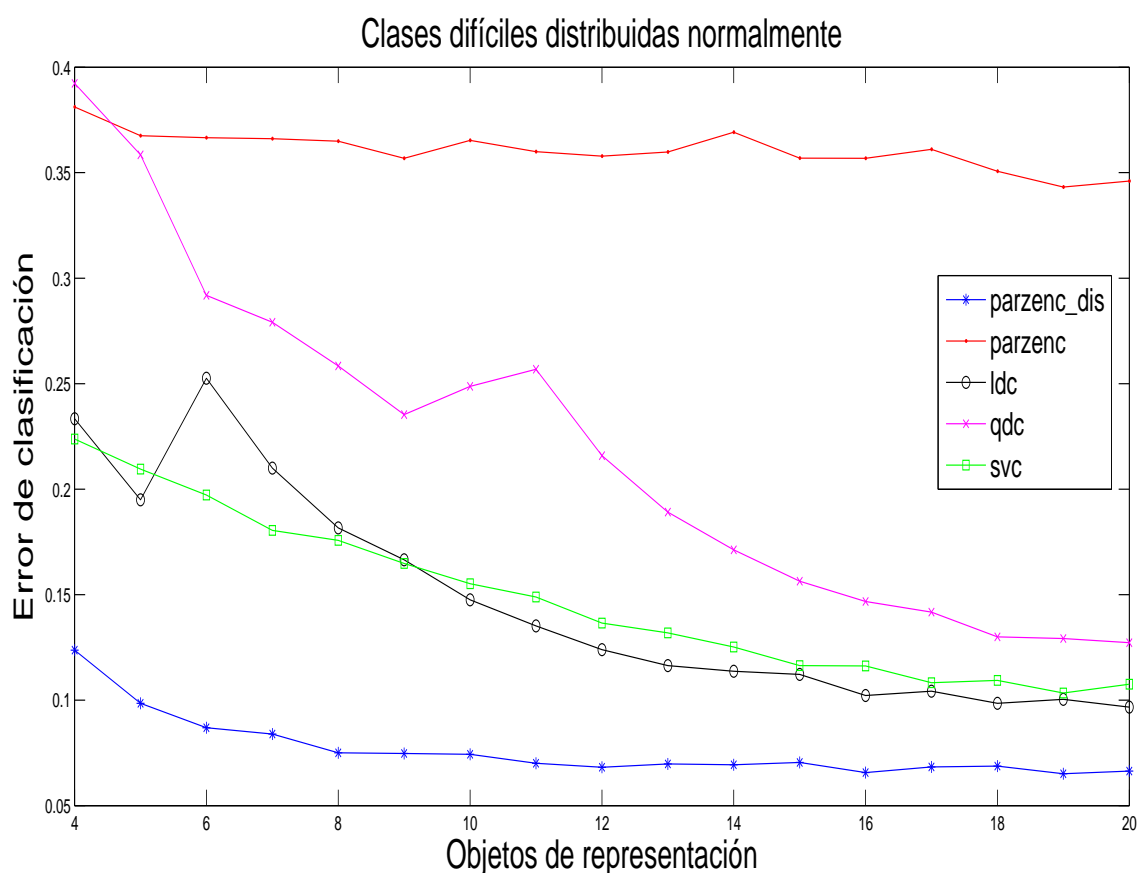
las 100 repeticiones, y para el clasificador *Parzenc\_dis* se muestra el promedio del valor del parámetro de suavizado obtenido de acuerdo al número de objetos de representación tomados (ver anexo B).

Al final de esta sección se presentan en las tablas 6-67 y 6-68, los resultados del desempeño de los clasificadores *ldc*, *qdc*, *svc*, *parzenc*, junto con el clasificador *knnc* en el espacio de características, con el fin de realizar una comparación del desempeño en el espacio generalizado de disimilitudes, en cada una de las tablas se encuentran los errores promedios, los intervalos de confianza para el error promedio y el tiempo de cómputo de cada clasificador, teniendo en cuenta que los conjuntos de entrenamiento y prueba han sido generados cada uno con el 70 % y 30 % de los elementos en el conjunto de datos respectivamente.

Se presentan ahora cada una de las diversas gráficas y tablas con los desempeños en cuanto al error promedio y tiempo de cómputo de los clasificadores:

### 6.1.1. Clases difíciles distribuidas normalmente

Mediante la rutina **gendatd** del toolbox de reconocimiento de patrones de Matlab (**PRTools**) [5], se genera el conjunto de clases difíciles distribuidas normalmente, en el cual el número de objetos por clase son: 150 objetos para la clase uno y 200 para la clase 2. Además, tomando el mismo número de objetos por clase, se genera un nuevo conjunto de clases difíciles distribuidas normalmente, variando su dimensión de 2 a 20, con el fin de mostrar el desempeño de los clasificadores con una base de datos de dimensión diferente a dos.



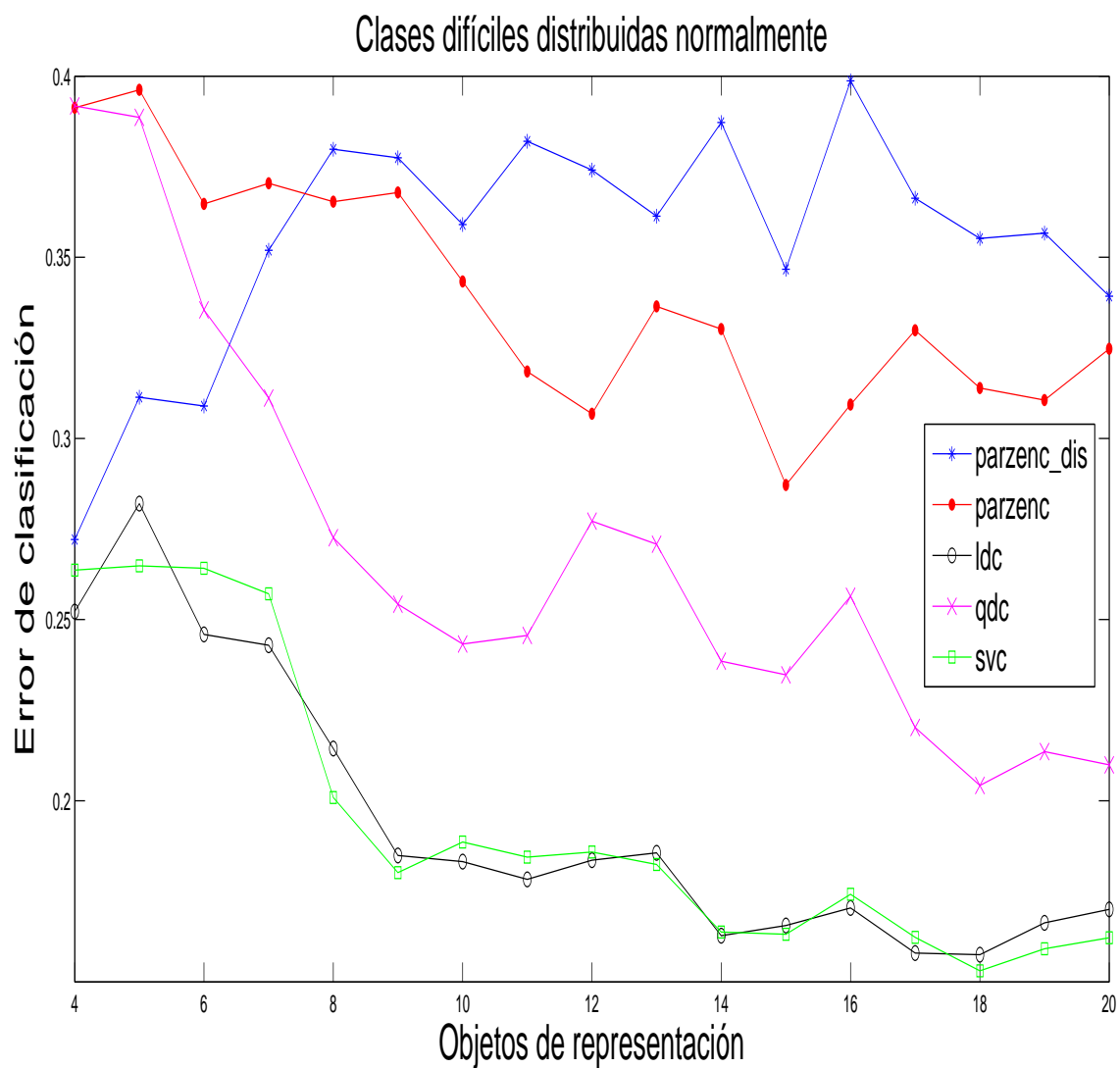
**Figura 6-1.:** Desempeño de cada uno de los clasificadores variando el número de objetos de representación por clase. Conjunto de clases difíciles distribuidas normalmente, con 150 objetos para la clase uno y 200 objetos para la clase 2. **Selección de las líneas de características de longitud mayor.**

**Tabla 6-1.:** Intervalos de confianza para el error promedio de cada clasificador de acuerdo al número de objetos de representación por clase (Resaltado se encuentra el mejor intervalo para cada clasificador).

Intervalos de confianza para el error promedio					
Objetos de representación	<i>parzenc_dis</i>	<i>parzenc</i>	<i>ldc</i>	<i>qdc</i>	<i>svc</i>
4	[0.1039,0.1433]	[0.3614,0.4008]	[0.21,0.2568]	[0.3682,0.4166]	[0.199,0.2484]
5	[0.0813,0.1155]	[0.3467,0.3883]	[0.1744,0.2154]	[0.3324,0.3846]	[0.1871,0.2319]
6	[0.0735,0.1003]	[0.3469,0.3863]	[0.2278,0.2772]	[0.2728,0.3108]	[0.1793,0.2151]
7	[0.0755,0.0923]	[0.3461,0.3861]	[0.1913,0.2287]	[0.2541,0.3041]	[0.1633,0.1975]
8	<b>[0.0676,0.0824]</b>	[0.347,0.3828]	[0.164,0.1992]	[0.2342,0.2826]	[0.1589,0.1925]
9	[0.0679,0.0815]	[0.3394,0.3742]	[0.1536,0.1794]	[0.2129,0.2577]	[0.1482,0.1814]
10	[0.0664,0.0822]	[0.345,0.3856]	[0.1314,0.1636]	[0.2269,0.2705]	[0.1397,0.1707]
11	[0.0637,0.0763]	[0.3415,0.3783]	[0.1228,0.1476]	[0.2328,0.2808]	[0.1357,0.1621]
12	[0.0623,0.0739]	[0.3413,0.3745]	[0.1121,0.1357]	[0.1983,0.2335]	[0.1236,0.1494]
13	[0.0634,0.076]	[0.3406,0.379]	[0.1068,0.1258]	[0.1728,0.2054]	[0.1205,0.1431]
14	[0.0646,0.074]	[0.3531,0.3853]	[0.1031,0.1241]	[0.1559,0.1865]	[0.113,0.1372]
15	[0.0651,0.0757]	[0.3422,0.3716]	[0.1024,0.1218]	[0.1437,0.1689]	[0.105,0.1276]
16	[0.061,0.0704]	[0.3415,0.3721]	[0.0925,0.1119]	[0.1356,0.1578]	[0.1057,0.1267]
17	[0.0617,0.0749]	[0.3455,0.3765]	[0.0953,0.1131]	[0.1301,0.1533]	[0.0979,0.1185]
18	[0.0629,0.0745]	[0.3341,0.3673]	[0.0903,0.1067]	[0.1199,0.1399]	[0.0994,0.1194]
19	<b>[0.0604,0.0698]</b>	<b>[0.325,0.3614]</b>	[0.0907,0.1101]	[0.1203,0.1381]	<b>[0.0946,0.112]</b>
20	[0.06,0.0726]	[0.3289,0.3631]	<b>[0.0888,0.1046]</b>	<b>[0.1169, 0.1375]</b>	[0.098,0.117]

**Tabla 6-2.:** Tabla del Error promedio, desviación estándar, tiempo de cómputo, de los clasificadores de acuerdo al número de objetos de representación por clase (resaltado se encuentra el mejor y el peor tiempo promedio de cómputo de cada clasificador).

Error promedio, desviación estándar, tiempo de cómputo de los clasificadores										
Objetos de representación	<i>parzenc_dis</i>		<i>parzenc</i>		<i>ldc</i>		<i>qdc</i>		<i>svc</i>	
	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)
4	$0.1236 \pm 0.075$	<b>0.1303</b>	$0.3811 \pm 0.075$	<b>0.033</b>	$0.2334 \pm 0.089$	<b>0.0298</b>	$0.3924 \pm 0.092$	<b>0.0225</b>	$0.224 \pm 0.094$	<b>0.0538</b>
5	$0.0984 \pm 0.065$	0.145	$0.3675 \pm 0.079$	<b>0.033</b>	$0.1949 \pm 0.078$	<b>0.0298</b>	$0.3585 \pm 0.099$	<b>0.0225</b>	$0.21 \pm 0.085$	0.0539
6	$0.0869 \pm 0.051$	0.1604	$0.3666 \pm 0.075$	0.0331	$0.2525 \pm 0.094$	0.0294	$0.2918 \pm 0.072$	0.0225	$0.197 \pm 0.068$	0.0561
7	$0.0839 \pm 0.032$	0.1759	$0.3661 \pm 0.076$	0.0332	$0.21 \pm 0.071$	0.0295	$0.2791 \pm 0.095$	0.0226	$0.18 \pm 0.065$	0.0576
8	$0.075 \pm 0.028$	0.1907	$0.3649 \pm 0.068$	0.0332	$0.1816 \pm 0.067$	0.0294	$0.2584 \pm 0.092$	0.0225	$0.176 \pm 0.064$	0.0582
9	$0.0747 \pm 0.026$	0.2066	$0.3568 \pm 0.066$	0.0331	$0.1665 \pm 0.049$	0.0295	$0.2353 \pm 0.085$	0.0226	$0.165 \pm 0.063$	0.0589
10	$0.0743 \pm 0.03$	0.222	$0.3653 \pm 0.077$	0.0333	$0.1475 \pm 0.061$	0.0295	$0.2487 \pm 0.083$	0.0225	$0.155 \pm 0.059$	0.0592
11	$0.07 \pm 0.024$	0.2374	$0.3599 \pm 0.07$	0.0332	$0.1352 \pm 0.047$	0.0295	$0.2568 \pm 0.091$	<b>0.0215</b>	$0.149 \pm 0.05$	0.0596
12	$0.0681 \pm 0.022$	0.2524	$0.3579 \pm 0.063$	0.0337	$0.1239 \pm 0.045$	0.03	$0.2159 \pm 0.067$	<b>0.0217</b>	$0.137 \pm 0.049$	0.059
13	$0.0697 \pm 0.024$	0.2681	$0.3598 \pm 0.073$	0.0347	$0.1163 \pm 0.036$	0.0312	$0.1891 \pm 0.062$	0.0224	$0.132 \pm 0.043$	0.0608
14	$0.0693 \pm 0.018$	0.2835	$0.3692 \pm 0.061$	0.0385	$0.1136 \pm 0.04$	0.0361	$0.1712 \pm 0.058$	0.025	$0.125 \pm 0.046$	0.0659
15	$0.0704 \pm 0.02$	0.2987	$0.3569 \pm 0.056$	0.0373	$0.1121 \pm 0.037$	0.0347	$0.1563 \pm 0.048$	0.0244	$0.116 \pm 0.043$	0.0662
16	$0.0657 \pm 0.018$	0.314	$0.3568 \pm 0.058$	0.0384	$0.1022 \pm 0.037$	0.0363	$0.1467 \pm 0.042$	0.0254	$0.116 \pm 0.04$	0.0676
17	$0.0683 \pm 0.025$	0.3302	$0.361 \pm 0.059$	0.0378	$0.1042 \pm 0.034$	0.0352	$0.1417 \pm 0.044$	0.0247	$0.108 \pm 0.039$	0.0673
18	$0.0687 \pm 0.022$	0.3459	$0.3507 \pm 0.063$	0.0387	$0.0985 \pm 0.031$	0.0366	$0.1299 \pm 0.038$	0.0256	$0.109 \pm 0.038$	0.0685
19	$0.0651 \pm 0.018$	0.3627	$0.3432 \pm 0.069$	0.0406	$0.1004 \pm 0.037$	0.0391	$0.1292 \pm 0.034$	0.0271	$0.103 \pm 0.033$	0.0726
20	$0.0663 \pm 0.024$	<b>0.3763</b>	$0.346 \pm 0.065$	<b>0.0392</b>	$0.0967 \pm 0.03$	<b>0.0369</b>	$0.1272 \pm 0.039$	<b>0.0255</b>	$0.108 \pm 0.036$	<b>0.0697</b>



**Figura 6-2.:** Desempeño de cada uno de los clasificadores variando el número de objetos de representación por clase. Conjunto de clases difíciles distribuidas normalmente, con 150 objetos para la clase uno y 200 objetos para la clase 2. **Selección de las líneas de características de longitud menor.**

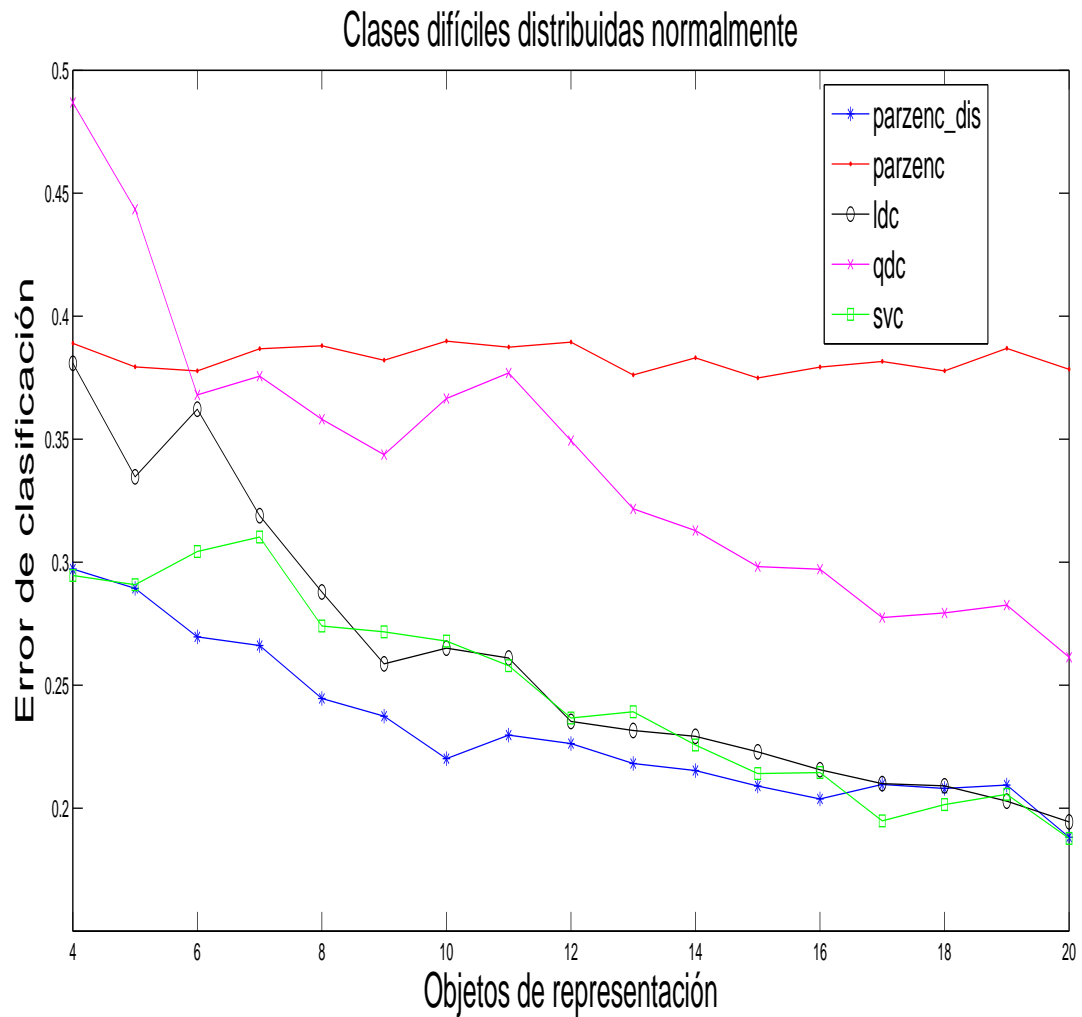


**Tabla 6-3.:** Intervalos de confianza para el error promedio de cada clasificador de acuerdo al número de objetos de representación por clase (Resaltado se encuentra el mejor y el peor intervalo para cada clasificador).

Intervalos de confianza para el error promedio					
Objetos de representación	<i>parzenc_dis</i>	<i>parzenc</i>	<i>ldc</i>	<i>qdc</i>	<i>svc</i>
4	<b>[0.2184,0.3258]</b>	<b>[0.3525,0.4300]</b>	[0.2129,0.2916]	<b>[0.3452,0.4383]</b>	[0.2200,0.3073]
5	[0.2433,0.3794]	[0.3454,0.4470]	<b>[0.2335,0.3306]</b>	[0.3356,0.4416]	[0.2049,0.3248]
6	[0.2741,0.3438]	[0.3290,0.4003]	[0.2067,0.2851]	[0.2936,0.3773]	<b>[0.2246,0.3037]</b>
7	[0.2738,0.4300]	[0.3217,0.4191]	[0.2053,0.2806]	[0.2733,0.3490]	[0.2178,0.2964]
8	[0.3031,0.4565]	[0.3288,0.4018]	[0.1827,0.2462]	[0.2298,0.3155]	[0.1748,0.2270]
9	[0.3138,0.4411]	[0.3353,0.4004]	[0.1588,0.2110]	[0.2084,0.3000]	[0.1504,0.2099]
10	[0.2941,0.4239]	[0.2998,0.3868]	[0.1564,0.2100]	[0.2055,0.2810]	[0.1563,0.2209]
11	[0.3143,0.4497]	[0.2835,0.3534]	[0.1610,0.1956]	[0.2072,0.2840]	[0.1614,0.2075]
12	[0.3047,0.4434]	[0.2746,0.3390]	[0.1557,0.2115]	[0.2262,0.3282]	[0.1492,0.2226]
13	[0.2786,0.4441]	[0.2964,0.3764]	[0.1595,0.2118]	[0.2362,0.3054]	[0.1585,0.2064]
14	[0.3221,0.4522]	[0.2986,0.3617]	[0.1455,0.1801]	[0.1987,0.2784]	[0.1475,0.1800]
15	[0.2830,0.4104]	<b>[0.2475,0.3267]</b>	[0.1455,0.1856]	[0.1982,0.2713]	[0.1463,0.1800]
16	<b>[0.3387,0.4587]</b>	[0.2728,0.3459]	[0.1515,0.1893]	[0.2069,0.3060]	[0.1506,0.1979]
17	[0.3156,0.4170]	[0.2984,0.3613]	<b>[0.1376,0.1783]</b>	[0.1756,0.2648]	[0.1302,0.1944]
18	[0.2897,0.4206]	[0.2825,0.3453]	[0.1393,0.1758]	<b>[0.1732,0.2353]</b>	<b>[0.1290,0.1772]</b>
19	[0.3100,0.4033]	[0.2816,0.3395]	[0.1486,0.1841]	[0.1797,0.2475]	[0.1362,0.1821]
20	[0.2710,0.4075]	[0.2869,0.3625]	[0.1556,0.1844]	[0.1743,0.2455]	[0.1438,0.1805]

**Tabla 6-4.:** Tabla del Error promedio, desviación estándar, tiempo de cómputo, de los clasificadores de acuerdo al número de objetos de representación por clase (resaltado se encuentra el mejor tiempo promedio de cómputo de los clasificadores *parzenc\_dis*) y *parzenc*.

Error promedio, desviación estándar, tiempo de cómputo de los clasificadores										
Objetos de representación	<i>parzenc_dis</i>		<i>parzenc</i>		<i>ldc</i>		<i>qdc</i>		<i>svc</i>	
	$\bar{e}_\infty \pm \sigma$	CPU(seg)	$\bar{e}_\infty \pm \sigma$	CPU(seg)	$\bar{e}_\infty \pm \sigma$	CPU(seg)	$\bar{e}_\infty \pm \sigma$	CPU(seg)	$\bar{e}_\infty \pm \sigma$	CPU(seg)
4	<b>0.2721±0.0209</b>	<b>0.1275</b>	<b>0.3912±0.0151</b>	<b>0.0368</b>	0.2522±0.0153	0.0322	0.3918±0.0181	0.0225	0.2636±0.0170	0.0570
5	0.3114±0.0265	0.1410	0.3962±0.0198	0.0325	0.2820±0.0189	0.0294	0.3886±0.0206	0.0223	0.2648±0.0233	0.0540
6	0.3090±0.0136	0.1555	0.3647±0.0139	0.0326	0.2459±0.0152	0.0293	0.3354±0.0163	0.0224	0.2642±0.0154	0.0554
7	0.3519±0.0304	0.1707	0.3704±0.0190	0.0327	0.2429±0.0146	0.0292	0.3111±0.0147	0.0224	0.2571±0.0153	0.0572
8	0.3798±0.0298	0.1865	0.3653±0.0142	0.0328	0.2144±0.0123	0.0307	0.2726±0.0167	0.0254	0.2009±0.0102	0.0637
9	0.3775±0.0248	0.2013	0.3679±0.0127	0.0331	0.1849±0.0101	0.0316	0.2542±0.0178	0.0269	0.1801±0.0116	0.0661
10	0.3590±0.0252	0.2164	0.3433±0.0169	0.0363	0.1832±0.0104	0.0345	0.2433±0.0147	0.0267	0.1886±0.0126	0.0708
11	0.3820±0.0263	0.2314	0.3185±0.0136	0.0366	0.1783±0.0067	0.0346	0.2456±0.0149	0.0272	0.1844±0.0090	0.0692
12	0.3741±0.0270	0.2500	0.3068±0.0125	0.0383	0.1836±0.0109	0.0375	0.2772±0.0199	0.0279	0.1859±0.0143	0.0707
13	0.3613±0.0322	0.2650	0.3364±0.0156	0.0349	0.1856±0.0102	0.0308	0.2708±0.0135	0.0236	0.1824±0.0093	0.0612
14	0.3872±0.0253	0.2763	0.3302±0.0123	0.0339	0.1628±0.0067	0.0301	0.2385±0.0155	0.0227	0.1638±0.0063	0.0599
15	0.3467±0.0248	0.2910	0.2871±0.0154	0.0337	0.1656±0.0078	0.0300	0.2348±0.0142	0.0222	0.1631±0.0065	0.0600
16	0.3987±0.0233	0.3065	0.3093±0.0142	0.0340	0.1704±0.0074	0.0301	0.2565±0.0193	0.0221	0.1742±0.0092	0.0605
17	0.3663±0.0197	0.3219	0.3299±0.0122	0.0366	0.1580±0.0079	0.0339	0.2202±0.0174	0.0246	0.1623±0.0125	0.0668
18	0.3552±0.0255	0.3381	0.3139±0.0122	0.0404	0.1575±0.0071	0.0387	0.2042±0.0121	0.0274	0.1531±0.0094	0.0791
19	0.3567±0.0182	0.3513	0.3106±0.0113	0.0353	0.1663±0.0069	0.0315	0.2136±0.0132	0.0230	0.1591±0.0089	0.0658
20	0.3393±0.0266	0.3661	0.3247±0.0147	0.0403	0.1700±0.0056	0.0385	0.2099±0.0139	0.0292	0.1622±0.0071	0.0793



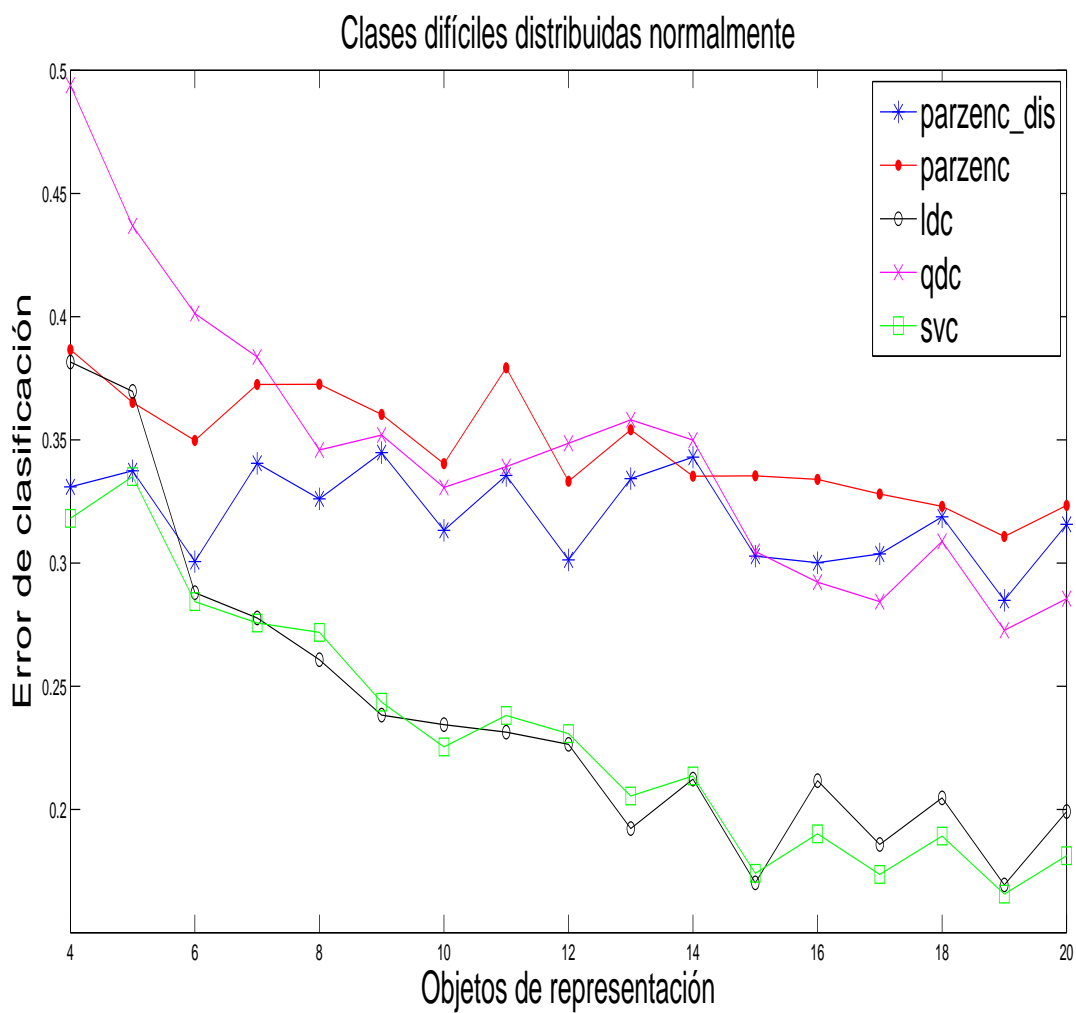
**Figura 6-3.:** Desempeño de cada uno de los clasificadores variando el número de objetos de representación por clase. Conjunto de clases difíciles distribuidas normalmente, con 150 objetos para la clase uno y 200 objetos para la clase 2 y dimensión del conjunto igual a 20. **Selección de las líneas de características de longitud mayor.**

**Tabla 6-5.:** Intervalos de confianza para el error promedio de cada clasificador de acuerdo al número de objetos de representación por clase (Resaltado se encuentra el mejor intervalo para cada clasificador).

Intervalos de confianza para el error promedio					
Objetos de representación	<i>parzenc_dis</i>	<i>parzenc</i>	<i>ldc</i>	<i>qdc</i>	<i>svc</i>
4	[0.2756,0.3188]	[0.3694,0.4084]	[0.3600,0.4016]	[0.4769,0.4969]	[0.2756,0.3136]
5	[0.2682,0.3104]	[0.3594,0.3994]	[0.3123,0.3571]	[0.4232,0.4638]	[0.2692,0.3124]
6	[0.2506,0.2886]	[0.3616,0.3938]	[0.3393,0.3851]	[0.3483,0.3877]	[0.2870,0.3218]
7	[0.2458,0.2864]	[0.3697,0.4039]	[0.2965,0.3413]	[0.3582,0.3930]	[0.2891,0.3313]
8	[0.2243,0.2649]	[0.3751,0.4009]	[0.2676,0.3082]	[0.3415,0.3747]	[0.2585,0.2895]
9	[0.2200,0.2548]	[0.3710,0.3932]	[0.2403,0.2771]	[0.3279,0.3595]	[0.2555,0.2877]
10	[0.2033,0.2369]	[0.3749,0.4049]	[0.2468,0.2832]	[0.3497,0.3833]	[0.2511,0.2847]
11	[0.2121,0.2473]	[0.3729,0.4019]	[0.2449,0.2771]	[0.3601,0.3937]	[0.2424,0.2734]
12	[0.2091,0.2433]	[0.3782,0.4008]	[0.2191,0.2513]	[0.3334,0.3656]	[0.2225,0.2509]
13	[0.2031,0.2331]	[0.3661,0.3861]	[0.2178,0.2452]	[0.3075,0.3359]	[0.2247,0.2537]
14	[0.2000,0.2306]	[0.3718,0.3944]	[0.2139,0.2445]	[0.2981,0.3275]	[0.2112,0.2402]
15	[0.1945,0.2235]	<b>[0.3652,0.3846]</b>	[0.2084,0.2374]	[0.2850,0.3114]	[0.2012,0.2270]
16	[0.1876,0.2198]	[0.3690,0.3896]	[0.2019,0.2293]	[0.2853,0.3089]	[0.2000,0.2290]
17	[0.1935,0.2257]	[0.3724,0.3908]	[0.1963,0.2237]	[0.2662,0.2888]	[0.1822,0.2074]
18	[0.1904,0.2256]	[0.3680,0.3874]	[0.1974,0.2206]	[0.2670,0.2918]	[0.1901,0.2127]
19	[0.1938,0.2248]	[0.3743,0.3995]	[0.1894,0.2162]	[0.2697,0.2955]	[0.1935,0.2177]
20	<b>[0.1747,0.2015]</b>	[0.3692,0.3876]	<b>[0.1831,0.2057]</b>	<b>[0.2488,0.2740]</b>	<b>[0.1743,0.2011]</b>

**Tabla 6-6.:** Tabla del Error promedio, desviación estándar, tiempo de cómputo, de los clasificadores de acuerdo al número de objetos de representación por clase (resaltado se encuentra el peor tiempo promedio de cómputo de cada clasificador).

Error promedio, desviación estándar, tiempo de cómputo de los clasificadores										
Objetos de representación	<i>parzenc_dis</i>		<i>parzenc</i>		<i>ldc</i>		<i>qdc</i>		<i>svc</i>	
	$\bar{e}_\infty \pm \sigma$	CPU(seg)	$\bar{e}_\infty \pm \sigma$	CPU(seg)	$\bar{e}_\infty \pm \sigma$	CPU(seg)	$\bar{e}_\infty \pm \sigma$	CPU(seg)	$\bar{e}_\infty \pm \sigma$	CPU(seg)
4	$0.2972 \pm 0.082$	<b>0.1303</b>	$0.3889 \pm 0.074$	0.0329	$0.3808 \pm 0.074$	0.0298	$0.4869 \pm 0.079$	0.0226	$0.2946 \pm 0.072$	0.0543
5	$0.2893 \pm 0.080$	0.1451	$0.3794 \pm 0.076$	0.0329	$0.3347 \pm 0.076$	0.0298	$0.4435 \pm 0.085$	0.0228	$0.2908 \pm 0.082$	0.0547
6	$0.2696 \pm 0.072$	0.1604	$0.3777 \pm 0.061$	0.033	$0.3622 \pm 0.061$	0.0295	$0.368 \pm 0.087$	0.0225	$0.3044 \pm 0.066$	0.0566
7	$0.2661 \pm 0.077$	0.1777	$0.3868 \pm 0.065$	0.0335	$0.3189 \pm 0.065$	0.0299	$0.3756 \pm 0.085$	0.0229	$0.3102 \pm 0.080$	0.0584
8	$0.2446 \pm 0.077$	0.1935	$0.388 \pm 0.049$	0.0332	$0.2879 \pm 0.049$	0.0296	$0.3581 \pm 0.077$	0.0229	$0.274 \pm 0.059$	0.0586
9	$0.2374 \pm 0.066$	0.2106	$0.3821 \pm 0.042$	0.0334	$0.2587 \pm 0.042$	0.0296	$0.3437 \pm 0.070$	0.0227	$0.2716 \pm 0.061$	0.0586
10	$0.2201 \pm 0.064$	0.2239	$0.3899 \pm 0.057$	0.0331	$0.265 \pm 0.057$	0.0295	$0.3665 \pm 0.069$	0.0223	$0.2679 \pm 0.064$	0.059
11	$0.2297 \pm 0.067$	0.2395	$0.3874 \pm 0.055$	0.0334	$0.261 \pm 0.055$	0.0296	$0.3769 \pm 0.061$	0.0215	$0.2579 \pm 0.059$	0.0588
12	$0.2262 \pm 0.065$	0.2539	$0.3895 \pm 0.043$	0.0336	$0.2352 \pm 0.043$	0.0301	$0.3495 \pm 0.061$	0.0218	$0.2367 \pm 0.054$	0.0594
13	$0.2181 \pm 0.057$	0.268	$0.3761 \pm 0.038$	0.0337	$0.2315 \pm 0.038$	0.03	$0.3217 \pm 0.052$	0.0218	$0.2392 \pm 0.055$	0.0594
14	$0.2153 \pm 0.058$	0.2855	$0.3831 \pm 0.043$	0.0338	$0.2292 \pm 0.043$	0.0302	$0.3128 \pm 0.058$	0.0219	$0.2257 \pm 0.055$	0.0598
15	$0.209 \pm 0.055$	0.2994	$0.3749 \pm 0.037$	0.034	$0.2229 \pm 0.037$	0.0302	$0.2982 \pm 0.055$	0.0219	$0.2141 \pm 0.049$	0.0601
16	$0.2037 \pm 0.061$	0.3185	$0.3793 \pm 0.039$	0.0343	$0.2156 \pm 0.039$	0.0303	$0.2971 \pm 0.052$	0.0219	$0.2145 \pm 0.055$	0.0603
17	$0.2096 \pm 0.061$	0.3327	$0.3816 \pm 0.035$	0.0345	$0.21 \pm 0.035$	0.0302	$0.2775 \pm 0.052$	0.0219	$0.1948 \pm 0.048$	0.0607
18	$0.208 \pm 0.067$	0.3501	$0.3777 \pm 0.037$	0.0347	$0.209 \pm 0.037$	0.0305	$0.2794 \pm 0.044$	0.0222	$0.2014 \pm 0.043$	0.0612
19	$0.2093 \pm 0.059$	0.3639	$0.3869 \pm 0.048$	0.0342	$0.2028 \pm 0.048$	0.0302	$0.2826 \pm 0.051$	0.0219	$0.2056 \pm 0.046$	0.061
20	$0.1881 \pm 0.051$	<b>0.3782</b>	$0.3784 \pm 0.035$	<b>0.0342</b>	$0.1944 \pm 0.035$	<b>0.0303</b>	$0.2614 \pm 0.043$	<b>0.0219</b>	$0.1877 \pm 0.051$	<b>0.061</b>



**Figura 6-4.:** Desempeño de cada uno de los clasificadores variando el número de objetos de representación por clase. Conjunto de clases difíciles distribuidas normalmente, con 150 objetos para la clase uno y 200 objetos para la clase 2 y dimensión del conjunto igual a 20. **Selección de las líneas de características de longitud menor.**

**Tabla 6-7.:** Intervalos de confianza para el error promedio de cada clasificador de acuerdo al número de objetos de representación por clase (Resaltado se encuentra el mejor y el peor intervalo para cada clasificador).

Intervalos de confianza para el error promedio					
Objetos de representación	<i>parzenc_dis</i>	<i>parzenc</i>	<i>ldc</i>	<i>qdc</i>	<i>svc</i>
4	[0.2728,0.3891]	[0.3365,0.4368]	<b>[0.3400,0.4232]</b>	<b>[0.4814,0.5063]</b>	<b>[0.2657,0.3705]</b>
5	[0.2700,0.4048]	[0.3212,0.4092]	[0.3142,0.4250]	[0.3951,0.4786]	[0.2842,0.3859]
6	[0.2692,0.3319]	[0.3144,0.3850]	[0.2528,0.3232]	[0.3764,0.4262]	[0.2514,0.3173]
7	[0.2827,0.3982]	[0.3376,0.4073]	[0.2266,0.3288]	[0.3392,0.4281]	[0.2306,0.3207]
8]	[0.2908,0.3613]	[0.3211,0.4240]	[0.2172,0.3043]	[0.2883,0.4034]	[0.2350,0.3087]
9]	[0.2908,0.3988]	[0.3234,0.3972]	[0.2110,0.2654]	[0.3074,0.3965]	[0.2024,0.2845]
10	[0.2642,0.3622]	[0.3145,0.3661]	[0.1925,0.2764]	[0.2888,0.3726]	[0.1818,0.2689]
11	[0.2679,0.4032]	<b>[0.3400,0.4185]</b>	[0.2001,0.2626]	[0.2992,0.3791]	[0.2078,0.2683]
12	[0.2564,0.3461]	[0.2932,0.3731]	[0.1956,0.2572]	[0.3118,0.3853]	[0.1944,0.2671]
13	[0.2764,0.3921]	[0.3257,0.3824]	[0.1572,0.2273]	[0.3218,0.3943]	[0.1751,0.2358]
14	<b>[0.2983,0.3876]</b>	[0.2952,0.3752]	[0.1718,0.2528]	[0.3110,0.3888]	[0.1774,0.2496]
15	[0.2601,0.3454]	[0.3017,0.3691]	[0.1296,0.2108]	[0.2707,0.3388]	[0.1402,0.2078]
16	[0.2562,0.3441]	[0.3109,0.3569]	[0.1828,0.2406]	[0.2722,0.3122]	[0.1672,0.2131]
17	[0.2697,0.3376]	[0.2957,0.3603]	[0.1483,0.2232]	[0.2592,0.3094]	[0.1421,0.2050]
18	[0.2905,0.3470]	[0.3016,0.3444]	[0.1645,0.2448]	[0.2791,0.3385]	[0.1539,0.2245]
19	<b>[0.2603,0.3095]</b>	<b>[0.2828,0.3387]</b>	[0.1431,0.1955]	<b>[0.2485,0.2968]</b>	<b>[0.1374,0.1937]</b>
20	[0.2917,0.3396]	[0.2973,0.3495]	<b>[0.1699,0.2283]</b>	[0.2583,0.3128]	[0.1522,0.2102]

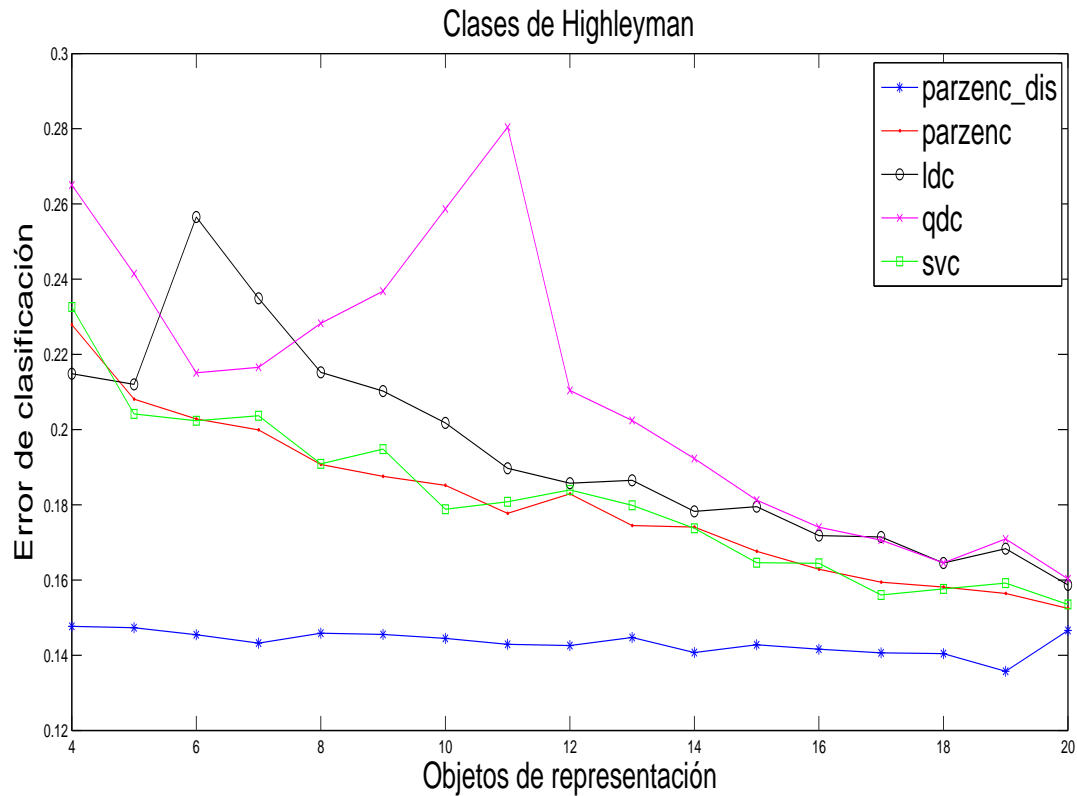
**Tabla 6-8.:** Tabla del Error promedio, desviación estándar, tiempo de cómputo, de los clasificadores de acuerdo al número de objetos de representación por clase (resaltado se encuentra el mejor tiempo promedio de cómputo de los clasificadores *parzenc\_dis*).

Error promedio, desviación estándar, tiempo de cómputo de los clasificadores										
Objetos de representación	<i>parzenc_dis</i>		<i>parzenc</i>		<i>ldc</i>		<i>qdc</i>		<i>svc</i>	
	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)
4	0.3309±0.0226	0.1305	0.3866±0.0195	0.0377	0.3816±0.0162	0.0335	0.4938±0.0048	0.0234	0.3181±0.0204	0.0618
5	0.3374±0.0262	0.3641	0.3652±0.0171	0.0353	0.3696±0.0216	0.0318	0.4368±0.0162	0.0235	0.3351±0.0198	0.0600
6	0.3006±0.0122	0.1591	0.3497±0.0138	0.0340	0.2880±0.0137	0.0305	0.4013±0.0097	0.0233	0.2843±0.0128	0.0620
7	0.3404±0.0225	0.1745	0.3724±0.0136	0.0340	0.2777±0.0199	0.0308	0.3836±0.0173	0.0238	0.2756±0.0175	0.0622
8	0.3260±0.0137	0.1907	0.3725±0.0200	0.0347	0.2608±0.0169	0.0312	0.3459±0.0224	0.0257	0.2718±0.0143	0.0660
9	0.3448±0.0210	0.2072	0.3603±0.0143	0.0349	0.2382±0.0106	0.0310	0.3520±0.0173	0.0239	0.2435±0.0160	0.0630
10	0.3132±0.0191	0.2223	0.3403±0.0100	0.0351	0.2344±0.0163	0.0309	0.3307±0.0163	0.0233	0.2254±0.0169	0.0615
11	0.3356±0.0263	0.2382	0.3792±0.0153	0.0350	0.2314±0.0122	0.0308	0.3391±0.0155	0.0232	0.2381±0.0118	0.0612
12	0.3013±0.0174	0.2533	0.3331±0.0155	0.0349	0.2264±0.0120	0.0311	0.3486±0.0143	0.0233	0.2307±0.0142	0.0632
13	0.3342±0.0225	23.315	0.3541±0.0110	0.0359	0.1922±0.0136	0.0322	0.3581±0.0141	0.0239	0.2055±0.0118	0.0639
14	<b>0.3429±0.0174</b>	<b>0.2846</b>	0.3352±0.0156	0.0350	0.2123±0.0158	0.0314	0.3499±0.0151	0.0233	0.2135±0.0140	0.0626
15	0.3028±0.0166	0.2982	0.3354±0.0131	0.0352	0.1702±0.0158	0.0315	0.3047±0.0132	0.0229	0.1740±0.0132	0.0625
16	0.3001±0.0171	0.3141	0.3339±0.0089	0.0350	0.2117±0.0112	0.0312	0.2922±0.0078	0.0229	0.1901±0.0089	0.0626
17	0.3037±0.0132	0.3310	0.3280±0.0126	0.0350	0.1857±0.0146	0.0313	0.2843±0.0098	0.0231	0.1736±0.0122	0.0634
18	0.3187±0.0110	0.3498	0.3230±0.0083	0.0371	0.2046±0.0156	0.0312	0.3088±0.0116	0.0230	0.1892±0.0137	0.0643
19	<b>0.2849±0.0096</b>	<b>0.3746</b>	0.3107±0.0109	0.0356	0.1693±0.0102	0.0317	0.2726±0.0094	0.0231	0.1656±0.0109	0.0641
20	0.3157±0.0093	0.3805	0.3234±0.0101	0.0402	0.1991±0.0114	0.0384	0.2856±0.0106	0.0282	0.1812±0.0113	0.0770



### 6.1.2. Clases de Highleyman

Mediante la rutina **gendath** del toolbox de reconocimiento de patrones de matlab (**PRTools**) [5], se genera el conjunto de clases de Highleyman (conjunto artificial), en el cual el número de objetos por clase son: 240 objetos para la clase uno y 200 para la clase 2.



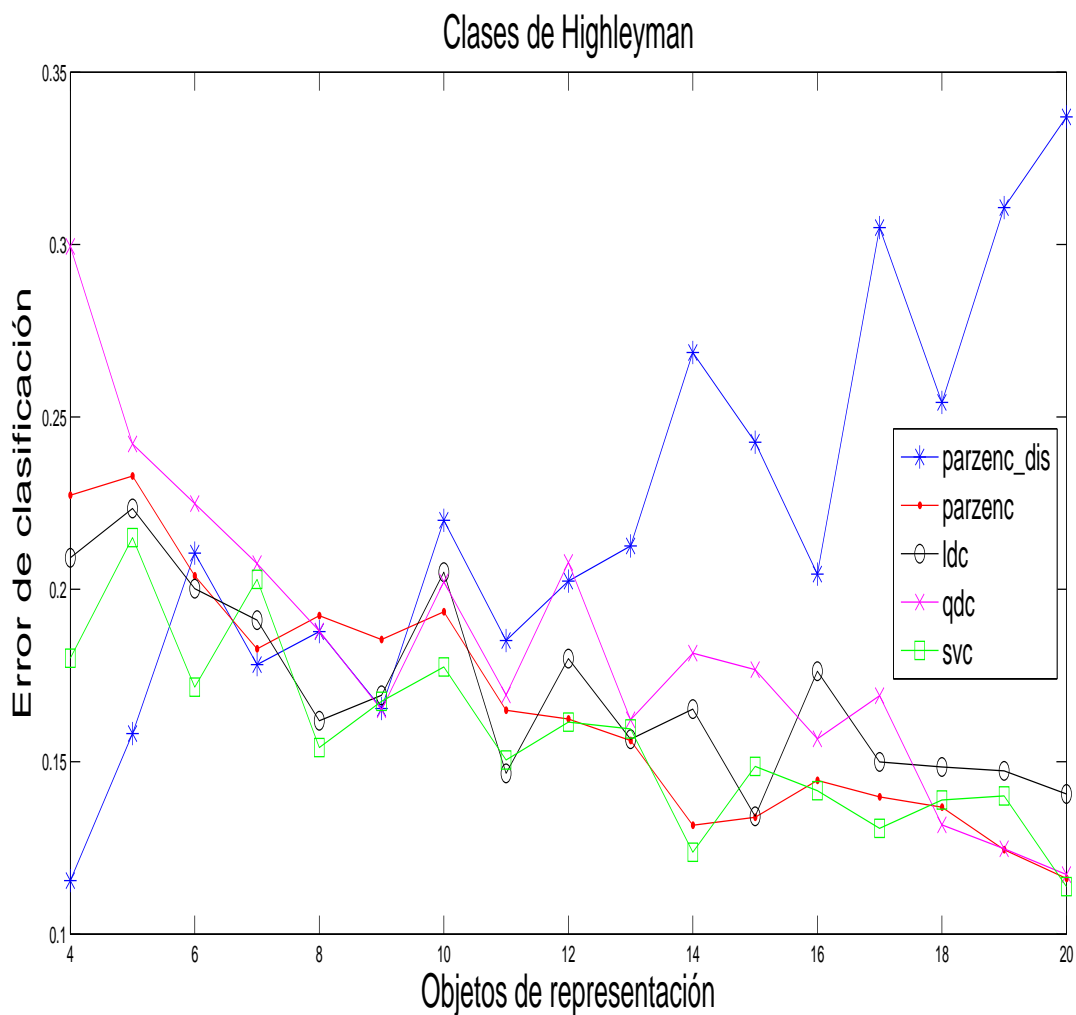
**Figura 6-5.:** Desempeño de cada uno de los clasificadores variando el número de objetos de representación por clase. Conjunto de clases de Highleyman, con 240 objetos para la clase uno y 200 objetos para la clase 2. **Selección de las líneas de características de longitud mayor.**

**Tabla 6-9.:** Intervalos de confianza para el error promedio de cada clasificador de acuerdo al número de objetos de representación por clase (Resaltado se encuentra el mejor intervalo para cada clasificador).

Intervalos de confianza para el error promedio					
Objetos de representación	<i>parzenc_dis</i>	<i>parzenc</i>	<i>ldc</i>	<i>qdc</i>	<i>svc</i>
4	[0.1380,0.1574]	[0.2116,0.2442]	[0.1966,0.2330]	[0.2450,0.2850]	[0.2126,0.2526]
5	[0.1399,0.1547]	[0.1970,0.2192]	[0.1928,0.2312]	[0.2199,0.2631]	[0.1871,0.2213]
6	[0.1383,0.1525]	[0.1904,0.2152]	[0.2370,0.2760]	[0.1954,0.2348]	[0.1887,0.2161]
7	[0.1369,0.1495]	[0.1894,0.2104]	[0.2170,0.2528]	[0.1979,0.2353]	[0.1884,0.2190]
8	[0.1388,0.1530]	[0.1810,0.2004]	[0.2007,0.2297]	[0.2101,0.2465]	[0.1751,0.2067]
9	[0.1379,0.1531]	[0.1781,0.1971]	[0.1963,0.2241]	[0.2181,0.2555]	[0.1809,0.2087]
10	[0.1371,0.1519]	[0.1757,0.1947]	[0.1892,0.2144]	[0.2334,0.2840]	[0.1659,0.1917]
11	[0.1366,0.1492]	[0.1685,0.1869]	[0.1773,0.2021]	[0.2580,0.3028]	[0.1671,0.1945]
12	[0.1358,0.1494]	[0.1750,0.1908]	[0.1746,0.1968]	[0.1933,0.2275]	[0.1702,0.1976]
13	[0.1373,0.1521]	[0.1653,0.1837]	[0.1749,0.1981]	[0.1887,0.2161]	[0.1673,0.1925]
14	[0.1349,0.1465]	[0.1654,0.1828]	[0.1679,0.1885]	[0.1799,0.2047]	[0.1620,0.1856]
15	[0.1370,0.1486]	[0.1576,0.1776]	[0.1684,0.1906]	[0.1692,0.1934]	[0.1525,0.1767]
16	[0.1350,0.1482]	[0.1547,0.1711]	[0.1629,0.1807]	[0.1628,0.1854]	[0.1532,0.1758]
17	[0.1345,0.1467]	[0.1505,0.1683]	[0.1623,0.1807]	[0.1598,0.1814]	[0.1449,0.1671]
18	[0.1330,0.1478]	[0.1497,0.1665]	[0.1556,0.1734]	[0.1551,0.1741]	[0.1469,0.1685]
19	<b>[0.1304,0.1410]</b>	[0.1482,0.1646]	[0.1588,0.1778]	[0.1612,0.1806]	[0.1476,0.1708]
20	[0.1392,0.1540]	<b>[0.1443,0.1607]</b>	<b>[0.1500,0.1674]</b>	<b>[0.1508,0.1698]</b>	<b>[0.1427,0.1643]</b>

**Tabla 6-10.:** Tabla del Error promedio, desviación estándar, tiempo de cómputo, de los clasificadores de acuerdo al número de objetos de representación por clase (resaltado se encuentra el mejor y el peor tiempo promedio de cómputo de cada clasificador).

Error promedio, desviación estándar, tiempo de cómputo de los clasificadores										
Objetos de representación	<i>parzenc_dis</i>		<i>parzenc</i>		<i>ldc</i>		<i>qdc</i>		<i>svc</i>	
	$\bar{e}_\infty \pm \sigma$	CPU(seg)	$\bar{e}_\infty \pm \sigma$	CPU(seg)	$\bar{e}_\infty \pm \sigma$	CPU(seg)	$\bar{e}_\infty \pm \sigma$	CPU(seg)	$\bar{e}_\infty \pm \sigma$	CPU(seg)
4	0.1477 $\pm$ 0.037	<b>0.131</b>	0.2279 $\pm$ 0.062	0.0336	0.2148 $\pm$ 0.069	<b>0.0301</b>	0.265 $\pm$ 0.076	0.0231	0.2326 $\pm$ 0.076	<b>0.0544</b>
5	0.1473 $\pm$ 0.028	0.1469	0.2081 $\pm$ 0.042	0.0338	0.212 $\pm$ 0.073	0.0307	0.2415 $\pm$ 0.082	0.0238	0.2042 $\pm$ 0.065	0.0552
6	0.1454 $\pm$ 0.027	0.1616	0.2028 $\pm$ 0.047	0.0335	0.2565 $\pm$ 0.074	0.0304	0.2151 $\pm$ 0.075	0.0236	0.2024 $\pm$ 0.052	0.0566
7	0.1432 $\pm$ 0.024	0.1742	0.1999 $\pm$ 0.040	0.0334	0.2349 $\pm$ 0.068	0.0301	0.2166 $\pm$ 0.071	0.0235	0.2037 $\pm$ 0.058	0.0585
8	0.1459 $\pm$ 0.027	0.1903	0.1907 $\pm$ 0.037	<b>0.0332</b>	0.2152 $\pm$ 0.055	0.0304	0.2283 $\pm$ 0.069	0.024	0.1909 $\pm$ 0.060	0.06
9	0.1455 $\pm$ 0.029	0.2044	0.1876 $\pm$ 0.036	0.0336	0.2102 $\pm$ 0.053	0.0302	0.2368 $\pm$ 0.071	0.0237	0.1948 $\pm$ 0.053	0.0587
10	0.1445 $\pm$ 0.028	0.2191	0.1852 $\pm$ 0.036	0.0338	0.2018 $\pm$ 0.048	0.0303	0.2587 $\pm$ 0.096	0.023	0.1788 $\pm$ 0.049	0.0586
11	0.1429 $\pm$ 0.024	0.2341	0.1777 $\pm$ 0.035	0.0341	0.1897 $\pm$ 0.047	0.0303	0.2804 $\pm$ 0.085	0.0221	0.1808 $\pm$ 0.052	0.0586
12	0.1426 $\pm$ 0.026	0.2647	0.1829 $\pm$ 0.030	0.0345	0.1857 $\pm$ 0.042	0.0307	0.2104 $\pm$ 0.065	0.0223	0.1839 $\pm$ 0.052	0.0592
13	0.1447 $\pm$ 0.028	0.2644	0.1745 $\pm$ 0.035	0.0344	0.1865 $\pm$ 0.044	0.0304	0.2024 $\pm$ 0.052	0.0221	0.1799 $\pm$ 0.048	0.0594
14	0.1407 $\pm$ 0.022	0.281	0.1741 $\pm$ 0.033	0.0343	0.1782 $\pm$ 0.039	0.0305	0.1923 $\pm$ 0.047	<b>0.0222</b>	0.1738 $\pm$ 0.045	0.0596
15	0.1428 $\pm$ 0.022	0.2944	0.1676 $\pm$ 0.038	0.0348	0.1795 $\pm$ 0.042	0.0304	0.1813 $\pm$ 0.046	<b>0.022</b>	0.1646 $\pm$ 0.046	0.0594
16	0.1416 $\pm$ 0.025	0.3098	0.1629 $\pm$ 0.031	0.0347	0.1718 $\pm$ 0.034	0.0304	0.1741 $\pm$ 0.043	<b>0.022</b>	0.1645 $\pm$ 0.043	0.0598
17	0.1406 $\pm$ 0.023	0.3248	0.1594 $\pm$ 0.034	0.0348	0.1715 $\pm$ 0.035	0.0304	0.1706 $\pm$ 0.041	<b>0.022</b>	0.156 $\pm$ 0.042	0.0599
18	0.1404 $\pm$ 0.028	0.3377	0.1581 $\pm$ 0.032	0.0348	0.1645 $\pm$ 0.034	0.0303	0.1646 $\pm$ 0.036	<b>0.022</b>	0.1577 $\pm$ 0.041	0.0601
19	0.1357 $\pm$ 0.020	<b>0.3524</b>	0.1564 $\pm$ 0.031	0.035	0.1683 $\pm$ 0.036	0.0305	0.1709 $\pm$ 0.037	0.0221	0.1592 $\pm$ 0.044	0.0606
20	0.1466 $\pm$ 0.028	0.3678	0.1525 $\pm$ 0.031	0.0352	0.1587 $\pm$ 0.033	0.0303	0.1603 $\pm$ 0.036	<b>0.022</b>	0.1535 $\pm$ 0.041	0.0607



**Figura 6-6.:** Desempeño de cada uno de los clasificadores variando el número de objetos de representación por clase. Conjunto de clases de Highleyman, con 240 objetos para la clase uno y 200 objetos para la clase 2. **Selección de las líneas de características de longitud menor.**

**Tabla 6-11.:** Intervalos de confianza para el error promedio de cada clasificador de acuerdo al número de objetos de representación por clase (Resaltado se encuentra el mejor y el peor intervalo para cada clasificador).

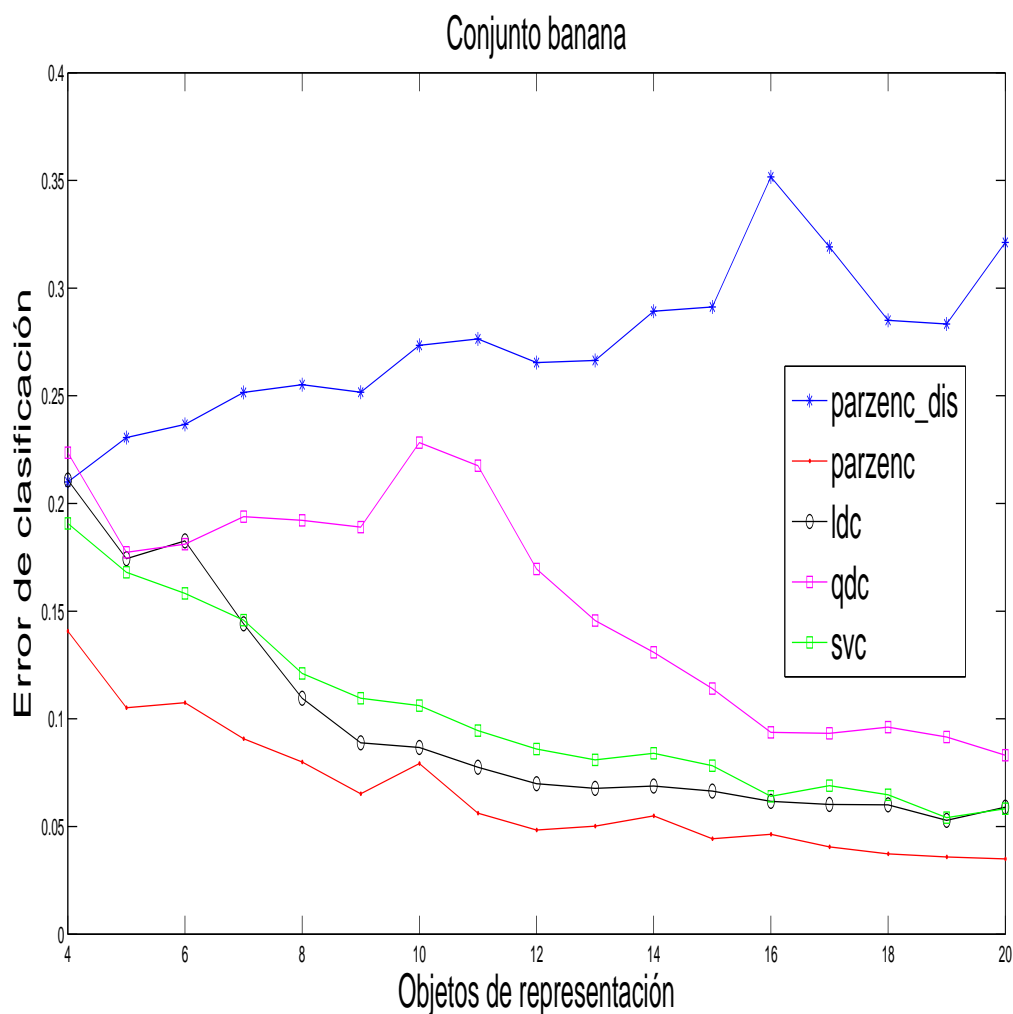
Intervalos de confianza para el error promedio					
Objetos de representación	<i>parzenc_dis</i>	<i>parzenc</i>	<i>ldc</i>	<i>qdc</i>	<i>svc</i>
4	<b>[0.0876,0.1434]</b>	[0.1451,0.3094]	[0.1312,0.2871]	<b>[0.2153,0.3836]</b>	[0.1142,0.2458]
5	[0.1143,0.2020]	<b>[0.1829,0.2827]</b>	[0.1617,0.2852]	[0.2023,0.2820]	<b>[0.1490,0.2810]</b>
6	[0.1112,0.3099]	[0.1641,0.2437]	[0.1482,0.2522]	[0.1875,0.2622]	[0.1249,0.2184]
7	[0.1338,0.2226]	[0.1470,0.2184]	[0.1437,0.2385]	[0.1736,0.2413]	[0.1445,0.2613]
8	[0.1461,0.2293]	[0.1479,0.2368]	[0.1258,0.1980]	[0.1285,0.2473]	[0.1118,0.1964]
9	[0.1237,0.2071]	[0.1441,0.2268]	[0.1214,0.2171]	[0.1087,0.2215]	[0.1172,0.2178]
10	[0.1641,0.2759]	[0.1608,0.2261]	<b>[0.1655,0.2445]</b>	[0.1706,0.2338]	[0.1295,0.2256]
11	[0.1442,0.2261]	[0.1215,0.2083]	[0.1106,0.1827]	[0.0989,0.2399]	[0.1103,0.1907]
12	[0.1327,0.2722]	[0.1291,0.1956]	[0.1354,0.2243]	[0.1628,0.2527]	[0.1000,0.2230]
13	[0.1388,0.2863]	[0.1211,0.1911]	[0.1100,0.2030]	[0.1117,0.2126]	[0.1075,0.2115]
14	[0.1685,0.3689]	[0.1112,0.1520]	[0.1326,0.1979]	[0.1324,0.2306]	<b>[0.0961,0.1514]</b>
15	[0.1901,0.2953]	[0.0944,0.1733]	[0.1062,0.1621]	[0.1072,0.2461]	[0.1060,0.1913]
16	[0.1625,0.2463]	[0.1135,0.1757]	[0.1321,0.2204]	[0.1118,0.2015]	[0.1085,0.1747]
17	[0.2007,0.4092]	[0.1040,0.1756]	[0.1097,0.1901]	[0.1071,0.2312]	[0.1083,0.1531]
18	[0.1756,0.3328]	[0.0885,0.1851]	<b>[0.1042,0.1927]</b>	[0.0974,0.1659]	[0.1039,0.1739]
19	[0.2291,0.3923]	[0.0975,0.1515]	[0.1114,0.1833]	[0.0917,0.1578]	[0.0977,0.1825]
20	<b>[0.2128,0.4612]</b>	<b>[0.0878,0.1444]</b>	[0.1176,0.1637]	<b>[0.0891,0.1454]</b>	[0.1028,0.1247]

**Tabla 6-12.:** Tabla del Error promedio, desviación estándar, tiempo de cómputo, de los clasificadores de acuerdo al número de objetos de representación por clase (resaltado se encuentra el mejor y el peor tiempo promedio de cómputo para el clasificador *parzenc\_dis*).

Error promedio, desviación estándar, tiempo de cómputo de los clasificadores										
Objetos de representación	<i>parzenc_dis</i>		<i>parzenc</i>		<i>ldc</i>		<i>qdc</i>		<i>svc</i>	
	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)
4	<b>0.1155±0.0109</b>	<b>0.1407</b>	0.2273±0.0320	0.0433	0.2092±0.0303	0.0374	0.2994±0.0327	0.0248	0.1800±0.0256	0.0658
5	0.1581±0.0171	0.1479	0.2328±0.0194	0.0351	0.2235±0.0240	0.0317	0.2421±0.0155	0.0234	0.2150±0.0257	0.0568
6	0.2105±0.0387	0.1620	0.2039±0.0155	0.0339	0.2002±0.0202	0.0308	0.2249±0.0145	0.0235	0.1716±0.0182	0.0580
7	0.1782±0.0173	0.1849	0.1827±0.0139	0.0345	0.1911±0.0184	0.0309	0.2074±0.0132	0.0236	0.2029±0.0227	0.0608
8	0.1877±0.0162	0.1923	0.1923±0.0173	0.0346	0.1619±0.0140	0.0312	0.1879±0.0231	0.0246	0.1541±0.0165	0.0619
9	0.1654±0.0162	0.2080	0.1855±0.0161	0.0345	0.1693±0.0186	0.0311	0.1651±0.0219	0.0248	0.1675±0.0196	0.0622
10	0.2200±0.0217	0.2229	0.1935±0.0127	0.0352	0.2050±0.0154	0.0313	0.2022±0.0123	0.0240	0.1775±0.0187	0.0611
11	0.1852±0.0159	0.2380	0.1649±0.0169	0.0350	0.1466±0.0140	0.0310	0.1694±0.0274	0.0236	0.1505±0.0156	0.0620
12	0.2024±0.0271	0.2526	0.1624±0.0129	0.0374	0.1799±0.0173	0.0347	0.2078±0.0175	0.0259	0.1615±0.0239	0.0675
13	0.2126±0.0287	0.2677	0.1561±0.0136	0.0351	0.1565±0.0181	0.0311	0.1621±0.0196	0.0235	0.1595±0.0202	0.0618
14	0.2687±0.0390	0.2855	0.1316±0.0079	0.0354	0.1652±0.0127	0.0316	0.1815±0.0191	0.0243	0.1238±0.0107	0.0635
15	0.2427±0.0205	0.3023	0.1339±0.0153	0.0477	0.1342±0.0109	0.0371	0.1767±0.0270	0.0238	0.1486±0.0166	0.0652
16	0.2044±0.0163	0.3139	0.1446±0.0121	0.0366	0.1762±0.0172	0.0333	0.1566±0.0175	0.0242	0.1416±0.0129	0.0655
17	0.3049±0.0406	0.3293	0.1398±0.0139	0.0387	0.1499±0.0156	0.0368	0.1691±0.0242	0.0267	0.1307±0.0087	0.0695
18	0.2542±0.0306	0.3445	0.1368±0.0188	0.0359	0.1485±0.0172	0.0312	0.1317±0.0133	0.0227	0.1389±0.0136	0.0625
19	0.3107±0.0317	0.3628	0.1245±0.0105	0.0363	0.1473±0.0140	0.0311	0.1248±0.0129	0.0227	0.1401±0.0165	0.0626
20	<b>0.3370±0.0483</b>	<b>0.3769</b>	0.1161±0.0110	0.0377	0.1407±0.0090	0.0355	0.1172±0.0110	0.0253	0.1138±0.0043	0.0711

### 6.1.3. Conjunto banana

Mediante la rutina **gendatb** del toolbox de reconocimiento de patrones de matlab (**PRTools**) [5], se genera el conjunto banana (conjunto artificial), en el cual el número de objetos por clase son: 1000 objetos para la clase uno y 1000 para la clase 2.



**Figura 6-7.:** Desempeño de cada uno de los clasificadores variando el número de objetos de representación por clase. Conjunto datos banana, con 1000 objetos para la clase uno y 1000 objetos para la clase 2. **Selección de las líneas de características de longitud mayor.**

**Tabla 6-13.:** Intervalos de confianza para el error promedio de cada clasificador de acuerdo al número de objetos de representación por clase (Resaltado se encuentra el mejor intervalo para cada clasificador).

Intervalos de confianza para el error promedio					
Objetos de representación	<i>parzenc_dis</i>	<i>parzenc</i>	<i>ldc</i>	<i>qdc</i>	<i>svc</i>
4	<b>[0.1920,0.2278]</b>	[0.1223,0.1591]	[0.1954,0.2264]	[0.2004,0.2468]	[0.1717,0.2097]
5	[0.2061,0.2551]	[0.0881,0.1223]	[0.1533,0.1955]	[0.1526,0.2020]	[0.1510,0.1852]
6	[0.2146,0.2588]	[0.0909,0.1241]	[0.1529,0.2123]	[0.1561,0.2061]	[0.1387,0.1777]
7	[0.2236,0.2794]	[0.0762,0.1052]	[0.1217,0.1665]	[0.1689,0.2189]	[0.1267,0.1651]
8	[0.2268,0.2836]	[0.0679,0.0921]	[0.0956,0.1234]	[0.1674,0.2168]	[0.1061,0.1361]
9	[0.2246,0.2788]	[0.0544,0.0760]	[0.0746,0.1030]	[0.1648,0.2132]	[0.0948,0.1242]
10	[0.2434,0.3034]	[0.0648,0.0938]	[0.0754,0.0980]	[0.2016,0.2548]	[0.0937,0.1185]
11	[0.2453,0.3075]	[0.0475,0.0649]	[0.0664,0.0886]	[0.1920,0.2430]	[0.0817,0.1075]
12	[0.2342,0.2968]	[0.0410,0.0558]	[0.0593,0.0803]	[0.1469,0.1921]	[0.0744,0.0976]
13	[0.2332,0.2996]	[0.0427,0.0575]	[0.0593,0.0761]	[0.1280,0.1632]	[0.0701,0.0917]
14	[0.2517,0.3269]	[0.0450,0.0650]	[0.0588,0.0788]	[0.1159,0.1459]	[0.0714,0.0966]
15	[0.2536,0.3288]	[0.0377,0.0509]	[0.0570,0.0760]	[0.1023,0.1259]	[0.0678,0.0888]
16	[0.3058,0.3974]	[0.0396,0.0532]	[0.0528,0.0706]	[0.0811,0.1063]	[0.0561,0.0719]
17	[0.2792,0.3592]	[0.0352,0.0458]	[0.0539,0.0665]	[0.0838,0.1028]	[0.0597,0.0781]
18	[0.2518,0.3182]	[0.0328,0.0418]	[0.0538,0.0664]	[0.0864,0.1058]	[0.0564,0.0732]
19	[0.2467,0.3199]	[0.0309,0.0409]	<b>[0.0476,0.0582]</b>	[0.0824,0.1008]	<b>[0.0481,0.0603]</b>
20	[0.2825,0.3599]	<b>[0.0307,0.0391]</b>	[0.0527,0.0653]	<b>[0.0739,0.0923]</b>	[0.0508,0.0656]

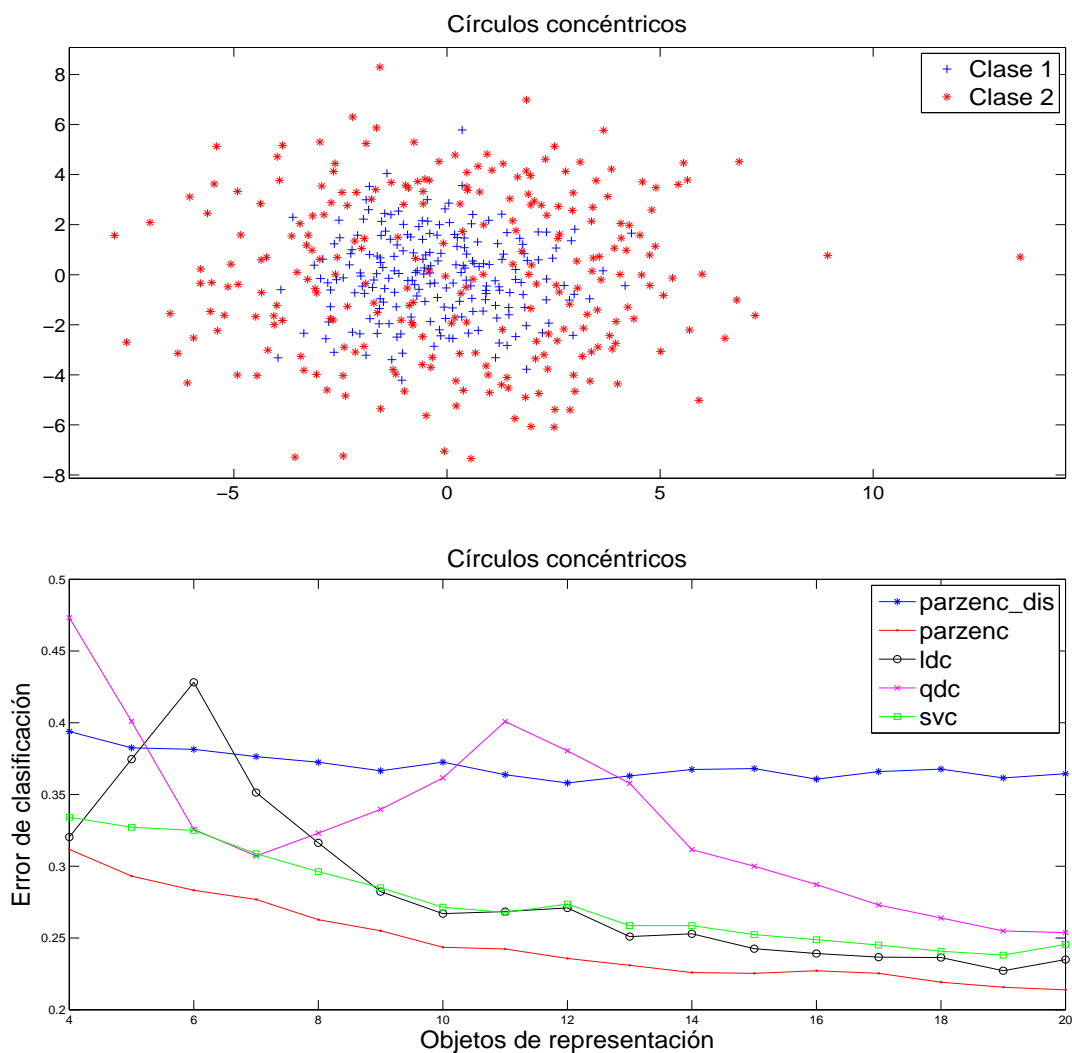


**Tabla 6-14.:** Tabla del Error promedio, desviación estándar, tiempo de cómputo, de los clasificadores de acuerdo al número de objetos de representación por clase (resaltado se encuentra el mejor tiempo promedio de cómputo de cada clasificador).

Error promedio, desviación estándar, tiempo de cómputo de los clasificadores										
Objetos de representación	<i>parzenc_dis</i>		<i>parzenc</i>		<i>ldc</i>		<i>qdc</i>		<i>svc</i>	
	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)
4	0.2099 $\pm$ 0.068	<b>0.1553</b>	0.1407 $\pm$ 0.070	0.0386	0.2109 $\pm$ 0.059	0.0373	0.2236 $\pm$ 0.088	0.0311	0.1907 $\pm$ 0.072	0.0679
5	0.2306 $\pm$ 0.093	0.1671	0.1052 $\pm$ 0.065	<b>0.0375</b>	0.1744 $\pm$ 0.080	0.0357	0.1773 $\pm$ 0.094	0.0310	0.1681 $\pm$ 0.065	0.0656
6	0.2367 $\pm$ 0.084	0.1801	0.1075 $\pm$ 0.063	0.0409	0.1826 $\pm$ 0.113	0.0373	0.1811 $\pm$ 0.095	0.0315	0.1582 $\pm$ 0.074	0.0672
7	0.2515 $\pm$ 0.106	0.1942	0.0907 $\pm$ 0.055	0.0457	0.1441 $\pm$ 0.085	0.0416	0.1939 $\pm$ 0.095	0.0359	0.1459 $\pm$ 0.073	0.0756
8	0.2552 $\pm$ 0.108	0.2084	0.0800 $\pm$ 0.046	0.0405	0.1095 $\pm$ 0.053	0.0372	0.1921 $\pm$ 0.094	0.0358	0.1211 $\pm$ 0.057	0.0704
9	0.2517 $\pm$ 0.103	0.2229	0.0652 $\pm$ 0.041	0.0375	0.0888 $\pm$ 0.054	<b>0.0319</b>	0.1890 $\pm$ 0.092	0.0321	0.1095 $\pm$ 0.056	<b>0.0616</b>
10	0.2734 $\pm$ 0.114	0.2424	0.0793 $\pm$ 0.055	0.0379	0.0867 $\pm$ 0.043	0.0320	0.2282 $\pm$ 0.101	0.0302	0.1061 $\pm$ 0.047	0.0624
11	0.2764 $\pm$ 0.118	0.2570	0.0562 $\pm$ 0.033	0.0384	0.0775 $\pm$ 0.042	0.0322	0.2175 $\pm$ 0.097	0.0253	0.0946 $\pm$ 0.049	0.0626
12	0.2655 $\pm$ 0.119	0.2720	0.0484 $\pm$ 0.028	0.0390	0.0698 $\pm$ 0.040	0.0324	0.1695 $\pm$ 0.086	0.0239	0.0860 $\pm$ 0.044	0.0636
13	0.2664 $\pm$ 0.126	0.2881	0.0501 $\pm$ 0.028	0.0391	0.0677 $\pm$ 0.032	0.0324	0.1456 $\pm$ 0.067	0.0238	0.0809 $\pm$ 0.041	0.0634
14	0.2893 $\pm$ 0.143	0.3057	0.0550 $\pm$ 0.038	0.0396	0.0688 $\pm$ 0.038	0.0327	0.1309 $\pm$ 0.057	0.0239	0.0840 $\pm$ 0.048	0.0646
15	0.2912 $\pm$ 0.143	0.3179	0.0443 $\pm$ 0.025	0.0409	0.0665 $\pm$ 0.036	0.0340	0.1141 $\pm$ 0.045	0.0249	0.0783 $\pm$ 0.040	0.0655
16	0.3516 $\pm$ 0.174	0.3340	0.0464 $\pm$ 0.026	0.0402	0.0617 $\pm$ 0.034	0.0324	0.0937 $\pm$ 0.048	0.0236	0.0640 $\pm$ 0.030	0.0634
17	0.3192 $\pm$ 0.152	0.3487	0.0405 $\pm$ 0.020	0.0433	0.0602 $\pm$ 0.024	0.0378	0.0933 $\pm$ 0.036	0.0278	0.0689 $\pm$ 0.035	0.0733
18	0.2850 $\pm$ 0.126	0.3622	0.0373 $\pm$ 0.017	0.0399	0.0601 $\pm$ 0.024	0.0323	0.0961 $\pm$ 0.037	0.0237	0.0648 $\pm$ 0.032	0.0643
19	0.2833 $\pm$ 0.139	0.3749	0.0359 $\pm$ 0.019	0.0403	0.0529 $\pm$ 0.020	0.0322	0.0916 $\pm$ 0.035	<b>0.0235</b>	0.0542 $\pm$ 0.023	0.0638
20	0.3212 $\pm$ 0.147	0.3904	0.0349 $\pm$ 0.016	0.0407	0.0590 $\pm$ 0.024	0.0328	0.0831 $\pm$ 0.035	0.0240	0.0582 $\pm$ 0.028	0.0656

### 6.1.4. Conjunto Círculos concéntricos

Mediante la rutina **gendatc** del toolbox de reconocimiento de patrones de matlab (**PRTools**) [5], se genera el conjunto círculos concéntricos (conjunto artificial), en el cual el número de objetos por clase son: 200 objetos para la clase uno y 250 para la clase 2. La dimensión del conjunto generado es 8, con el fin de mostrar el desempeño de los clasificadores con una base de datos de dimensión diferente a dos.



**Figura 6-8.:** Grafico de dispersión del conjunto mediante escalamiento multidimensional y desempeño de cada uno de los clasificadores variando el número de objetos de representación por clase. Conjunto de datos círculos concéntricos. **Selección de las líneas de características de longitud mayor.**

**Tabla 6-15.:** Intervalos de confianza para el error promedio de cada clasificador de acuerdo al número de objetos de representación por clase (Resaltado se encuentra el mejor intervalo para cada clasificador).

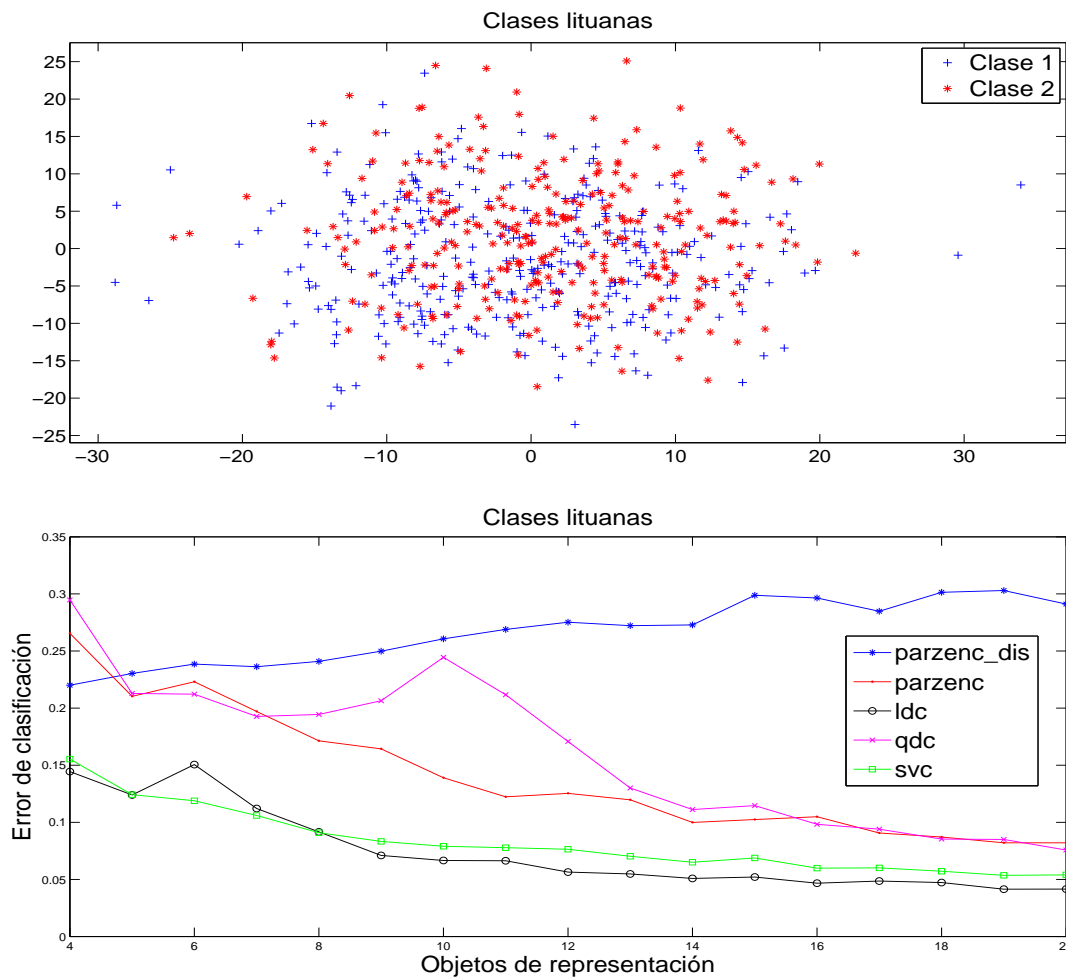
Intervalos de confianza para el error promedio					
Objetos de representación	<i>parzenc_dis</i>	<i>parzenc</i>	<i>ldc</i>	<i>qdc</i>	<i>svc</i>
4	[0.3828,0.4050]	[0.2959,0.3275]	[0.3025,0.3383]	[0.4598,0.4866]	[0.3100,0.3584]
5	[0.3714,0.3936]	[0.2784,0.3078]	[0.3397,0.4097]	[0.3776,0.4244]	[0.3064,0.3480]
6	[0.3709,0.3919]	[0.2696,0.2970]	[0.3916,0.4648]	[0.3041,0.3473]	[0.3032,0.3468]
7	[0.3656,0.3872]	[0.2631,0.2905]	[0.3198,0.3830]	[0.2900,0.3242]	[0.2900,0.3274]
8	[0.3625,0.3825]	[0.2488,0.2766]	[0.2884,0.3442]	[0.3039,0.3423]	[0.2767,0.3157]
9	[0.3560,0.3770]	[0.2430,0.2672]	[0.2598,0.3050]	[0.3226,0.3568]	[0.2721,0.2979]
10	[0.3608,0.3844]	[0.2324,0.2546]	[0.2459,0.2881]	[0.3428,0.3802]	[0.2562,0.2868]
11	[0.3517,0.3759]	[0.2331,0.2515]	[0.2506,0.2864]	[0.3808,0.4208]	[0.2557,0.2799]
12	<b>[0.3464,0.3696]</b>	[0.2260,0.2454]	[0.2499,0.2921]	[0.3604,0.4004]	[0.2594,0.2878]
13	[0.3524,0.3734]	[0.2215,0.2405]	[0.2381,0.2639]	[0.3394,0.3762]	[0.2481,0.2691]
14	[0.3556,0.3792]	[0.2191,0.2327]	[0.2397,0.2661]	[0.2963,0.3269]	[0.2483,0.2689]
15	[0.3573,0.3789]	[0.2180,0.2328]	[0.2333,0.2517]	[0.2879,0.3121]	[0.2419,0.2629]
16	[0.3504,0.3710]	[0.2201,0.2343]	[0.2284,0.2500]	[0.2751,0.2993]	[0.2393,0.2583]
17	[0.3534,0.3786]	[0.2178,0.2330]	[0.2288,0.2446]	[0.2614,0.2846]	[0.2364,0.2538]
18	[0.3564,0.3790]	[0.2124,0.2260]	[0.2282,0.2446]	[0.2521,0.2757]	[0.2308,0.2508]
19	[0.3508,0.3724]	[0.2094,0.2220]	<b>[0.2222,0.2322]</b>	[0.2444,0.2654]	<b>[0.2297,0.2465]</b>
20	[0.3529,0.3761]	<b>[0.2075,0.2201]</b>	[0.2289,0.2411]	<b>[0.2433,0.2643]</b>	[0.2355,0.2555]

**Tabla 6-16.:** Tabla del Error promedio, desviación estándar, tiempo de cómputo, de los clasificadores de acuerdo al número de objetos de representación por clase (resaltado se encuentra el mejor tiempo promedio de cómputo de cada clasificador).

Error promedio, desviación estándar, tiempo de cómputo de los clasificadores										
Objetos de representación	<i>parzenc_dis</i>		<i>parzenc</i>		<i>ldc</i>		<i>qdc</i>		<i>svc</i>	
	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)
4	$0.3939 \pm 0.042$	<b>0.1382</b>	$0.3117 \pm 0.060$	0.0363	$0.3204 \pm 0.068$	0.0325	$0.4732 \pm 0.051$	0.0236	$0.3342 \pm 0.092$	<b>0.0573</b>
5	$0.3825 \pm 0.042$	0.1543	$0.2931 \pm 0.056$	0.0349	$0.3747 \pm 0.133$	0.0314	$0.401 \pm 0.089$	0.0249	$0.3272 \pm 0.079$	0.0577
6	$0.3814 \pm 0.040$	0.166	$0.2833 \pm 0.052$	0.0342	$0.4282 \pm 0.139$	0.0312	$0.3257 \pm 0.082$	0.0243	$0.325 \pm 0.083$	0.0609
7	$0.3764 \pm 0.041$	0.1786	$0.2768 \pm 0.052$	<b>0.0342</b>	$0.3514 \pm 0.120$	0.0317	$0.3071 \pm 0.065$	0.0245	$0.3087 \pm 0.071$	0.0629
8	$0.3725 \pm 0.038$	0.1938	$0.2627 \pm 0.053$	0.0349	$0.3163 \pm 0.106$	0.034	$0.3231 \pm 0.073$	0.0282	$0.2962 \pm 0.074$	0.0722
9	$0.3665 \pm 0.040$	0.2104	$0.2551 \pm 0.046$	0.0355	$0.2824 \pm 0.086$	0.0351	$0.3397 \pm 0.065$	0.0292	$0.285 \pm 0.049$	0.0695
10	$0.3726 \pm 0.045$	0.2234	$0.2435 \pm 0.042$	<b>0.0342</b>	$0.267 \pm 0.080$	0.0307	$0.3615 \pm 0.071$	0.023	$0.2715 \pm 0.058$	0.0596
11	$0.3638 \pm 0.046$	0.2387	$0.2423 \pm 0.035$	0.0345	$0.2685 \pm 0.068$	0.0308	$0.4008 \pm 0.076$	0.0227	$0.2678 \pm 0.046$	0.06
12	$0.358 \pm 0.044$	0.255	$0.2357 \pm 0.037$	0.0345	$0.271 \pm 0.080$	0.031	$0.3804 \pm 0.076$	0.0226	$0.2736 \pm 0.054$	0.0602
13	$0.3629 \pm 0.040$	0.2697	$0.231 \pm 0.036$	0.0343	$0.251 \pm 0.049$	<b>0.0308</b>	$0.3578 \pm 0.070$	0.0225	$0.2586 \pm 0.040$	0.0602
14	$0.3674 \pm 0.045$	0.2856	$0.2259 \pm 0.026$	0.0348	$0.2529 \pm 0.050$	0.0312	$0.3116 \pm 0.058$	0.0228	$0.2586 \pm 0.039$	0.0613
15	$0.3681 \pm 0.041$	0.3012	$0.2254 \pm 0.028$	0.0349	$0.2425 \pm 0.035$	0.0313	$0.3 \pm 0.046$	0.0228	$0.2524 \pm 0.040$	0.0613
16	$0.3607 \pm 0.039$	0.3153	$0.2272 \pm 0.027$	0.0349	$0.2392 \pm 0.041$	0.0311	$0.2872 \pm 0.046$	0.0227	$0.2488 \pm 0.036$	0.0616
17	$0.366 \pm 0.048$	0.3311	$0.2254 \pm 0.029$	0.0351	$0.2367 \pm 0.030$	0.0313	$0.273 \pm 0.044$	0.0229	$0.2451 \pm 0.033$	0.0622
18	$0.3677 \pm 0.043$	0.3463	$0.2192 \pm 0.026$	0.0348	$0.2364 \pm 0.031$	0.0309	$0.2639 \pm 0.045$	<b>0.0226</b>	$0.2408 \pm 0.038$	0.0615
19	$0.3616 \pm 0.041$	0.3611	$0.2157 \pm 0.024$	0.036	$0.2272 \pm 0.019$	0.0319	$0.2549 \pm 0.040$	0.0233	$0.2381 \pm 0.032$	0.0639
20	$0.3645 \pm 0.044$	0.3755	$0.2138 \pm 0.024$	0.0356	$0.235 \pm 0.023$	0.0316	$0.2538 \pm 0.040$	0.0231	$0.2455 \pm 0.038$	0.0636

### 6.1.5. Conjunto clases lituanas

Mediante la rutina **gendatl** del toolbox de reconocimiento de patrones de matlab (**PRTools**) [5], se genera el conjunto de clases difíciles distribuidas normalmente (conjunto artificial), en el cual el número de objetos por clase son: 340 objetos para la clase uno y 300 para la clase 2.



**Figura 6-9.:** Gráfico de dispersión del conjunto de datos proyectado en 2D mediante escalamiento multidimensional y Desempeño de cada uno de los clasificadores variando el número de objetos de representación por clase. Conjunto de datos clases lituanas. **Selección de las líneas de características de longitud mayor.**

**Tabla 6-17.:** Intervalos de confianza para el error promedio de cada clasificador de acuerdo al número de objetos de representación por clase (Resaltado se encuentra el mejor intervalo para cada clasificador).

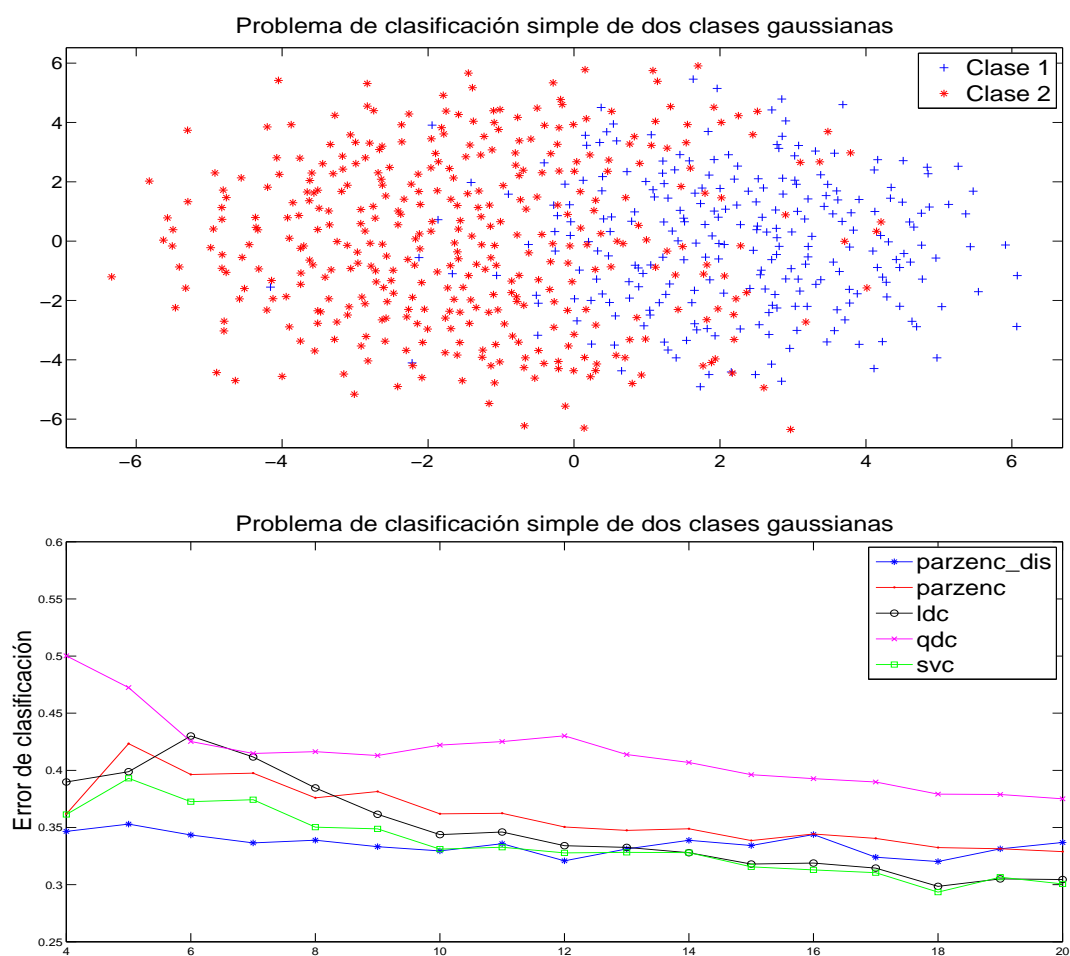
Intervalos de confianza para el error promedio					
Objetos de representación	<i>parzenc_dis</i>	<i>parzenc</i>	<i>ldc</i>	<i>qdc</i>	<i>svc</i>
4	<b>[0.2021,0.2379]</b>	[0.2439,0.2871]	[0.1226,0.1662]	[0.2647,0.3247]	[0.1314,0.1794]
5	[0.2075,0.2533]	[0.1888,0.2320]	[0.1016,0.1464]	[0.1868,0.2390]	[0.1031,0.1453]
6	[0.2160,0.2612]	[0.1994,0.2468]	[0.1244,0.1766]	[0.1854,0.2390]	[0.0984,0.1394]
7	[0.2197,0.2529]	[0.1760,0.2182]	[0.0937,0.1305]	[0.1686,0.2170]	[0.0912,0.1206]
8	[0.2201,0.2617]	[0.1514,0.1914]	[0.0780,0.1054]	[0.1711,0.2179]	[0.0780,0.1038]
9	[0.2291,0.2707]	[0.1459,0.1827]	[0.0584,0.0836]	[0.1799,0.2331]	[0.0722,0.0944]
10	[0.2388,0.2824]	[0.1219,0.1561]	[0.0566,0.0766]	[0.2163,0.2727]	[0.0678,0.0904]
11	[0.2444,0.2934]	[0.1091,0.1355]	[0.0574,0.0752]	[0.1822,0.2412]	[0.0670,0.0886]
12	[0.2560,0.2944]	[0.1108,0.1398]	[0.0494,0.0636]	[0.1499,0.1915]	[0.0656,0.0872]
13	[0.2551,0.2893]	[0.1034,0.1360]	[0.0480,0.0616]	[0.1118,0.1482]	[0.0608,0.0798]
14	[0.2555,0.2903]	[0.0883,0.1115]	[0.0459,0.0559]	[0.0983,0.1241]	[0.0568,0.0732]
15	[0.2735,0.3241]	[0.0909,0.1141]	[0.0460,0.0582]	[0.1013,0.1281]	[0.0599,0.0777]
16	[0.2724,0.3204]	[0.0912,0.1186]	[0.0420,0.0514]	[0.0850,0.1114]	[0.0533,0.0665]
17	[0.2640,0.3056]	[0.0817,0.0995]	[0.0428,0.0544]	[0.0824,0.1056]	[0.0533,0.0669]
18	[0.2843,0.3185]	[0.0763,0.0979]	[0.0426,0.0520]	[0.0746,0.0962]	[0.0511,0.0633]
19	[0.2855,0.3203]	[0.0750,0.0892]	[0.0375,0.0453]	[0.0753,0.0947]	<b>[0.0474,0.0596]</b>
20	[0.2721,0.3101]	<b>[0.0729,0.0913]</b>	<b>[0.0370,0.0460]</b>	<b>[0.0668,0.0846]</b>	[0.0477,0.0603]

**Tabla 6-18.:** Tabla del Error promedio, desviación estándar, tiempo de cómputo, de los clasificadores de acuerdo al número de objetos de representación por clase (resaltado se encuentra el mejor tiempo promedio de cómputo de cada clasificador).

Error promedio, desviación estándar, tiempo de cómputo de los clasificadores										
Objetos de representación	<i>parzenc_dis</i>		<i>parzenc</i>		<i>ldc</i>		<i>qdc</i>		<i>svc</i>	
	$\bar{e}_\infty \pm \sigma$	CPU(seg)	$\bar{e}_\infty \pm \sigma$	CPU(seg)	$\bar{e}_\infty \pm \sigma$	CPU(seg)	$\bar{e}_\infty \pm \sigma$	CPU(seg)	$\bar{e}_\infty \pm \sigma$	CPU(seg)
4	0.22 $\pm$ 0.068	<b>0.1333</b>	0.2655 $\pm$ 0.082	0.0342	0.1444 $\pm$ 0.083	0.0316	0.2947 $\pm$ 0.114	0.024	0.1554 $\pm$ 0.091	0.0572
5	0.2304 $\pm$ 0.087	0.1476	0.2104 $\pm$ 0.082	<b>0.0341</b>	0.124 $\pm$ 0.085	0.031	0.2129 $\pm$ 0.099	0.0233	0.1242 $\pm$ 0.080	<b>0.0561</b>
6	0.2386 $\pm$ 0.086	0.1628	0.2231 $\pm$ 0.090	0.0351	0.1505 $\pm$ 0.099	0.0312	0.2122 $\pm$ 0.102	0.0233	0.1189 $\pm$ 0.078	0.0586
7	0.2363 $\pm$ 0.063	0.1792	0.1971 $\pm$ 0.080	0.0349	0.1121 $\pm$ 0.070	0.031	0.1928 $\pm$ 0.092	0.0233	0.1059 $\pm$ 0.056	0.0601
8	0.2409 $\pm$ 0.079	0.1929	0.1714 $\pm$ 0.076	0.035	0.0917 $\pm$ 0.052	0.031	0.1945 $\pm$ 0.089	0.0232	0.0909 $\pm$ 0.049	0.0601
9	0.2499 $\pm$ 0.079	0.2086	0.1643 $\pm$ 0.070	0.0349	0.071 $\pm$ 0.048	<b>0.0309</b>	0.2065 $\pm$ 0.101	0.0232	0.0833 $\pm$ 0.042	0.0602
10	0.2606 $\pm$ 0.083	0.225	0.139 $\pm$ 0.065	0.0351	0.0666 $\pm$ 0.038	0.031	0.2445 $\pm$ 0.107	0.0231	0.0791 $\pm$ 0.043	0.0606
11	0.2689 $\pm$ 0.093	0.2393	0.1223 $\pm$ 0.050	0.0354	0.0663 $\pm$ 0.034	0.0311	0.2117 $\pm$ 0.112	0.0226	0.0778 $\pm$ 0.041	0.0607
12	0.2752 $\pm$ 0.073	0.2555	0.1253 $\pm$ 0.055	0.0354	0.0565 $\pm$ 0.027	<b>0.0309</b>	0.1707 $\pm$ 0.079	<b>0.0225</b>	0.0764 $\pm$ 0.041	0.0606
13	0.2722 $\pm$ 0.065	0.2708	0.1197 $\pm$ 0.062	0.0355	0.0548 $\pm$ 0.026	<b>0.0309</b>	0.13 $\pm$ 0.069	<b>0.0225</b>	0.0703 $\pm$ 0.036	0.061
14	0.2729 $\pm$ 0.066	0.286	0.0999 $\pm$ 0.044	0.0359	0.0509 $\pm$ 0.019	0.031	0.1112 $\pm$ 0.049	<b>0.0225</b>	0.065 $\pm$ 0.031	0.0611
15	0.2988 $\pm$ 0.096	0.3014	0.1025 $\pm$ 0.044	0.0359	0.0521 $\pm$ 0.023	0.031	0.1147 $\pm$ 0.051	<b>0.0225</b>	0.0688 $\pm$ 0.034	0.0614
16	0.2964 $\pm$ 0.091	0.3161	0.1049 $\pm$ 0.052	0.0359	0.0467 $\pm$ 0.018	<b>0.0309</b>	0.0982 $\pm$ 0.050	<b>0.0225</b>	0.0599 $\pm$ 0.025	0.0617
17	0.2848 $\pm$ 0.079	0.3316	0.0906 $\pm$ 0.034	0.0359	0.0486 $\pm$ 0.022	0.031	0.094 $\pm$ 0.044	<b>0.0225</b>	0.0601 $\pm$ 0.026	0.062
18	0.3014 $\pm$ 0.065	0.3472	0.0871 $\pm$ 0.041	0.0361	0.0473 $\pm$ 0.018	0.031	0.0854 $\pm$ 0.041	0.0227	0.0572 $\pm$ 0.023	0.0624
19	0.3029 $\pm$ 0.066	0.363	0.0821 $\pm$ 0.027	0.0361	0.0414 $\pm$ 0.015	0.031	0.085 $\pm$ 0.037	<b>0.0225</b>	0.0535 $\pm$ 0.023	0.0623
20	0.2911 $\pm$ 0.072	0.3778	0.0821 $\pm$ 0.035	0.0364	0.0415 $\pm$ 0.017	0.031	0.0757 $\pm$ 0.034	<b>0.0225</b>	0.054 $\pm$ 0.024	0.0626

### 6.1.6. Problema de clasificación simple de dos clases gaussianas

Mediante la rutina **gendats** del toolbox de reconocimiento de patrones de matlab (**PRTools**) [5], se genera el conjunto problema de clasificación simple de dos clases gaussianas de dimensión 17, el número de objetos por clase es 250 para la clase 1 y 380 para la clase 2.



**Figura 6-10.:** Grafico de dispersión del conjunto de datos proyectado en 2D mediante escalamiento multidimensional y desempeño de cada uno de los clasificadores variando el número de objetos de representación por clase. Conjunto de datos problema de clasificación simple de dos clases gaussianas. **Selección de las líneas de características de longitud mayor.**



**Tabla 6-19.:** Intervalos de confianza para el error promedio de cada clasificador de acuerdo al número de objetos de representación por clase (Resaltado se encuentra el mejor intervalo para cada clasificador).

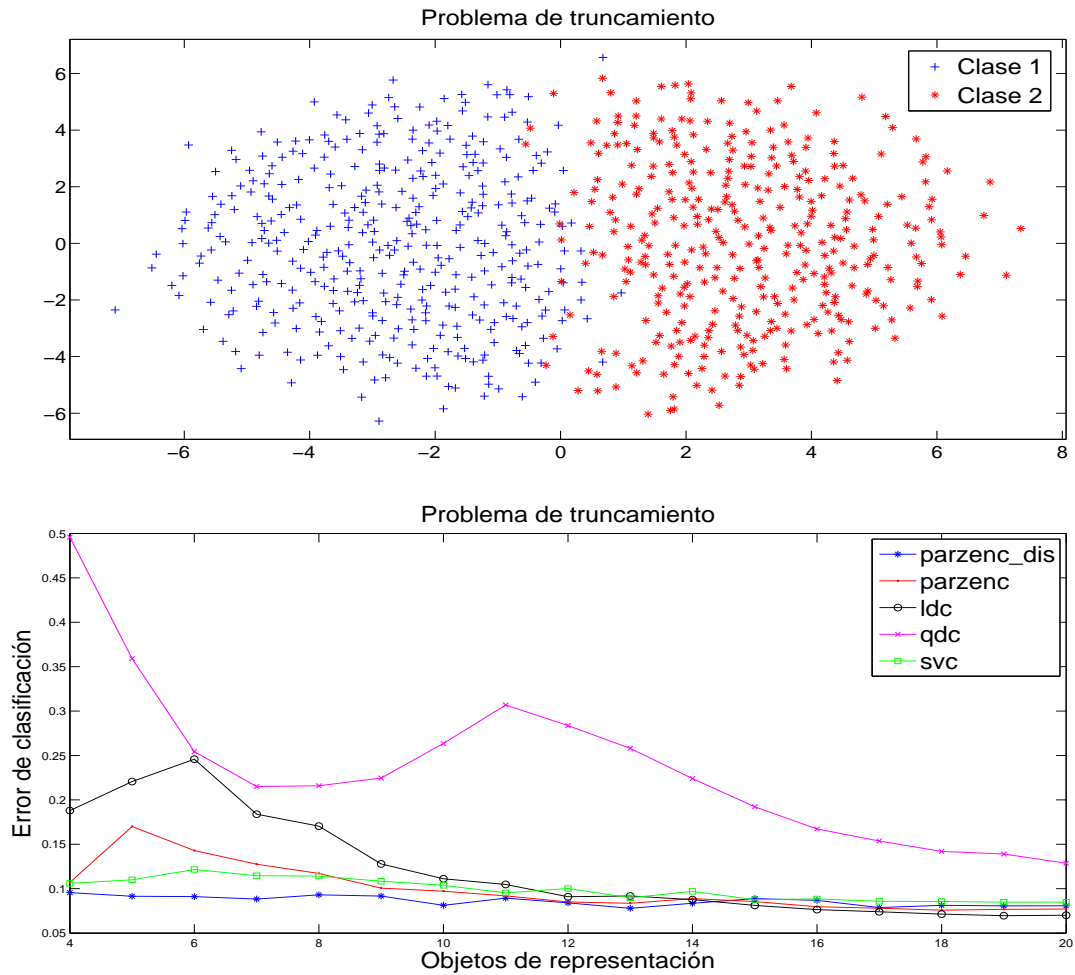
Intervalos de confianza para el error promedio					
Objetos de representación	<i>parzenc_dis</i>	<i>parzenc</i>	<i>ldc</i>	<i>qdc</i>	<i>svc</i>
4	[0.3282,0.3650]	[0.3469,0.3769]	[0.3711,0.4085]	[0.4997,0.5007]	[0.3451,0.3777]
5	[0.3354,0.3706]	[0.4063,0.4405]	[0.3798,0.4178]	[0.4590,0.4858]	[0.3759,0.4101]
6	[0.3243,0.3627]	[0.3800,0.4126]	[0.4138,0.4464]	[0.4107,0.4397]	[0.3554,0.3896]
7	[0.3223,0.3507]	[0.3809,0.4141]	[0.3946,0.4288]	[0.4003,0.4293]	[0.3577,0.3909]
8	[0.3227,0.3549]	[0.3577,0.3941]	[0.3691,0.4001]	[0.4026,0.4300]	[0.3336,0.3668]
9	[0.3164,0.3500]	[0.3669,0.3959]	[0.3460,0.3770]	[0.3998,0.4262]	[0.3330,0.3646]
10	[0.3129,0.3461]	[0.3456,0.3782]	[0.3301,0.3575]	[0.4092,0.4350]	[0.3174,0.3448]
11	[0.3197,0.3519]	[0.3487,0.3761]	[0.3302,0.3618]	[0.4107,0.4397]	[0.3190,0.3464]
12	[0.3045,0.3371]	[0.3344,0.3666]	[0.3174,0.3506]	[0.4184,0.4420]	[0.3119,0.3435]
13	[0.3167,0.3457]	[0.3332,0.3616]	[0.3181,0.3471]	[0.4014,0.4262]	[0.3138,0.3428]
14	[0.3209,0.3567]	[0.3362,0.3614]	[0.3129,0.3429]	[0.3930,0.4208]	[0.3143,0.3417]
15	[0.3159,0.3527]	[0.3238,0.3532]	[0.3048,0.3312]	[0.3833,0.4091]	[0.3035,0.3277]
16	[0.3275,0.3601]	[0.3330,0.3556]	[0.3078,0.3300]	[0.3814,0.4040]	[0.3026,0.3232]
17	[0.3077,0.3403]	[0.3278,0.3530]	[0.3007,0.3281]	[0.3785,0.4011]	[0.2976,0.3234]
18	<b>[0.3049,0.3355]</b>	[0.3190,0.3458]	[0.2884,0.3084]	[0.3689,0.3895]	[0.2823,0.3045]
19	[0.3155,0.3471]	[0.3198,0.3430]	[0.2929,0.3171]	[0.3689,0.3889]	[0.2947,0.3183]
20	[0.3206,0.3532]	<b>[0.3168,0.3410]</b>	<b>[0.2931,0.3157]</b>	<b>[0.3646,0.3856]</b>	<b>[0.2895,0.3117]</b>

**Tabla 6-20.:** Tabla del Error promedio, desviación estándar, tiempo de cómputo, de los clasificadores de acuerdo al número de objetos de representación por clase (resaltado se encuentra el mejor tiempo promedio de cómputo de cada clasificador).

Error promedio, desviación estándar, tiempo de cómputo de los clasificadores										
Objetos de representación	<i>parzenc_dis</i>		<i>parzenc</i>		<i>ldc</i>		<i>qdc</i>		<i>svc</i>	
	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)
4	$0.3466 \pm 0.070$	<b>0.1311</b>	$0.3619 \pm 0.057$	<b>0.0337</b>	$0.3898 \pm 0.071$	0.0308	$0.5002 \pm 0.002$	0.0231	$0.3614 \pm 0.062$	<b>0.0555</b>
5	$0.353 \pm 0.067$	0.1458	$0.4234 \pm 0.065$	0.0339	$0.3988 \pm 0.072$	<b>0.0307</b>	$0.4724 \pm 0.051$	0.0231	$0.393 \pm 0.065$	0.0561
6	$0.3435 \pm 0.073$	0.1611	$0.3963 \pm 0.062$	0.0346	$0.4301 \pm 0.062$	0.0308	$0.4252 \pm 0.055$	0.0232	$0.3725 \pm 0.065$	0.0583
7	$0.3365 \pm 0.054$	0.1762	$0.3975 \pm 0.063$	0.0346	$0.4117 \pm 0.065$	0.0308	$0.4148 \pm 0.055$	0.0232	$0.3743 \pm 0.063$	0.0592
8	$0.3388 \pm 0.061$	0.1920	$0.3759 \pm 0.069$	0.0347	$0.3846 \pm 0.059$	0.0309	$0.4163 \pm 0.052$	0.0233	$0.3502 \pm 0.063$	0.0597
9	$0.3332 \pm 0.064$	0.2077	$0.3814 \pm 0.055$	0.0348	$0.3615 \pm 0.059$	0.0310	$0.413 \pm 0.050$	0.0233	$0.3488 \pm 0.060$	0.0601
10	$0.3295 \pm 0.063$	0.2232	$0.3619 \pm 0.062$	0.0350	$0.3438 \pm 0.052$	0.0312	$0.4221 \pm 0.049$	0.0233	$0.3311 \pm 0.052$	0.0602
11	$0.3358 \pm 0.061$	0.2382	$0.3624 \pm 0.052$	0.0352	$0.346 \pm 0.060$	0.0309	$0.4252 \pm 0.055$	0.0230	$0.3327 \pm 0.052$	0.0601
12	$0.3208 \pm 0.062$	0.2532	$0.3505 \pm 0.061$	0.0351	$0.334 \pm 0.063$	0.0309	$0.4302 \pm 0.045$	0.0227	$0.3277 \pm 0.060$	0.0599
13	$0.3312 \pm 0.055$	0.2686	$0.3474 \pm 0.054$	0.0352	$0.3326 \pm 0.055$	0.0308	$0.4138 \pm 0.047$	<b>0.0225</b>	$0.3283 \pm 0.055$	0.0601
14	$0.3388 \pm 0.068$	0.2831	$0.3488 \pm 0.048$	0.0353	$0.3279 \pm 0.057$	0.0309	$0.4069 \pm 0.053$	<b>0.0225</b>	$0.328 \pm 0.052$	0.0602
15	$0.3343 \pm 0.070$	0.2985	$0.3385 \pm 0.056$	0.0354	$0.318 \pm 0.050$	0.0309	$0.3962 \pm 0.049$	0.0225	$0.3156 \pm 0.046$	0.0602
16	$0.3438 \pm 0.062$	0.3137	$0.3443 \pm 0.043$	0.0356	$0.3189 \pm 0.042$	0.0309	$0.3927 \pm 0.043$	0.0225	$0.3129 \pm 0.039$	0.0604
17	$0.324 \pm 0.062$	0.3286	$0.3404 \pm 0.048$	0.0356	$0.3144 \pm 0.052$	0.0309	$0.3898 \pm 0.043$	0.0225	$0.3105 \pm 0.049$	0.0607
18	$0.3202 \pm 0.058$	0.3441	$0.3324 \pm 0.051$	0.0353	$0.2984 \pm 0.038$	0.0309	$0.3792 \pm 0.039$	0.0225	$0.2934 \pm 0.042$	0.0609
19	$0.3313 \pm 0.060$	0.3603	$0.3314 \pm 0.044$	0.0355	$0.305 \pm 0.046$	0.0309	$0.3789 \pm 0.038$	0.0225	$0.3065 \pm 0.045$	0.0611
20	$0.3369 \pm 0.062$	0.3748	$0.3289 \pm 0.046$	0.0355	$0.3044 \pm 0.043$	0.0309	$0.3751 \pm 0.040$	0.0225	$0.3006 \pm 0.042$	0.0615

### 6.1.7. Conjunto problema de truncamiento

Mediante la rutina **gentrunk** del toolbox de reconocimiento de patrones de matlab (**PRTools**) [5], se genera el conjunto de problema de truncamiento de dimensión 17, el número de objetos por clase es 400 para la clase 1 y 358 para la clase 2.



**Figura 6-11.:** Grafico de dispersión del conjunto de datos proyectado en 2D mediante escalamiento multidimensional y desempeño de cada uno de los clasificadores variando el número de objetos de representación por clase. Conjunto de datos problema de truncamiento de dos clases. Selección de las líneas de características de longitud mayor.

**Tabla 6-21.:** Intervalos de confianza para el error promedio de cada clasificador de acuerdo al número de objetos de representación por clase (Resaltado se encuentra el mejor intervalo para cada clasificador).

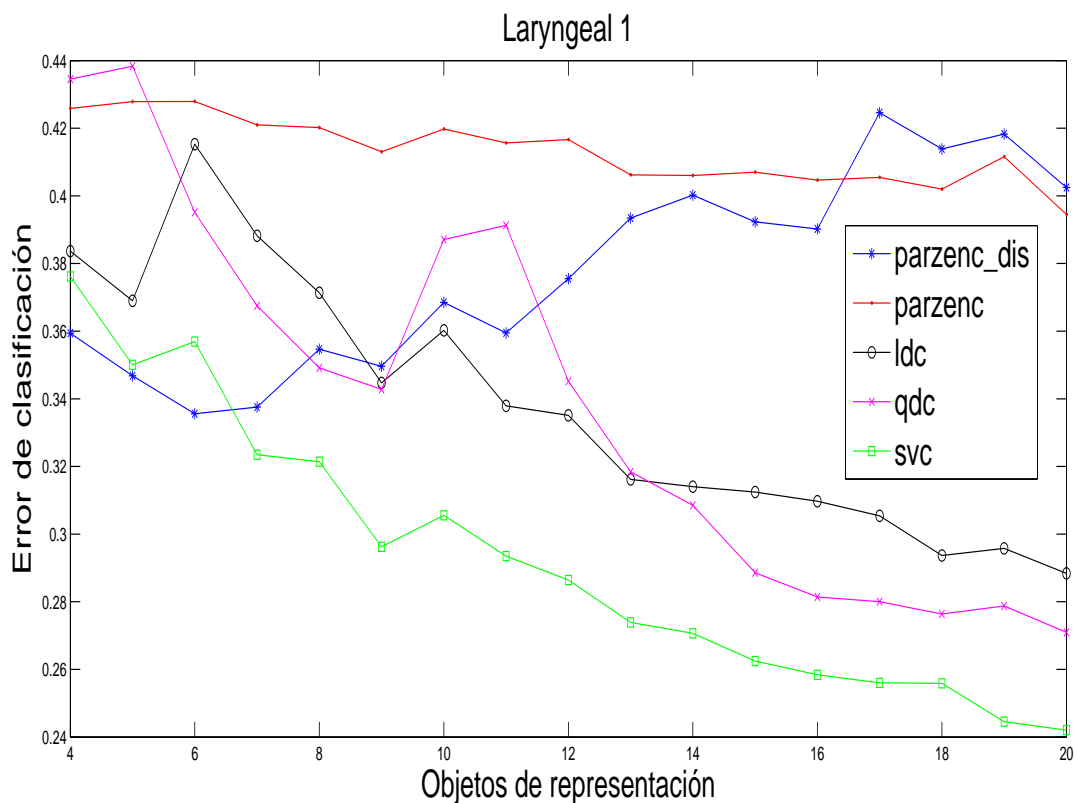
Intervalos de confianza para el error promedio					
Objetos de representación	<i>parzenc_dis</i>	<i>parzenc</i>	<i>ldc</i>	<i>qdc</i>	<i>svc</i>
4	[0.0856,0.1056]	[0.0958,0.1184]	[0.1632,0.2126]	[0.4874,0.5042]	[0.0927,0.1195]
5	[0.0845,0.0987]	[0.1438,0.1964]	[0.1874,0.2538]	[0.3198,0.3982]	[0.0950,0.1250]
6	[0.0811,0.1011]	[0.1238,0.1622]	[0.2092,0.2824]	[0.2225,0.2861]	[0.1055,0.1377]
7	[0.0818,0.0950]	[0.1108,0.1444]	[0.1557,0.2121]	[0.1909,0.2393]	[0.0994,0.1300]
8	[0.0833,0.1027]	[0.1019,0.1325]	[0.1451,0.1957]	[0.1913,0.2407]	[0.1000,0.1284]
9	[0.0836,0.1000]	[0.0886,0.1128]	[0.1106,0.1454]	[0.2014,0.2478]	[0.0945,0.1223]
10	[0.0748,0.0880]	[0.0877,0.1067]	[0.0956,0.1266]	[0.2391,0.2881]	[0.0917,0.1159]
11	[0.0818,0.0970]	[0.0844,0.0996]	[0.0916,0.1180]	[0.2768,0.3368]	[0.0865,0.1043]
12	[0.0761,0.0919]	[0.0798,0.0904]	[0.0815,0.1005]	[0.2561,0.3113]	[0.0903,0.1103]
13	[0.0716,0.0842]	[0.0765,0.0907]	[0.0802,0.1034]	[0.2325,0.2835]	[0.0806,0.0990]
14	[0.0773,0.0899]	[0.0818,0.0960]	[0.0775,0.0981]	[0.2037,0.2443]	[0.0862,0.1078]
15	[0.0810,0.0968]	[0.0779,0.0931]	[0.0733,0.0891]	[0.1741,0.2105]	[0.0794,0.0962]
16	[0.0790,0.0948]	[0.0744,0.0850]	[0.0689,0.0841]	[0.1512,0.1834]	[0.0802,0.0966]
17	<b>[0.0720,0.0856]</b>	[0.0717,0.0843]	[0.0677,0.0803]	[0.1398,0.1676]	[0.0785,0.0933]
18	[0.0747,0.0879]	<b>[0.0711,0.0805]</b>	[0.0661,0.0767]	[0.1299,0.1541]	[0.0788,0.0924]
19	[0.0728,0.0886]	[0.0714,0.0820]	<b>[0.0643,0.0749]</b>	[0.1280,0.1502]	<b>[0.0765,0.0933]</b>
20	[0.0745,0.0871]	[0.0716,0.0826]	[0.0648,0.0754]	<b>[0.1187,0.1387]</b>	[0.0781,0.0917]

**Tabla 6-22.:** Tabla del Error promedio, desviación estándar, tiempo de cómputo, de los clasificadores de acuerdo al número de objetos de representación por clase (resaltado se encuentra el mejor tiempo promedio de cómputo de cada clasificador).

Error promedio, desviación estándar, tiempo de cómputo de los clasificadores										
Objetos de representación	<i>parzenc_dis</i>		<i>parzenc</i>		<i>ldc</i>		<i>qdc</i>		<i>svc</i>	
	$\bar{e}_\infty \pm \sigma$	CPU(seg)	$\bar{e}_\infty \pm \sigma$	CPU(seg)	$\bar{e}_\infty \pm \sigma$	CPU(seg)	$\bar{e}_\infty \pm \sigma$	CPU(seg)	$\bar{e}_\infty \pm \sigma$	CPU(seg)
4	0.0956 $\pm$ 0.038	<b>0.1409</b>	0.1071 $\pm$ 0.043	0.0356	0.1879 $\pm$ 0.094	0.0349	0.4958 $\pm$ 0.032	0.0274	0.1061 $\pm$ 0.051	0.0631
5	0.0916 $\pm$ 0.027	0.1531	0.1701 $\pm$ 0.100	<b>0.0344</b>	0.2206 $\pm$ 0.126	0.0329	0.359 $\pm$ 0.149	0.0284	0.11 $\pm$ 0.057	0.0646
6	0.0911 $\pm$ 0.038	0.1678	0.143 $\pm$ 0.073	0.0408	0.2458 $\pm$ 0.139	0.0410	0.2543 $\pm$ 0.121	0.0344	0.1216 $\pm$ 0.061	0.0759
7	0.0884 $\pm$ 0.025	0.1832	0.1276 $\pm$ 0.064	0.0374	0.1839 $\pm$ 0.107	0.0347	0.2151 $\pm$ 0.092	0.0308	0.1147 $\pm$ 0.058	0.0645
8	0.093 $\pm$ 0.030	0.1989	0.1172 $\pm$ 0.058	0.0359	0.1704 $\pm$ 0.096	0.0323	0.216 $\pm$ 0.094	0.0244	0.1142 $\pm$ 0.054	0.0619
9	0.0918 $\pm$ 0.031	0.2140	0.1007 $\pm$ 0.046	0.0354	0.128 $\pm$ 0.066	0.0314	0.2246 $\pm$ 0.088	0.0236	0.1084 $\pm$ 0.053	<b>0.0604</b>
10	0.0814 $\pm$ 0.025	0.2302	0.0972 $\pm$ 0.036	0.0356	0.1111 $\pm$ 0.059	0.0313	0.2636 $\pm$ 0.093	0.0235	0.1038 $\pm$ 0.046	0.0607
11	0.0894 $\pm$ 0.029	0.2471	0.092 $\pm$ 0.029	0.0356	0.1048 $\pm$ 0.050	0.0313	0.3068 $\pm$ 0.114	0.0231	0.0954 $\pm$ 0.034	0.0609
12	0.084 $\pm$ 0.030	0.2641	0.0851 $\pm$ 0.020	0.0359	0.091 $\pm$ 0.036	0.0314	0.2837 $\pm$ 0.105	0.0230	0.1003 $\pm$ 0.038	0.0613
13	0.0779 $\pm$ 0.024	0.2801	0.0836 $\pm$ 0.027	0.0359	0.0918 $\pm$ 0.044	0.0313	0.258 $\pm$ 0.097	0.0228	0.0898 $\pm$ 0.035	0.0613
14	0.0836 $\pm$ 0.024	0.2953	0.0889 $\pm$ 0.027	0.0361	0.0878 $\pm$ 0.039	0.0313	0.224 $\pm$ 0.077	<b>0.0228</b>	0.097 $\pm$ 0.041	0.0615
15	0.0889 $\pm$ 0.030	0.3118	0.0855 $\pm$ 0.029	0.0356	0.0812 $\pm$ 0.030	0.0313	0.1923 $\pm$ 0.069	<b>0.0228</b>	0.0878 $\pm$ 0.032	0.0617
16	0.0869 $\pm$ 0.030	0.3293	0.0797 $\pm$ 0.020	0.0359	0.0765 $\pm$ 0.029	0.0315	0.1673 $\pm$ 0.061	0.0229	0.0884 $\pm$ 0.031	0.0625
17	0.0788 $\pm$ 0.026	0.3430	0.078 $\pm$ 0.024	0.0358	0.074 $\pm$ 0.024	<b>0.0312</b>	0.1537 $\pm$ 0.053	0.0228	0.0859 $\pm$ 0.028	0.0623
18	0.0813 $\pm$ 0.025	0.3617	0.0758 $\pm$ 0.018	0.0360	0.0714 $\pm$ 0.020	0.0313	0.142 $\pm$ 0.046	0.0228	0.0856 $\pm$ 0.026	0.0628
19	0.0807 $\pm$ 0.030	0.3789	0.0767 $\pm$ 0.020	0.0361	0.0696 $\pm$ 0.020	0.0313	0.1391 $\pm$ 0.042	0.0228	0.0849 $\pm$ 0.032	0.0630
20	0.0808 $\pm$ 0.024	0.3957	0.0771 $\pm$ 0.021	0.0365	0.0701 $\pm$ 0.020	0.0314	0.1287 $\pm$ 0.038	0.0228	0.0849 $\pm$ 0.026	0.0634

### 6.1.8. Conjunto Laryngeal 1

Conjunto de datos del mundo real, con dos clases, voces normales y patológicas, tomado de la academia búlgara de ciencias [16]. La dimensión del conjunto es 16 y se presenta el desempeño de los clasificadores a continuación.



**Figura 6-12.:** Desempeño de cada uno de los clasificadores variando el número de objetos de representación por clase. Conjunto de datos laryngeal 1, con 81 objetos para la clase de voces normales y 132 objetos para la clase de voces patológicas y dimensión del conjunto igual a 16. **Selección de las líneas de características de longitud mayor.**

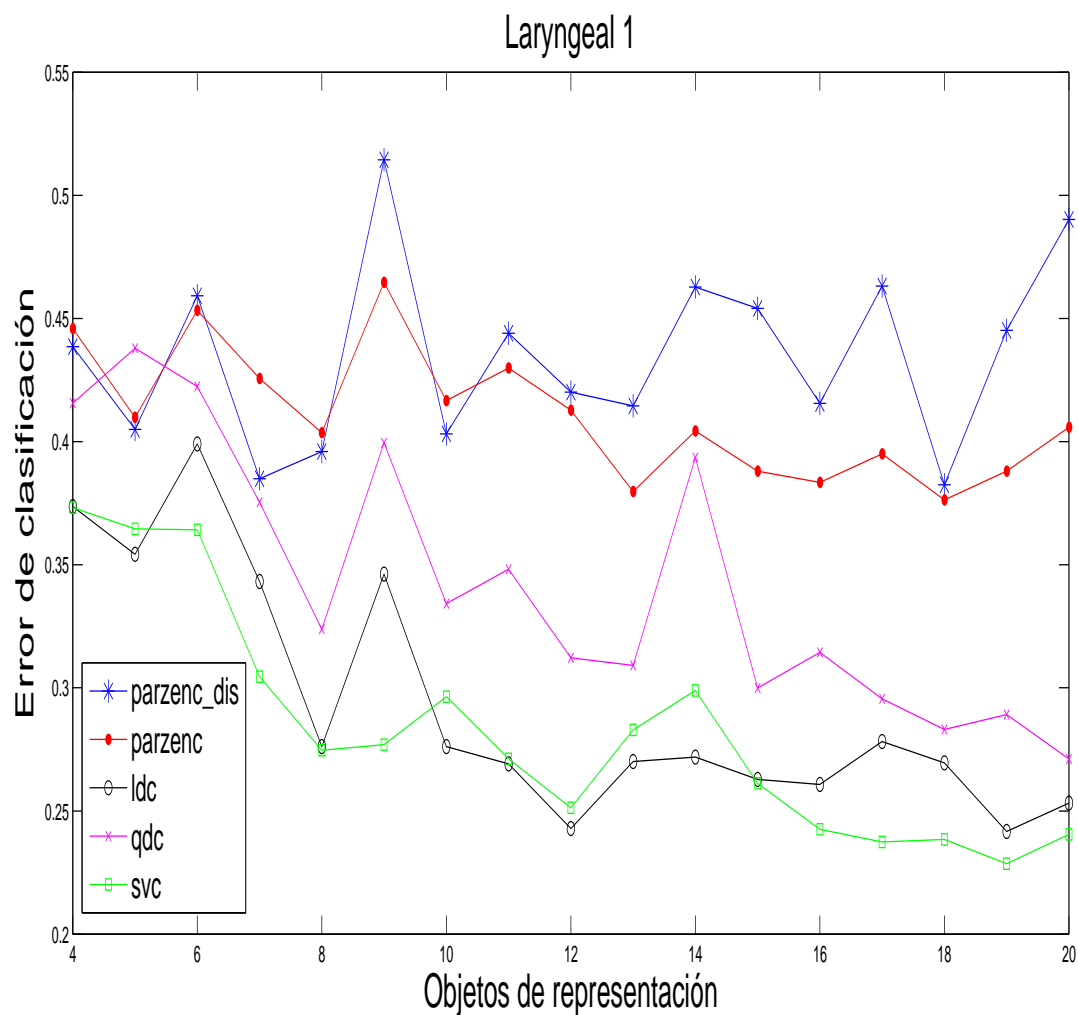
**Tabla 6-23.:** Intervalos de confianza para el error promedio de cada clasificador de acuerdo al número de objetos de representación por clase (Resaltado se encuentra el mejor intervalo para cada clasificador).

Intervalos de confianza para el error promedio					
Objetos de representación	<i>parzenc_dis</i>	<i>parzenc</i>	<i>ldc</i>	<i>qdc</i>	<i>svc</i>
4	[0.3339,0.3849]	[0.4080,0.4438]	[0.3634,0.4040]	[0.4074,0.4616]	[0.3544,0.3980]
5	[0.3202,0.3734]	[0.4095,0.4463]	[0.3479,0.3901]	[0.4104,0.4662]	[0.3313,0.3687]
6	[0.3088,0.3624]	[0.4084,0.4474]	[0.3927,0.4379]	[0.3735,0.4167]	[0.3379,0.3759]
7	<b>[0.3091,0.3659]</b>	[0.4034,0.4386]	[0.3674,0.4090]	[0.3472,0.3878]	[0.3030,0.3440]
8	[0.3281,0.3813]	[0.4039,0.4365]	[0.3506,0.3922]	[0.3294,0.3688]	[0.3030,0.3398]
9	[0.3238,0.3754]	[0.3975,0.4285]	[0.3273,0.3621]	[0.3231,0.3625]	[0.2813,0.3113]
10	[0.3432,0.3938]	[0.4027,0.4369]	[0.3402,0.3802]	[0.3639,0.4103]	[0.2865,0.3245]
11	[0.3334,0.3856]	[0.3991,0.4323]	[0.3218,0.3540]	[0.3689,0.4137]	[0.2761,0.3109]
12	[0.3476,0.4034]	[0.3995,0.4337]	[0.3185,0.3517]	[0.3270,0.3634]	[0.2682,0.3046]
13	[0.3658,0.4210]	[0.3901,0.4223]	[0.3014,0.3308]	[0.3023,0.3345]	[0.2594,0.2884]
14	[0.3736,0.4268]	[0.3921,0.4199]	[0.3008,0.3272]	[0.2930,0.3240]	[0.2559,0.2853]
15	[0.3628,0.4218]	[0.3915,0.4225]	[0.2979,0.3269]	[0.2741,0.3031]	[0.2501,0.2749]
16	[0.3628,0.4176]	[0.3904,0.4188]	[0.2965,0.3229]	[0.2696,0.2932]	[0.2466,0.2702]
17	[0.3996,0.4496]	[0.3918,0.4192]	[0.2896,0.3212]	[0.2679,0.2921]	[0.2423,0.2697]
18	[0.3823,0.4455]	[0.3865,0.4175]	[0.2798,0.3076]	[0.2656,0.2872]	[0.2443,0.2675]
19	[0.3878,0.4488]	[0.3969,0.4263]	[0.2821,0.3095]	[0.2645,0.2929]	[0.2322,0.2570]
20	[0.3746,0.4304]	<b>[0.3803,0.4087]</b>	<b>[0.2760,0.3008]</b>	<b>[0.2596,0.2822]</b>	<b>[0.2307,0.2533]</b>

**Tabla 6-24.:** Tabla del Error promedio, desviación estándar, tiempo de cómputo, de los clasificadores de acuerdo al número de objetos de representación por clase (resaltado se encuentra el mejor tiempo promedio de cómputo de cada clasificador).

Error promedio, desviación estándar, tiempo de cómputo de los clasificadores										
Objetos de representación	<i>parzenc_dis</i>		<i>parzenc</i>		<i>ldc</i>		<i>qdc</i>		<i>svc</i>	
	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)
4	$0.3594 \pm 0.097$	<b>0.1292</b>	$0.4259 \pm 0.068$	<b>0.0342</b>	$0.3837 \pm 0.077$	0.0312	$0.4345 \pm 0.103$	0.0233	$0.3762 \pm 0.083$	<b>0.0552</b>
5	$0.3468 \pm 0.101$	0.1439	$0.4279 \pm 0.070$	0.0344	$0.369 \pm 0.080$	0.0313	$0.4383 \pm 0.106$	0.0233	$0.35 \pm 0.071$	0.0553
6	$0.3356 \pm 0.102$	0.1592	$0.4279 \pm 0.074$	0.0343	$0.4153 \pm 0.086$	<b>0.0308</b>	$0.3951 \pm 0.082$	0.0233	$0.3569 \pm 0.072$	0.0578
7	$0.3375 \pm 0.108$	0.1749	$0.421 \pm 0.067$	0.0344	$0.3882 \pm 0.079$	0.0309	$0.3675 \pm 0.077$	0.0233	$0.3235 \pm 0.078$	0.0597
8	$0.3547 \pm 0.101$	0.1902	$0.4202 \pm 0.062$	0.0344	$0.3714 \pm 0.079$	0.0309	$0.3491 \pm 0.075$	0.0233	$0.3214 \pm 0.070$	0.0600
9	$0.3496 \pm 0.098$	0.2078	$0.413 \pm 0.059$	0.0348	$0.3447 \pm 0.066$	0.0313	$0.3428 \pm 0.075$	0.0237	$0.2963 \pm 0.057$	0.0610
10	$0.3685 \pm 0.096$	0.2227	$0.4198 \pm 0.065$	0.0357	$0.3602 \pm 0.076$	0.0320	$0.3871 \pm 0.088$	0.0243	$0.3055 \pm 0.072$	0.0622
11	$0.3595 \pm 0.099$	0.2384	$0.4157 \pm 0.063$	0.0360	$0.3379 \pm 0.061$	0.0323	$0.3913 \pm 0.085$	0.0239	$0.2935 \pm 0.066$	0.0624
12	$0.3755 \pm 0.106$	0.2534	$0.4166 \pm 0.065$	0.0361	$0.3351 \pm 0.063$	0.0324	$0.3452 \pm 0.069$	0.0240	$0.2864 \pm 0.069$	0.0624
13	$0.3934 \pm 0.105$	0.2681	$0.4062 \pm 0.061$	0.0361	$0.3161 \pm 0.056$	0.0325	$0.3184 \pm 0.061$	0.0244	$0.2739 \pm 0.055$	0.0624
14	$0.4002 \pm 0.101$	0.2825	$0.406 \pm 0.053$	0.0349	$0.314 \pm 0.050$	0.0313	$0.3085 \pm 0.059$	0.0231	$0.2706 \pm 0.056$	0.0611
15	$0.3923 \pm 0.112$	0.2981	$0.407 \pm 0.059$	0.0348	$0.3124 \pm 0.055$	0.0311	$0.2886 \pm 0.055$	<b>0.0229</b>	$0.2625 \pm 0.047$	0.0611
16	$0.3902 \pm 0.104$	0.3131	$0.4046 \pm 0.054$	0.0349	$0.3097 \pm 0.050$	0.0313	$0.2814 \pm 0.045$	0.0231	$0.2584 \pm 0.045$	0.0616
17	$0.4246 \pm 0.095$	0.3295	$0.4055 \pm 0.052$	0.0349	$0.3054 \pm 0.060$	0.0311	$0.28 \pm 0.046$	0.0230	$0.256 \pm 0.052$	0.0617
18	$0.4139 \pm 0.120$	0.3447	$0.402 \pm 0.059$	0.0351	$0.2937 \pm 0.053$	0.0314	$0.2764 \pm 0.041$	0.0232	$0.2559 \pm 0.044$	0.0620
19	$0.4183 \pm 0.116$	0.3598	$0.4116 \pm 0.056$	0.0354	$0.2958 \pm 0.052$	0.0314	$0.2787 \pm 0.054$	0.0232	$0.2446 \pm 0.047$	0.0621
20	$0.4025 \pm 0.106$	0.3752	$0.3945 \pm 0.054$	0.0350	$0.2884 \pm 0.047$	0.0313	$0.2709 \pm 0.043$	0.0231	$0.242 \pm 0.043$	0.0637





**Figura 6-13.:** Desempeño de cada uno de los clasificadores variando el número de objetos de representación por clase. Conjunto de datos laryngeal 1, con 81 objetos para la clase de voces normales y 132 objetos para la clase de voces patológicas y dimensión del conjunto igual a 16. **Selección de las líneas de características de longitud menor.**

**Tabla 6-25.:** Intervalos de confianza para el error promedio de cada clasificador de acuerdo al número de objetos de representación por clase (Resaltado se encuentra el mejor y el peor intervalo para el clasificador *parzenc\_dis*).

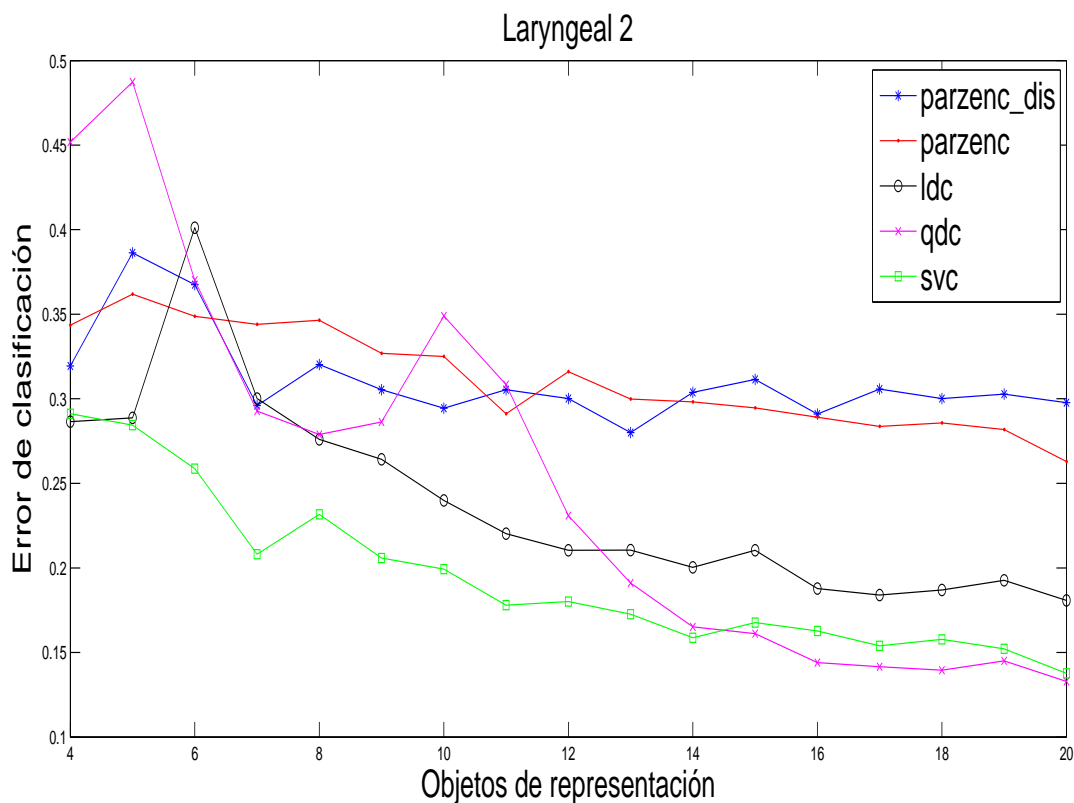
Intervalos de confianza para el error promedio					
Objetos de representación	<i>parzenc_dis</i>	<i>parzenc</i>	<i>ldc</i>	<i>qdc</i>	<i>svc</i>
4	[0.3624,0.5147]	[0.3899,0.5018]	[0.3243,0.4230]	[0.3610,0.4702]	[0.3610,0.4702]
5	[0.3330,0.4769]	[0.3527,0.4670]	[0.3071,0.4013]	[0.3300,0.5458]	[0.3300,0.5458]
6	[0.3976,0.5208]	[0.4230,0.4835]	[0.3355,0.4626]	[0.3355,0.5093]	[0.3355,0.5093]
7	[0.3348,0.4351]	[0.3439,0.5074]	[0.3042,0.3822]	[0.2753,0.4754]	[0.2753,0.4754]
8	[0.3274,0.4645]	[0.3609,0.4462]	[0.2388,0.3135]	[0.2811,0.3666]	[0.2811,0.3666]
9	<b>[0.4318,0.5969]</b>	[0.3680,0.5613]	[0.2709,0.4214]	[0.3514,0.4475]	[0.3514,0.4475]
10	[0.3502,0.4560]	[0.3740,0.4591]	[0.2367,0.3157]	[0.2887,0.3796]	[0.2887,0.3796]
11	[0.3549,0.5330]	[0.3691,0.4906]	[0.2328,0.3054]	[0.2899,0.4064]	[0.2899,0.4064]
12	[0.3480,0.4922]	[0.3669,0.4585]	[0.2268,0.2589]	[0.2314,0.3929]	[0.2314,0.3929]
13	[0.3438,0.4851]	[0.3183,0.4411]	[0.2355,0.3046]	[0.2577,0.3605]	[0.2577,0.3605]
14	[0.3873,0.5381]	[0.3693,0.4394]	[0.2300,0.3138]	[0.2919,0.4951]	[0.2919,0.4951]
15	[0.3880,0.5202]	[0.3230,0.4529]	[0.2360,0.2897]	[0.2640,0.3360]	[0.2640,0.3360]
16	[0.3451,0.4859]	[0.3401,0.4268]	[0.2154,0.3062]	[0.2443,0.3844]	[0.2443,0.3844]
17	[0.3606,0.5656]	[0.3496,0.4404]	[0.2249,0.3315]	[0.2443,0.3468]	[0.2443,0.3468]
18	<b>[0.3118,0.4532]</b>	[0.3356,0.4169]	[0.2266,0.3124]	[0.2453,0.3208]	[0.2453,0.3208]
19	[0.3434,0.5469]	[0.3420,0.4340]	[0.2036,0.2798]	[0.2357,0.3426]	[0.2357,0.3426]
20	[0.3950,0.5854]	[0.3487,0.4628]	[0.2121,0.2942]	[0.2373,0.3049]	[0.2373,0.3049]

**Tabla 6-26.:** Tabla del Error promedio, desviación estándar, tiempo de cómputo, de los clasificadores de acuerdo al número de objetos de representación por clase (resaltado se encuentra el mejor tiempo promedio de cómputo para el clasificador *parzenc\_dis*).

Error promedio, desviación estándar, tiempo de cómputo de los clasificadores										
Objetos de representación	<i>parzenc_dis</i>		<i>parzenc</i>		<i>ldc</i>		<i>qdc</i>		<i>svc</i>	
	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)
4	0.4385±0.0296	0.0433	0.4459±0.0218	0.0368	0.3737±0.0192	0.0368	0.4156±0.0212	0.0248	0.3732±0.0224	0.0636
5	0.4049±0.0280	0.0345	0.4099±0.0222	0.0325	0.3542±0.0183	0.0314	0.4379±0.0420	0.0236	0.3645±0.0314	0.0561
6	0.4592±0.0240	0.0353	0.4532±0.0118	0.0326	0.3990±0.0247	0.0314	0.4224±0.0338	0.0237	0.3642±0.0234	0.0601
7	0.3849±0.0195	0.0350	0.4256±0.0318	0.0327	0.3432±0.0152	0.0315	0.3754±0.0389	0.0243	0.3045±0.0199	0.0601
8	0.3959±0.0267	0.0519	0.4036±0.0166	0.0328	0.2761±0.0145	0.0334	0.3239±0.0166	0.0256	0.2746±0.0165	0.0649
9	<b>0.5144±0.0321</b>	<b>0.0368</b>	0.4646±0.0376	0.0331	0.3462±0.0293	0.0338	0.3995±0.0187	0.0264	0.2769±0.0126	0.0651
10	0.4031±0.0206	0.0358	0.4166±0.0166	0.0363	0.2762±0.0154	0.0321	0.3342±0.0177	0.0258	0.2964±0.0189	0.0641
11	0.4440±0.0346	0.0367	0.4298±0.0237	0.0366	0.2691±0.0141	0.0329	0.3482±0.0227	0.0268	0.2712±0.0165	0.0666
12	0.4201±0.0281	0.0355	0.4127±0.0178	0.0383	0.2429±0.0062	0.0318	0.3122±0.0314	0.0245	0.2513±0.0109	0.0622
13	0.4144±0.0275	0.0381	0.3797±0.0239	0.0349	0.2701±0.0134	0.0334	0.3091±0.0200	0.0248	0.2829±0.0143	0.0652
14	0.4627±0.0293	0.0369	0.4043±0.0136	0.0339	0.2719±0.0163	0.0332	0.3935±0.0395	0.0258	0.2989±0.0153	0.0634
15	0.4541±0.0257	0.0378	0.3880±0.0253	0.0337	0.2628±0.0104	0.0346	0.3000±0.0140	0.0268	0.2612±0.0129	0.0651
16	<b>0.4155±0.0274</b>	<b>0.0381</b>	0.3834±0.0169	0.0340	0.2608±0.0177	0.0341	0.3144±0.0273	0.0259	0.2425±0.0115	0.0646
17	0.4631±0.0399	0.0378	0.3950±0.0177	0.0366	0.2782±0.0207	0.0345	0.2955±0.0199	0.0258	0.2374±0.0170	0.0642
18	0.3825±0.0275	0.0432	0.3763±0.0158	0.0404	0.2695±0.0167	0.0430	0.2831±0.0147	0.0317	0.2384±0.0122	0.0715
19	0.4451±0.0396	0.0426	0.3880±0.0179	0.0353	0.2417±0.0148	0.0406	0.2891±0.0208	0.0285	0.2286±0.0100	0.0715
20	0.4902±0.0370	0.0445	0.4058±0.0222	0.0403	0.2532±0.0160	0.0438	0.2711±0.0132	0.0328	0.2405±0.0140	0.0731

### 6.1.9. Conjunto Laryngeal 2

Conjunto de datos del mundo real, con dos clases, voces normales y patológicas, tomado de la academia búlgara de ciencias [16]. La dimensión del conjunto es 16 y se presenta el desempeño de los clasificadores a continuación.



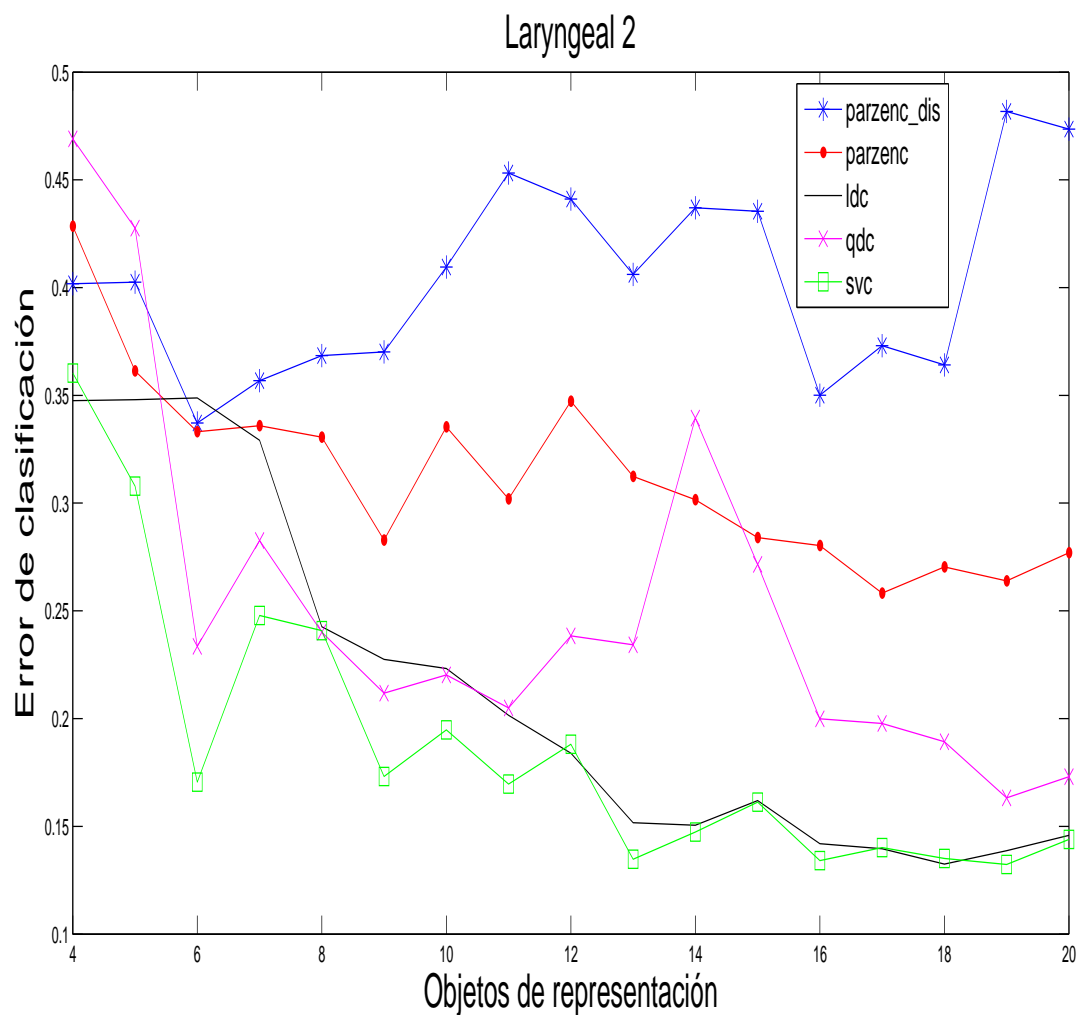
**Figura 6-14.:** Desempeño de cada uno de los clasificadores variando el número de objetos de representación por clase. Conjunto de datos laryngeal 2, con 53 objetos para la clase de voces normales y 639 objetos para la clase de voces patológicas y dimensión del conjunto igual a 16. **Selección de las líneas de características de longitud mayor.**

**Tabla 6-27.:** Intervalos de confianza para el error promedio de cada clasificador de acuerdo al número de objetos de representación por clase (Resaltado se encuentra el mejor intervalo para cada clasificador).

Intervalos de confianza para el error promedio					
Objetos de representación	<i>parzenc_dis</i>	<i>parzenc</i>	<i>ldc</i>	<i>qdc</i>	<i>svc</i>
4	[0.2683,0.3705]	[0.3117,0.3753]	[0.2514,0.3214]	[0.3896,0.5138]	[0.2536,0.3288]
5	[0.3229,0.4497]	[0.3243,0.3995]	[0.2555,0.3219]	[0.4232,0.5516]	[0.2476,0.3212]
6	[0.3080,0.4270]	[0.3151,0.3825]	[0.3598,0.4424]	[0.3090,0.4312]	[0.2184,0.2990]
7	[0.2494,0.3426]	[0.3143,0.3737]	[0.2711,0.3291]	[0.2425,0.3431]	[0.1784,0.2378]
8	[0.2702,0.3702]	[0.3151,0.3777]	[0.2476,0.3044]	[0.2326,0.3252]	[0.2014,0.2620]
9	[0.2540,0.3566]	[0.3001,0.3537]	[0.2363,0.2921]	[0.2473,0.3253]	[0.1772,0.2346]
10	[0.2452,0.3436]	[0.2950,0.3550]	[0.2144,0.2654]	[0.2949,0.4029]	[0.1703,0.2283]
11	[0.2520,0.3588]	[0.2672,0.3152]	[0.1969,0.2437]	[0.2552,0.3616]	[0.1553,0.2005]
12	[0.2530,0.3472]	[0.2916,0.3406]	[0.1901,0.2307]	[0.1946,0.2672]	[0.1593,0.2009]
13	[0.2332,0.3268]	[0.2778,0.3220]	[0.1915,0.2295]	[0.1628,0.2192]	[0.1530,0.1924]
14	[0.2572,0.3504]	[0.2766,0.3198]	[0.1837,0.2169]	[0.1446,0.1856]	[0.1408,0.1766]
15	[0.2628,0.3602]	[0.2733,0.3159]	[0.1928,0.2280]	[0.1398,0.1824]	[0.1466,0.1888]
16	[0.2468,0.3352]	[0.2667,0.3115]	[0.1715,0.2041]	[0.1287,0.1593]	[0.1456,0.1798]
17	[0.2600,0.3516]	[0.2613,0.3061]	[0.1690,0.1990]	[0.1287,0.1545]	[0.1378,0.1700]
18	[0.2555,0.3449]	[0.2671,0.3045]	[0.1702,0.2038]	[0.1267,0.1525]	[0.1396,0.1760]
19	[0.2594,0.3462]	[0.2652,0.2984]	[0.1764,0.2090]	[0.1321,0.1579]	[0.1355,0.1687]
20	<b>[0.2511,0.3443]</b>	<b>[0.2455,0.2803]</b>	<b>[0.1658,0.1958]</b>	<b>[0.1224,0.1434]</b>	<b>[0.1244,0.1508]</b>

**Tabla 6-28.:** Tabla del Error promedio, desviación estándar, tiempo de cómputo, de los clasificadores de acuerdo al número de objetos de representación por clase (resaltado se encuentra el mejor tiempo promedio de cómputo de cada clasificador).

Error promedio, desviación estándar, tiempo de cómputo de los clasificadores										
Objetos de representación	<i>parzenc_dis</i>		<i>parzenc</i>		<i>ldc</i>		<i>qdc</i>		<i>svc</i>	
	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)
4	$0.3194 \pm 0.194$	<b>0.1319</b>	$0.3435 \pm 0.0121$	0.0360	$0.2864 \pm 0.133$	0.0328	$0.4517 \pm 0.236$	0.0245	$0.2912 \pm 0.143$	0.0573
5	$0.3863 \pm 0.241$	0.1459	$0.3619 \pm 0.0143$	0.0352	$0.2887 \pm 0.126$	0.0320	$0.4874 \pm 0.244$	0.0243	$0.2844 \pm 0.140$	<b>0.0569</b>
6	$0.3675 \pm 0.226$	0.1604	$0.3488 \pm 0.0128$	0.0358	$0.4011 \pm 0.157$	0.0319	$0.3701 \pm 0.232$	0.0243	$0.2587 \pm 0.153$	0.0598
7	$0.296 \pm 0.177$	0.1771	$0.344 \pm 0.0113$	0.0363	$0.3001 \pm 0.110$	0.0323	$0.2928 \pm 0.191$	0.0246	$0.2081 \pm 0.113$	0.0618
8	$0.3202 \pm 0.190$	0.1895	$0.3464 \pm 0.0119$	<b>0.0358</b>	$0.276 \pm 0.108$	0.0320	$0.2789 \pm 0.176$	0.0243	$0.2317 \pm 0.115$	0.0614
9	$0.3053 \pm 0.195$	0.2043	$0.3269 \pm 0.0102$	0.0360	$0.2642 \pm 0.106$	0.0320	$0.2863 \pm 0.148$	0.0243	$0.2059 \pm 0.109$	0.0618
10	$0.2944 \pm 0.187$	0.2198	$0.325 \pm 0.0114$	0.0361	$0.2399 \pm 0.097$	0.0321	$0.3489 \pm 0.205$	0.0241	$0.1993 \pm 0.110$	0.0622
11	$0.3054 \pm 0.203$	0.2336	$0.2912 \pm 0.0091$	0.0360	$0.2203 \pm 0.089$	0.0320	$0.3084 \pm 0.202$	<b>0.0235</b>	$0.1779 \pm 0.086$	0.0621
12	$0.3001 \pm 0.179$	0.2478	$0.3161 \pm 0.0093$	0.0359	$0.2104 \pm 0.077$	0.0318	$0.2309 \pm 0.138$	0.0235	$0.1801 \pm 0.079$	0.0620
13	$0.28 \pm 0.178$	0.2619	$0.2999 \pm 0.0084$	0.0360	$0.2105 \pm 0.072$	<b>0.0318</b>	$0.191 \pm 0.107$	0.0235	$0.1727 \pm 0.075$	0.0622
14	$0.3038 \pm 0.177$	0.2771	$0.2982 \pm 0.0082$	0.0405	$0.2003 \pm 0.063$	0.0384	$0.1651 \pm 0.078$	0.0282	$0.1587 \pm 0.068$	0.0768
15	$0.3115 \pm 0.185$	0.2933	$0.2946 \pm 0.0081$	0.0386	$0.2104 \pm 0.067$	0.0349	$0.1611 \pm 0.081$	0.0257	$0.1677 \pm 0.080$	0.0677
16	$0.291 \pm 0.168$	0.3065	$0.2891 \pm 0.0085$	0.0360	$0.1878 \pm 0.062$	0.0319	$0.144 \pm 0.058$	0.0235	$0.1627 \pm 0.065$	0.0629
17	$0.3058 \pm 0.174$	0.3215	$0.2837 \pm 0.0085$	0.0365	$0.184 \pm 0.057$	0.0319	$0.1416 \pm 0.049$	0.0235	$0.1539 \pm 0.061$	0.0631
18	$0.3002 \pm 0.170$	0.3371	$0.2858 \pm 0.0071$	0.0367	$0.187 \pm 0.064$	0.0322	$0.1396 \pm 0.049$	0.0237	$0.1578 \pm 0.069$	0.0637
19	$0.3028 \pm 0.165$	0.3508	$0.2818 \pm 0.0063$	0.0367	$0.1927 \pm 0.062$	0.0320	$0.145 \pm 0.049$	0.0236	$0.1521 \pm 0.063$	0.0638
20	$0.2977 \pm 0.177$	0.3662	$0.2629 \pm 0.0066$	0.0366	$0.1808 \pm 0.057$	0.0320	$0.1329 \pm 0.040$	0.0236	$0.1376 \pm 0.050$	0.0640



**Figura 6-15.:** Desempeño de cada uno de los clasificadores variando el número de objetos de representación por clase. Conjunto de datos laryngeal 2, con 53 objetos para la clase de voces normales y 639 objetos para la clase de voces patológicas y dimensión del conjunto igual a 16. **Selección de las líneas de características de longitud media.**

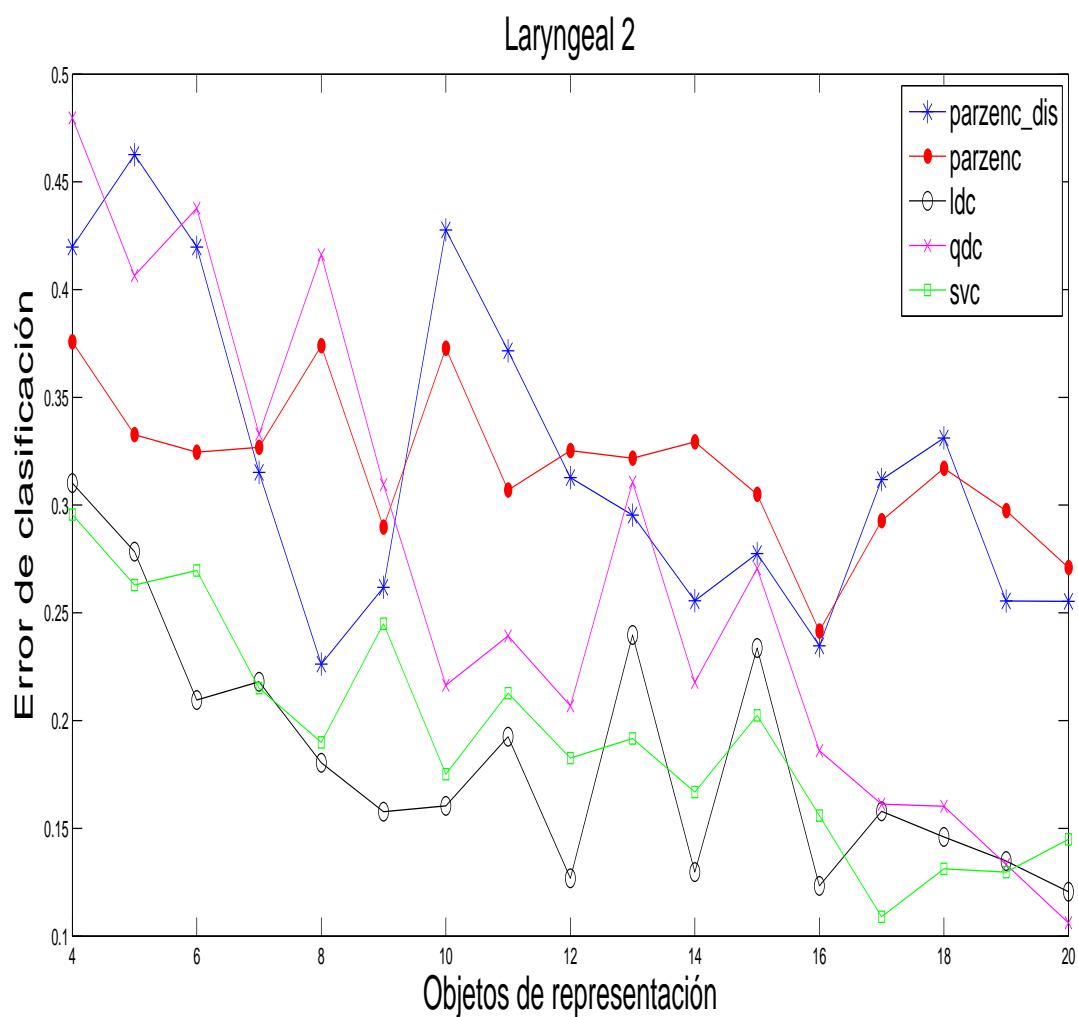
**Tabla 6-29.:** Intervalos de confianza para el error promedio de cada clasificador de acuerdo al número de objetos de representación por clase (Resaltado se encuentra el mejor y el peor intervalo para el clasificador *parzenc\_dis*).

Intervalos de confianza para el error promedio					
Objetos de representación	<i>parzenc_dis</i>	<i>parzenc</i>	<i>ldc</i>	<i>qdc</i>	<i>svc</i>
4	[0.3021,0.5015]	[0.3274,0.5296]	[0.2360,0.4591]	[0.3184,0.6196]	[0.2531,0.4675]
5	[0.2774,0.5276]	[0.2747,0.4478]	[0.2399,0.4562]	[0.2951,0.5600]	[0.2016,0.4141]
6	<b>[0.2207,0.4536]</b>	[0.2691,0.3972]	[0.2247,0.4729]	[0.1315,0.3355]	[0.1161,0.2250]
7	[0.2489,0.4646]	[0.2778,0.3940]	[0.2392,0.4191]	[0.1849,0.3804]	[0.1806,0.3152]
8	[0.2671,0.4697]	[0.2715,0.3897]	[0.1708,0.3144]	[0.1561,0.3240]	[0.1838,0.2978]
9	[0.2548,0.4856]	[0.2243,0.3413]	[0.1622,0.2929]	[0.1109,0.3127]	[0.1227,0.2236]
10	[0.2929,0.5263]	[0.2624,0.4082]	[0.1647,0.2819]	[0.1084,0.3321]	[0.1379,0.2517]
11	[0.3162,0.5903]	[0.2420,0.3619]	[0.1544,0.2487]	[0.1317,0.2782]	[0.1283,0.2110]
12	[0.3483,0.5339]	[0.2817,0.4129]	[0.1543,0.2136]	[0.1693,0.3075]	[0.1411,0.2351]
13	[0.2702,0.5421]	[0.2686,0.3562]	[0.1309,0.1724]	[0.1706,0.2978]	[0.1049,0.1646]
14	[0.2994,0.5747]	[0.2438,0.3594]	[0.1137,0.1873]	[0.1875,0.4914]	[0.1086,0.1861]
15	[0.3245,0.5464]	[0.2350,0.3330]	[0.1343,0.1896]	[0.1535,0.3895]	[0.1178,0.2047]
16	[0.2251,0.4751]	[0.2368,0.3238]	[0.1209,0.1630]	[0.1336,0.2662]	[0.1071,0.1610]
17	[0.2627,0.4833]	[0.2123,0.3041]	[0.1098,0.1693]	[0.1528,0.2427]	[0.1043,0.1761]
18	[0.2579,0.4703]	[0.2274,0.3134]	[0.0953,0.1696]	[0.1334,0.2451]	[0.1011,0.1691]
19	[0.3539,0.6095]	[0.2277,0.3001]	[0.1068,0.1706]	[0.1388,0.1877]	[0.1010,0.1637]
20	<b>[0.3727,0.5743]</b>	[0.2254,0.3286]	[0.1243,0.1675]	[0.1355,0.2105]	[0.1138,0.1740]



**Tabla 6-30.:** Tabla del Error promedio, desviación estándar, tiempo de cómputo, de los clasificadores de acuerdo al número de objetos de representación por clase (resaltado se encuentra el mejor y el peor tiempo promedio de cómputo para el clasificador *parzenc\_dis*).

Error promedio, desviación estándar, tiempo de cómputo de los clasificadores										
Objetos de representación	<i>parzenc_dis</i>		<i>parzenc</i>		<i>ldc</i>		<i>qdc</i>		<i>svc</i>	
	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)
4	0.4018±0.0388	224.025	0.4285±0.0393	0.0354	0.3475±0.0434	0.0332	0.4690±0.0586	0.0261	0.3603±0.0417	0.0601
5	0.4025±0.0487	206.894	0.3613±0.0337	0.0346	0.3480±0.0421	0.0334	0.4276±0.0515	0.0278	0.3078±0.0414	0.0592
6	<b>0.3371±0.0453</b>	<b>176.653</b>	0.3332±0.0249	0.0444	0.3488±0.0483	0.0326	0.2335±0.0397	0.0285	0.1705±0.0212	14.712
7	0.3568±0.0420	189.073	0.3359±0.0226	0.0356	0.3291±0.0350	0.0316	0.2827±0.0380	0.0244	0.2479±0.0262	0.0596
8	0.3684±0.0394	198.696	0.3306±0.0230	0.0392	0.2426±0.0280	0.0342	0.2401±0.0327	0.0265	0.2408±0.0222	0.0619
9	0.3702±0.0449	199.934	0.2828±0.0228	0.0371	0.2275±0.0254	0.0333	0.2118±0.0393	0.0274	0.1731±0.0196	0.0646
10	0.4096±0.0454	215.617	0.3353±0.0284	0.0378	0.2233±0.0228	0.0340	0.2202±0.0435	0.0281	0.1948±0.0221	0.0639
11	0.4532±0.0533	210.227	0.3019±0.0233	0.0385	0.2016±0.0184	0.0351	0.2049±0.0285	0.0294	0.1696±0.0161	0.0642
12	0.4411±0.0361	198.707	0.3473±0.0255	0.0392	0.1840±0.0115	0.0359	0.2384±0.0269	0.0296	0.1881±0.0183	0.0636
13	0.4062±0.0529	210.641	0.3124±0.0171	0.0396	0.1517±0.0081	0.0357	0.2342±0.0247	0.0294	0.1348±0.0116	0.0642
14	0.4370±0.0536	194.589	0.3016±0.0225	0.0405	0.1505±0.0143	0.0365	0.3395±0.0591	0.0286	0.1474±0.0151	0.0646
15	0.4354±0.0432	189.992	0.2840±0.0191	0.0401	0.1619±0.0107	0.0368	0.2715±0.0459	0.0284	0.1613±0.0169	0.0651
16	0.3501±0.0486	189.122	0.2803±0.0169	0.0421	0.1420±0.0082	0.0374	0.1999±0.0258	0.0292	0.1341±0.0105	0.0664
17	0.3730±0.0429	199.686	0.2582±0.0179	0.0419	0.1396±0.0116	0.0377	0.1977±0.0175	0.0292	0.1402±0.0140	0.0660
18	0.3641±0.0413	189.534	0.2704±0.0167	0.0453	0.1325±0.0145	0.0437	0.1893±0.0217	0.0349	0.1351±0.0132	0.0799
19	0.4817±0.0497	214.616	0.2639±0.0141	0.0456	0.1387±0.0124	0.0420	0.1632±0.0095	0.0340	0.1323±0.0122	0.0761
20	<b>0.4735±0.0392</b>	<b>202.656</b>	0.2770±0.0201	0.0446	0.1459±0.0084	0.0396	0.1730±0.0146	0.0313	0.1439±0.0117	0.0712



**Figura 6-16.:** Desempeño de cada uno de los clasificadores variando el número de objetos de representación por clase. Conjunto de datos laryngeal 2, con 53 objetos para la clase de voces normales y 639 objetos para la clase de voces patológicas y dimensión del conjunto igual a 16. **Selección de las líneas de características de longitud menor.**

**Tabla 6-31.:** Intervalos de confianza para el error promedio de cada clasificador de acuerdo al número de objetos de representación por clase (Resaltado se encuentra el mejor y el peor intervalo para el clasificador *parzenc\_dis*).

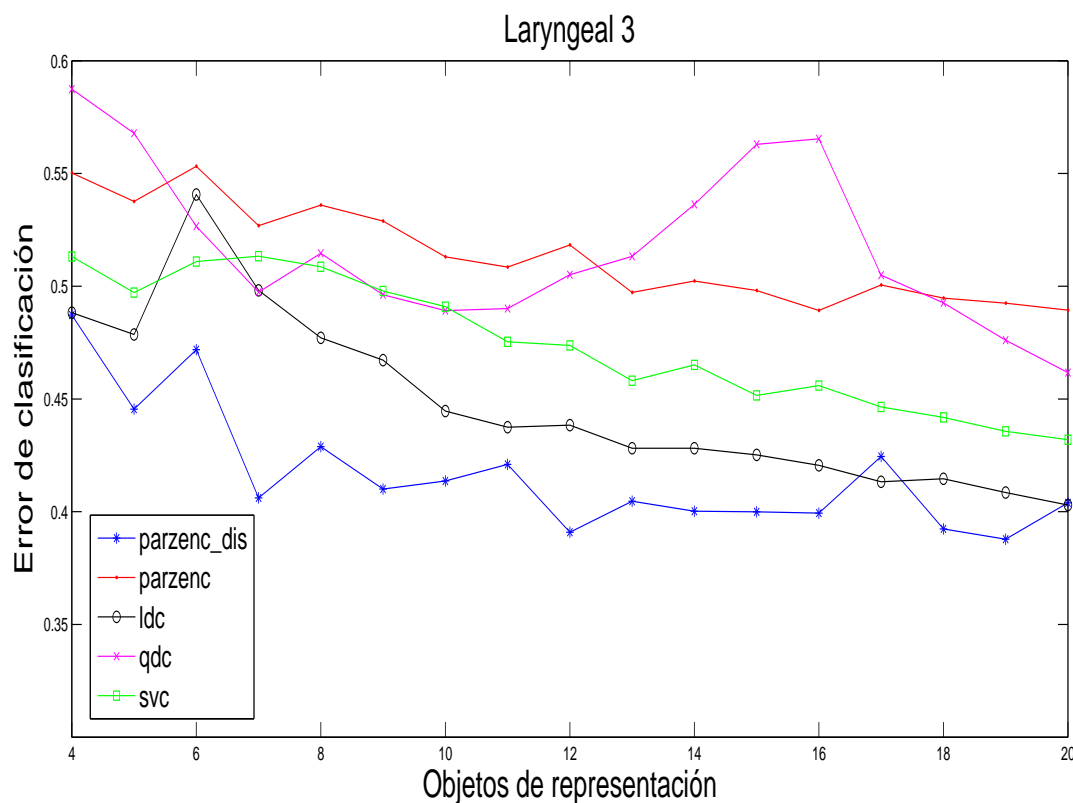
Intervalos de confianza para el error promedio					
Objetos de representación	<i>parzenc_dis</i>	<i>parzenc</i>	<i>ldc</i>	<i>qdc</i>	<i>svc</i>
4	[0.2769,0.5626]	[0.2640,0.4874]	[0.1933,0.4272]	[0.2795,0.6799]	[0.1811,0.4101]
5	<b>[0.2904,0.6348]</b>	[0.2419,0.4235]	[0.1424,0.4145]	[0.2061,0.6068]	[0.1451,0.3807]
6	[0.3005,0.5389]	[0.2399,0.4093]	[0.0995,0.3196]	[0.2297,0.6459]	[0.1674,0.3721]
7	[0.1688,0.4616]	[0.2107,0.4429]	[0.0871,0.3489]	[0.1564,0.5094]	[0.1233,0.3068]
8	<b>[0.1241,0.3283]</b>	[0.2505,0.4974]	[0.1228,0.2381]	[0.2874,0.5449]	[0.1117,0.2682]
9	[0.1029,0.4209]	[0.2350,0.3446]	[0.0949,0.2205]	[0.1996,0.4194]	[0.1689,0.3210]
10	[0.2768,0.5786]	[0.2810,0.4645]	[0.1273,0.1935]	[0.1065,0.3262]	[0.1234,0.2269]
11	[0.2702,0.4731]	[0.2566,0.3574]	[0.1305,0.2545]	[0.1202,0.3583]	[0.1142,0.3111]
12	[0.2204,0.4050]	[0.2410,0.4096]	[0.0858,0.1678]	[0.0573,0.3565]	[0.1050,0.2603]
13	[0.1863,0.4044]	[0.2234,0.4201]	[0.1607,0.3189]	[0.1868,0.4345]	[0.1157,0.2678]
14	[0.1736,0.3376]	[0.2420,0.4168]	[0.1011,0.1582]	[0.1379,0.2976]	[0.1218,0.2119]
15	[0.1871,0.3678]	[0.2238,0.3862]	[0.1538,0.3136]	[0.0834,0.4577]	[0.1391,0.2657]
16	[0.1031,0.3663]	[0.1980,0.2851]	[0.0758,0.1708]	[0.0669,0.3052]	[0.0978,0.2140]
17	[0.1815,0.4422]	[0.2410,0.3444]	[0.1092,0.2066]	[0.1018,0.2207]	[0.0816,0.1363]
18	[0.2408,0.4214]	[0.2473,0.3869]	[0.1045,0.1876]	[0.0927,0.2277]	[0.0780,0.1845]
19	[0.1670,0.3440]	[0.2359,0.3589]	[0.1026,0.1672]	[0.0962,0.1701]	[0.0858,0.1735]
20	[0.1294,0.3813]	[0.2136,0.3284]	[0.0964,0.1447]	[0.0580,0.1542]	[0.1049,0.1850]

**Tabla 6-32.:** Tabla del Error promedio, desviación estándar, tiempo de cómputo, de los clasificadores de acuerdo al número de objetos de representación por clase (resaltado se encuentra el mejor tiempo promedio de cómputo para el clasificador *parzenc\_dis*).

Error promedio, desviación estándar, tiempo de cómputo de los clasificadores										
Objetos de representación	<i>parzenc_dis</i>		<i>parzenc</i>		<i>ldc</i>		<i>qdc</i>		<i>svc</i>	
	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)
4	0.4197±0.0556	0.1312	0.3757±0.0435	0.0354	0.3102±0.0455	0.0323	0.4797±0.0779	0.0247	0.2956±0.0445	0.0569
5	<b>0.4626±0.0670</b>	<b>0.1456</b>	0.3327±0.0353	0.0353	0.2784±0.0529	0.0335	0.4065±0.0780	0.0283	0.2629±0.0458	0.0663
6	0.4197±0.0464	0.1608	0.3246±0.0330	0.0421	0.2096±0.0428	0.0412±0.4378	0.0810	0.0325	0.2697±0.0398	0.0761
7	0.3152±0.0570	0.1867	0.3268±0.0452	0.0414	0.2180±0.0509	0.0384	0.3329±0.0687	0.0299	0.2150±0.0357	0.5383
8	<b>0.2262±0.0397</b>	<b>0.1916</b>	0.3740±0.0480	0.0368	0.1805±0.0224	0.0329	0.4161±0.0501	0.0260	0.1899±0.0305	0.0635
9	0.2619±0.0619	0.2057	0.2898±0.0213	0.0425	0.1577±0.0244	0.0419	0.3095±0.0427	0.0322	0.2450±0.0296	0.0846
10	0.4277±0.0587	0.2237	0.3728±0.0357	0.0403	0.1604±0.0129	0.0376	0.2164±0.0427	0.0291	0.1751±0.0201	0.0733
11	0.3716±0.0395	0.2375	0.3070±0.0196	0.0365	0.1925±0.0241	0.0322	0.2393±0.0463	0.0246	0.2127±0.0383	0.0630
12	0.3127±0.0359	0.2555	0.3253±0.0328	0.0366	0.1268±0.0159	0.0321	0.2069±0.0582	0.0249	0.1826±0.0302	0.0629
13	0.2953±0.0424	0.2660	0.3218±0.0383	0.0363	0.2398±0.0308	0.0321	0.3107±0.0482	0.0244	0.1917±0.0296	0.0634
14	0.2556±0.0319	0.2805	0.3294±0.0340	0.0360	0.1297±0.0111	0.0324	0.2178±0.0311	0.0240	0.1669±0.0175	0.0642
15	0.2775±0.0352	0.2961	0.3050±0.0316	0.0368	0.2337±0.0311	0.0321	0.2705±0.0728	0.0238	0.2024±0.0246	0.0635
16	0.2347±0.0512	0.3165	0.2415±0.0169	0.0375	0.1233±0.0185	0.0321	0.1861±0.0464	0.0240	0.1559±0.0226	0.0641
17	0.3119±0.0507	0.3258	0.2927±0.0201	0.0367	0.1579±0.0190	0.0324	0.1612±0.0231	0.0239	0.1090±0.0106	0.0635
18	0.3311±0.0351	0.3419	0.3171±0.0272	0.0368	0.1460±0.0162	0.0321	0.1602±0.0263	0.0238	0.1313±0.0207	0.0640
19	0.2555±0.0344	0.3567	0.2974±0.0239	0.0367	0.1349±0.0126	0.0323	0.1332±0.0144	0.0252	0.1297±0.0171	0.0641
20	0.2554±0.0490	0.3714	0.2710±0.0223	0.0369	0.1206±0.0094	0.0322	0.1061±0.0187	0.0241	0.1449±0.0156	0.0653

### 6.1.10. Conjunto Laryngeal 3

Conjunto de datos del mundo real, con tres clases, voces normales, voces patológicas orgánicas y voces patológicas funcionales, tomado de la academia búlgara de ciencias [16]. La dimensión del conjunto es 16 y se presenta el desempeño de los clasificadores a continuación.



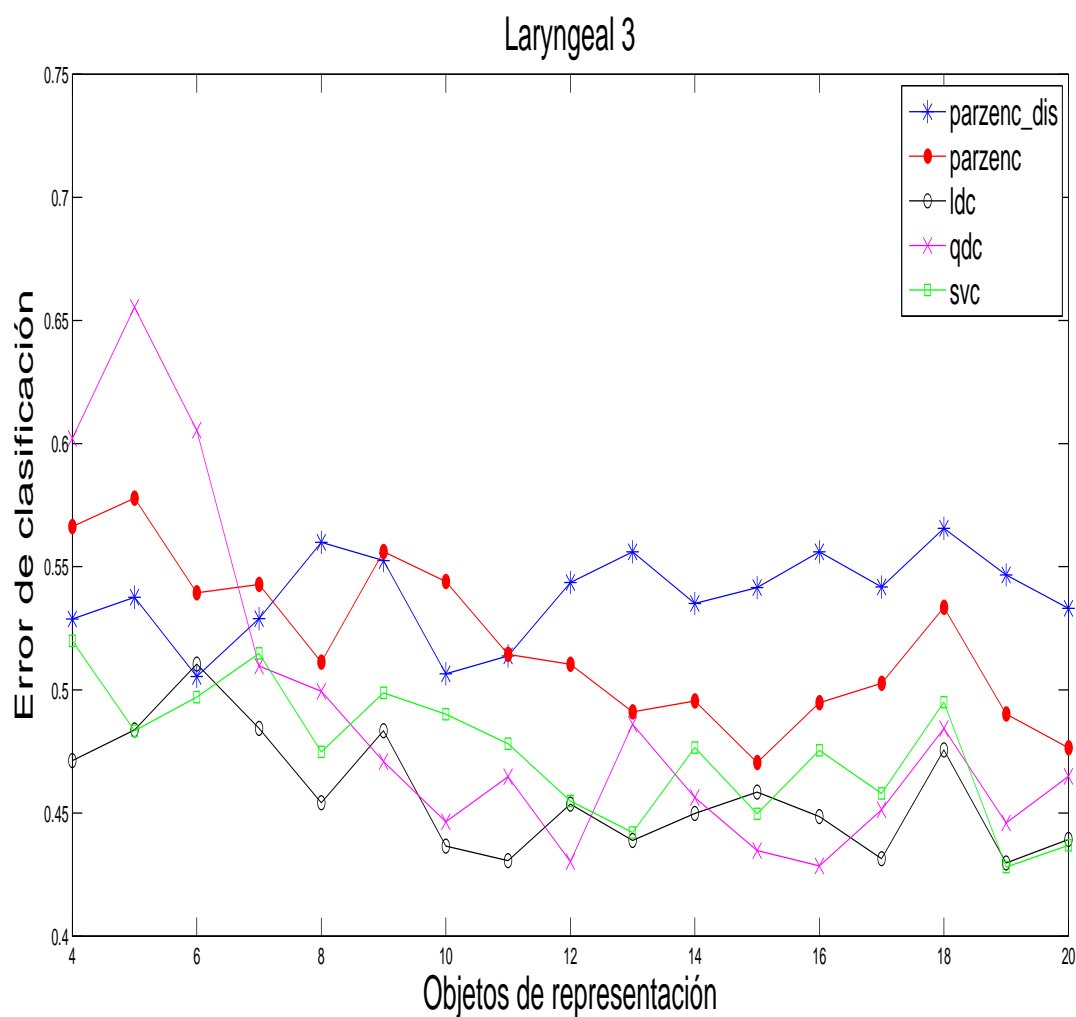
**Figura 6-17.:** Desempeño de cada uno de los clasificadores variando el número de objetos de representación por clase. Conjunto de datos laryngeal 3, con 53 objetos para la clase de voces normales, 218 para la clase de voces patológicas orgánicas y 82 objetos para la clase de voces patológicas funcionales y dimensión del conjunto igual a 16. **Selección de las líneas de características de longitud mayor.**

**Tabla 6-33.:** Intervalos de confianza para el error promedio de cada clasificador de acuerdo al número de objetos de representación por clase (Resaltado se encuentra el mejor intervalo para cada clasificador).

Intervalos de confianza para el error promedio					
Objetos de representación	<i>parzenc_dis</i>	<i>parzenc</i>	<i>ldc</i>	<i>qdc</i>	<i>svc</i>
4	[0.4569,0.5175]	[0.5302,0.5702]	[0.4688,0.5078]	[0.5466,0.6282]	[0.4948,0.5316]
5	[0.4150,0.4760]	[0.5173,0.5579]	[0.4536,0.5036]	[0.5307,0.6049]	[0.4764,0.5180]
6	[0.4372,0.5066]	[0.5335,0.5729]	[0.5144,0.5670]	[0.4937,0.5595]	[0.4912,0.5306]
7	[0.3766,0.4356]	[0.5069,0.5469]	[0.4798,0.5166]	[0.4719,0.5229]	[0.4949,0.5317]
8	[0.3976,0.4602]	[0.5155,0.5565]	[0.4584,0.4958]	[0.4888,0.5404]	[0.4873,0.5299]
9	[0.3826,0.4374]	[0.5099,0.5479]	[0.4490,0.4854]	[0.4751,0.5173]	[0.4795,0.5163]
10	[0.3841,0.4431]	[0.4940,0.5320]	[0.4272,0.4620]	[0.4650,0.5134]	[0.4712,0.5106]
11	[0.3884,0.4536]	[0.4898,0.5272]	[0.4191,0.4559]	[0.4675,0.5127]	[0.4582,0.4924]
12	[0.3656,0.4162]	[0.4999,0.5367]	[0.4218,0.4550]	[0.4811,0.5291]	[0.4593,0.4883]
13	[0.3770,0.4322]	[0.4791,0.5155]	[0.4128,0.4434]	[0.4864,0.5400]	[0.4426,0.4736]
14	[0.3757,0.4247]	[0.4810,0.5236]	[0.4119,0.4445]	[0.5125,0.5599]	[0.4475,0.4827]
15	[0.3737,0.4263]	[0.4820,0.5142]	[0.4089,0.4415]	[0.5369,0.5891]	[0.4369,0.4663]
16	<b>[0.3731,0.4257]</b>	[0.4710,0.5074]	[0.4045,0.4367]	[0.5359,0.5949]	[0.4375,0.4743]
17	[0.3958,0.4532]	[0.4822,0.5190]	[0.3957,0.4309]	[0.4802,0.5296]	[0.4283,0.4647]
18	[0.3684,0.4164]	[0.4765,0.5129]	[0.3991,0.4301]	[0.4718,0.5134]	[0.4236,0.4600]
19	[0.3610,0.4146]	[0.4762,0.5088]	[0.3964,0.4206]	[0.4585,0.4937]	[0.4203,0.4509]
20	[0.3739,0.4339]	<b>[0.4712,0.5076]</b>	<b>[0.3891,0.4169]</b>	<b>[0.4462,0.4772]</b>	<b>[0.4161,0.4477]</b>

**Tabla 6-34.:** Tabla del Error promedio, desviación estándar, tiempo de cómputo, de los clasificadores de acuerdo al número de objetos de representación por clase (resaltado se encuentra el mejor tiempo promedio de cómputo de cada clasificador).

Error promedio, desviación estándar, tiempo de cómputo de los clasificadores										
Objetos de representación	<i>parzenc_dis</i>		<i>parzenc</i>		<i>ldc</i>		<i>qdc</i>		<i>svc</i>	
	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)
4	$0.4872 \pm 0.115$	<b>0.2061</b>	$0.5502 \pm 0.076$	0.0451	$0.4883 \pm 0.074$	0.0456	$0.5874 \pm 0.155$	0.0390	$0.5132 \pm 0.070$	0.2335
5	$0.4455 \pm 0.116$	0.2367	$0.5376 \pm 0.077$	0.0460	$0.4786 \pm 0.095$	0.0495	$0.5678 \pm 0.141$	0.0504	$0.4972 \pm 0.079$	0.2417
6	$0.4719 \pm 0.132$	0.2723	$0.5532 \pm 0.075$	0.0476	$0.5407 \pm 0.100$	0.0520	$0.5266 \pm 0.125$	0.0593	$0.5109 \pm 0.075$	0.2620
7	$0.4061 \pm 0.112$	0.3058	$0.5269 \pm 0.076$	0.0504	$0.4982 \pm 0.070$	0.0568	$0.4974 \pm 0.097$	0.0718	$0.5133 \pm 0.070$	0.2723
8	$0.4289 \pm 0.119$	0.3395	$0.536 \pm 0.078$	0.0492	$0.4771 \pm 0.071$	0.0562	$0.5146 \pm 0.098$	0.0693	$0.5086 \pm 0.081$	0.2732
9	$0.41 \pm 0.104$	0.3713	$0.5289 \pm 0.072$	0.0419	$0.4672 \pm 0.069$	0.0433	$0.4962 \pm 0.080$	0.0371	$0.4979 \pm 0.070$	0.2590
10	$0.4136 \pm 0.112$	0.4053	$0.513 \pm 0.072$	0.0492	$0.4446 \pm 0.066$	0.0484	$0.4892 \pm 0.092$	0.0349	$0.4909 \pm 0.075$	0.2485
11	$0.421 \pm 0.124$	0.4390	$0.5085 \pm 0.071$	0.0510	$0.4375 \pm 0.070$	0.0479	$0.4901 \pm 0.086$	0.0381	$0.4753 \pm 0.065$	0.2631
12	$0.3909 \pm 0.096$	0.4735	$0.5183 \pm 0.07$	0.0519	$0.4384 \pm 0.063$	0.0526	$0.5051 \pm 0.091$	0.0369	$0.4738 \pm 0.055$	0.2640
13	$0.4046 \pm 0.105$	0.5079	$0.4973 \pm 0.069$	0.0469	$0.4281 \pm 0.058$	0.0448	$0.5132 \pm 0.102$	0.0327	$0.4581 \pm 0.059$	0.2388
14	$0.4002 \pm 0.093$	0.5430	$0.5023 \pm 0.081$	0.0469	$0.4282 \pm 0.062$	0.0444	$0.5362 \pm 0.090$	0.0328	$0.4651 \pm 0.067$	0.2402
15	$0.4 \pm 0.100$	0.5768	$0.4981 \pm 0.061$	0.0506	$0.4252 \pm 0.062$	0.0487	$0.563 \pm 0.099$	0.0358	$0.4516 \pm 0.056$	0.2663
16	$0.3994 \pm 0.100$	0.6111	$0.4892 \pm 0.069$	0.0459	$0.4206 \pm 0.061$	0.0432	$0.5654 \pm 0.112$	0.0296	$0.4559 \pm 0.070$	0.2370
17	$0.4245 \pm 0.109$	0.6450	$0.5006 \pm 0.07$	<b>0.0434</b>	$0.4133 \pm 0.067$	0.0393	$0.5049 \pm 0.094$	<b>0.0270</b>	$0.4465 \pm 0.069$	0.2257
18	$0.3924 \pm 0.091$	0.6802	$0.4947 \pm 0.069$	0.0437	$0.4146 \pm 0.059$	<b>0.0393</b>	$0.4926 \pm 0.079$	0.0270	$0.4418 \pm 0.069$	0.2286
19	$0.3878 \pm 0.102$	0.7151	$0.4925 \pm 0.062$	0.0441	$0.4085 \pm 0.046$	0.0393	$0.4761 \pm 0.067$	<b>0.0270</b>	$0.4356 \pm 0.058$	<b>0.2320</b>
20	$0.4039 \pm 0.114$	0.7497	$0.4894 \pm 0.069$	0.0441	$0.403 \pm 0.053$	0.0394	$0.4617 \pm 0.059$	0.0270	$0.4319 \pm 0.060$	0.2339



**Figura 6-18.:** Desempeño de cada uno de los clasificadores variando el número de objetos de representación por clase. Conjunto de datos laryngeal 3, con 53 objetos para la clase de voces normales, 218 para la clase de voces patológicas orgánicas y 82 objetos para la clase de voces patológicas funcionales y dimensión del conjunto igual a 16. **Selección de las líneas de características de longitud menor.**

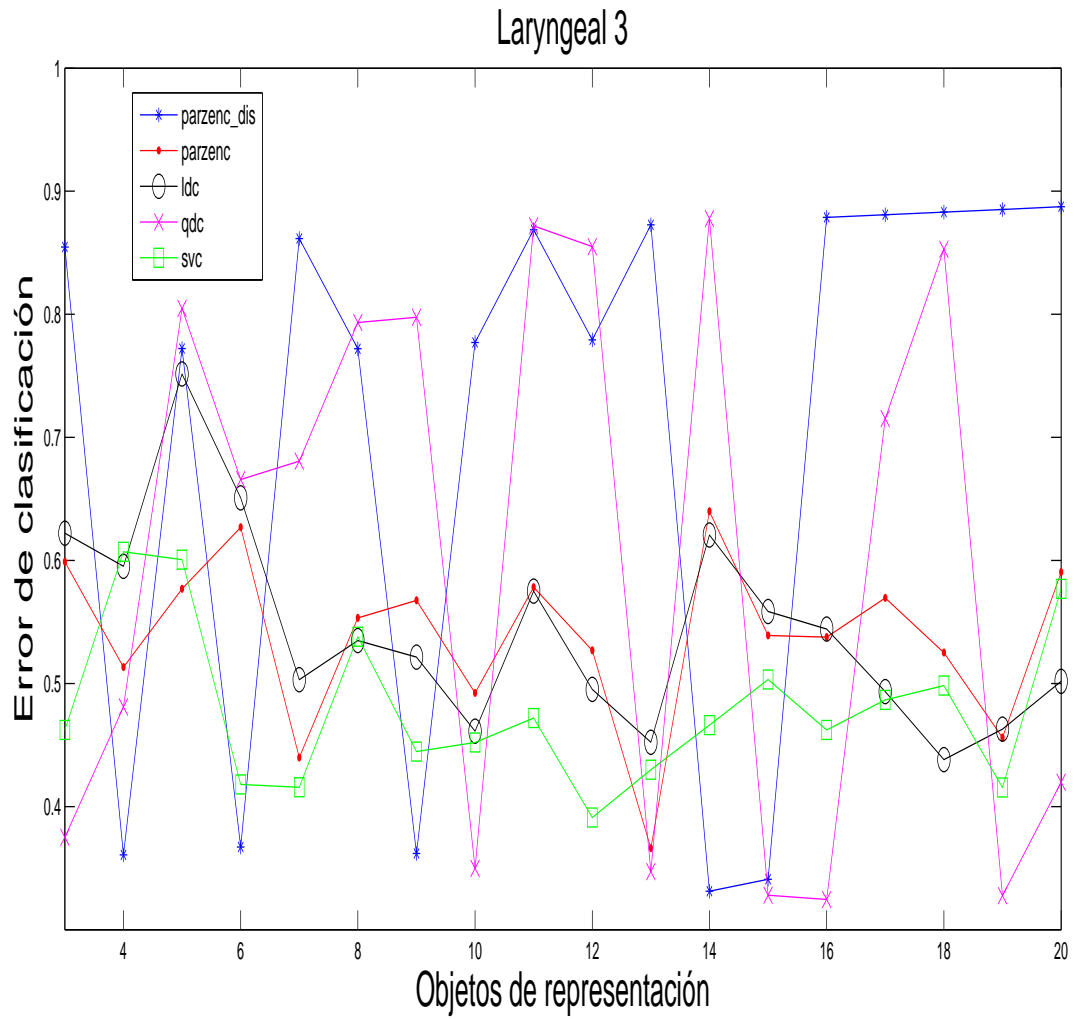


**Tabla 6-35.:** Intervalos de confianza para el error promedio de cada clasificador de acuerdo al número de objetos de representación por clase (Resaltado se encuentra el mejor y el peor intervalo para el clasificador *parzenc\_dis*).

Intervalos de confianza para el error promedio					
Objetos de representación	<i>parzenc_dis</i>	<i>parzenc</i>	<i>ldc</i>	<i>qdc</i>	<i>svc</i>
4	[0.4480,0.6095]	[0.4919,0.6407]	[0.3775,0.5650]	[0.4636,0.7405]	[0.4653,0.5745]
5	[0.4875,0.5877]	[0.5160,0.6397]	[0.4484,0.5191]	[0.5135,0.7972]	[0.4217,0.5452]
6	<b>[0.4192,0.5916]</b>	[0.4932,0.5856]	[0.3770,0.6439]	[0.4889,0.7219]	[0.4187,0.5754]
7	[0.4543,0.6035]	[0.4911,0.5945]	[0.4025,0.5661]	[0.4086,0.6107]	[0.4486,0.5810]
8	<b>[0.4947,0.6251]</b>	[0.4476,0.5749]	[0.3904,0.5178]	[0.3839,0.6149]	[0.4160,0.5335]
9	[0.4578,0.6471]	[0.4893,0.6230]	[0.4285,0.5383]	[0.4081,0.5336]	[0.4488,0.5487]
10	[0.4215,0.5915]	[0.4819,0.6060]	[0.4097,0.4633]	[0.4017,0.4912]	[0.4238,0.5564]
11	[0.4558,0.5717]	[0.4521,0.5766]	[0.3775,0.4838]	[0.3932,0.5362]	[0.4223,0.5340]
12	[0.4607,0.6264]	[0.4700,0.5508]	[0.4068,0.5004]	[0.3934,0.4672]	[0.3963,0.5135]
13	[0.4462,0.6659]	[0.4069,0.5753]	[0.4002,0.4775]	[0.4228,0.5492]	[0.3583,0.5258]
14	[0.4475,0.6226]	[0.4271,0.5639]	[0.3934,0.5063]	[0.3986,0.5140]	[0.4278,0.5253]
15	[0.4569,0.6263]	[0.4261,0.5148]	[0.4203,0.4966]	[0.3732,0.4962]	[0.4029,0.4965]
16	[0.4499,0.6623]	[0.4527,0.5368]	[0.4368,0.4603]	[0.3754,0.4816]	[0.4279,0.5229]
17	[0.4608,0.6226]	[0.4564,0.5489]	[0.3872,0.4757]	[0.4053,0.4973]	[0.4224,0.4934]
18	[0.4731,0.6580]	[0.4863,0.5806]	[0.4407,0.5105]	[0.3993,0.5692]	[0.4578,0.5322]
19	[0.4289,0.6643]	[0.4214,0.5590]	[0.3839,0.4755]	[0.3797,0.5122]	[0.3795,0.4766]
20	[0.4488,0.6174]	[0.4290,0.5239]	[0.3973,0.4812]	[0.3812,0.5485]	[0.4019,0.4718]

**Tabla 6-36.:** Tabla del Error promedio, desviación estándar, tiempo de cómputo, de los clasificadores de acuerdo al número de objetos de representación por clase (resaltado se encuentra el mejor tiempo promedio de cómputo de cada clasificador).

Error promedio, desviación estándar, tiempo de cómputo para el clasificador <i>parzenc_dis</i>										
Objetos de representación	<i>parzenc_dis</i>		<i>parzenc</i>		<i>ldc</i>		<i>qdc</i>		<i>svc</i>	
	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)
4	<b>0.5287±0.0314</b>	<b>0.1981</b>	0.5663±0.0289	0.0498	0.4713±0.0365	0.0790	0.6021±0.0539	0.0327	0.5199±0.0212	0.3149
5	0.5376±0.0195	0.2366	0.5778±0.0241	0.0415	0.4837±0.0138	0.0391	0.6553±0.0552	0.0289	0.4834±0.0240	0.2003
6	0.5054±0.0335	0.2617	0.5394±0.0180	0.0417	0.5104±0.0519	0.0386	0.6054±0.0453	0.0295	0.4970±0.0305	0.2138
7	0.5289±0.0290	0.2942	0.5428±0.0201	0.0423	0.4843±0.0318	0.0391	0.5096±0.0393	0.0299	0.5148±0.0258	0.2161
8	0.5599±0.0254	0.3276	0.5112±0.0248	0.0429	0.4541±0.0248	0.0397	0.4994±0.0449	0.0310	0.4748±0.0229	0.2196
9	0.5525±0.0368	0.3609	0.5561±0.0260	0.0433	0.4834±0.0214	0.0399	0.4709±0.0244	0.0313	0.4988±0.0194	0.2154
10	0.5065±0.0331	0.3946	0.5440±0.0241	0.0447	0.4365±0.0104	0.0413	0.4464±0.0174	0.0349	0.4901±0.0258	0.2393
11	0.5138±0.0225	0.4278	0.5144±0.0242	0.0448	0.4306±0.0207	0.0411	0.4647±0.0278	0.0334	0.4781±0.0217	0.2269
12	0.5435±0.0322	0.4616	0.5104±0.0157	0.0464	0.4536±0.0182	0.0419	0.4303±0.0144	0.0356	0.4549±0.0228	0.2373
13	0.5561±0.0427	0.4946	0.4911±0.0328	0.0462	0.4389±0.0151	0.0421	0.4860±0.0246	0.0354	0.4420±0.0326	0.2331
14	0.5350±0.0341	0.5381	0.4955±0.0266	0.0457	0.4498±0.0220	0.0418	0.4563±0.0224	0.0352	0.4765±0.0190	0.2476
15	0.5416±0.0330	0.5691	0.4705±0.0173	0.0493	0.4584±0.0149	0.0479	0.4347±0.0239	0.0357	0.4497±0.0182	0.2505
16	0.5561±0.0413	0.5967	0.4948±0.0164	0.0577	0.4485±0.0046	0.0528	0.4285±0.0207	0.0392	0.4754±0.0185	0.3103
17	0.5417±0.0315	0.6374	0.5026±0.0180	0.0472	0.4315±0.0172	0.0419	0.451±0.0179	0.0295	0.4579±0.0138	0.2469
18	0.5656±0.0360	0.6728	0.5334±0.0183	0.0599	0.4756±0.0136	0.0539	0.4843±0.0330	0.0405	0.4950±0.0145	0.3108
19	0.5466±0.0458	0.7015	0.4902±0.0268	0.0604	0.4297±0.0178	0.0526	0.4459±0.0258	0.0388	0.4280±0.0189	0.3121
20	0.5331±0.0328	0.7267	0.4765±0.0185	0.0617	0.4392±0.0163	0.0554	0.4648±0.0325	0.0392	0.4369±0.0136	0.3176



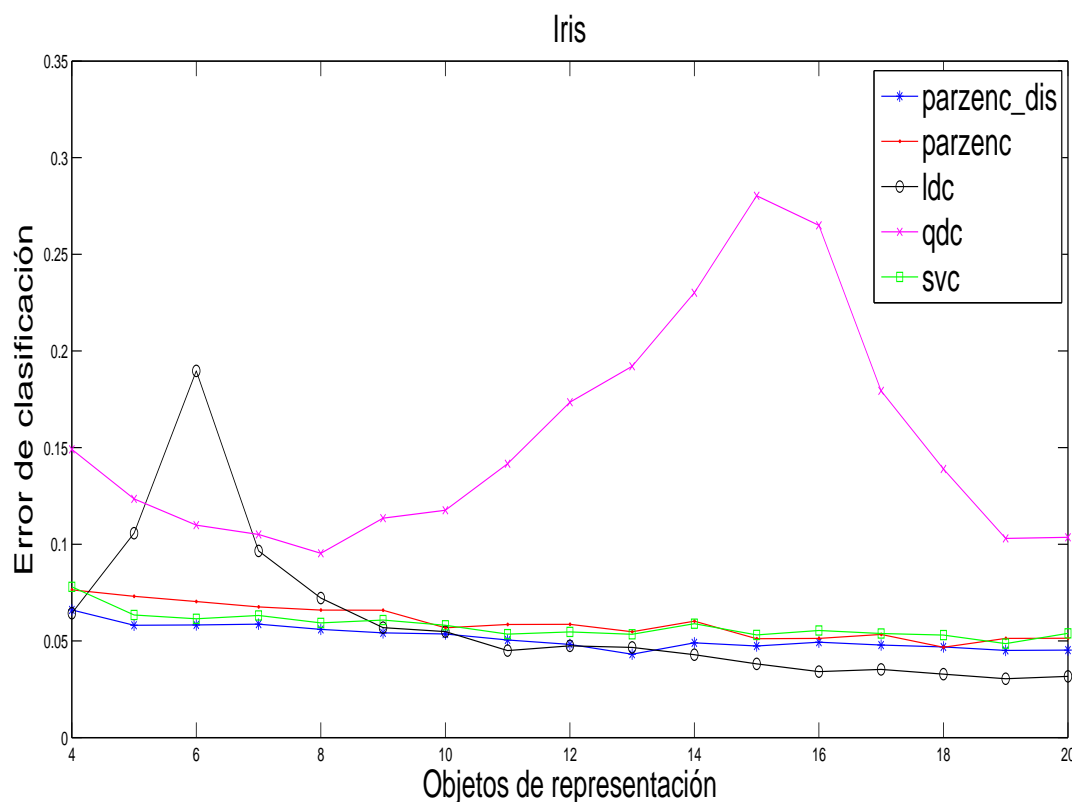
**Figura 6-19.:** Desempeño de cada uno de los clasificadores variando el número de objetos de representación por clase. Conjunto laryngeal 3. **Selección de las líneas de características mediante el algoritmo genético (GA).**

**Tabla 6-37.:** Intervalos de confianza para el error promedio de cada clasificador de acuerdo al número de objetos de representación por clase. **Tiempo de cómputo de la prueba igual a un día, doce horas, 32 minutos y 28 segundos.**

	<i>parzenc_dis</i>		<i>parzenc</i>		<i>ldc</i>		<i>qdc</i>		<i>svc</i>	
Objetos de representación	$\bar{e}_\infty$	CPU(seg)	$\bar{e}_\infty$	CPU(seg)	$\bar{e}_\infty$	CPU(seg)	$\bar{e}_\infty$	CPU(seg)	$\bar{e}_\infty$	CPU(seg)
3	0.8547	0.1118	0.5988	0.0439	0.6221	0.0397	0.3750	0.0274	0.4622	0.2092
4	0.3607	0.1591	0.5132	0.0434	0.5953	0.0393	0.4809	0.0269	0.6070	0.2167
5	0.7722	0.2352	0.5769	0.0428	0.7515	0.0391	0.8047	0.0272	0.6006	0.2004
6	0.3672	0.4845	0.6269	0.0448	0.6507	0.0417	0.6657	0.0276	0.4179	0.2068
7	0.8614	0.6937	0.4398	0.0433	0.5030	0.0398	0.6807	0.0275	0.4157	0.2026
8	0.7720	0.9259	0.5532	0.0434	0.5350	0.0398	0.7933	0.0278	0.5380	0.2040
9	0.3620	13.837	0.5675	0.0436	0.5215	0.0395	0.7975	0.0280	0.4448	0.2042
10	0.7771	18.138	0.4923	0.0440	0.4613	0.0401	0.3498	0.0282	0.4520	0.2052
11	0.8688	24.713	0.5781	0.0533	0.5750	0.0513	0.8719	0.0388	0.4719	0.2628
12	0.7792	29.656	0.5268	0.0448	0.4953	0.0415	0.8549	0.0297	0.3912	0.2089
13	0.8726	43.426	0.3662	0.0449	0.4522	0.0409	0.3471	0.0293	0.4299	0.2086
14	0.3312	51.764	0.6399	0.0447	0.6206	0.0411	0.8778	0.0291	0.4662	0.2078
15	0.3409	67.118	0.5390	0.0451	0.5584	0.0417	0.3279	0.0305	0.5032	0.2098
16	0.8787	79.298	0.5377	0.0588	0.5443	0.0552	0.3246	0.0408	0.4623	0.2858
17	0.8808	96.408	0.5695	0.0461	0.4934	0.0416	0.7152	0.0317	0.4868	0.2078
18	0.8829	122.763	0.5251	0.0596	0.4381	0.0584	0.8528	0.0446	0.4983	0.2920
19	0.8851	137.098	0.4561	0.0477	0.4628	0.0434	0.3277	0.0335	0.4155	0.2132
20	0.8874	164.255	0.5904	0.0573	0.5017	0.0580	0.4198	0.0334	0.5768	0.2194

### 6.1.11. Conjunto iris

Esta es quizás la mejor base de datos conocida que se encuentran en la literatura de reconocimiento de patrones. El conjunto de datos consta de 3 clases de 50 objetos cada una, donde cada una de ellas se refiere a un tipo de planta iris. Una clase es linealmente separable de las otras 2, mientras que las restantes no son linealmente separables una de otra. El desempeño de los clasificadores se presenta a continuación.



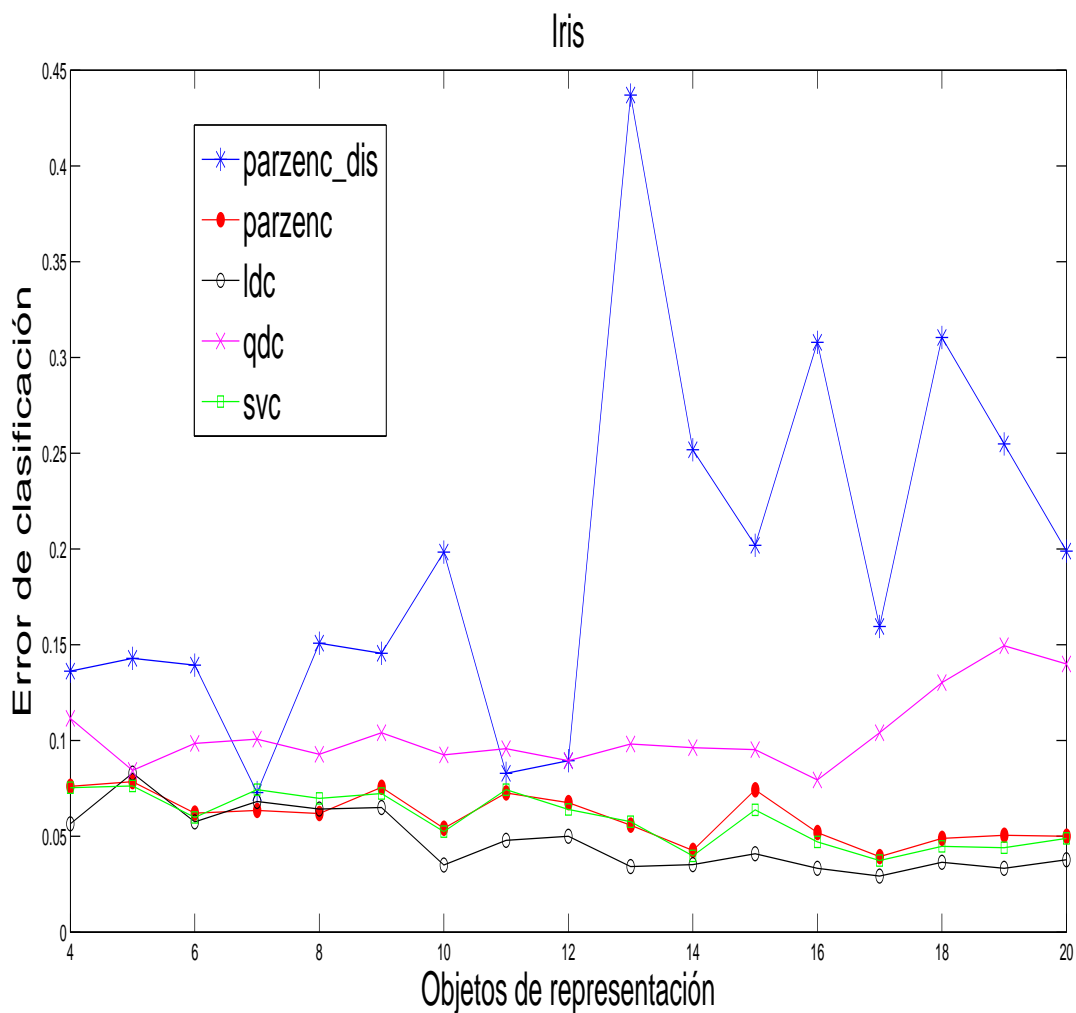
**Figura 6-20.:** Desempeño de cada uno de los clasificadores variando el número de objetos de representación por clase. Conjunto de datos iris, con 50 objetos para cada una de las tres clases y dimensión del conjunto igual a 4. **Selección de las líneas de características de longitud mayor.**

**Tabla 6-38.:** Intervalos de confianza para el error promedio de cada clasificador de acuerdo al número de objetos de representación por clase (Resaltado se encuentra el mejor intervalo para cada clasificador).

Intervalos de confianza para el error promedio					
Objetos de representación	<i>parzenc_dis</i>	<i>parzenc</i>	<i>ldc</i>	<i>qdc</i>	<i>svc</i>
4	[0.0565,0.0755]	[0.0685,0.0843]	[0.0556,0.0730]	[0.1267,0.1715]	[0.0698,0.0862]
5	[0.0492,0.0670]	[0.0648,0.0812]	[0.0791,0.1323]	[0.1053,0.1417]	[0.0560,0.0708]
6	[0.0486,0.0680]	[0.0630,0.0778]	[0.1547,0.2247]	[0.0937,0.1259]	[0.0546,0.0682]
7	[0.0490,0.0684]	[0.0610,0.0742]	[0.0807,0.1123]	[0.0887,0.1213]	[0.0553,0.0711]
8	[0.0465,0.0655]	[0.0589,0.0731]	[0.0613,0.0829]	[0.0827,0.1079]	[0.0511,0.0675]
9	[0.0454,0.0628]	[0.0593,0.0725]	[0.0489,0.0647]	[0.1001,0.1269]	[0.0532,0.0684]
10	[0.0449,0.0623]	[0.0510,0.0626]	[0.0464,0.0632]	[0.1031,0.1321]	[0.0510,0.0652]
11	[0.0438,0.0570]	[0.0524,0.0646]	[0.0395,0.0505]	[0.1242,0.1590]	[0.0474,0.0596]
12	[0.0414,0.0550]	[0.0518,0.0654]	[0.0401,0.0549]	[0.1527,0.1943 ]	[0.0485,0.0607]
13	[0.0368,0.0494]	[0.0484,0.0610]	[0.0400,0.0532]	[0.1692,0.2150]	[0.0471,0.0597]
14	[0.0411,0.0569]	[0.0536,0.0668]	[0.0379,0.0479]	[0.2035,0.2567]	[0.0521,0.0657]
15	[0.0384,0.0562]	[0.0458,0.0564]	[0.0331,0.0431]	[0.2503,0.3103]	[0.0477,0.0583]
16	[0.0422,0.0564]	[0.0458,0.0568]	[0.0304,0.0378]	[0.2305,0.2995]	[0.0491,0.0617]
17	[0.0405,0.0553]	[0.0475,0.0591]	[0.0306,0.0400]	[0.1581,0.2007]	[0.0472,0.0604]
18	[0.0393,0.0545]	[0.0420,0.0514]	[0.0291,0.0365]	[0.1229,0.1551]	[0.0472,0.0588]
19	[0.0385,0.0517]	[0.0458,0.0568]	[0.0267,0.0341]	<b>[0.0922,0.1138]</b>	<b>[0.0425,0.0547]</b>
20	<b>[0.0384,0.0520]</b>	<b>[0.0450,0.0576]</b>	<b>[0.0278,0.0356]</b>	[0.0931,0.1141]	[0.0473,0.0605]

**Tabla 6-39.:** Tabla del Error promedio, desviación estándar, tiempo de cómputo, de los clasificadores de acuerdo al número de objetos de representación por clase (resaltado se encuentra el mejor tiempo promedio de cómputo de cada clasificador).

Error promedio, desviación estándar, tiempo de cómputo de los clasificadores										
Objetos de representación	<i>parzenc_dis</i>		<i>parzenc</i>		<i>ldc</i>		<i>qdc</i>		<i>svc</i>	
	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)
4	0.066 $\pm$ 0.036	<b>0.1935</b>	0.0764 $\pm$ 0.03	0.0412	0.0643 $\pm$ 0.033	0.0383	0.1491 $\pm$ 0.085	0.0267	0.078 $\pm$ 0.031	0.1982
5	0.0581 $\pm$ 0.034	0.2256	0.073 $\pm$ 0.031	<b>0.0402</b>	0.1057 $\pm$ 0.101	0.0376	0.1235 $\pm$ 0.069	0.0265	0.0634 $\pm$ 0.028	<b>0.1920</b>
6	0.0583 $\pm$ 0.037	0.2708	0.0704 $\pm$ 0.028	0.0406	0.1897 $\pm$ 0.133	<b>0.0372</b>	0.1098 $\pm$ 0.061	0.0266	0.0614 $\pm$ 0.026	0.2060
7	0.0587 $\pm$ 0.037	0.3118	0.0676 $\pm$ 0.025	0.0408	0.0965 $\pm$ 0.060	0.0373	0.105 $\pm$ 0.062	0.0267	0.0632 $\pm$ 0.030	0.2092
8	0.056 $\pm$ 0.036	0.3457	0.066 $\pm$ 0.027	0.0407	0.0721 $\pm$ 0.041	<b>0.0372</b>	0.0953 $\pm$ 0.048	0.0265	0.0593 $\pm$ 0.031	0.2091
9	0.0541 $\pm$ 0.033	0.3791	0.0659 $\pm$ 0.025	0.0411	0.0568 $\pm$ 0.030	<b>0.0371</b>	0.1135 $\pm$ 0.051	0.0264	0.0608 $\pm$ 0.029	0.2094
10	0.0536 $\pm$ 0.033	0.4127	0.0568 $\pm$ 0.022	0.0411	0.0548 $\pm$ 0.032	<b>0.0371</b>	0.1176 $\pm$ 0.055	0.0264	0.0581 $\pm$ 0.027	0.2101
11	0.0504 $\pm$ 0.025	0.4472	0.0585 $\pm$ 0.023	0.0414	0.045 $\pm$ 0.021	0.0371	0.1416 $\pm$ 0.066	0.0264	0.0535 $\pm$ 0.023	0.2111
12	0.0482 $\pm$ 0.026	0.4840	0.0586 $\pm$ 0.026	0.0417	0.0475 $\pm$ 0.028	0.0372	0.1735 $\pm$ 0.079	0.0265	0.0546 $\pm$ 0.023	0.2126
13	0.0431 $\pm$ 0.024	0.5188	0.0547 $\pm$ 0.024	0.0418	0.0466 $\pm$ 0.025	0.0372	0.1921 $\pm$ 0.087	0.0265	0.0534 $\pm$ 0.024	0.2139
14	0.049 $\pm$ 0.030	0.5533	0.0602 $\pm$ 0.025	0.0420	0.0429 $\pm$ 0.019	0.0373	0.2301 $\pm$ 0.101	0.0265	0.0589 $\pm$ 0.026	0.2152
15	0.0473 $\pm$ 0.034	0.5880	0.0511 $\pm$ 0.02	0.0423	0.0381 $\pm$ 0.019	0.0371	0.2803 $\pm$ 0.114	0.0262	0.053 $\pm$ 0.020	0.2161
16	0.0493 $\pm$ 0.027	0.6233	0.0513 $\pm$ 0.021	0.0423	0.0341 $\pm$ 0.014	0.0371	0.265 $\pm$ 0.131	<b>0.0254</b>	0.0554 $\pm$ 0.024	0.2170
17	0.0479 $\pm$ 0.028	0.6578	0.0533 $\pm$ 0.022	0.0424	0.0353 $\pm$ 0.018	0.0371	0.1794 $\pm$ 0.081	<b>0.0254</b>	0.0538 $\pm$ 0.025	0.2185
18	0.0469 $\pm$ 0.029	0.6921	0.0467 $\pm$ 0.018	0.0493	0.0328 $\pm$ 0.014	0.0422	0.139 $\pm$ 0.061	0.0278	0.053 $\pm$ 0.022	0.2545
19	0.0451 $\pm$ 0.025	0.7266	0.0513 $\pm$ 0.021	0.0558	0.0304 $\pm$ 0.014	0.0457	0.103 $\pm$ 0.041	0.0290	0.0486 $\pm$ 0.023	0.2851
20	0.0452 $\pm$ 0.026	0.7623	0.0513 $\pm$ 0.024	0.0564	0.0317 $\pm$ 0.015	0.0457	0.1036 $\pm$ 0.040	0.0287	0.0539 $\pm$ 0.025	0.2940



**Figura 6-21.:** Desempeño de cada uno de los clasificadores variando el número de objetos de representación por clase. Conjunto de datos iris, con 50 objetos para cada una de las tres clases y dimensión del conjunto igual a 4. **Selección de las líneas de características de longitud menor.**

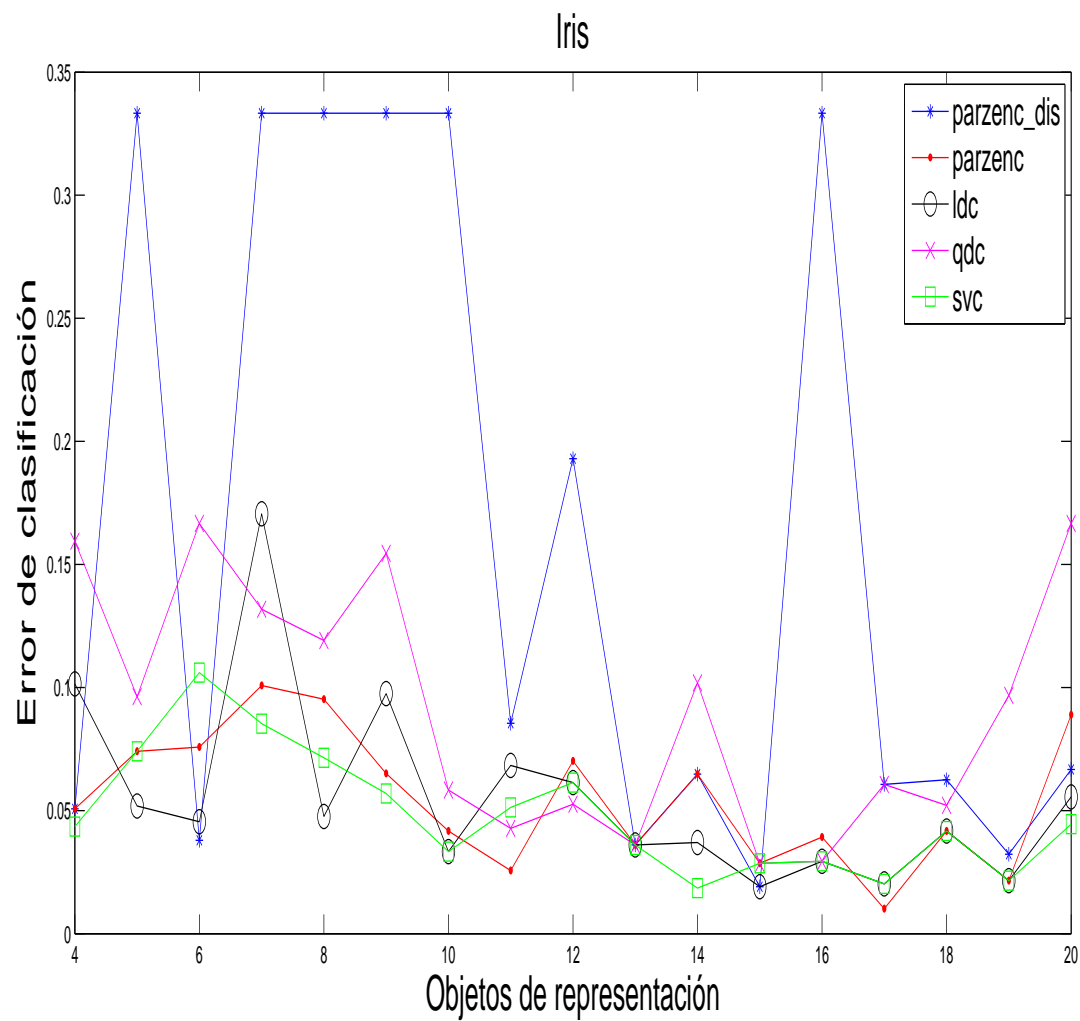


**Tabla 6-40.:** Intervalos de confianza para el error promedio de cada clasificador de acuerdo al número de objetos de representación por clase (Resaltado se encuentra el mejor y el peor intervalo para el clasificador *parzenc\_dis*).

Intervalos de confianza para el error promedio					
Objetos de representación	<i>parzenc_dis</i>	<i>parzenc</i>	<i>ldc</i>	<i>qdc</i>	<i>svc</i>
4	[0.0179,0.2904]	[0.0426,0.1095]	[0.0177,0.0954]	[0.0445,0.1787]	[0.0385,0.1122]
5	<b>[0.0090,0.2949]</b>	[0.0565,0.1005]	[0.0313,0.1347]	[0.0519,0.1170]	[0.0501,0.1025]
6	[0.0136,0.2924]	[0.0372,0.0871]	[0.0294,0.0857]	[0.0664,0.1305]	[0.0330,0.0867]
7	[0.0552,0.0905]	[0.0441,0.0830]	[0.0453,0.0912]	[0.0617,0.1398]	[0.0514,0.0975]
8	[0.0014,0.3002]	[0.0390,0.0848]	[0.0478,0.0808]	[0.0579,0.1278]	[0.0507,0.0890]
9	[0.0075,0.2985]	[0.0464,0.1048]	[0.0372,0.0929]	[0.0527,0.1554]	[0.0393,0.1054]
10	[0.0034,0.4001]	[0.0349,0.0734]	[0.0236,0.0464]	[0.0503,0.1347]	[0.0316,0.0734]
11	[0.0595,0.1063]	[0.0430,0.1023]	[0.0274,0.0684]	[0.0581,0.1333]	[0.0445,0.1042]
12	[0.0660,0.1130]	[0.0501,0.0850]	[0.0332,0.0668]	[0.0631,0.1159]	[0.0417,0.0863]
13	<b>[0.1930,0.6809]</b>	[0.0394,0.0723]	[0.0185,0.0500]	[0.0701,0.1263]	[0.0337,0.0816]
14	[0.0185,0.4852]	[0.0313,0.0539]	[0.0292,0.0411]	[0.0533,0.1393]	[0.0286,0.0511]
15	[0.0015,0.4023]	[0.0458,0.1027]	[0.0283,0.0536]	[0.0551,0.1354]	[0.0314,0.0963]
16	[0.0554,0.5603]	[0.0288,0.0751]	[0.0165,0.0502]	[0.0410,0.1178]	[0.0225,0.0716]
17	[0.0118,0.3074]	[0.0247,0.0541]	[0.0180,0.0405]	[0.0480,0.1601]	[0.0209,0.0538]
18	[0.0600,0.5608]	[0.0376,0.0603]	[0.0265,0.0464]	[0.0870,0.1734]	[0.0304,0.0592]
19	[0.0224,0.4873]	[0.0340,0.0670]	[0.0237,0.0430]	[0.0505,0.2485]	[0.0269,0.0613]
20	[0.0051,0.4028]	[0.0318,0.0682]	[0.0264,0.0492]	[0.0767,0.2033]	[0.0323,0.0655]

**Tabla 6-41.:** Tabla del Error promedio, desviación estándar, tiempo de cómputo, de los clasificadores de acuerdo al número de objetos de representación por clase (resaltado se encuentra el mejor tiempo promedio de cómputo para el clasificador *parzenc\_dis*).

Error promedio, desviación estándar, tiempo de cómputo de los clasificadores										
Objetos de representación	<i>parzenc_dis</i>		<i>parzenc</i>		<i>ldc</i>		<i>qdc</i>		<i>svc</i>	
	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)
4	<b>0.1362±0.0600</b>	<b>0.2176</b>	0.0761±0.0130	0.0428	0.0565±0.0151	0.0402	0.1116±0.0261	0.0288	0.0754±0.0143	0.2950
5	0.1430±0.0591	0.2444	0.0785±0.0086	0.0410	0.0830±0.0201	0.0379	0.0844±0.0126	0.0274	0.0763±0.0102	0.2042
6	0.1394±0.0595	0.2802	0.0621±0.0097	0.0418	0.0576±0.0110	0.0381	0.0985±0.0125	0.0274	0.0598±0.0104	0.2018
7	0.0729±0.0069	0.3179	0.0636±0.0076	0.0414	0.0682±0.0089	0.0383	0.1008±0.0152	0.0280	0.0744±0.0090	0.2092
8	0.1508±0.0581	0.3533	0.0619±0.0089	0.0422	0.0643±0.0064	0.0382	0.0929±0.0136	0.0287	0.0698±0.0075	0.2120
9	0.1455±0.0595	0.3880	0.0756±0.0114	0.0419	0.0650±0.0108	0.0379	0.1041±0.0200	0.0276	0.0724±0.0129	0.2113
10	0.1983±0.0785	0.4240	0.0542±0.0075	0.0421	0.0350±0.0044	0.0382	0.0925±0.0164	0.0278	0.0525±0.0082	0.2129
11	0.0829±0.0091	0.4619	0.0726±0.0116	0.0415	0.0479±0.0080	0.0384	0.0957±0.0146	0.0277	0.0744±0.0116	0.2137
12	0.0895±0.0091	0.4961	0.0675±0.0068	0.0421	0.0500±0.0065	0.0377	0.0895±0.0103	0.0275	0.0640±0.0087	0.2140
13	0.4369±0.0949	0.5270	0.0559±0.0064	0.0425	0.0342±0.0061	0.0380	0.0982±0.0109	0.0284	0.0577±0.0093	0.2132
14	0.2519±0.0908	0.5703	0.0426±0.0044	0.0427	0.0352±0.0023	0.0379	0.0963±0.0167	0.0276	0.0398±0.0044	0.2154
15	0.2019±0.0780	0.6049	0.0743±0.0111	0.0424	0.0410±0.0049	0.0379	0.0952±0.0156	0.0274	0.0638±0.0126	0.2218
16	0.3078±0.0982	0.6512	0.0520±0.0090	0.0449	0.0333±0.0066	0.0388	0.0794±0.0149	0.0288	0.0471±0.0096	0.2244
17	0.1596±0.0575	0.6864	0.0394±0.0057	0.0458	0.0293±0.0044	0.0381	0.1040±0.0218	0.0271	0.0374±0.0064	0.2271
18	0.3104±0.0974	0.7099	0.0490±0.0044	0.0463	0.0365±0.0039	0.0383	0.1302±0.0168	0.0270	0.0448±0.0056	0.2272
19	0.2548±0.0904	0.7589	0.0505±0.0064	0.0450	0.0333±0.0037	0.0388	0.1495±0.0385	0.0273	0.0441±0.0067	0.2318
20	0.1989±0.0794	0.8078	0.0500±0.0071	0.0480	0.0378±0.0044	0.0407	0.1400±0.0246	0.0292	0.0489±0.0065	0.2486



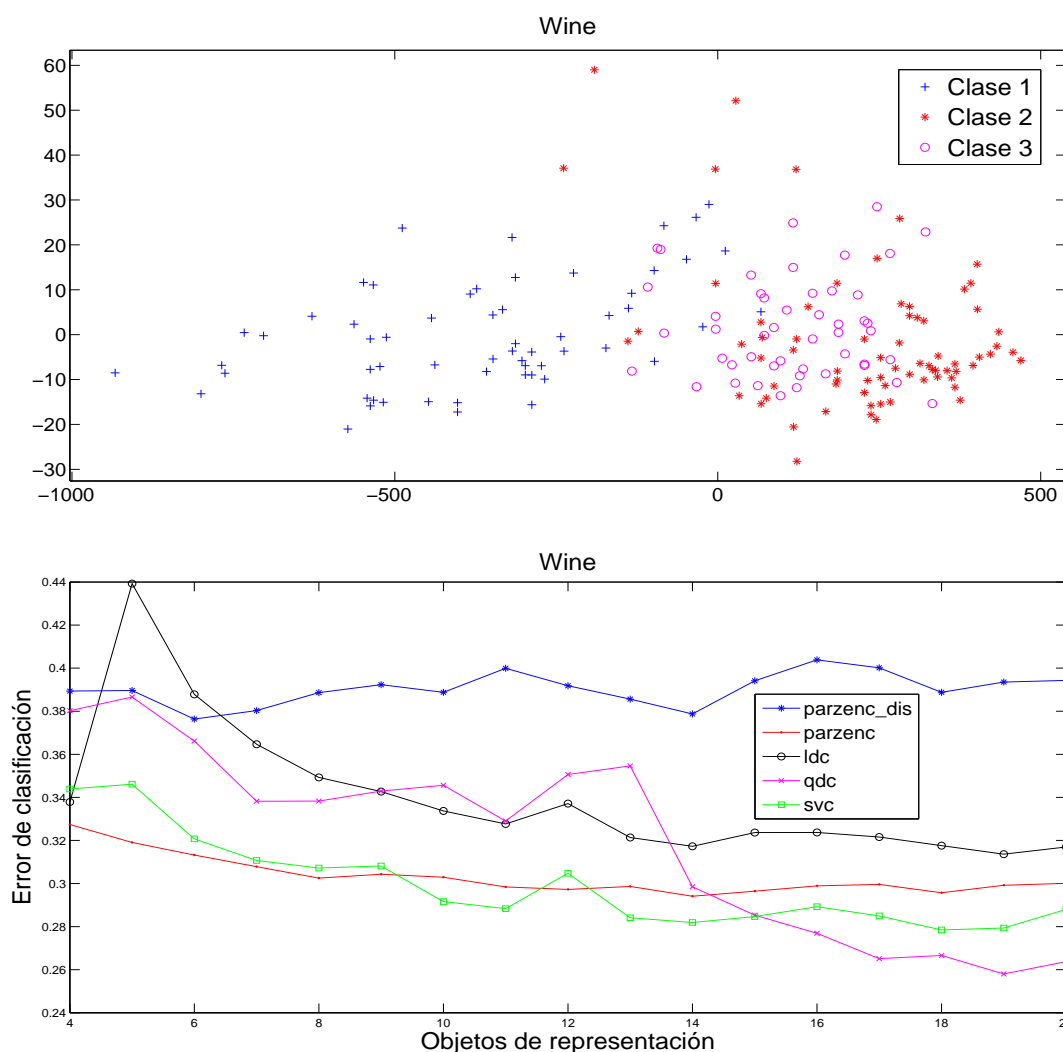
**Figura 6-22.:** Desempeño de cada uno de los clasificadores variando el número de objetos de representación por clase. Conjunto Iris. **Selección de las líneas de características mediante el algoritmo genético (GA).**

**Tabla 6-42.:** Intervalos de confianza para el error promedio de cada clasificador de acuerdo al número de objetos de representación por clase (Resaltado se encuentra el mejor intervalo para cada clasificador). **Tiempo de cómputo de la prueba igual a 8 horas, 25 minutos y 8 segundos.**

	<i>parzenc_dis</i>		<i>parzenc</i>		<i>ldc</i>		<i>qdc</i>		<i>svc</i>	
Objetos de representación	$\bar{e}_\infty$	CPU(seg)	$\bar{e}_\infty$	CPU(seg)	$\bar{e}_\infty$	CPU(seg)	$\bar{e}_\infty$	CPU(seg)	$\bar{e}_\infty$	CPU(seg)
4	0.0507	0.1876	0.0507	0.0436	0.1014	0.0395	0.1594	0.0280	0.0435	0.1976
5	0.3333	0.2949	0.0741	0.0424	0.0519	0.0390	0.0963	0.0269	0.0741	0.1970
6	0.0379	0.5520	0.0758	0.0445	0.0455	0.0402	0.1667	0.0283	0.1061	0.1971
7	0.3333	0.7156	0.1008	0.0415	0.1705	0.0391	0.1318	0.0282	0.0853	0.1988
8	0.3333	12.403	0.0952	0.0425	0.0476	0.0395	0.1190	0.0287	0.0714	0.1988
9	0.3333	14.144	0.0650	0.0408	0.0976	0.0393	0.1545	0.0287	0.0569	0.2025
10	0.3333	18.861	0.0417	0.0456	0.0333	0.0397	0.0583	0.0304	0.0333	0.2060
11	0.0855	26.588	0.0256	0.0428	0.0684	0.0445	0.0427	0.0391	0.0513	0.1968
12	0.1930	36.998	0.0702	0.0559	0.0614	0.0546	0.0526	0.0414	0.0614	0.2766
13	0.0360	47.412	0.0360	0.0445	0.0360	0.0405	0.0360	0.0307	0.0360	0.2016
14	0.0648	56.160	0.0648	0.0423	0.0370	0.0405	0.1019	0.0312	0.0185	0.2032
15	0.0190	65.144	0.0286	0.0548	0.0190	0.0565	0.0286	0.0375	0.0286	0.2048
16	0.3333	113.765	0.0392	0.0436	0.0294	0.0413	0.0294	0.0323	0.0294	0.2047
17	0.0606	110.402	0.0101	0.0731	0.0202	0.0657	0.0606	0.0486	0.0202	0.3096
18	0.0625	124.040	0.0417	0.0578	0.0417	0.0574	0.0521	0.0470	0.0417	0.2829
19	0.0323	159.117	0.0215	0.0555	0.0215	0.0581	0.0968	0.0476	0.0215	0.3450
20	0.0667	176.573	0.0889	0.0560	0.0556	0.0587	0.1667	0.0464	0.0444	0.2729

### 6.1.12. Conjunto wine

Estos datos son los resultados de un análisis químico de los vinos cultivados en la misma región en Italia, pero derivada de tres cultivares diferentes. El análisis determina las cantidades de 13 componentes que se encuentran en cada uno de los tres tipos de vinos. El desempeño de los clasificadores se presenta a continuación.



**Figura 6-23.:** Gráfico de dispersión del conjunto de datos proyectado en 2D mediante escalamiento multidimensional y desempeño de cada uno de los clasificadores variando el número de objetos de representación por clase. Conjunto de datos wine. **Selección de las líneas de características de longitud mayor.**

**Tabla 6-43.:** Intervalos de confianza para el error promedio de cada clasificador de acuerdo al número de objetos de representación por clase (Resaltado se encuentra el mejor intervalo para cada clasificador).

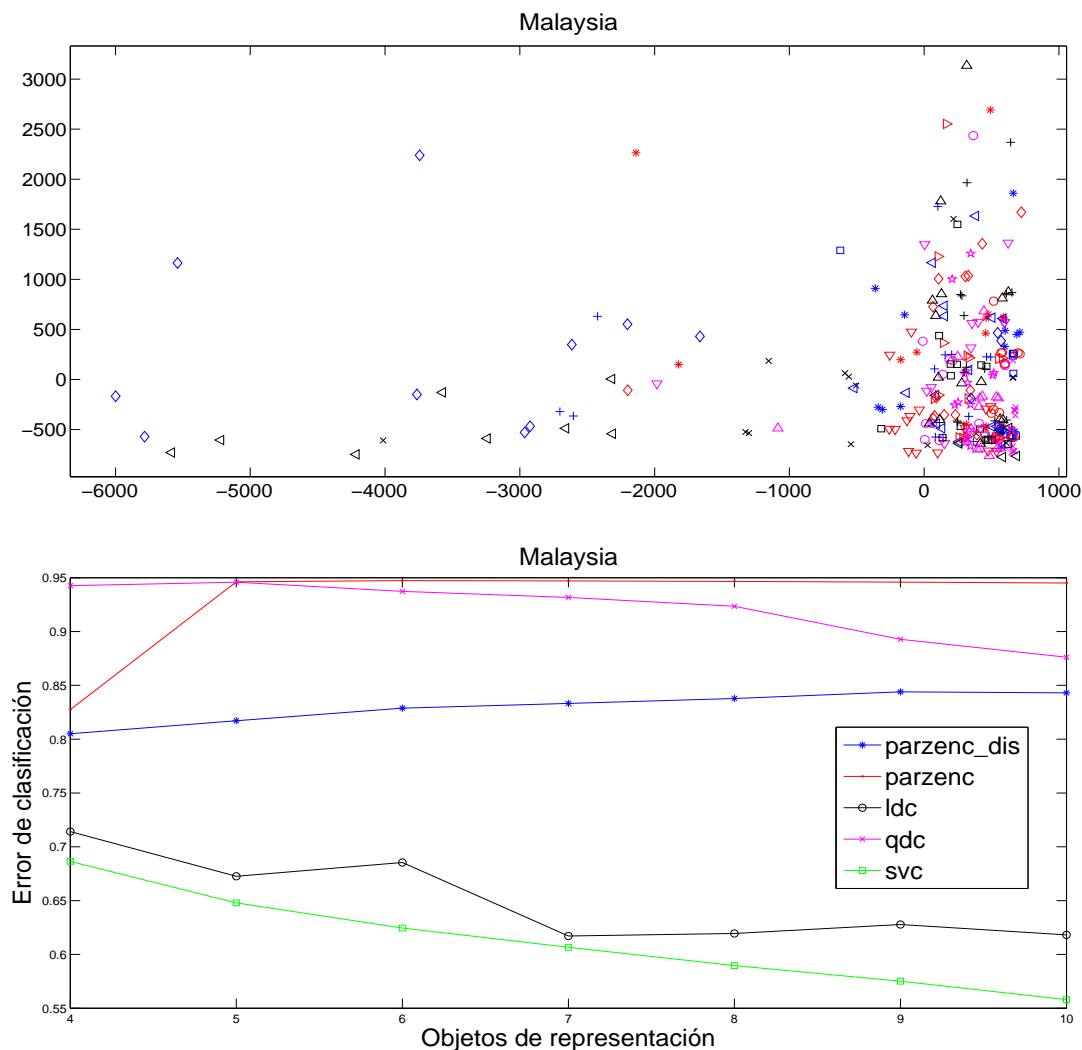
Intervalos de confianza para el error promedio					
Objetos de representación	<i>parzenc_dis</i>	<i>parzenc</i>	<i>ldc</i>	<i>qdc</i>	<i>svc</i>
4	[0.3632,0.4154]	[0.3116,0.3432]	[0.3207,0.3549]	[0.3562,0.4042]	[0.3257,0.3621]
5	[0.3617,0.4175]	[0.3054,0.3328]	[0.4185,0.4601]	[0.3676,0.4056]	[0.3287,0.3635]
6	[0.3530,0.3998]	[0.2979,0.3285]	[0.3704,0.4052]	[0.3464,0.3858]	[0.3047,0.3369]
7	[0.3598,0.4008]	[0.2967,0.3189]	[0.3491,0.3801]	[0.3229,0.3535]	[0.2965,0.3249]
8	[0.3657,0.4115]	[0.2943,0.3107]	[0.3354,0.3632]	[0.3241,0.3525]	[0.2922,0.3222]
9	[0.3705,0.4141]	[0.2938,0.3148]	[0.3277,0.3577]	[0.3263,0.3595]	[0.2934,0.3228]
10	[0.3670,0.4106]	[0.2937,0.3121]	[0.3221,0.3453]	[0.3288,0.3624]	[0.2795,0.3037]
11	[0.3770,0.4228]	[0.2902,0.3066]	[0.3143,0.3411]	[0.3153,0.3427]	[0.2754,0.3012]
12	[0.3689,0.4147]	[0.2897,0.3049]	[0.3253,0.3489]	[0.3311,0.3701]	[0.2909,0.3187]
13	[0.3614,0.4098]	[0.2915,0.3057]	[0.3096,0.3332]	[0.3325,0.3767]	[0.2730,0.2952]
14	<b>[0.3585,0.3991]</b>	[0.2873,0.3009]	[0.3060,0.3286]	[0.2824,0.3146]	[0.2701,0.2937]
15	[0.3762,0.4120]	[0.2889,0.3041]	[0.3124,0.3350]	[0.2719,0.2987]	[0.2739,0.2955]
16	[0.3843,0.4233]	[0.2921,0.3057]	[0.3132,0.3342]	[0.2637,0.2901]	[0.2774,0.3010]
17	[0.3791,0.4213]	[0.2930,0.3062]	[0.3127,0.3305]	[0.2538,0.2764]	[0.2732,0.2968]
18	[0.3688,0.4088]	<b>[0.2870,0.3044]</b>	[0.3058,0.3294]	[0.2563,0.2769]	[0.2661,0.2909]
19	[0.3736,0.4136]	[0.2905,0.3079]	<b>[0.3031,0.3241]</b>	<b>[0.2477,0.2683]</b>	<b>[0.2664,0.2922]</b>
20	[0.3722,0.4164]	[0.2930,0.3072]	[0.3064,0.3274]	[0.2519,0.2755]	[0.2751,0.3009]

**Tabla 6-44.:** Tabla del Error promedio, desviación estándar, tiempo de cómputo, de los clasificadores de acuerdo al número de objetos de representación por clase (resaltado se encuentra el mejor tiempo promedio de cómputo de cada clasificador).

Error promedio, desviación estándar, tiempo de cómputo de los clasificadores										
Objetos de representación	<i>parzenc_dis</i>		<i>parzenc</i>		<i>ldc</i>		<i>qdc</i>		<i>svc</i>	
	$\bar{e}_\infty \pm \sigma$	CPU(seg)	$\bar{e}_\infty \pm \sigma$	CPU(seg)	$\bar{e}_\infty \pm \sigma$	CPU(seg)	$\bar{e}_\infty \pm \sigma$	CPU(seg)	$\bar{e}_\infty \pm \sigma$	CPU(seg)
4	$0.3893 \pm 0.099$	<b>0.1726</b>	$0.3274 \pm 0.06$	0.0420	$0.3378 \pm 0.065$	0.0395	$0.3802 \pm 0.091$	0.0280	$0.3439 \pm 0.069$	<b>0.1985</b>
5	$0.3896 \pm 0.106$	0.1974	$0.3191 \pm 0.052$	0.0411	$0.4393 \pm 0.079$	0.0382	$0.3866 \pm 0.072$	0.0273	$0.3461 \pm 0.066$	0.2069
6	$0.3764 \pm 0.089$	0.2265	$0.3132 \pm 0.058$	0.0408	$0.3878 \pm 0.066$	0.0381	$0.3661 \pm 0.075$	0.0272	$0.3208 \pm 0.061$	0.2134
7	$0.3803 \pm 0.078$	0.2502	$0.3078 \pm 0.042$	0.0406	$0.3646 \pm 0.059$	0.0377	$0.3382 \pm 0.058$	0.0268	$0.3107 \pm 0.054$	0.2115
8	$0.3886 \pm 0.087$	0.2758	$0.3025 \pm 0.031$	<b>0.0405</b>	$0.3493 \pm 0.053$	0.0376	$0.3383 \pm 0.054$	0.0266	$0.3072 \pm 0.057$	0.2124
9	$0.3923 \pm 0.083$	0.4102	$0.3043 \pm 0.04$	0.0409	$0.3427 \pm 0.057$	0.0378	$0.3429 \pm 0.063$	0.0267	$0.3081 \pm 0.056$	0.2141
10	$0.3888 \pm 0.083$	0.3289	$0.3029 \pm 0.035$	0.0409	$0.3337 \pm 0.044$	0.0381	$0.3456 \pm 0.064$	0.0269	$0.2916 \pm 0.046$	0.2140
11	$0.3999 \pm 0.087$	0.3584	$0.2984 \pm 0.031$	0.0408	$0.3277 \pm 0.051$	0.0388	$0.329 \pm 0.052$	0.0268	$0.2883 \pm 0.049$	0.2154
12	$0.3918 \pm 0.087$	0.3804	$0.2973 \pm 0.029$	0.0406	$0.3371 \pm 0.045$	0.0376	$0.3506 \pm 0.074$	0.0266	$0.3048 \pm 0.053$	0.2156
13	$0.3856 \pm 0.092$	0.4109	$0.2986 \pm 0.027$	0.0411	$0.3214 \pm 0.045$	0.0381	$0.3546 \pm 0.084$	0.0260	$0.2841 \pm 0.042$	0.2192
14	$0.3788 \pm 0.077$	0.4318	$0.2941 \pm 0.026$	0.0406	$0.3173 \pm 0.043$	0.0375	$0.2985 \pm 0.061$	0.0255	$0.2819 \pm 0.045$	0.2176
15	$0.3941 \pm 0.068$	0.4568	$0.2965 \pm 0.029$	<b>0.0405</b>	$0.3237 \pm 0.043$	<b>0.0374</b>	$0.2853 \pm 0.051$	0.0255	$0.2847 \pm 0.041$	0.2185
16	$0.4038 \pm 0.074$	0.4834	$0.2989 \pm 0.026$	0.0406	$0.3237 \pm 0.040$	<b>0.0374</b>	$0.2769 \pm 0.05$	0.0255	$0.2892 \pm 0.045$	0.2203
17	$0.4002 \pm 0.080$	0.5102	$0.2996 \pm 0.025$	0.0407	$0.3216 \pm 0.034$	<b>0.0374</b>	$0.2651 \pm 0.043$	<b>0.0254</b>	$0.285 \pm 0.045$	0.2215
18	$0.3888 \pm 0.076$	0.6206	$0.2957 \pm 0.033$	0.0423	$0.3176 \pm 0.045$	0.0385	$0.2666 \pm 0.039$	0.0263	$0.2785 \pm 0.047$	0.2309
19	$0.3936 \pm 0.076$	0.5997	$0.2992 \pm 0.033$	0.0422	$0.3136 \pm 0.040$	0.0386	$0.258 \pm 0.039$	0.0264	$0.2793 \pm 0.049$	0.2344
20	$0.3943 \pm 0.084$	0.6130	$0.3001 \pm 0.027$	0.0550	$0.3169 \pm 0.040$	0.0470	$0.2637 \pm 0.045$	0.0302	$0.288 \pm 0.049$	0.3014

### 6.1.13. Conjunto Malaysia

Conjunto de datos que se refiere a medidas simples en los segmentos de los símbolos de servicios públicos de malasia, la dimensión de este conjunto de datos es 8 y consta de 20 clases. El desempeño de los clasificadores se presenta a continuación.



**Figura 6-24.:** Grafico de dispersión del conjunto de datos proyectado en 2D mediante escalamiento multidimensional y desempeño de cada uno de los clasificadores variando el número de objetos de representación por clase. Conjunto de datos Malaysia. **Selección de las líneas de características de longitud mayor.**



**Tabla 6-45.:** Intervalos de confianza para el error promedio de cada clasificador de acuerdo al número de objetos de representación por clase (Resaltado se encuentra el mejor intervalo para cada clasificador).

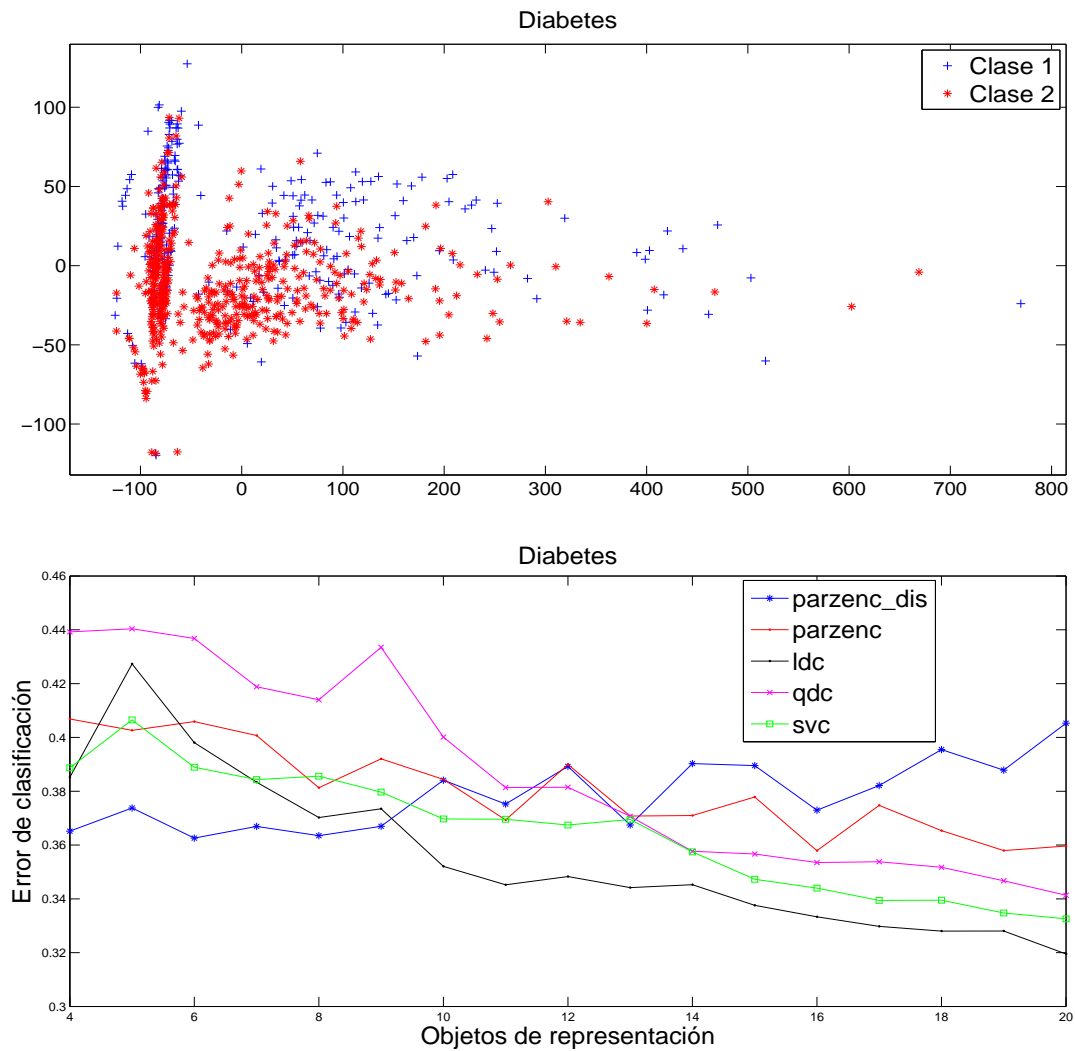
Intervalos de confianza para el error promedio					
Objetos de representación	<i>parzenc_dis</i>	<i>parzenc</i>	<i>ldc</i>	<i>qdc</i>	<i>svc</i>
4	<b>[0.7980,0.8122]</b>	<b>[0.8212,0.8338]</b>	[0.7062,0.7220]	[0.9390,0.9464]	[0.6777,0.6951]
5	[0.8100,0.8242]	[0.9423,0.9501]	[0.6642,0.6810]	[0.9434,0.9482]	[0.6395,0.6563]
6	[0.8209,0.8367]	[0.9474,0.9474]	[0.6683,0.7025]	[0.9339,0.9407]	[0.6156,0.6334]
7	[0.8264,0.8400]	[0.9470,0.9470]	[0.6069,0.6275]	[0.9272,0.9362]	[0.5969,0.6163]
8	[0.8302,0.8454]	[0.9466,0.9466]	[0.6095,0.6295]	[0.9190,0.9280]	[0.5807,0.5985]
9	[0.8358,0.8522]	[0.9459,0.9459]	[0.6166,0.6388]	[0.8865,0.8991]	[0.5633,0.5869]
10	[0.8347,0.8515]	[0.9451,0.9451]	<b>[0.6068,0.6294]</b>	<b>[0.8686,0.8838]</b>	<b>[0.5459,0.5701]</b>

**Tabla 6-46.:** Tabla del Error promedio, desviación estándar, tiempo de cómputo, de los clasificadores de acuerdo al número de objetos de representación por clase (resaltado se encuentra el mejor tiempo promedio de cómputo de cada clasificador).

Error promedio, desviación estándar, tiempo de cómputo de los clasificadores										
Objetos de representación	<i>parzenc_dis</i>		<i>parzenc</i>		<i>ldc</i>		<i>qdc</i>		<i>svc</i>	
	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)
4	$0.8051 \pm 0.027$	<b>55.873</b>	$0.8275 \pm 0.024$	0.1780	$0.7141 \pm 0.030$	0.2000	$0.9427 \pm 0.014$	<b>0.2654</b>	$0.6864 \pm 0.033$	<b>17.004</b>
5	$0.8171 \pm 0.027$	69.898	$0.9462 \pm 0.015$	<b>0.1740</b>	$0.6726 \pm 0.032$	0.1932	$0.9458 \pm 0.009$	0.3289	$0.6479 \pm 0.032$	24.133
6	$0.8288 \pm 0.030$	83.631	$0.9474 \pm 0.013$	0.1721	$0.6854 \pm 0.065$	0.1783	$0.9373 \pm 0.013$	0.3136	$0.6245 \pm 0.034$	28.242
7	$0.8332 \pm 0.026$	98.238	$0.947 \pm 0.016$	0.2096	$0.6172 \pm 0.039$	0.2183	$0.9317 \pm 0.017$	0.4022	$0.6066 \pm 0.037$	42.450
8	$0.8378 \pm 0.029$	112.440	$0.9466 \pm 0.018$	0.2035	$0.6195 \pm 0.038$	0.2104	$0.9235 \pm 0.017$	0.3915	$0.5896 \pm 0.034$	44.393
9	$0.844 \pm 0.031$	127.455	$0.9459 \pm 0.015$	0.2123	$0.6277 \pm 0.042$	0.2198	$0.8928 \pm 0.024$	0.4065	$0.5751 \pm 0.045$	55.453
10	$0.8431 \pm 0.032$	141.906	$0.9451 \pm 0.012$	0.1762	$0.6181 \pm 0.043$	0.1822	$0.8762 \pm 0.029$	0.3287	$0.558 \pm 0.046$	60.069

### 6.1.14. Conjunto diabetes

Conjunto del mundo real, tomado de muestras de pacientes con diabetes, en él se describe la fisiología básica y la fisiopatología de la diabetes mellitus y su tratamiento. Este es un conjunto con 2 clases de dimensión 8 y el número de objetos por clase es 268 y 500 respectivamente. El desempeño de los clasificadores se presenta a continuación.



**Figura 6-25.:** Grafico de dispersión del conjunto de datos proyectado en 2D mediante escalamiento multidimensional y desempeño de cada uno de los clasificadores variando el número de objetos de representación por clase. Conjunto de datos diabetes. **Selección de las líneas de características de longitud mayor.**

**Tabla 6-47.:** Intervalos de confianza para el error promedio de cada clasificador de acuerdo al número de objetos de representación por clase (Resaltado se encuentra el mejor intervalo para cada clasificador).

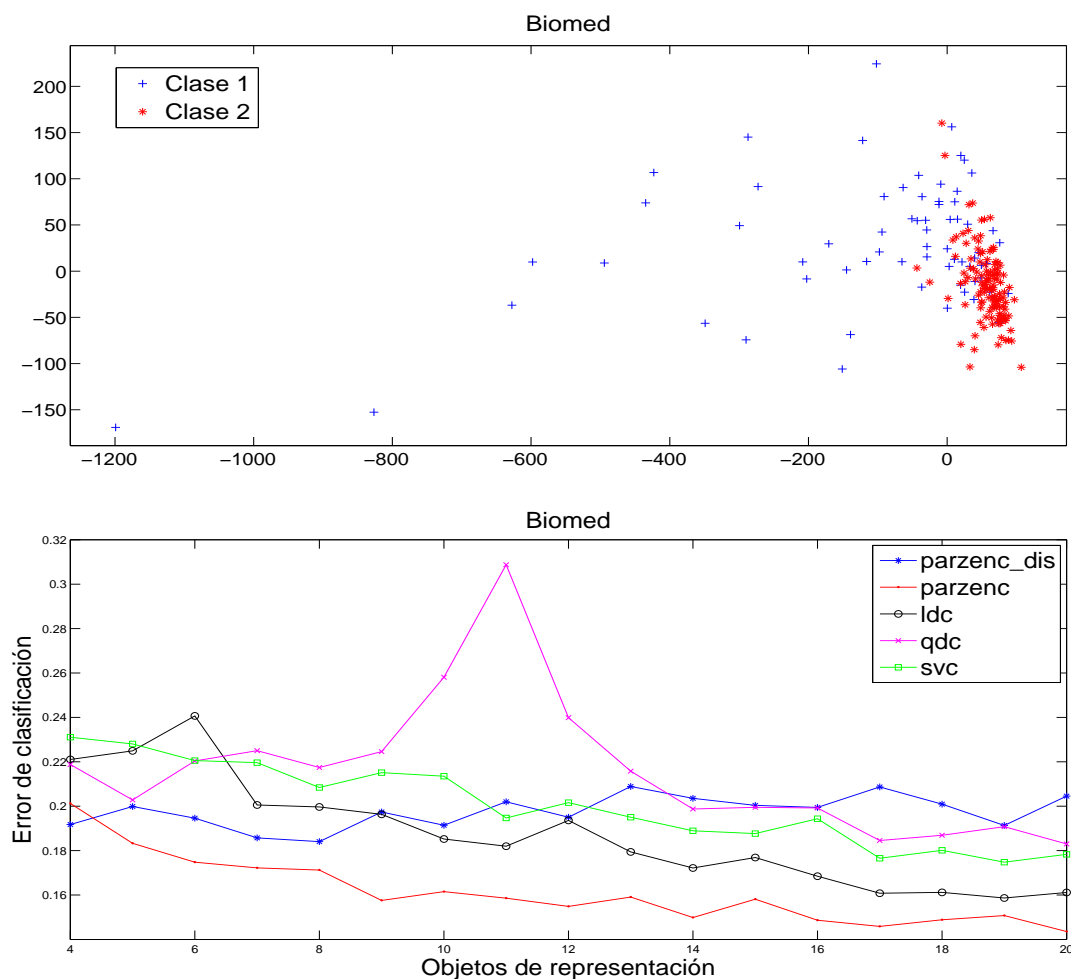
Intervalos de confianza para el error promedio					
Objetos de representación	<i>parzenc_dis</i>	<i>parzenc</i>	<i>ldc</i>	<i>qdc</i>	<i>svc</i>
4	[0.3428,0.3876]	[0.3816,0.4322]	[0.3647,0.4057]	[0.4039,0.4745]	[0.3697,0.4077]
5	[0.3472,0.4004]	[0.3800,0.4252]	[0.4065,0.4481]	[0.4120,0.4688]	[0.3865,0.4265]
6	<b>[0.3402,0.3850]</b>	[0.3835,0.4283]	[0.3772,0.4188]	[0.4126,0.4610]	[0.3681,0.4097]
7	[0.3408,0.3930]	[0.3789,0.4225]	[0.3651,0.4015]	[0.3967,0.4409]	[0.3653,0.4033]
8	[0.3414,0.3856]	[0.3618,0.4008]	[0.3534,0.3870]	[0.3877,0.4403]	[0.3664,0.4048]
9	[0.3423,0.3917]	[0.3725,0.4115]	[0.3572,0.3898]	[0.4069,0.4601]	[0.3629,0.3965]
10	[0.3549,0.4133]	[0.3631,0.4057]	[0.3350,0.3692]	[0.3761,0.4241]	[0.3536,0.3858]
11	[0.3486,0.4018]	[0.3503,0.3883]	[0.3326,0.3578]	[0.3643,0.3985]	[0.3538,0.3854]
12	[0.3621,0.4163]	[0.3670,0.4128]	[0.3354,0.3612]	[0.3610,0.4020]	[0.3527,0.3821]
13	[0.3430,0.3920]	[0.3542,0.3874]	[0.3303,0.3581]	[0.3551,0.3861]	[0.3545,0.3845]
14	[0.3606,0.4200]	[0.3523,0.3897]	[0.3300,0.3606]	[0.3424,0.3730]	[0.3433,0.3717]
15	[0.3579,0.4211]	[0.3591,0.3965]	[0.3255,0.3497]	[0.3420,0.3714]	[0.3339,0.3607]
16	[0.3437,0.4021]	[0.3432,0.3726]	[0.3217,0.3449]	[0.3367,0.3703]	[0.3316,0.3564]
17	[0.3534,0.4108]	[0.3553,0.3943]	[0.3172,0.3424]	[0.3372,0.3704]	[0.3268,0.3520]
18	[0.3637,0.4273]	[0.3478,0.3830]	[0.3172,0.3388]	[0.3372,0.3662]	[0.3279,0.3511]
19	[0.3591,0.4165]	<b>[0.3408,0.3750]</b>	[0.3178,0.3384]	[0.3314,0.3620]	[0.3250,0.3444]
20	[0.3724,0.4382]	[0.3435,0.3757]	<b>[0.3111,0.3279]</b>	<b>[0.3281,0.3545]</b>	<b>[0.3223,0.3429]</b>

**Tabla 6-48.:** Tabla del Error promedio, desviación estándar, tiempo de cómputo, de los clasificadores de acuerdo al número de objetos de representación por clase (resaltado se encuentra el mejor tiempo promedio de cómputo de cada clasificador).

Error promedio, desviación estándar, tiempo de cómputo de los clasificadores										
Objetos de representación	<i>parzenc_dis</i>		<i>parzenc</i>		<i>ldc</i>		<i>qdc</i>		<i>svc</i>	
	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)
4	$0.3652 \pm 0.085$	<b>0.1245</b>	$0.4069 \pm 0.096$	<b>0.0352</b>	$0.3852 \pm 0.078$	0.0337	$0.4392 \pm 0.134$	0.0276	$0.3887 \pm 0.072$	<b>0.0624</b>
5	$0.3738 \pm 0.101$	0.1381	$0.4026 \pm 0.086$	0.0358	$0.4273 \pm 0.079$	0.0328	$0.4404 \pm 0.108$	0.0262	$0.4065 \pm 0.076$	0.0626
6	$0.3626 \pm 0.085$	0.1551	$0.4059 \pm 0.085$	0.0410	$0.398 \pm 0.079$	0.0379	$0.4368 \pm 0.092$	0.0301	$0.3889 \pm 0.079$	0.0705
7	$0.3669 \pm 0.099$	0.1684	$0.4007 \pm 0.083$	0.0395	$0.3833 \pm 0.069$	0.0354	$0.4188 \pm 0.084$	0.0275	$0.3843 \pm 0.072$	0.0682
8	$0.3635 \pm 0.084$	0.1977	$0.3813 \pm 0.074$	0.0443	$0.3702 \pm 0.064$	0.0390	$0.414 \pm 0.1$	0.0295	$0.3856 \pm 0.073$	0.0718
9	$0.367 \pm 0.094$	0.1980	$0.392 \pm 0.074$	0.0430	$0.3735 \pm 0.062$	0.0395	$0.4335 \pm 0.101$	0.0300	$0.3797 \pm 0.064$	0.0744
10	$0.3841 \pm 0.111$	0.2064	$0.3844 \pm 0.081$	0.0378	$0.3521 \pm 0.065$	0.0346	$0.4001 \pm 0.091$	0.0259	$0.3697 \pm 0.061$	0.0635
11	$0.3752 \pm 0.101$	0.2119	$0.3693 \pm 0.072$	0.0375	$0.3452 \pm 0.048$	0.0331	$0.3814 \pm 0.065$	0.0245	$0.3696 \pm 0.060$	0.0628
12	$0.3892 \pm 0.103$	0.2316	$0.3899 \pm 0.087$	0.0374	$0.3483 \pm 0.049$	0.0329	$0.3815 \pm 0.078$	0.0243	$0.3674 \pm 0.056$	0.0631
13	$0.3675 \pm 0.093$	0.2388	$0.3708 \pm 0.063$	0.0387	$0.3442 \pm 0.053$	0.0356	$0.3706 \pm 0.059$	0.0263	$0.3695 \pm 0.057$	0.0677
14	$0.3903 \pm 0.113$	0.2497	$0.371 \pm 0.071$	0.0416	$0.3453 \pm 0.058$	0.0404	$0.3577 \pm 0.058$	0.0298	$0.3575 \pm 0.054$	0.0760
15	$0.3895 \pm 0.120$	0.2621	$0.3778 \pm 0.071$	0.0411	$0.3376 \pm 0.046$	0.0383	$0.3567 \pm 0.056$	0.0285	$0.3473 \pm 0.051$	0.0739
16	$0.3729 \pm 0.111$	0.2763	$0.3579 \pm 0.056$	0.0372	$0.3333 \pm 0.044$	0.0327	$0.3535 \pm 0.064$	0.0240	$0.344 \pm 0.047$	0.0627
17	$0.3821 \pm 0.109$	0.2868	$0.3748 \pm 0.074$	0.0411	$0.3298 \pm 0.048$	0.0407	$0.3538 \pm 0.063$	0.0285	$0.3394 \pm 0.048$	0.0742
18	$0.3955 \pm 0.121$	0.3014	$0.3654 \pm 0.067$	0.0397	$0.328 \pm 0.041$	0.0363	$0.3517 \pm 0.055$	0.0271	$0.3395 \pm 0.044$	0.0710
19	$0.3878 \pm 0.109$	0.3199	$0.3579 \pm 0.065$	0.0395	$0.3281 \pm 0.039$	0.0350	$0.3467 \pm 0.058$	0.0262	$0.3347 \pm 0.037$	0.0683
20	$0.4053 \pm 0.125$	0.3235	$0.3596 \pm 0.061$	0.0367	$0.3195 \pm 0.032$	<b>0.0321</b>	$0.3413 \pm 0.05$	<b>0.0237</b>	$0.3326 \pm 0.039$	0.0632

### 6.1.15. Conjunto biomed

El conjunto de datos de biomedicina desarrolla métodos de detección para identificar a los portadores de un trastorno genético poco común. Debido a que la enfermedad es rara, hay sólo unos pocos portadores de la enfermedad de la que se dispone de datos. El conjunto posee dos clases y su dimensión es 5. El número de objetos por clase es de 67 y 127 respectivamente. El desempeño de los clasificadores se presenta a continuación.



**Figura 6-26.:** Grafico de dispersión del conjunto de datos proyectado en 2D mediante escalamiento multidimensional y desempeño de cada uno de los clasificadores variando el número de objetos de representación por clase. Conjunto de datos biomed. **Selección de las líneas de características de longitud mayor.**

**Tabla 6-49.:** Intervalos de confianza para el error promedio de cada clasificador de acuerdo al número de objetos de representación por clase (Resaltado se encuentra el mejor intervalo para cada clasificador).

Intervalos de confianza para el error promedio					
Objetos de representación	<i>parzenc_dis</i>	<i>parzenc</i>	<i>ldc</i>	<i>qdc</i>	<i>svc</i>
4	[0.1773,0.2063]	[0.1809,0.2215]	[0.2008,0.2414]	[0.1891,0.2485]	[0.2108,0.2514]
5	[0.1844,0.2154]	[0.1694,0.1972]	[0.2044,0.2454]	[0.1881,0.2175]	[0.2090,0.2470]
6	[0.1775,0.2117]	[0.1632,0.1864]	[0.2165,0.2649]	[0.2017,0.2391]	[0.2008,0.2402]
7	<b>[0.1733,0.1981]</b>	[0.1633,0.1811]	[0.1845,0.2167]	[0.2018,0.2482]	[0.2030,0.2362]
8	[0.1745,0.1935]	[0.1578,0.1846]	[0.1826,0.2168]	[0.1948,0.2400]	[0.1926,0.2242]
9	[0.1813,0.2135]	[0.1489,0.1663]	[0.1790,0.2138]	[0.2038,0.2454]	[0.1951,0.2351]
10	[0.1793,0.2035]	[0.1516,0.1716]	[0.1720,0.1984]	[0.2344,0.2818]	[0.1964,0.2306]
11	[0.1873,0.2167]	[0.1504,0.1668]	[0.1694,0.1946]	[0.2724,0.3450]	[0.1813,0.2081]
12	[0.1809,0.2093]	[0.1486,0.1612]	[0.1797,0.2075]	[0.2191,0.2607]	[0.1869,0.2163]
13	[0.1892,0.2286]	[0.1449,0.1733]	[0.1673,0.1915]	[0.2000,0.2316]	[0.1814,0.2088]
14	[0.1846,0.2226]	[0.1441,0.1557]	[0.1617,0.1827]	[0.1861,0.2113]	[0.1752,0.2026]
15	[0.1825,0.2183]	[0.1507,0.1655]	[0.1664,0.1874]	[0.1863,0.2127]	[0.1738,0.2016]
16	[0.1849,0.2139]	[0.1415,0.1557]	[0.1572,0.1798]	[0.1860,0.2124]	[0.1797,0.2091]
17	[0.1911,0.2263]	[0.1393,0.1525]	[0.1505,0.1711]	[0.1741,0.1951]	[0.1650,0.1882]
18	[0.1859,0.2159]	[0.1421,0.1557]	[0.1520,0.1704]	[0.1753,0.1985]	[0.1669,0.1933]
19	[0.1787,0.2039]	[0.1445,0.1571]	<b>[0.1490,0.1684]</b>	[0.1790,0.2026]	<b>[0.1640,0.1856]</b>
20	[0.1884,0.2206]	<b>[0.1378,0.1494]</b>	[0.1525,0.1699]	<b>[0.1738,0.1922]</b>	[0.1667,0.1899]

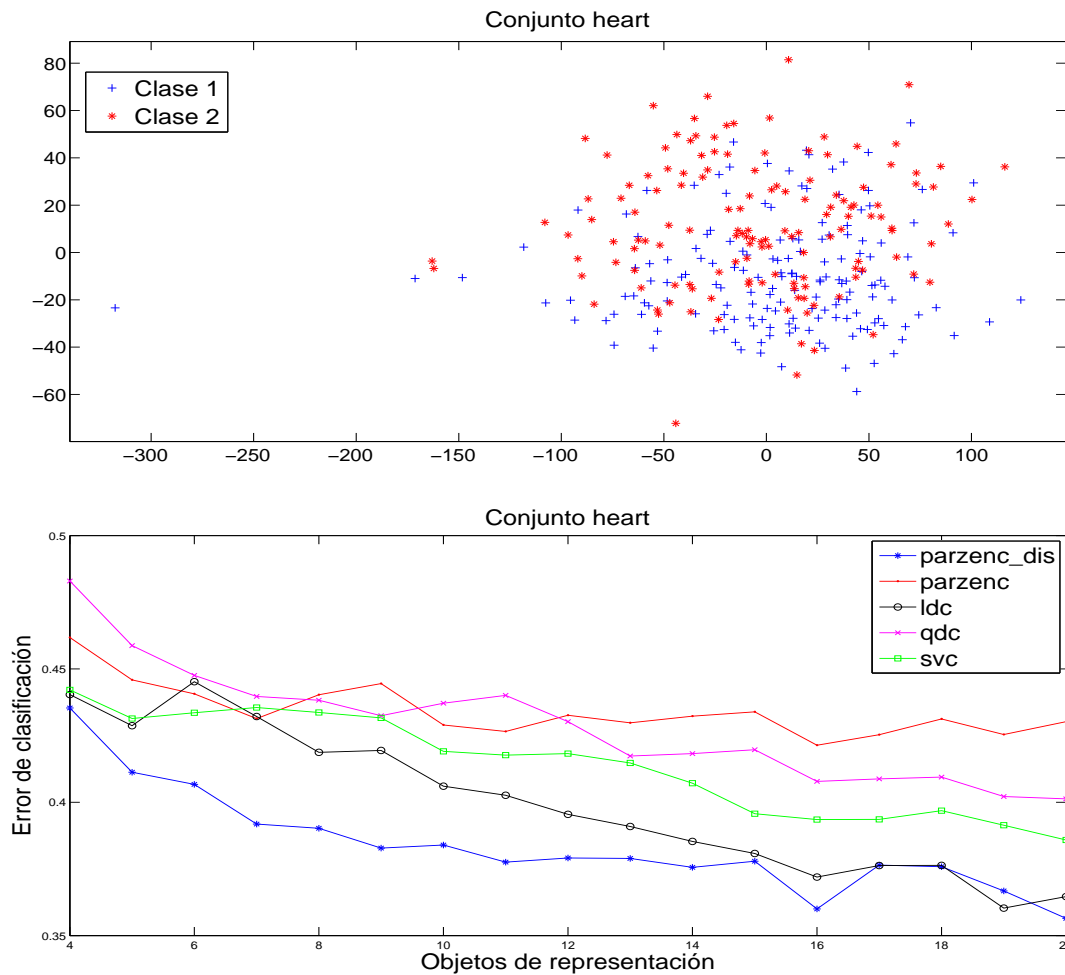
**Tabla 6-50.:** Tabla del Error promedio, desviación estándar, tiempo de cómputo, de los clasificadores de acuerdo al número de objetos de representación por clase (resaltado se encuentra el mejor tiempo promedio de cómputo de cada clasificador).

Error promedio, desviación estándar, tiempo de cómputo de los clasificadores										
Objetos de representación	<i>parzenc_dis</i>		<i>parzenc</i>		<i>ldc</i>		<i>qdc</i>		<i>svc</i>	
	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)
4	0.1918 $\pm$ 0.055	<b>0.1320</b>	0.2012 $\pm$ 0.077	0.0344	0.2211 $\pm$ 0.077	0.0314	0.2188 $\pm$ 0.113	0.0235	0.2311 $\pm$ 0.077	<b>0.0560</b>
5	0.1999 $\pm$ 0.059	0.1459	0.1833 $\pm$ 0.053	<b>0.0341</b>	0.2249 $\pm$ 0.078	0.0312	0.2028 $\pm$ 0.056	0.0234	0.228 $\pm$ 0.072	0.0582
6	0.1946 $\pm$ 0.065	0.1609	0.1748 $\pm$ 0.044	0.0343	0.2407 $\pm$ 0.092	<b>0.0309</b>	0.2204 $\pm$ 0.071	0.0234	0.2205 $\pm$ 0.075	0.0603
7	0.1857 $\pm$ 0.047	0.1762	0.1722 $\pm$ 0.034	0.0345	0.2006 $\pm$ 0.061	<b>0.0309</b>	0.225 $\pm$ 0.088	0.0234	0.2196 $\pm$ 0.063	0.0608
8	0.184 $\pm$ 0.036	0.1921	0.1712 $\pm$ 0.051	0.0343	0.1997 $\pm$ 0.065	<b>0.0309</b>	0.2174 $\pm$ 0.086	0.0234	0.2084 $\pm$ 0.060	0.0612
9	0.1974 $\pm$ 0.061	0.2083	0.1576 $\pm$ 0.033	0.0345	0.1964 $\pm$ 0.066	0.0310	0.2246 $\pm$ 0.079	0.0235	0.2151 $\pm$ 0.076	0.0613
10	0.1914 $\pm$ 0.046	0.2232	0.1616 $\pm$ 0.038	0.0348	0.1852 $\pm$ 0.050	0.0310	0.2581 $\pm$ 0.09	0.0236	0.2135 $\pm$ 0.065	0.0614
11	0.202 $\pm$ 0.056	0.2389	0.1586 $\pm$ 0.031	0.0347	0.182 $\pm$ 0.048	0.0310	0.3087 $\pm$ 0.138	0.0229	0.1947 $\pm$ 0.051	0.0619
12	0.1951 $\pm$ 0.054	0.2551	0.1549 $\pm$ 0.024	0.0347	0.1936 $\pm$ 0.053	0.0310	0.2399 $\pm$ 0.079	0.0229	0.2016 $\pm$ 0.056	0.0619
13	0.2089 $\pm$ 0.075	0.2699	0.1591 $\pm$ 0.054	0.0347	0.1794 $\pm$ 0.046	0.0310	0.2158 $\pm$ 0.06	0.0229	0.1951 $\pm$ 0.052	0.0622
14	0.2036 $\pm$ 0.072	0.2863	0.1499 $\pm$ 0.022	0.0348	0.1722 $\pm$ 0.040	0.0311	0.1987 $\pm$ 0.048	0.0229	0.1889 $\pm$ 0.052	0.0623
15	0.2004 $\pm$ 0.068	0.3019	0.1581 $\pm$ 0.028	0.0350	0.1769 $\pm$ 0.040	0.0310	0.1995 $\pm$ 0.05	0.0229	0.1877 $\pm$ 0.053	0.0622
16	0.1994 $\pm$ 0.055	0.3162	0.1486 $\pm$ 0.027	0.0348	0.1685 $\pm$ 0.043	0.0310	0.1992 $\pm$ 0.05	<b>0.0228</b>	0.1944 $\pm$ 0.056	0.0626
17	0.2087 $\pm$ 0.067	0.3321	0.1459 $\pm$ 0.025	0.0349	0.1608 $\pm$ 0.039	0.0310	0.1846 $\pm$ 0.04	0.0229	0.1766 $\pm$ 0.044	0.0628
18	0.2009 $\pm$ 0.057	0.3488	0.1489 $\pm$ 0.026	0.0350	0.1612 $\pm$ 0.035	0.0311	0.1869 $\pm$ 0.044	<b>0.0228</b>	0.1801 $\pm$ 0.050	0.0631
19	0.1913 $\pm$ 0.048	0.3635	0.1508 $\pm$ 0.024	0.0352	0.1587 $\pm$ 0.037	0.0311	0.1908 $\pm$ 0.045	0.0229	0.1748 $\pm$ 0.041	0.0635
20	0.2045 $\pm$ 0.061	0.3810	0.1436 $\pm$ 0.022	0.0351	0.1612 $\pm$ 0.033	0.0311	0.183 $\pm$ 0.035	0.0229	0.1783 $\pm$ 0.044	0.0632



### 6.1.16. Conjunto heart

Esta base de datos contiene 13 características y 303 objetos. La base de datos de heart-Cleveland es utilizada por investigadores con el "objetivo" de encontrar la presencia de una enfermedad cardíaca en pacientes. Contiene valores enteros desde 0 (sin presencia) a 4. Los experimentos con la base de datos de heart-Cleveland se han concentrado en tratar simplemente de distinguir la presencia de los valores 1, 2, 3, 4 en la ausencia (valor 0).



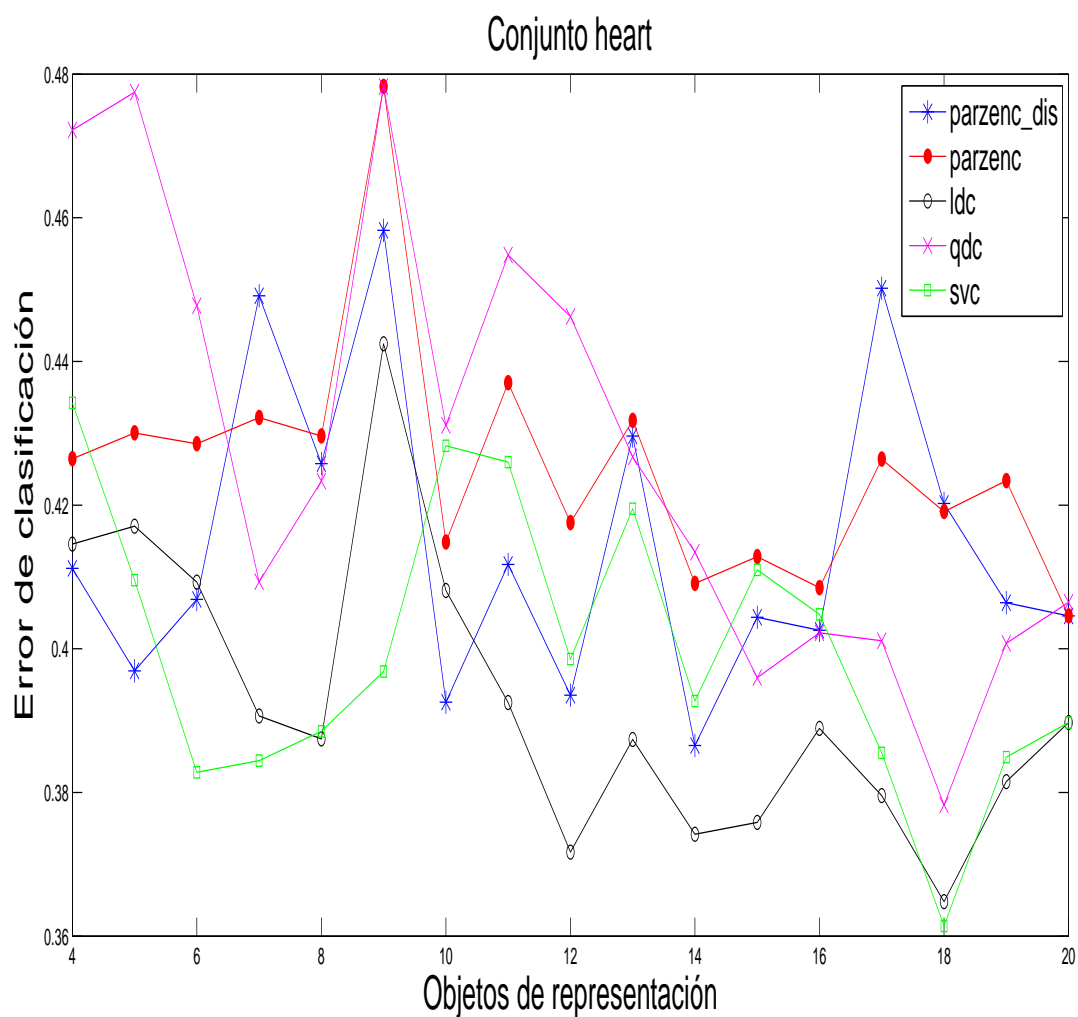
**Figura 6-27.:** Grafico de dispersión del conjunto de datos proyectado en 2D mediante escalamiento multidimensional y desempeño de cada uno de los clasificadores variando el número de objetos de representación por clase. Conjunto de datos heart. **Selección de las líneas de características de longitud mayor.**

**Tabla 6-51.:** Intervalos de confianza para el error promedio de cada clasificador de acuerdo al número de objetos de representación por clase (Resaltado se encuentra el mejor intervalo para cada clasificador).

Intervalos de confianza para el error promedio					
Objetos de representación	<i>parzenc_dis</i>	<i>parzenc</i>	<i>ldc</i>	<i>qdc</i>	<i>svc</i>
4	[0.4149,0.4559]	[0.4444,0.4792]	[0.4207,0.4601]	[0.4680,0.4980]	[0.4229,0.4613]
5	[0.3939,0.4287]	[0.4325,0.4593]	[0.4116,0.4458]	[0.4450,0.4724]	[0.4151,0.4477]
6	[0.3901,0.4233]	[0.4268,0.4546]	[0.4313,0.4591]	[0.4355,0.4597]	[0.4175,0.4497]
7	[0.3760,0.4076]	[0.4190,0.4438]	[0.4184,0.4458]	[0.4268,0.4526]	[0.4210,0.4500]
8	[0.3736,0.4068]	[0.4266,0.4540]	[0.4045,0.4329]	[0.4246,0.4520]	[0.4190,0.4484]
9	[0.3665,0.3991]	[0.4306,0.4584]	[0.4047,0.4341]	[0.4191,0.4459]	[0.4148,0.4484]
10	[0.3666,0.4014]	[0.4164,0.4416]	[0.3923,0.4197]	[0.4242,0.4500]	[0.4044,0.4338]
11	[0.3633,0.3917]	[0.4154,0.4376]	[0.3900,0.4152]	[0.4277,0.4525]	[0.4048,0.4306]
12	[0.3633,0.3949]	[0.4200,0.4452]	[0.3839,0.4071]	[0.4176,0.4428]	[0.4056,0.4308]
13	[0.3616,0.3964]	[0.4177,0.4419]	[0.3775,0.4043]	[0.4049,0.4297]	[0.4008,0.4286]
14	[0.3627,0.3885]	[0.4189,0.4457]	[0.3735,0.3971]	[0.4054,0.4312]	[0.3940,0.4204]
15	[0.3611,0.3947]	[0.4221,0.4457]	[0.3700,0.3916]	[0.4079,0.4315]	[0.3846,0.4068]
16	[0.3455,0.3745]	<b>[0.4101,0.4327]</b>	[0.3633,0.3807]	[0.3981,0.4175]	[0.3824,0.4046]
17	[0.3577,0.3951]	[0.4146,0.4362]	[0.3642,0.3884]	[0.3967,0.4209]	[0.3820,0.4052]
18	[0.3584,0.3932]	[0.4194,0.4430]	[0.3668,0.3858]	[0.4002,0.4186]	[0.3853,0.4085]
19	[0.3505,0.3831]	[0.4146,0.4362]	<b>[0.3521,0.3685]</b>	[0.3925,0.4119]	[0.3814,0.4014]
20	<b>[0.3431,0.3699]</b>	[0.4176,0.4428]	[0.3554,0.3738]	<b>[0.3918,0.4108]</b>	<b>[0.3762,0.3956]</b>

**Tabla 6-52.:** Tabla del Error promedio, desviación estándar, tiempo de cómputo, de los clasificadores de acuerdo al número de objetos de representación por clase (resaltado se encuentra el mejor tiempo promedio de cómputo de cada clasificador).

Error promedio, desviación estándar, tiempo de cómputo de los clasificadores										
Objetos de representación	<i>parzenc_dis</i>		<i>parzenc</i>		<i>ldc</i>		<i>qdc</i>		<i>svc</i>	
	$\bar{e}_\infty \pm \sigma$	CPU(seg)	$\bar{e}_\infty \pm \sigma$	CPU(seg)	$\bar{e}_\infty \pm \sigma$	CPU(seg)	$\bar{e}_\infty \pm \sigma$	CPU(seg)	$\bar{e}_\infty \pm \sigma$	CPU(seg)
4	0.4354 $\pm$ 0.078	<b>0.1323</b>	0.4618 $\pm$ 0.066	<b>0.0343</b>	0.4404 $\pm$ 0.075	0.0312	0.483 $\pm$ 0.057	0.0233	0.4421 $\pm$ 0.073	<b>0.0552</b>
5	0.4113 $\pm$ 0.066	0.1474	0.4459 $\pm$ 0.051	0.0345	0.4287 $\pm$ 0.065	0.0312	0.4587 $\pm$ 0.052	0.0234	0.4314 $\pm$ 0.062	0.0559
6	0.4067 $\pm$ 0.063	0.1634	0.4407 $\pm$ 0.053	0.0345	0.4452 $\pm$ 0.053	<b>0.0309</b>	0.4476 $\pm$ 0.046	0.0234	0.4336 $\pm$ 0.061	0.058
7	0.3918 $\pm$ 0.060	0.3865	0.4314 $\pm$ 0.047	0.0349	0.4321 $\pm$ 0.052	0.0313	0.4397 $\pm$ 0.049	0.0236	0.4355 $\pm$ 0.055	0.0604
8	0.3902 $\pm$ 0.063	0.1953	0.4403 $\pm$ 0.052	0.0348	0.4187 $\pm$ 0.054	0.031	0.4383 $\pm$ 0.052	0.0235	0.4337 $\pm$ 0.056	0.0605
9	0.3828 $\pm$ 0.062	0.2112	0.4445 $\pm$ 0.053	0.0349	0.4194 $\pm$ 0.056	0.031	0.4325 $\pm$ 0.051	0.0235	0.4316 $\pm$ 0.064	0.0609
10	0.384 $\pm$ 0.066	0.2272	0.429 $\pm$ 0.048	0.0348	0.406 $\pm$ 0.052	0.031	0.4371 $\pm$ 0.049	0.0233	0.4191 $\pm$ 0.056	0.0608
11	0.3775 $\pm$ 0.054	0.2428	0.4265 $\pm$ 0.042	0.0349	0.4026 $\pm$ 0.048	0.0311	0.4401 $\pm$ 0.047	<b>0.0228</b>	0.4177 $\pm$ 0.049	0.061
12	0.3791 $\pm$ 0.060	0.2578	0.4326 $\pm$ 0.048	0.0348	0.3955 $\pm$ 0.044	<b>0.0309</b>	0.4302 $\pm$ 0.048	0.0228	0.4182 $\pm$ 0.048	0.0608
13	0.379 $\pm$ 0.066	0.2733	0.4298 $\pm$ 0.046	0.0349	0.3909 $\pm$ 0.051	0.031	0.4173 $\pm$ 0.047	<b>0.0228</b>	0.4147 $\pm$ 0.053	0.061
14	0.3756 $\pm$ 0.049	0.2888	0.4323 $\pm$ 0.051	0.0348	0.3853 $\pm$ 0.045	<b>0.0309</b>	0.4183 $\pm$ 0.049	0.0228	0.4072 $\pm$ 0.050	0.0612
15	0.3779 $\pm$ 0.064	0.3047	0.4339 $\pm$ 0.045	0.035	0.3808 $\pm$ 0.041	0.031	0.4197 $\pm$ 0.045	0.0228	0.3957 $\pm$ 0.042	0.0613
16	0.36 $\pm$ 0.055	0.3213	0.4214 $\pm$ 0.043	0.0355	0.372 $\pm$ 0.033	0.0315	0.4078 $\pm$ 0.037	0.0231	0.3935 $\pm$ 0.042	0.0621
17	0.3764 $\pm$ 0.071	0.3372	0.4254 $\pm$ 0.041	0.0355	0.3763 $\pm$ 0.046	0.0315	0.4088 $\pm$ 0.046	0.0231	0.3936 $\pm$ 0.044	0.0629
18	0.3758 $\pm$ 0.066	0.3528	0.4312 $\pm$ 0.045	0.0357	0.3763 $\pm$ 0.036	0.0317	0.4094 $\pm$ 0.035	0.0232	0.3969 $\pm$ 0.044	0.0627
19	0.3668 $\pm$ 0.062	0.3682	0.4254 $\pm$ 0.041	0.0357	0.3603 $\pm$ 0.031	0.0317	0.4022 $\pm$ 0.037	0.0233	0.3914 $\pm$ 0.038	0.063
20	0.3565 $\pm$ 0.051	0.3853	0.4302 $\pm$ 0.048	0.0358	0.3646 $\pm$ 0.035	0.0318	0.4013 $\pm$ 0.036	0.0233	0.3859 $\pm$ 0.037	0.0633



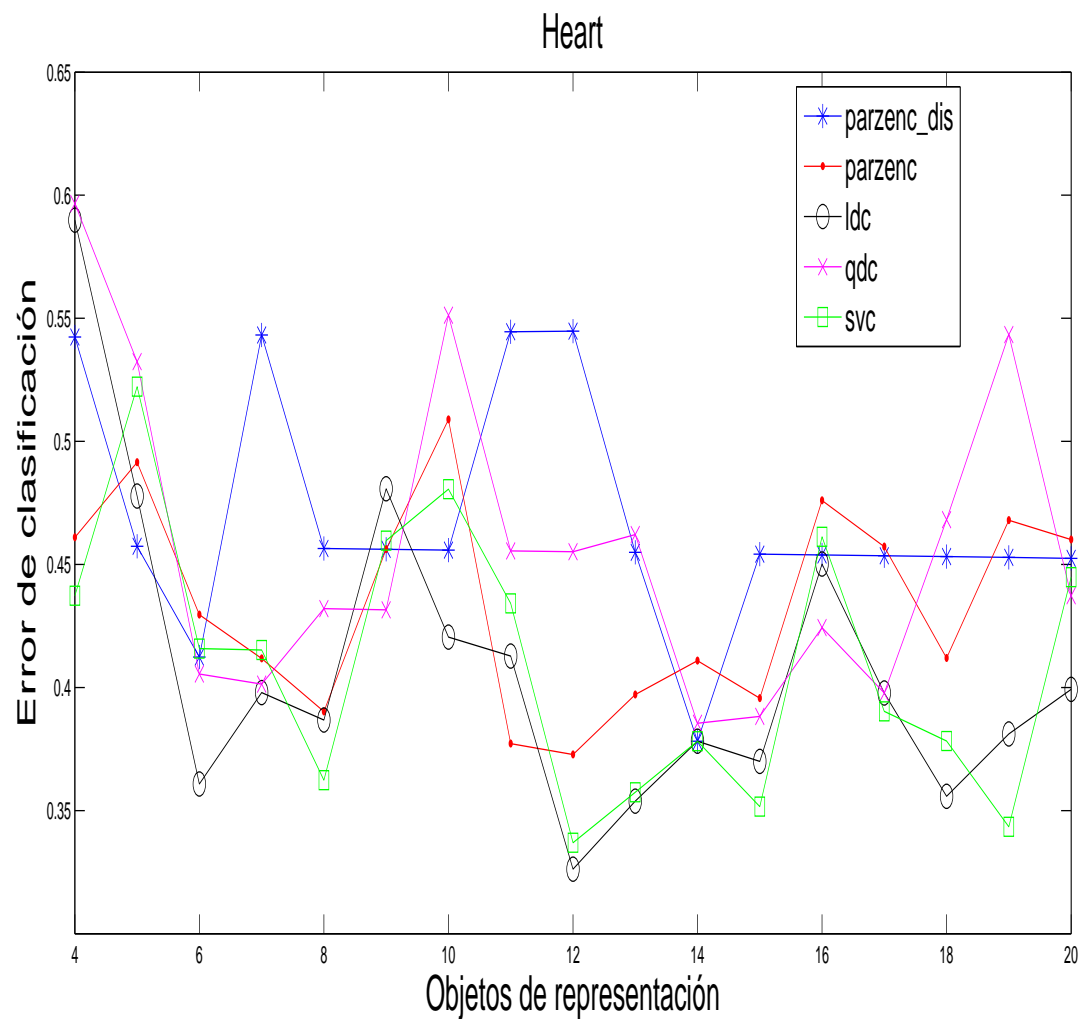
**Figura 6-28.:** Grafico de dispersión del conjunto de datos proyectado en 2D mediante escalamiento multidimensional y desempeño de cada uno de los clasificadores variando el número de objetos de representación por clase. Conjunto de datos heart. **Selección de las líneas de características de longitud media.**

**Tabla 6-53.:** Intervalos de confianza para el error promedio de cada clasificador de acuerdo al número de objetos de representación por clase (Resaltado se encuentra el mejor intervalo para el clasificador *parzenc\_dis*).

Intervalos de confianza para el error promedio					
Objetos de representación	<i>parzenc_dis</i>	<i>parzenc</i>	<i>ldc</i>	<i>qdc</i>	<i>svc</i>
4	[0.3743,0.4480]	[0.3857,0.4672]	[0.3616,0.4676]	[0.4401,0.5043]	[0.3790,0.4894]
5	[0.3550,0.4388]	[0.3929,0.4671]	[0.3701,0.4640]	[0.4281,0.5269]	[0.3674,0.4517]
6	[0.3696,0.4441]	[0.4040,0.4530]	[0.3755,0.4430]	[0.4148,0.4808]	[0.3589,0.4068]
7	[0.3928,0.5054]	[0.3917,0.4727]	[0.3501,0.4312]	[0.3785,0.4402]	[0.3578,0.4110]
8	[0.3947,0.4569]	[0.4065,0.4527]	[0.3560,0.4189]	[0.3768,0.4699]	[0.3436,0.4334]
9	[0.3917,0.5248]	[0.4305,0.5260]	[0.3860,0.4989]	[0.4254,0.5311]	[0.3444,0.4493]
10	[0.3722,0.4130]	[0.3923,0.4374]	[0.3642,0.4521]	[0.3829,0.4793]	[0.3854,0.4712]
11	[0.3442,0.4793]	[0.3951,0.4789]	[0.3751,0.4100]	[0.4166,0.4930]	[0.3879,0.4640]
12	[0.3505,0.4366]	[0.3904,0.4447]	[0.3575,0.3859]	[0.4081,0.4844]	[0.3648,0.4324]
13	<b>[0.4022,0.4570]</b>	[0.3959,0.4676]	[0.3507,0.4240]	[0.3934,0.4600]	[0.3576,0.4814]
14	[0.3692,0.4039]	[0.3859,0.4323]	[0.3609,0.3875]	[0.3785,0.4485]	[0.3539,0.4315]
15	<b>[0.3366,0.4722]</b>	[0.3760,0.4497]	[0.3543,0.3973]	[0.3597,0.4323]	[0.3787,0.4433]
16	[0.3687,0.4365]	[0.3778,0.4392]	[0.3565,0.4214]	[0.3770,0.4275]	[0.3644,0.4452]
17	[0.3913,0.5091]	[0.3935,0.4593]	[0.3533,0.4058]	[0.3654,0.4368]	[0.3659,0.4051]
18	[0.3819,0.4586]	[0.3959,0.4423]	[0.3492,0.3804]	[0.3496,0.4069]	[0.3483,0.3746]
19	[0.3638,0.4491]	[0.3828,0.4640]	[0.3656,0.3974]	[0.3837,0.4178]	[0.3602,0.4097]
20	[0.3644,0.4447]	[0.3766,0.4325]	[0.3566,0.4229]	[0.3844,0.4286]	[0.3602,0.4193]

**Tabla 6-54.:** Tabla del Error promedio, desviación estándar, tiempo de cómputo, de los clasificadores de acuerdo al número de objetos de representación por clase (resaltado se encuentra el mejor tiempo promedio de cómputo de cada clasificador).

Error promedio, desviación estándar, tiempo de cómputo de los clasificadores										
Objetos de representación	<i>parzenc_dis</i>		<i>parzenc</i>		<i>ldc</i>		<i>qdc</i>		<i>svc</i>	
	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)
4	0.4112±0.0143	0.1276	0.4264±0.0158	0.0338	0.4146±0.0206	0.0309	0.4722±0.0125	0.0234	0.4342±0.0215	0.0550
5	0.3969±0.0163	0.1441	0.4300±0.0144	0.0341	0.4171±0.0183	0.0306	0.4775±0.0192	0.0232	0.4096±0.0164	0.0568
6	0.4069±0.0145	0.1578	0.4285±0.0095	0.0342	0.4093±0.0131	0.0304	0.4478±0.0128	0.0232	0.3828±0.0093	0.0574
7	0.4491±0.0219	0.1729	0.4322±0.0158	0.0342	0.3907±0.0158	0.0305	0.4093±0.0120	0.0232	0.3844±0.0103	0.0603
8	0.4258±0.0121	0.1881	0.4296±0.0090	0.0345	0.3875±0.0122	0.0306	0.4233±0.0181	0.0233	0.3885±0.0175	0.0603
9	0.4582±0.0259	0.2034	0.4782±0.0186	0.0346	0.4425±0.0220	0.0307	0.4782±0.0206	0.0234	0.3968±0.0204	0.0609
10	0.3926±0.0079	0.2188	0.4148±0.0088	0.0344	0.4081±0.0171	0.0308	0.4311±0.0188	0.0236	0.4283±0.0167	0.0614
11	0.4117±0.0263	0.2340	0.4370±0.0163	0.0349	0.3925±0.0068	0.0311	0.4548±0.0149	0.0239	0.4260±0.0148	0.0609
12	0.3935±0.0168	0.2490	0.4176±0.0106	0.0360	0.3717±0.0055	0.0328	0.4462±0.0148	0.0277	0.3986±0.0132	0.0651
13	0.4296±0.0107	0.2639	0.4318±0.0139	0.0352	0.3874±0.0143	0.0313	0.4267±0.0130	0.0243	0.4195±0.0241	0.0623
14	0.3865±0.0067	0.2792	0.4091±0.0090	0.0353	0.3742±0.0052	0.0315	0.4135±0.0136	0.0238	0.3927±0.0151	0.0625
15	0.4044±0.0264	0.2946	0.4128±0.0143	0.0355	0.3758±0.0084	0.0314	0.3960±0.0141	0.0236	0.4110±0.0126	0.0627
16	0.4026±0.0132	0.3098	0.4085±0.0119	0.0357	0.3889±0.0126	0.0315	0.4022±0.0098	0.0234	0.4048±0.0157	0.0624
17	0.4502±0.0229	0.3316	0.4264±0.0128	0.0395	0.3796±0.0102	0.0374	0.4011±0.0139	0.0279	0.3855±0.0076	0.0749
18	0.4202±0.0149	0.3415	0.4191±0.0090	0.0417	0.3648±0.0061	0.0418	0.3783±0.0112	0.0310	0.3614±0.0051	0.0832
19	0.4064±0.0166	0.3558	0.4234±0.0158	0.0392	0.3815±0.0062	0.0373	0.4008±0.0066	0.0271	0.3849±0.0096	0.0720
20	0.4046±0.0156	0.3721	0.4046±0.0109	0.0412	0.3897±0.0129	0.0401	0.4065±0.0086	0.0296	0.3897±0.0115	0.0776



**Figura 6-29.:** Desempeño de cada uno de los clasificadores variando el número de objetos de representación por clase. Conjunto Heart. **Selección de las líneas de características mediante el algoritmo genético (GA).**

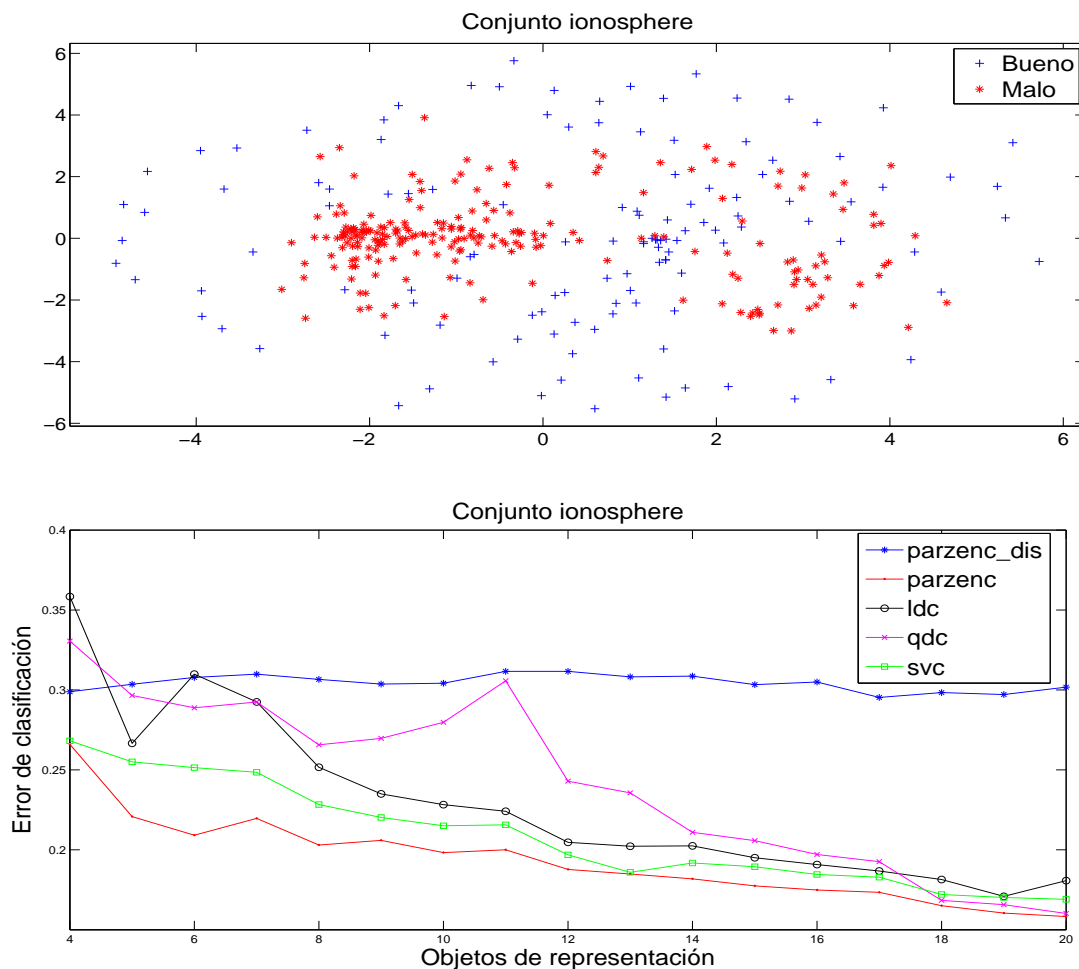
**Tabla 6-55.:** Intervalos de confianza para el error promedio de cada clasificador de acuerdo al número de objetos de representación por clase. **Tiempo de cómputo de la prueba igual a 15 horas, 15 minutos y 18 segundos.**

	<i>parzenc_dis</i>		<i>parzenc</i>		<i>ldc</i>		<i>qdc</i>		<i>svc</i>	
Objetos de representación	$\bar{e}_\infty$	CPU(seg)	$\bar{e}_\infty$	CPU(seg)	$\bar{e}_\infty$	CPU(seg)	$\bar{e}_\infty$	CPU(seg)	$\bar{e}_\infty$	CPU(seg)
4	0.5424	0.1176	0.4610	0.0346	0.5898	0.0310	0.5966	0.0235	0.4373	0.0568
5	0.4573	0.1568	0.4915	0.0343	0.4778	0.0304	0.5324	0.0230	0.5222	0.0547
6	0.4124	0.2613	0.4296	0.0350	0.3608	0.0314	0.4055	0.0239	0.4158	0.0560
7	0.5433	0.3321	0.4118	0.0350	0.3979	0.0363	0.4014	0.0333	0.4152	0.0622
8	0.4564	0.3800	0.3902	0.0356	0.3868	0.0362	0.4321	0.0285	0.3624	0.0565
9	0.4561	0.6866	0.4561	0.0394	0.4807	0.0389	0.4316	0.0358	0.4596	0.0921
10	0.4558	0.8998	0.5088	0.0354	0.4205	0.0324	0.5512	0.0257	0.4806	0.0565
11	0.5445	11.447	0.3772	0.0361	0.4128	0.0324	0.4555	0.0305	0.4342	0.0565
12	0.5448	14.005	0.3728	0.0363	0.3262	0.0327	0.4552	0.0260	0.3369	0.0566
13	0.4549	17.398	0.3971	0.0354	0.3538	0.0327	0.4621	0.0258	0.3574	0.0565
14	0.3782	23.122	0.4109	0.0424	0.3782	0.0442	0.3855	0.0342	0.3782	0.0752
15	0.4542	25.737	0.3956	0.0400	0.3700	0.0360	0.3883	0.0338	0.3516	0.0770
16	0.4539	35.957	0.4760	0.0365	0.4502	0.0329	0.4244	0.0264	0.4613	0.0568
17	0.4535	42.831	0.4572	0.0355	0.3978	0.0325	0.3978	0.0254	0.3903	0.0563
18	0.4532	45.923	0.4120	0.0365	0.3558	0.0334	0.4682	0.0367	0.3783	0.0569
19	0.4528	59.653	0.4679	0.0359	0.3811	0.0328	0.5434	0.0258	0.3434	0.0564
20	0.4525	71.566	0.4601	0.0364	0.3992	0.0335	0.4373	0.0269	0.4449	0.0568



### 6.1.17. Conjunto ionosphere

Estos datos fueron recogidos por un sistema de radar. Este sistema consta de una matriz por etapas de 16 antenas de alta frecuencia. Los objetivos eran los electrones libres en la ionosfera. Es "Bueno" si los ecos del radar son los que muestran evidencia de algún tipo de estructura en la ionosfera. Es "Malo" si los rendimientos son los que no lo hacen, sus señales pasan a través de la ionosfera. Las señales recibidas se procesaron utilizando una función de auto correlación, cuyos argumentos son el momento de un pulso y el número de impulsos.



**Figura 6-30.:** Grafico de dispersión del conjunto de datos proyectado en 2D mediante escalamiento multidimensional y desempeño de cada uno de los clasificadores variando el número de objetos de representación por clase. Conjunto de datos ionosphere. **Selección de las líneas de características de longitud mayor.**

**Tabla 6-56.:** Intervalos de confianza para el error promedio de cada clasificador de acuerdo al número de objetos de representación por clase (Resaltado se encuentra el mejor intervalo para cada clasificador).

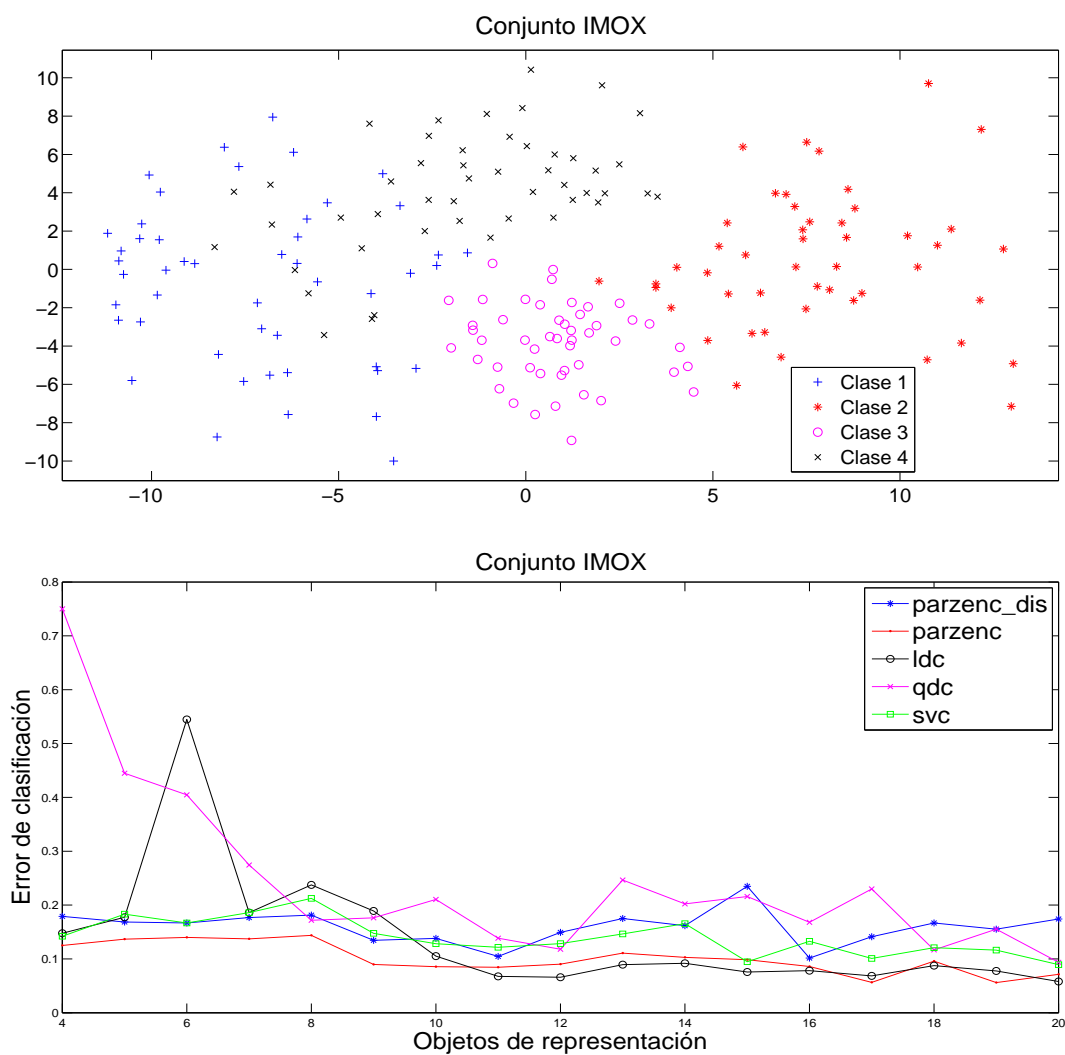
Intervalos de confianza para el error promedio					
Objetos de representación	<i>parzenc_dis</i>	<i>parzenc</i>	<i>ldc</i>	<i>qdc</i>	<i>svc</i>
4	<b>[0.2821,0.3157]</b>	[0.2464,0.2854]	[0.3260,0.3908]	[0.3119,0.3493]	[0.2471,0.2893]
5	[0.2906,0.3164]	[0.2074,0.2342]	[0.2440,0.2892]	[0.2794,0.3136]	[0.2379,0.2721]
6	[0.2936,0.3220]	[0.1975,0.2207]	[0.2838,0.3360]	[0.2667,0.3109]	[0.2324,0.2704]
7	[0.2981,0.3217]	[0.2021,0.2373]	[0.2717,0.3133]	[0.2719,0.3129]	[0.2301,0.2669]
8	[0.2919,0.3213]	[0.1914,0.2146]	[0.2345,0.2687]	[0.2420,0.2894]	[0.2130,0.2436]
9	[0.2940,0.3134]	[0.1914,0.2204]	[0.2163,0.2537]	[0.2429,0.2965]	[0.2040,0.2366]
10	[0.2950,0.3134]	[0.1854,0.2112]	[0.2098,0.2466]	[0.2532,0.3064]	[0.1979,0.2321]
11	[0.3034,0.3198]	[0.1889,0.2111]	[0.2078,0.2404]	[0.2762,0.3352]	[0.2019,0.2293]
12	[0.3032,0.3200]	[0.1751,0.2003]	[0.1902,0.2192]	[0.2187,0.2671]	[0.1853,0.2085]
13	[0.2971,0.3193]	[0.1731,0.1967]	[0.1885,0.2159]	[0.2130,0.2582]	[0.1727,0.1991]
14	[0.3000,0.3174]	[0.1708,0.1930]	[0.1900,0.2148]	[0.1889,0.2331]	[0.1802,0.2034]
15	[0.2965,0.3101]	[0.1659,0.1891]	[0.1818,0.2082]	[0.1834,0.2282]	[0.1773,0.2015]
16	[0.2929,0.3171]	[0.1631,0.1867]	[0.1782,0.2034]	[0.1779,0.2163]	[0.1720,0.1972]
17	[0.2845,0.3061]	[0.1626,0.1842]	[0.1731,0.2005]	[0.1718,0.2134]	[0.1708,0.1950]
18	[0.2899,0.3067]	[0.1550,0.1750]	[0.1682,0.1946]	[0.1526,0.1842]	[0.1608,0.1834]
19	[0.2884,0.3058]	[0.1491,0.1717]	<b>[0.1604,0.1814]</b>	[0.1518,0.1796]	[0.1581,0.1823]
20	[0.2879,0.3157]	<b>[0.1486,0.1680]</b>	[0.1678,0.1936]	<b>[0.1473,0.1731]</b>	<b>[0.1579,0.1801]</b>

**Tabla 6-57.:** Tabla del Error promedio, desviación estándar, tiempo de cómputo, de los clasificadores de acuerdo al número de objetos de representación por clase (resaltado se encuentra el mejor tiempo promedio de cómputo de cada clasificador).

Error promedio, desviación estándar, tiempo de cómputo de los clasificadores										
Objetos de representación	<i>parzenc_dis</i>		<i>parzenc</i>		<i>ldc</i>		<i>qdc</i>		<i>svc</i>	
	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)
4	0.2989 $\pm$ 0.064	<b>0.1311</b>	0.2659 $\pm$ 0.074	<b>0.0346</b>	0.3584 $\pm$ 0.123	0.0314	0.3306 $\pm$ 0.071	0.0236	0.2682 $\pm$ 0.080	<b>0.0558</b>
5	0.3035 $\pm$ 0.049	0.1467	0.2208 $\pm$ 0.051	0.0348	0.2666 $\pm$ 0.086	0.0314	0.2965 $\pm$ 0.065	0.0237	0.255 $\pm$ 0.065	0.0566
6	0.3078 $\pm$ 0.054	0.1621	0.2091 $\pm$ 0.044	0.0349	0.3099 $\pm$ 0.099	<b>0.0312</b>	0.2888 $\pm$ 0.084	0.0237	0.2514 $\pm$ 0.072	0.0587
7	0.3099 $\pm$ 0.045	0.1780	0.2197 $\pm$ 0.067	0.0349	0.2925 $\pm$ 0.079	<b>0.0312</b>	0.2924 $\pm$ 0.078	0.0237	0.2485 $\pm$ 0.070	0.0598
8	0.3066 $\pm$ 0.056	0.1935	0.203 $\pm$ 0.044	0.0349	0.2516 $\pm$ 0.065	0.0313	0.2657 $\pm$ 0.090	0.0238	0.2283 $\pm$ 0.058	0.0600
9	0.3037 $\pm$ 0.037	0.2088	0.2059 $\pm$ 0.055	0.0348	0.235 $\pm$ 0.071	0.0312	0.2697 $\pm$ 0.102	0.0237	0.2203 $\pm$ 0.062	0.0603
10	0.3042 $\pm$ 0.035	0.2241	0.1983 $\pm$ 0.049	0.0348	0.2282 $\pm$ 0.070	0.0312	0.2798 $\pm$ 0.101	0.0234	0.215 $\pm$ 0.065	0.0603
11	0.3116 $\pm$ 0.031	0.2398	0.2 $\pm$ 0.042	0.0349	0.2241 $\pm$ 0.062	0.0312	0.3057 $\pm$ 0.112	<b>0.0231</b>	0.2156 $\pm$ 0.052	0.0604
12	0.3116 $\pm$ 0.032	0.2551	0.1877 $\pm$ 0.048	0.0351	0.2047 $\pm$ 0.055	0.0313	0.2429 $\pm$ 0.092	<b>0.0231</b>	0.1969 $\pm$ 0.044	0.0607
13	0.3082 $\pm$ 0.042	0.2707	0.1849 $\pm$ 0.045	0.0351	0.2022 $\pm$ 0.052	0.0313	0.2356 $\pm$ 0.086	<b>0.0231</b>	0.1859 $\pm$ 0.050	0.0609
14	0.3087 $\pm$ 0.033	0.2857	0.1819 $\pm$ 0.042	0.0355	0.2024 $\pm$ 0.047	0.0317	0.211 $\pm$ 0.084	0.0233	0.1918 $\pm$ 0.044	0.0614
15	0.3033 $\pm$ 0.026	0.3018	0.1775 $\pm$ 0.044	0.0372	0.195 $\pm$ 0.050	0.0318	0.2058 $\pm$ 0.085	0.0234	0.1894 $\pm$ 0.046	0.0617
16	0.305 $\pm$ 0.046	0.3159	0.1749 $\pm$ 0.045	0.0357	0.1908 $\pm$ 0.048	0.0320	0.1971 $\pm$ 0.073	0.0236	0.1846 $\pm$ 0.048	0.0620
17	0.2953 $\pm$ 0.041	0.3334	0.1734 $\pm$ 0.041	0.0359	0.1868 $\pm$ 0.052	0.0319	0.1926 $\pm$ 0.079	0.0235	0.1829 $\pm$ 0.046	0.0622
18	0.2983 $\pm$ 0.032	0.3484	0.165 $\pm$ 0.038	0.0360	0.1814 $\pm$ 0.050	0.0320	0.1684 $\pm$ 0.060	0.0235	0.1721 $\pm$ 0.043	0.0625
19	0.2971 $\pm$ 0.033	0.3642	0.1604 $\pm$ 0.043	0.0361	0.1709 $\pm$ 0.040	0.0319	0.1657 $\pm$ 0.053	0.0235	0.1702 $\pm$ 0.046	0.0627
20	0.3018 $\pm$ 0.053	0.3795	0.1583 $\pm$ 0.037	0.0363	0.1807 $\pm$ 0.049	0.0319	0.1602 $\pm$ 0.049	0.0235	0.169 $\pm$ 0.042	0.0632

### 6.1.18. Conjunto IMOX

Las características son las distancias desde un cuadro de límite a los píxeles de un conjunto de caracteres escritos a mano 'i', 'M', 'O', y 'X', medidos desde las esquinas a lo largo de las diagonales y desde los puntos medios del borde a lo largo de la horizontal y vertical de los ejes centrales



**Figura 6-31.:** Grafico de dispersión del conjunto de datos proyectado en 2D mediante escalamiento multidimensional y desempeño de cada uno de los clasificadores variando el número de objetos de representación por clase. Conjunto de datos IMOX. **Selección de las líneas de características de longitud mayor.**

**Tabla 6-58.:** Intervalos de confianza para el error promedio de cada clasificador de acuerdo al número de objetos de representación por clase (Resaltado se encuentra el mejor intervalo para cada clasificador).

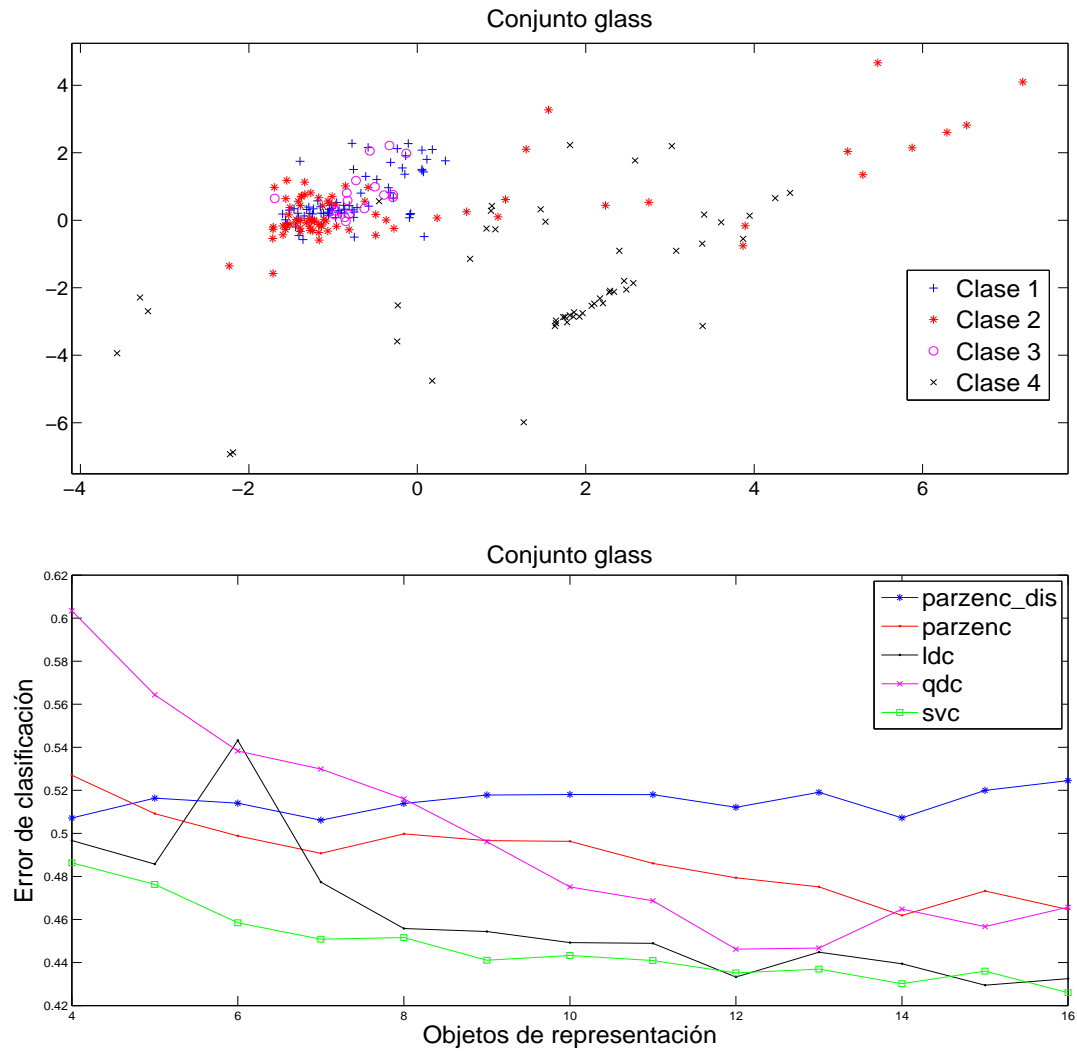
Intervalos de confianza para el error promedio					
Objetos de representación	<i>parzenc_dis</i>	<i>parzenc</i>	<i>ldc</i>	<i>qdc</i>	<i>svc</i>
4	[0.1416,0.2164]	[0.1100,0.1400]	[0.1327,0.1627]	[0.7500,0.7500]	[0.0823,0.2017]
5	[0.0309,0.3063]	[0.1137,0.1595]	[0.0778,0.2768]	[0.3914,0.4982]	[0.0836,0.2826]
6	[0.1196,0.2138]	[0.1165,0.1633]	[0.4114,0.6778]	[0.1229,0.6867]	[0.1196,0.2138]
7	[0.1607,0.1929]	[0.1132,0.1612]	[0.1781,0.1939]	[0.2423,0.3065]	[0.1620,0.2100]
8	[0.1646,0.1978]	[0.0121,0.2753]	[0.1059,0.3691]	[0.1308,0.2130]	[0.2125,0.2125]
9	[0.1346,0.1346]	[0.0897,0.0897]	[0.1807,0.1975]	[0.1679,0.1847]	[0.1137,0.1811]
10	[0.0863,0.1901]	[0.0681,0.1029]	[0.0879,0.1227]	[0.1758,0.2452]	<b>[0.0015,0.2581]</b>
11	[0.0602,0.1492]	[0.0400,0.1290]	[0.0142,0.1210]	[0.0230,0.2540]	[0.0861,0.1571]
12	[0.0306,0.2680]	[0.0356,0.1450]	[0.0202,0.1118]	[0.1181,0.1181]	[0.1011,0.1559]
13	[0.1279,0.2221]	[0.0262,0.1952]	[0.0422,0.1364]	[0.1243,0.3685]	[0.0243,0.2685]
14	[0.1618,0.1618]	[0.0447,0.1611]	[0.0435,0.1403]	[0.0572,0.3472]	[0.1557,0.1751]
15	[0.1551,0.3145]	[0.0985,0.0985]	[0.0358,0.1158]	[0.1462,0.2856]	[0.0847,0.1047]
16	<b>[0.0605,0.1427]</b>	<b>[0.0038,0.1680]</b>	<b>[0.0040,0.1602]</b>	<b>[0.0138,0.3222]</b>	[0.0917,0.1739]
17	[0.1306,0.1516]	[0.0141,0.0989]	[0.0057,0.1427]	[0.1766,0.2830]	[0.0476,0.1540]
18	[0.1009,0.2325]	[0.0847,0.1069]	[0.0546,0.1204]	[0.0949,0.1385]	[0.0879,0.1537]
19	[0.0717,0.3821]	[0.0220,0.0900]	[0.0776,0.0776]	[0.0264,0.3368]	[0.0824,0.1504]
20	[0.0683,0.2799]	[0.0243,0.1185]	[0.0462,0.0698]	[0.0820,0.1056]	[0.0659,0.1127]

**Tabla 6-59.:** Tabla del Error promedio, desviación estándar, tiempo de cómputo, de los clasificadores de acuerdo al número de objetos de representación por clase (resaltado se encuentra el mejor tiempo promedio de cómputo de cada clasificador).

Error promedio, desviación estándar, tiempo de cómputo de los clasificadores										
Objetos de representación	<i>parzenc_dis</i>		<i>parzenc</i>		<i>ldc</i>		<i>qdc</i>		<i>svc</i>	
	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)
4	0.179 $\pm$ 0.142	<b>0.2946</b>	0.125 $\pm$ 0.057	0.0476	0.1477 $\pm$ 0.057	0.0459	0.75 $\pm$ 0.000	0.0316	0.142 $\pm$ 0.227	<b>0.2602</b>
5	0.1686 $\pm$ 0.523	0.35	0.1366 $\pm$ 0.087	0.0474	0.1773 $\pm$ 0.378	0.046	0.4448 $\pm$ 0.203	0.0318	0.1831 $\pm$ 0.378	0.2725
6	0.1667 $\pm$ 0.179	0.4054	0.1399 $\pm$ 0.089	0.0472	0.5446 $\pm$ 0.506	<b>0.0453</b>	0.4048 $\pm$ 1.071	0.032	0.1667 $\pm$ 0.179	0.2864
7	0.1768 $\pm$ 0.061	0.4639	0.1372 $\pm$ 0.091	0.0474	0.186 $\pm$ 0.030	0.0454	0.2744 $\pm$ 0.122	<b>0.0315</b>	0.186 $\pm$ 0.091	0.2857
8	0.1812 $\pm$ 0.063	0.5241	0.1437 $\pm$ 0.5	0.0481	0.2375 $\pm$ 0.500	0.0455	0.1719 $\pm$ 0.156	0.0317	0.2125 $\pm$ 0.000	0.2893
9	0.1346 $\pm$ 0.000	0.5795	0.0897 $\pm$ 0	0.048	0.1891 $\pm$ 0.032	0.0454	0.1763 $\pm$ 0.032	0.0319	0.1474 $\pm$ 0.128	0.2927
10	0.1382 $\pm$ 0.197	0.6405	0.0855 $\pm$ 0.066	<b>0.0474</b>	0.1053 $\pm$ 0.066	0.0456	0.2105 $\pm$ 0.132	0.0318	0.1283 $\pm$ 0.493	0.2951
11	0.1047 $\pm$ 0.169	0.6983	0.0845 $\pm$ 0.169	0.0483	0.0676 $\pm$ 0.203	0.0454	0.1385 $\pm$ 0.439	0.0317	0.1216 $\pm$ 0.135	0.2964
12	0.1493 $\pm$ 0.451	0.7577	0.0903 $\pm$ 0.208	0.0495	0.066 $\pm$ 0.174	0.046	0.1181 $\pm$ 0.000	0.0317	0.1285 $\pm$ 0.104	0.3057
13	0.175 $\pm$ 0.179	0.815	0.1107 $\pm$ 0.321	0.0493	0.0893 $\pm$ 0.179	0.0457	0.2464 $\pm$ 0.464	0.0318	0.1464 $\pm$ 0.464	0.3078
14	0.1618 $\pm$ 0.000	0.8732	0.1029 $\pm$ 0.221	0.0503	0.0919 $\pm$ 0.184	0.0461	0.2022 $\pm$ 0.551	0.0319	0.1654 $\pm$ 0.037	0.3102
15	0.2348 $\pm$ 0.303	0.9332	0.0985 $\pm$ 0	0.0493	0.0758 $\pm$ 0.152	0.0458	0.2159 $\pm$ 0.265	0.0317	0.0947 $\pm$ 0.038	0.3122
16	0.1016 $\pm$ 0.156	0.991	0.0859 $\pm$ 0.312	0.0513	0.0781 $\pm$ 0.312	0.046	0.168 $\pm$ 0.586	0.0318	0.1328 $\pm$ 0.156	0.3185
17	0.1411 $\pm$ 0.040	10.603	0.0565 $\pm$ 0.161	0.051	0.0685 $\pm$ 0.282	0.0462	0.2298 $\pm$ 0.202	0.0319	0.1008 $\pm$ 0.202	0.3233
18	0.1667 $\pm$ 0.250	11.068	0.0958 $\pm$ 0.042	0.0515	0.0875 $\pm$ 0.125	0.0456	0.1167 $\pm$ 0.083	0.0318	0.1208 $\pm$ 0.125	0.3311
19	0.1552 $\pm$ 0.862	11.726	0.056 $\pm$ 0.129	0.0519	0.0776 $\pm$ 0.000	0.0461	0.1552 $\pm$ 0.690	0.0318	0.1164 $\pm$ 0.129	0.337
20	0.1741 $\pm$ 0.402	12.291	0.0714 $\pm$ 0.179	0.0518	0.058 $\pm$ 0.045	0.046	0.0938 $\pm$ 0.045	0.032	0.0893 $\pm$ 0.089	0.3456

### 6.1.19. Conjunto glass

Conjunto de datos que describe los componentes químicos en varios tipos de gafas. Consiste de 214 objetos divididos en 4 clases, cada objeto tiene 9 características.



**Figura 6-32.:** Gráfico de dispersión del conjunto de datos proyectado en 2D mediante escalamiento multidimensional y desempeño de cada uno de los clasificadores variando el número de objetos de representación por clase. Conjunto de datos ionosphere. **Selección de las líneas de características de longitud mayor.**

**Tabla 6-60.:** Intervalos de confianza para el error promedio de cada clasificador de acuerdo al número de objetos de representación por clase (Resaltado se encuentra el mejor intervalo para cada clasificador).

Intervalos de confianza para el error promedio					
Objetos de representación	<i>parzenc_dis</i>	<i>parzenc</i>	<i>ldc</i>	<i>qdc</i>	<i>svc</i>
4	[0.4869,0.5275]	[0.5075,0.5465]	[0.4796,0.5138]	[0.4680,0.5048]	[0.4680,0.5048]
5	[0.4943,0.5385]	[0.4889,0.5295]	[0.4646,0.5068]	[0.4566,0.4960]	[0.4566,0.4960]
6	[0.4970,0.5312]	[0.4814,0.5162]	[0.5240,0.5624]	[0.4403,0.4767]	[0.4403,0.4767]
7	<b>[0.4858,0.5264]</b>	[0.4721,0.5095]	[0.4621,0.4927]	[0.4335,0.4683]	[0.4335,0.4683]
8	[0.4934,0.5344]	[0.4801,0.5195]	[0.4419,0.4697]	[0.4350,0.4682]	[0.4350,0.4682]
9	[0.4970,0.5386]	[0.4761,0.5171]	[0.4397,0.4691]	[0.4224,0.4598]	[0.4224,0.4598]
10	[0.4959,0.5401]	[0.4768,0.5158]	[0.4354,0.4632]	[0.4249,0.4617]	[0.4249,0.4617]
11	[0.4980,0.5380]	[0.4679,0.5043]	[0.4365,0.4613]	[0.4246,0.4572]	[0.4246,0.4572]
12	[0.4895,0.5347]	[0.4618,0.4970]	[0.4225,0.4441]	[0.4199,0.4505]	[0.4199,0.4505]
13	[0.4954,0.5428]	[0.4568,0.4936]	[0.4332,0.4564]	[0.4212,0.4528]	[0.4212,0.4528]
14	[0.4859,0.5285]	<b>[0.4458,0.4780]</b>	[0.4283,0.4505]	[0.4131,0.4473]	[0.4131,0.4473]
15	[0.4942,0.5458]	[0.4566,0.4898]	<b>[0.4195,0.4395]</b>	[0.4207,0.4513]	[0.4207,0.4513]
16	[0.4963,0.5527]	[0.4473,0.4821]	[0.4201,0.4449]	<b>[0.4086,0.4434]</b>	<b>[0.4086,0.4434]</b>



**Tabla 6-61.:** Tabla del Error promedio, desviación estándar, tiempo de cómputo, de los clasificadores de acuerdo al número de objetos de representación por clase (resaltado se encuentra el mejor tiempo promedio de cómputo de cada clasificador).

Error promedio, desviación estándar, tiempo de cómputo de los clasificadores										
Objetos de representación	<i>parzenc_dis</i>		<i>parzenc</i>		<i>ldc</i>		<i>qdc</i>		<i>svc</i>	
	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)
4	$0.5072 \pm 0.077$	<b>0.2825</b>	$0.527 \pm 0.074$	<b>0.0472</b>	$0.4967 \pm 0.065$	0.0455	$0.6034 \pm 0.6310$	<b>0.0312</b>	$0.4864 \pm 0.070$	0.27
5	$0.5164 \pm 0.084$	0.3386	$0.5092 \pm 0.077$	0.0473	$0.4857 \pm 0.080$	0.0456	$0.5644 \pm 0.5879$	0.0315	$0.4763 \pm 0.075$	<b>0.2563</b>
6	$0.5141 \pm 0.065$	0.3943	$0.4988 \pm 0.066$	0.0475	$0.5432 \pm 0.073$	0.0451	$0.5383 \pm 0.5590$	0.0316	$0.4585 \pm 0.069$	0.2783
7	$0.5061 \pm 0.077$	0.4508	$0.4908 \pm 0.071$	0.0476	$0.4774 \pm 0.058$	0.0451	$0.5299 \pm 0.5492$	0.0315	$0.4509 \pm 0.066$	0.2804
8	$0.5139 \pm 0.078$	0.5073	$0.4998 \pm 0.075$	0.0475	$0.4558 \pm 0.053$	<b>0.0450</b>	$0.5160 \pm 0.5351$	0.0314	$0.4516 \pm 0.063$	0.2819
9	$0.5178 \pm 0.079$	0.5647	$0.4966 \pm 0.078$	0.0478	$0.4544 \pm 0.056$	0.0451	$0.4961 \pm 0.5132$	0.0314	$0.4411 \pm 0.071$	0.2833
10	$0.518 \pm 0.084$	0.6212	$0.4963 \pm 0.074$	0.0477	$0.4493 \pm 0.053$	0.0451	$0.4751 \pm 0.4935$	0.0314	$0.4433 \pm 0.070$	0.2851
11	$0.518 \pm 0.076$	0.6783	$0.4861 \pm 0.069$	0.0481	$0.4489 \pm 0.047$	0.0452	$0.4687 \pm 0.4864$	0.0315	$0.4409 \pm 0.062$	0.2879
12	$0.5121 \pm 0.086$	0.7376	$0.4794 \pm 0.067$	0.0498	$0.4333 \pm 0.041$	0.0457	$0.4462 \pm 0.4628$	0.0315	$0.4352 \pm 0.058$	0.2935
13	$0.5191 \pm 0.090$	0.7964	$0.4752 \pm 0.07$	0.0501	$0.4448 \pm 0.044$	0.0457	$0.4467 \pm 0.4599$	0.0315	$0.437 \pm 0.060$	0.2981
14	$0.5072 \pm 0.081$	0.8511	$0.4619 \pm 0.061$	0.0502	$0.4394 \pm 0.042$	0.0457	$0.4648 \pm 0.4814$	0.0315	$0.4302 \pm 0.065$	0.3008
15	$0.52 \pm 0.098$	0.9070	$0.4732 \pm 0.063$	0.0500	$0.4295 \pm 0.038$	0.0456	$0.4567 \pm 0.4715$	0.0314	$0.436 \pm 0.058$	0.305
16	$0.5245 \pm 0.107$	0.9640	$0.4647 \pm 0.066$	0.0502	$0.4325 \pm 0.047$	0.0458	$0.4657 \pm 0.4837$	0.0316	$0.426 \pm 0.066$	0.3107

### 6.1.20. Conjunto de imágenes ORL

El conjunto de imágenes ORL, es un conjunto de rostros utilizada en reconocimiento facial. Se utiliza en publicaciones como un ejemplo de un conjunto sencillo de imágenes dado que, las imágenes son tomadas bajo condiciones muy similares. Este conjunto contiene 400 imágenes de rostros con 40 clases, es decir cada clase esta formada por un conjunto de diez caras de una persona. En este conjunto, los pixeles de cada imagen se interpretan como características y como cada imagen tiene un tamaño de  $112 \times 92$  entonces, los vectores de características tienen una longitud de 10304 [6-34](#). En este caso para realizar las pruebas se hace una reducción del conjunto mediante el análisis de componentes principales (*PCA*), y se calcula las primeras 24 componentes, llamadas *eigenrostros* o rostros propios [6-33](#). A continuación se presentan gráficos de este conjunto y un gráfico de dispersión construido mediante los dos primeros rostros propios.



**Figura 6-33.:** Rostros propios del seleccionados del conjunto de datos ORL.

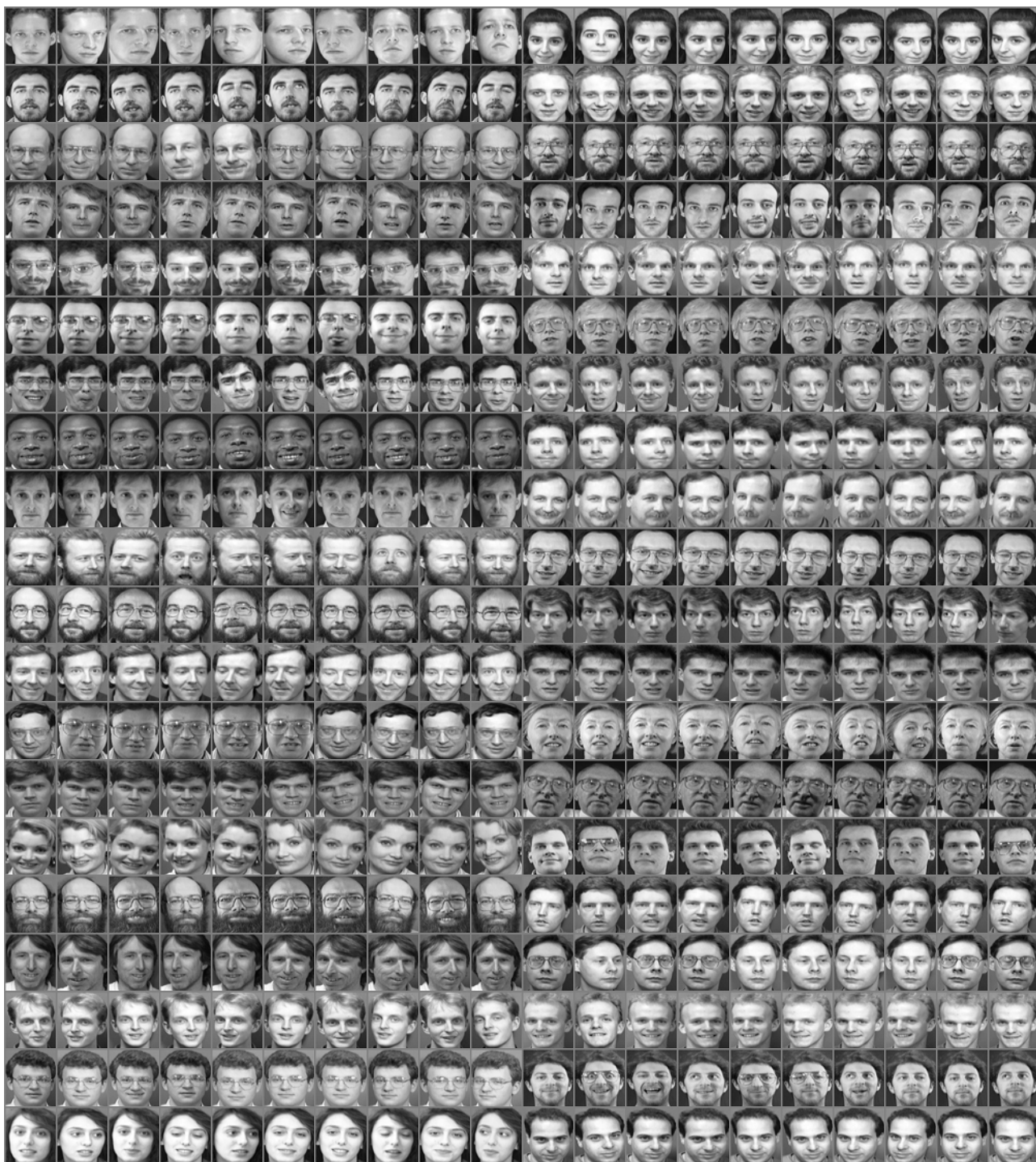
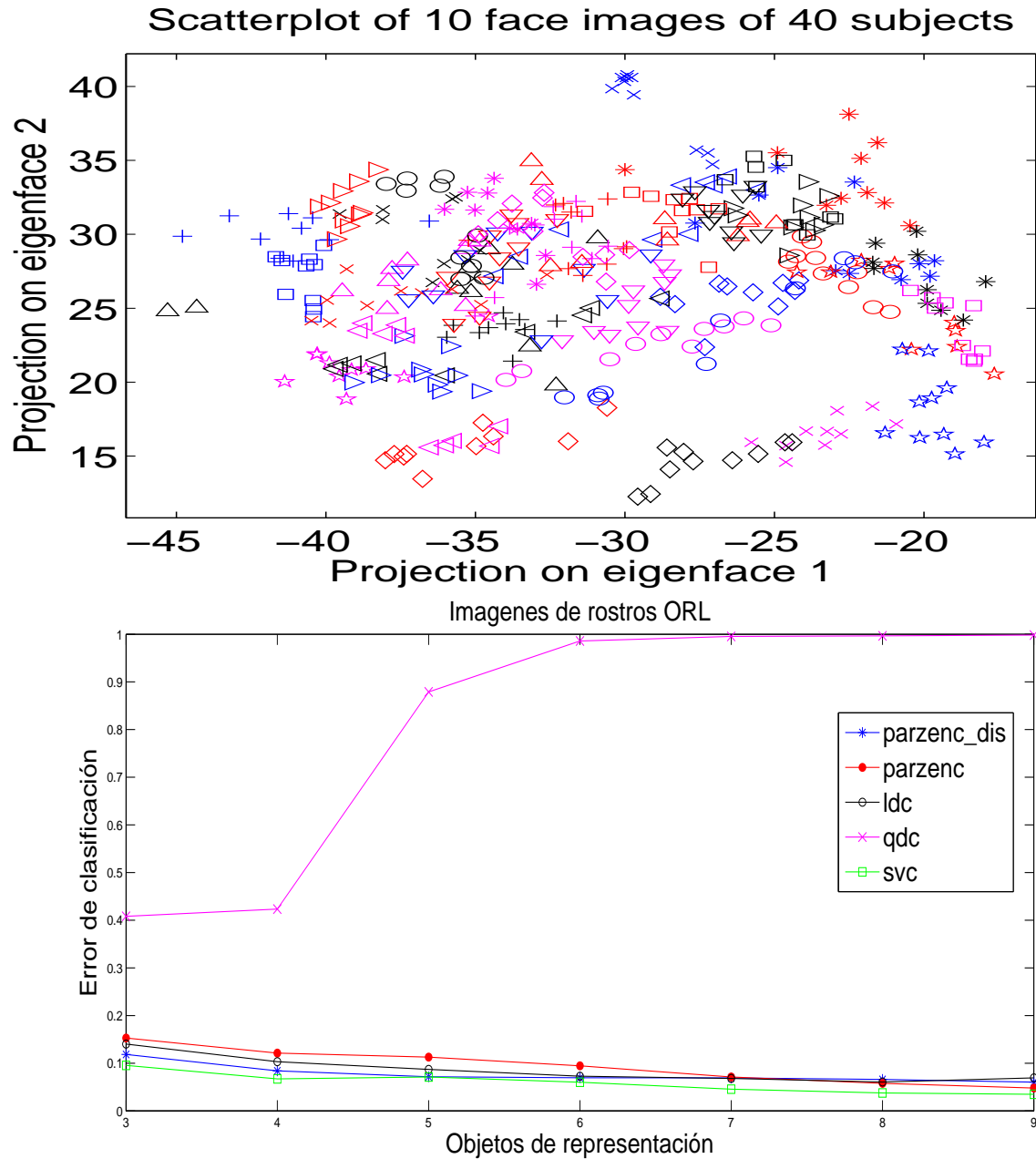


Figura 6-34.: Rostros del conjunto de datos ORL.



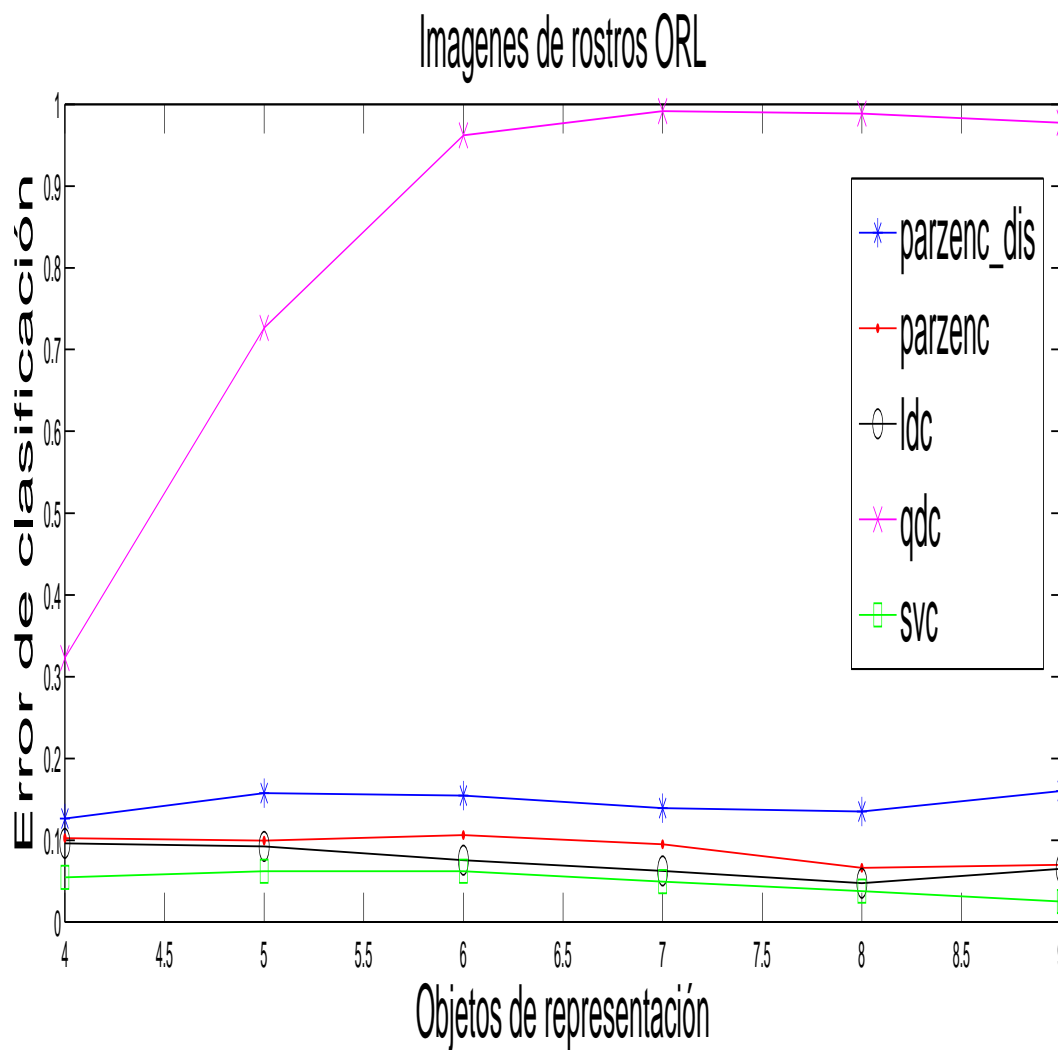
**Figura 6-35.:** Gráfico de dispersión del conjunto de datos proyectado en 2D mediante *PCA* y Desempeño de cada uno de los clasificadores variando el número de objetos de representación por clase. Conjunto de imágenes ORL. **Selección de las líneas de características de longitud mayor.**

**Tabla 6-62.:** Intervalos de confianza para el error promedio de cada clasificador de acuerdo al número de objetos de representación por clase. (Resaltado se encuentra el mejor intervalo para cada clasificador).

Intervalos de confianza para el error promedio					
Objetos de representación	<i>parzenc_dis</i>	<i>parzenc</i>	<i>ldc</i>	<i>qdc</i>	<i>svc</i>
3	[0.1123,0.1245]	[0.1461,0.1596]	[0.1335,0.1468]	<b>[0.3862,0.4300]</b>	[0.0896,0.1013]
4	[0.0785,0.0893]	[0.1144,0.1280]	[0.0967,0.1095]	[0.3730,0.4740]	[0.0619,0.0725]
5	[0.0672,0.0761]	[0.1066,0.1189]	[0.0813,0.0927]	[0.8487,0.9093]	[0.0658,0.0766]
6	[0.0648,0.0739]	[0.0887,0.1004]	[0.0669,0.0783]	[0.9782,0.9933]	[0.0554,0.0648]
7	[0.0643,0.0741]	[0.0645,0.0777]	[0.0621,0.0739]	[0.9930,0.9977]	[0.0408,0.0504]
8	[0.0591,0.0726]	[0.0512,0.0640]	[0.0547,0.0666]	<b>[0.9946,0.9984]</b>	[0.0324,0.0426]
9	<b>[0.0513,0.0692]</b>	<b>[0.0410,0.0550]</b>	[0.0605,0.0780]	[0.9966,0.9999]	<b>[0.0286,0.0409]</b>

**Tabla 6-63.:** Tabla del Error promedio, desviación estándar, tiempo de cómputo, de los clasificadores de acuerdo al número de objetos de representación por clase (resaltado se encuentra el mejor tiempo promedio de cómputo de cada clasificador).

Error promedio, desviación estándar, tiempo de cómputo de los clasificadores										
Objetos de representación	<i>parzenc_dis</i>		<i>parzenc</i>		<i>ldc</i>		<i>qdc</i>		<i>svc</i>	
	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)
3	0.1184±0.0024	<b>132.855</b>	0.1528±0.0026	48.694	0.1402±0.0026	0.4601	0.4081±0.0085	<b>0.4029</b>	0.0955±0.0023	<b>169.508</b>
4	0.0839±0.0021	177.749	0.1212±0.0026	46.852	0.1031±0.0025	0.4761	0.4235±0.0196	0.5748	0.0672±0.0021	218.821
5	0.0717±0.0017	225.358	0.1128±0.0024	46.864	0.0870±0.0022	<b>0.4423</b>	0.8790±0.0118	0.5820	0.0712±0.0021	288.419
6	0.0694±0.0018	274.114	0.0946±0.0023	46.778	0.0726±0.0022	0.4663	0.9858±0.0029	0.6158	0.0601±0.0018	396.478
7	0.0692±0.0019	324.522	0.0711±0.0026	46.411	0.0680±0.0023	0.4664	0.9953±0.0009	0.6156	0.0456±0.0019	524.720
8	0.0659±0.0026	377.320	0.0576±0.0025	<b>43.153</b>	0.0606±0.0023	0.4599	0.9965±0.0007	0.6078	0.0375±0.0020	705.715
9	0.0602±0.0035	430.619	0.0480±0.0027	44.129	0.0692±0.0034	0.4549	0.9983±0.0006	0.6027	0.0347±0.0024	931.628



**Figura 6-36.:** Desempeño de cada uno de los clasificadores variando el número de objetos de representación por clase. Conjunto de imágenes ORL. **Selección de las líneas de características de longitud menor.**

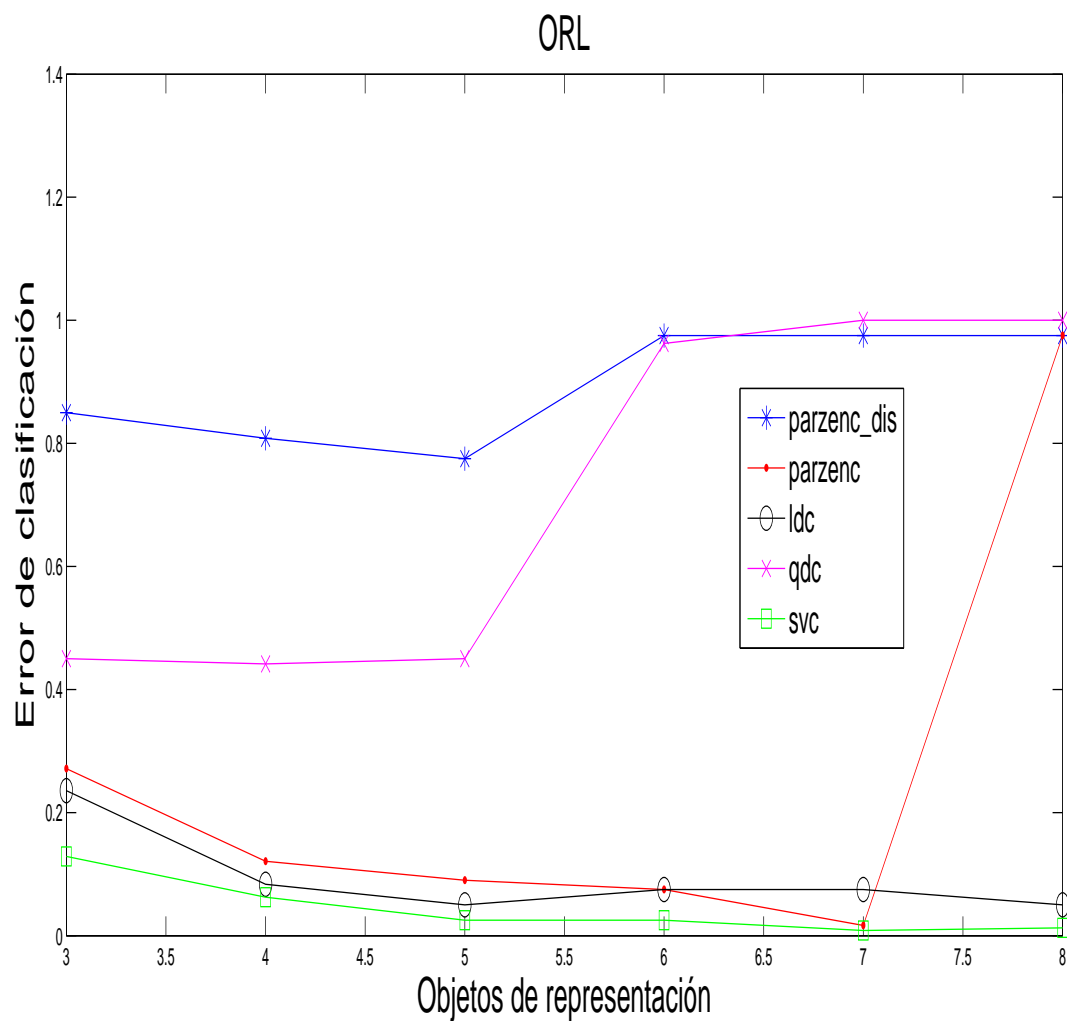
**Tabla 6-64.:** Intervalos de confianza para el error promedio de cada clasificador de acuerdo al número de objetos de representación por clase (Resaltado se encuentra el mejor intervalo para el clasificador *parzenc<sub>dis</sub>*).

Intervalos de confianza para el error promedio					
Objetos de representación	<i>parzenc_dis</i>	<i>parzenc</i>	<i>ldc</i>	<i>qdc</i>	<i>svc</i>
4	[0.1127,0.1398]	[0.0906,0.1144]	[0.0816,0.1109]	[0.1437,0.5013]	[0.0429,0.0663]
5	[0.1324,0.1826]	[0.0818,0.1172]	[0.0677,0.1173]	[0.5432,0.9098]	[0.0473,0.0767]
6	[0.1357,0.1731]	[0.0865,0.1260]	[0.0473,0.1039]	[0.9124,10.114]	[0.0439,0.0799]
7	[0.1157,0.1626]	[0.0838,0.1062]	[0.0468,0.0782]	[0.9816,10.018]	[0.0327,0.0656]
8	[0.1160,0.1540]	[0.0527,0.0798]	[0.0333,0.0617]	[0.9712,10.063]	[0.0268,0.0482]
9	<b>[0.1110,0.2090]</b>	[0.0433,0.0967]	[0.0412,0.0888]	[0.9331,10.219]	[0.0015,0.0485]



**Tabla 6-65.:** Tabla del Error promedio, desviación estándar, tiempo de cómputo, de los clasificadores de acuerdo al número de objetos de representación por clase (resaltado se encuentra el mejor tiempo promedio de cómputo para el clasificador  $parzenc_{dis}$ ).

Error promedio, desviación estándar, tiempo de cómputo de los clasificadores										
Objetos de representación	$parzenc_{dis}$		$parzenc$		$ldc$		$qdc$		$svc$	
	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)	$\bar{e}_{\infty} \pm \sigma$	CPU(seg)
4	0.1263±0.0053	<b>240.940</b>	0.1025±0.0046	50.037	0.0962±0.0057	0.4618	0.3225±0.0696	0.5721	0.0546±0.0045	199.213
5	0.1575±0.0098	304.851	0.0995±0.0069	49.386	0.0925±0.0096	0.4500	0.7265±0.0713	0.7243	0.0620±0.0057	272.473
6	0.1544±0.0073	371.191	0.1062±0.0077	53.483	0.0756±0.0110	0.5155	0.9619±0.0193	0.8777	0.0619±0.0070	349.280
7	0.1392±0.0091	442.253	0.0950±0.0043	48.191	0.0625±0.0061	0.4793	0.9917±0.0039	0.9337	0.0492±0.0064	499.431
8	0.1350±0.0074	509.515	0.0663±0.0053	47.962	0.0475±0.0055	0.5096	0.9888±0.0068	10.540	0.0375±0.0042	840.778
9	0.1600±0.0191	579.359	0.0700±0.0104	47.024	0.0650±0.0093	0.5029	0.9775±0.0173	11.190	0.0250±0.0091	12757.50



**Figura 6-37.:** Desempeño de cada uno de los clasificadores variando el número de objetos de representación por clase. Conjunto de imágenes ORL. **Selección de las líneas de características mediante el algoritmo genético (GA).**

**Tabla 6-66.:** Intervalos de confianza para el error promedio de cada clasificador de acuerdo al número de objetos de representación por clase. Tiempo de cómputo de la prueba igual a un día, diez horas, 4 minutos y 37 segundos.

Objetos de representa- ción	<i>parzenc_dis</i>		<i>parzenc</i>		<i>ldc</i>		<i>qdc</i>		<i>svc</i>	
	$\bar{e}_\infty$	CPU(seg)	$\bar{e}_\infty$	CPU(seg)	$\bar{e}_\infty$	CPU(seg)	$\bar{e}_\infty$	CPU(seg)	$\bar{e}_\infty$	CPU(seg)
3	0.8500	62.026	0.2714	119.456	0.2357	0.3570	0.4500	0.2596	0.1286	199.953
4	0.8083	184.955	0.1208	42.105	0.0833	0.4001	0.4417	0.5323	0.0625	208.799
5	0.7750	389.687	0.0900	41.353	0.0500	0.5362	0.4500	0.9388	0.0250	234.207
6	0.9750	803.773	0.0750	44.804	0.0750	0.6270	0.9625	14.040	0.0250	409.497
7	0.9750	1.361.057	0.0167	37.210	0.0750	0.5966	10.000	19.223	0.0083	428.033
8	0.9750	2.077.098	0.9750	36.387	0.0500	0.5592	10.000	19.173	0.0125	307.792

Las tablas 6-67 y 6-68 muestran el desempeño de los clasificadores *ldc*, *qdc*, *svc* y *parzenc*, junto con el clasificador *knn* en el espacio de características, se presentan con el motivo fundamental de verificar si existe gran diferencia entre los desempeños de los clasificadores en el espacio generalizado de disimilitudes y en el espacio de características y además verificar si el desempeño del clasificador *parzenc\_dis* es competitivo con los desempeños de los clasificadores en el espacio de características, competitivo respecto a el error y al costo computacional.

**Tabla 6-67.:** Error promedio de clasificación en el espacio de características

Errores promedio					
Bases de datos	<i>parzenc</i>	<i>ldc</i>	<i>qdc</i>	<i>svc</i>	<i>knn</i>
Clases Difíciles	0.074	0.0595	0.0575	0.0672	0.0865
Clases Difíciles	0.1091	0.0827	0.0956	0.0974	0.1166
Clases de Highleyman	0.0768	0.1596	0.0651	0.1113	0.0805
Conjunto banana	0.0168	0.1561	0.1563	0.1569	0.017
Círculos concéntricos	0.2065	0.5081	0.1884	0.5016	0.2042
Clases lituanas	0.0235	0.171	0.085	0.1721	0.0255
Clases distribuidas de manera gaussiana	0.192	0.1427	0.1737	0.1527	0.19
Truncamiento de dos clases gaussianas	0.0484	0.029	0.0386	0.0365	0.0394
Laryngeal 1	0.2586	0.1783	0.2163	0.174	0.2497
Laryngeal 2	0.0448	0.0469	0.2037	0.0321	0.052
Laryngeal 3	0.3188	0.2785	0.4054	0.2649	0.3072
Iris	0.0411	0.0211	0.0278	0.0304	0.042
Wine	0.2871	0.0158	0.0131	0.0417	0.2925
Malaysia	0.8417	0.3296	0.7294	0.5254	0.4519
Diabetes	0.2697	0.2284	0.2558	0.229	0.2665
IMOX	0.0561	0.0852	0.0482	0.1062	0.0648
Biomed	0.124	0.1152	0.0919	0.1005	0.1205
Heart	0.345	0.1712	0.1917	0.178	0.35
Glass	0.2875	0.3484	0.3973	0.3927	0.2703
Ionosphere	0.1502	0.1364	0.1079	0.1292	0.1456
ORL	0.9750	0.0444	0.2993	0.0198	0.365

**Tabla 6-68.:** Intervalos de confianza para el error promedio y tiempo promedio de cómputo en el espacio de características

Intervalos de confianza para el error promedio y tiempo promedio de cómputo										
Bases de datos	<i>parzenc</i>		<i>ldc</i>		<i>qdc</i>		<i>svc</i>		<i>knn</i>	
	I.C.	CPU(s)	I.C.	CPU(s)	I.C.	CPU(s)	I.C.	CPU(s)	I.C.	CPU(s)
Clases Difíciles	[0.0682,0.0799]	0.174	[0.0541,0.0648]	0.0332	[0.0522,0.0629]	0.0236	[0.0615,0.0729]	0.3611	[0.0797,0.0933]	0.0362
Clases Difíciles	[ 0.102,0.1163]	0.3205	[0.0766,0.0888]	0.0444	[0.0866,0.1046]	0.032	[0.0901,0.1047]	0.6472	[0.1092, 0.124]	0.05
Clases de Highleyman	[0.0716,0.0821]	0.1794	[0.1518,0.1673]	0.0401	[0.0608,0.0695]	0.0283	[0.1059,0.1167]	0.6572	[0.0751,0.0859]	0.0527
Conjunto banana	[0.0155, 0.018]	10.731	[0.1526,0.1595]	0.0413	[0.1529,0.1597]	0.0295	[0.1533,0.1604]	469.551	[0.0156,0.0184]	0.4006
Círculos concéntricos	[0.1987,0.2143]	0.1602	[0.4979,0.5183]	0.0408	[0.1801,0.1967]	0.0287	[0.4976,0.5056]	0.3638	[0.1971,0.2112]	0.05
Clases lituanas	[0.0212,0.0259]	0.201	[0.1646,0.1774]	0.0363	[0.0791,0.0908]	0.0257	[0.1655,0.1786]	12.591	[0.0228,0.0282]	0.0656
Clases distribuidas de manera gaussiana	[0.1841,0.1998]	0.1381	[0.1367,0.1487]	0.0387	[0.1669,0.1805]	0.0278	[0.1469,0.1585]	16.559	[0.1823,0.1976]	0.0628
Truncamiento de dos clases gaussianas	[0.0446,0.0523]	0.1789	[0.0263,0.0317]	0.0423	[0.0358,0.0413]	0.0315	[0.0333,0.0397]	36.284	[0.0363,0.0425]	0.0809
Laryngeal 1	[0.2461,0.2711]	0.1629	[0.1669,0.1896]	0.0469	[0.2049,0.2278]	0.0347	[0.1629,0.1851]	0.2047	[0.2376,0.2618]	0.0494
Laryngeal 2	[0.0419,0.0477]	0.3344	[0.0435,0.0503]	0.0479	[0.1924,0.2151]	0.0364	[0.0296,0.0346]	17.685	[0.0486,0.0555]	0.0835
Laryngeal 3	[0.3125,0.3252]	0.293	[0.2685,0.2884]	0.0629	[0.3935,0.4173]	0.0442	[0.2573,0.2725]	18.132	[0.2989,0.3156]	0.0568
Iris	[0.0352,0.0471]	0.2789	[ 0.016,0.0262]	0.0643	[0.0216, 0.034]	0.044	[0.0246,0.0362]	0.6323	[0.0354,0.0486]	0.0479
Wine	[0.2736,0.3007]	0.4213	[0.0118,0.0197]	0.0635	[0.0086,0.0175]	0.0417	[0.0364,0.0471]	0.8537	[0.2782,0.3068]	0.0549
Malaysia	[0.8329,0.8505]	0.4858	[0.3196,0.3396]	0.1946	[0.7182,0.7406]	0.1021	[0.5167,0.5341]	42.371	[0.4386,0.4652]	0.114
Diabetes	[0.2636,0.2758]	0.4471	[0.2229, 0.234]	0.0502	[0.2491,0.2625]	0.0361	[0.2236,0.2343]	26.055	[0.2615,0.2715]	0.1015
IMOX	[ 0.049,0.0632]	0.1787	[0.0765,0.0939]	0.0772	[0.0414, 0.055]	0.048	[0.0964,0.1161]	27.253	[0.0578,0.0718]	0.0528
Biomed	[0.1159,0.1321]	0.215	[ 0.106,0.1243]	0.0485	[0.0831,0.1006]	0.0348	[0.0914,0.1096]	0.206	[0.1119,0.1291]	0.0544
Heart	[0.3334,0.3566]	0.1807	[0.1622,0.1802]	0.0515	[0.1818,0.2015]	0.0359	[0.1682,0.1878]	0.3148	[0.3393,0.3607]	0.0511
Glass	[ 0.275,0.2999]	0.242	[0.3353,0.3615]	0.0813	[0.3849,0.4097]	0.052	[0.3747,0.4107]	38.367	[0.2561,0.2845]	0.0574
Ionosphere	[0.1432,0.1572]	0.115	[0.1289, 0.144]	0.048	[0.1011,0.1146]	0.0386	[0.1219,0.1365]	0.5377	[0.1372,0.1539]	0.0483
Ionosphere	[0.9750,0.9750]	16.8552	[0.0395,0.0494]	0.6770	[0.2894,0.3093]	3.1435	[0.0168,0.0229]	23.9364	[0.0326,0.0404]	0.1787

## 6.2. Discusiones

En términos generales el desempeño del clasificador de Parzen para el espacio generalizado de disimilitudes no es el mejor, dado que en la mayoría de las bases de datos utilizadas el desempeño, con respecto a los demás clasificadores, no fue el mejor y además, el tiempo de cómputo siempre es el peor de los 5 clasificadores. La anterior afirmación se puede verificar en los intervalos de confianza para el error promedio de los clasificadores y las tablas para el error promedio, la desviación estándar y el tiempo de cómputo referenciadas en las tablas 6-1 a 6-49. Ahora, para algunos conjuntos de datos el desempeño del clasificador *parzenc\_dis* resulta ser competitivo o mejor en cuanto al error promedio, en comparación a los demás clasificadores, como se analiza a continuación:

Para el conjunto de Clases difíciles distribuidas normalmente de dimensión 2 (ver gráfica 6-1), variando el número de objetos de representación por clase, el desempeño del clasificador de Parzen para el espacio generalizado de disimilitudes, resulta ser mejor que el de los demás clasificadores, en cada uno de los casos tomados (número de objetos de representación), según se observa en la tabla 6-1, en la cual se muestran los intervalos de confianza para el error promedio de cada clasificador. En esta, se encuentra resaltado el mejor intervalo de confianza para cada uno de los clasificadores, notando que el clasificador *parzenc\_dis* tiene el mejor desempeño de los 5 clasificadores.

A medida que el número de objetos de representación aumenta, el desempeño de los clasificadores *ldc*, *qdc*, *svc*, *parzenc* mejora de manera constante, al igual para el clasificador *parzenc\_dis*. Se puede notar también en la tabla 6-1 que, los intervalos de confianza a partir de la selección de 8 objetos de representación por clase para el clasificador *parzenc\_dis*, se encuentran entre el 6 % y el 8 %. Esto permite afirmar que, el clasificador de Parzen para el espacio generalizado de disimilitudes no requiere de un conjunto de entrenamiento extenso, evitando caer en problemas tales como, el sobre entrenamiento, que es conocido como uno de los inconvenientes en clasificación de patrones. Además, obtener buenos resultados a partir de conjuntos de entrenamiento pequeños muestra las bondades del método propuesto en cuanto a la caracterización del conjunto de datos original, logrando enriquecerlo a partir de pocos objetos de representación tomados.

Por otra parte, aunque el desempeño del clasificador de Parzen para el espacio generalizado de disimilitudes es el mejor, el tiempo de cómputo de este clasificador es mayor que el de los demás, como se puede notar en el promedio de los tiempos mostrados consignado en la tabla 6-2, el tiempo de cómputo aumenta de manera constante y continua de acuerdo al número de objetos de representación, como se encuentra resaltado en esta tabla, y aunque el promedio del tiempo para los demás clasificadores también aumenta, no es tan notable como en el caso del clasificador *parzenc\_dis*, e incluso, para el clasificador *qdc*, el tiempo de cómputo disminuye. Aunque también es importante hacer mención que la diferencia al final entre los promedios de tiempos de cómputo de los clasificadores es de unos pocos segundos, lo que implica que, aunque el costo computacional aumenta, el error de clasificación mejora

respecto a los demás.

Por otra parte, variando la metodología de selección de las líneas, para este conjunto de datos, los resultados para la selección de líneas de longitud media no se incluyen debido a que la diferencia respecto a la selección de las líneas de características mas largas no es significativa comparando los intervalos de confianza para el error promedio. Además, teniendo en cuenta que las líneas de características mas largas abarcan mayor cantidad de objetos, se obtiene una mejor caracterización del conjunto de datos. La selección de las líneas mas cortas provoca que el desempeño del clasificador decaiga de manera considerable como se puede notar en la gráfica 6-2 y en las tablas 6-3 y 6-4, comparando los desempeños y los intervalos de confianza. Esto evidencia que la selección de las líneas mas cortas, no permite realizar una caracterización adecuada del conjunto de datos, provocando que el desempeño del clasificador no sea el mejor. Además, los resultados para este método de selección, de acuerdo a la longitud de los intervalos de confianza, son variables y no permite realizar una buena comparación de los resultados obtenidos.

Para el conjunto de Clases difíciles distribuidas normalmente de dimensión 20 (ver gráfica 6-3), el desempeño del clasificador de Parzen para el espacio generalizado de disimilitudes, es mejor que el de los demás clasificadores, en la mayoría de los casos tomados (número de objetos de representación), según se puede verificar en la tabla 6-5. Aunque en este caso, en comparación con el desempeño para el conjunto de clases difíciles distribuidas normalmente de dimensión 2, es menos competitivo, ya que el clasificador *svc* logra igualarlo e incluso superarlo de acuerdo a los intervalos de confianza de la tabla 6-2, lo que evidencia que el cambio de dimensión del conjunto afecta de manera considerable el desempeño del clasificador *parzenc\_dis*. Al igual que en la prueba anterior, a medida que el número de objetos de representación aumenta, el desempeño de los clasificadores *ldc*, *qdc*, *svc*, *parzenc* mejora de manera constante, lo que pasa también para el clasificador *parzenc\_dis*. El desempeño de los clasificadores *parzenc\_dis* y *svc* es básicamente el mismo, de acuerdo al número de objetos de representación tomados, mientras que respecto a los restantes clasificadores, el desempeño es notablemente mejor.

Ahora, el desempeño del clasificador de Parzen para el espacio generalizado de disimilitudes es el mejor, pero el tiempo de cómputo de este clasificador, es mayor que el de los demás, reflejado en el promedio de los tiempos mostrados en la tabla 6-6.

El tiempo de cómputo aumenta de manera constante y continua de acuerdo al número de objetos de representación, como se encuentra resaltado en esta tabla, y aunque el promedio del tiempo para los demás clasificadores también aumenta, no es tan notable como lo es para el clasificador *parzenc\_dis*.

Los resultados para la selección de líneas de longitud media no se incluyen para este conjunto de datos dado que, la diferencia en cuanto al desempeño respecto a las líneas de características mas largas no es significativa, esta conclusión se obtuvo comparando los intervalos

de confianza para el error promedio, y teniendo en cuenta que las líneas de características mas largas logran una mejor caracterización del conjunto de datos. Ahora, la selección de las líneas mas cortas provoca que el desempeño del clasificador decaiga de manera considerable como se puede notar en la grafica 6-4 y en las tablas 6-7 y 6-8, comparando los desempeños y los intervalos de confianza, esto evidencia que la selección de las líneas mas cortas, no permite realizar una caracterización adecuada del conjunto de datos y además, esto provoca que el desempeño del clasificador no sea el mejor.

Para el conjunto de Clases de Highleyman (ver gráfica 6-5), el desempeño del clasificador de Parzen para el espacio generalizado de disimilitudes es, respecto a los demás clasificadores, es mejor según se nota en la tabla 6-9, en la cual se muestran los intervalos de confianza para el error promedio de cada clasificador. En este caso el desempeño del clasificador *parzenc\_dis* no es muy variable y se mantiene, según cada intervalo de confianza de la tabla, de manera constante entre el 13 % y el 14 %, y aunque su desempeño es mejor que el de los demás clasificadores, es bastante similar cuando el número de objetos de representación aumenta. Como ha sucedido anteriormente, el desempeño del clasificador *parzenc\_dis* es el mejor, pero el tiempo de cómputo de este clasificador es mayor que el de los demás, como se observa en el promedio de los tiempos de la tabla 6-2, este promedio aumenta de manera constante y continua de acuerdo al número de objetos de representación, como se encuentra resaltado en esta tabla, y aunque el promedio del tiempo para los demás clasificadores también aumenta, no es tan notable como en el caso del clasificador *parzenc\_dis*. Pero es importante observar que la diferencia al final entre los promedios de tiempos de cómputo de los clasificadores, es de unos pocos segundos, lo que implica que, aunque el costo computacional aumenta un poco, el error de clasificación mejora respecto a los demás, aunque en este caso dado que, los desempeños son tan parejos, queda la duda si sacrificar el desempeño por el tiempo.

Para este conjunto de datos, los resultados para la selección de líneas de longitud media no se incluyen por no ser, respecto al desempeño del clasificador, significativamente diferente comparando los intervalos de confianza para el error promedio con el método de selección de las líneas de longitud mayor. Por otra parte, la selección de las líneas mas cortas provoca que el desempeño del clasificador decaiga de manera considerable, como se puede notar en la grafica 6-2 y en las tablas 6-11 y 6-12, comparando los desempeños y los intervalos de confianza. Esto evidencia que la selección de las líneas mas cortas, no permite realizar una caracterización adecuada del conjunto de datos y esto provoca que el desempeño del clasificador no sea el mejor. Además, los resultados para este método de selección, de acuerdo a la longitud de los intervalos de confianza, son variables y no permite realizar una buena comparación de los resultados obtenidos.

Para el conjunto banana (ver gráfica 6-7), el desempeño del clasificador de Parzen para el espacio generalizado de disimilitudes es el peor, no es competitivo dado que la diferencia, comparando los intervalos de confianza para el error promedio de la tabla 6-13 y los errores promedio de la tabla 6-14, son claramente mayores y en cuanto al tiempo de cómputo, como en los casos anteriores, no es el mejor. Estos resultados son esperados, dado que de acuerdo



al grafico de dispersión del conjunto, mostrado en la gráfica 5-14, si se toman dos elementos, por ejemplo de la clase 1, en lados opuestos, según la dispersión de los objetos de esta clase, la construcción de una línea de característica, implicaría que esta pase sobre la zona en donde los elementos de la clase 2 se acercan a los de la clase 1, lo que implica, según el método de clasificación de Parzen, que los elementos de la clase 2, en esta zona, sean clasificados de manera errónea, dado que estarán muy cerca de esta línea de características.

Para el conjunto de círculos concéntricos (ver gráfica 6-8), el desempeño del clasificador de Parzen para el espacio generalizado de disimilitudes es, al igual que para el conjunto banana, el peor, no es competitivo, comparando los intervalos de confianza para el error promedio de la tabla 6-15 y los errores promedio de la tabla 6-16, son claramente mayores y en cuanto al tiempo de cómputo, como en los casos anteriores no es el mejor. Estos resultados son esperados, tomando como referencia el gráfico de dispersión en la figura 6-8, si se toman dos elementos, por ejemplo de la clase 2, en lados opuestos según la dispersión de los objetos de esta clase, la construcción de una línea de característica, implicaría que esta pase sobre la zona en donde están los elementos de la clase 1, ocasionando, según el método de clasificación de Parzen, que los elementos de la clase 1 en esta zona sean clasificados de manera errónea, por la cercanía a esta línea de característica.

Para el conjunto de clases lituanas (ver gráfica 6-9), el desempeño del clasificador de Parzen para el espacio generalizado de disimilitudes es el peor, no es competitivo dado que la diferencia, comparando los intervalos de confianza para el error promedio de la tabla 6-17 y los errores promedio de la tabla 6-18, son claramente mayores y en cuanto al tiempo de cómputo, como en los casos anteriores no es el mejor. Estos resultados son esperados, dado que como se evidencia en la grafica 6-9, las dos clases del conjunto están totalmente traslapadas, provocando que, al momento de construir una línea de característica por ejemplo para la clase 1, elementos de la clase 2 queden mal clasificados.

Para el conjunto, problema de clasificación simple de dos clases gaussianas (ver gráfica 6-10), el desempeño del clasificador de Parzen para el espacio generalizado de disimilitudes, al igual que el de los demás clasificadores no es bueno, aunque en este caso es sensiblemente mejor como se puede verificar en los intervalos de confianza para el error promedio de la tabla 6-19 y los errores promedio de la tabla 6-20. En términos generales, comparando los resultados de estas tablas se puede notar que los desempeños de los clasificadores son muy parecidos pero no son buenos. Estos resultados se deben indudablemente a que los objetos de las clases 1 y 2 se encuentra casi que en su totalidad traslapados, lo que inevitablemente complica la tarea de clasificación y provoca que los desempeños de los clasificadores no sean buenos. Como ha sucedido en ocasiones anteriores el tiempo de cómputo del clasificador *parzenc\_dis* es el peor de los 5 clasificadores y de acuerdo al desempeño obtenido no es, en términos ge-

nerales el mejor de los clasificadores por costo computacional y por desempeño.

Para el conjunto, problema de truncamiento (ver gráfica 6-11), el desempeño del clasificador de Parzen para el espacio generalizado de disimilitudes es bueno y en ciertos momentos mejor, como se puede verificar en los intervalos de confianza para el error promedio de la tabla 6-21. En general el desempeño de los 5 clasificadores es bastante parecido a excepción del clasificador *qdc*, lo que evidencia que el clasificador *parzenc\_dis* es competitivo, en cuanto al desempeño, pero computacionalmente, el costo del clasificador es el peor de los 5 clasificadores, y es inevitablemente una desventaja respecto a los demás, aunque el tiempo no es excesivo, solo de unos pocos segundos.

Ahora, para el conjunto laryngeal 1 (ver gráfica 6-12), aunque el desempeño del clasificador de Parzen para el espacio generalizado de disimilitudes no es el peor, en general, el desempeño de los 5 clasificadores no es bueno, como se puede notar en los intervalos de confianza para el error promedio de la tabla 6-23, pero, para el clasificador *svc* a medida de que el número de objetos de representación aumenta, su desempeño mejora notablemente respecto a los demás, contrario al desempeño del clasificador *parzenc\_dis*, el cual empeora. Esto permite afirmar que el clasificador *parzenc\_dis* para este conjunto de datos, no es el mejor y más aun, es el peor, dado que los tiempos de cómputo de este, una vez más, son los peores, como se puede notar en la tabla 6-24.

Por otra parte, los resultados para la selección de líneas de longitud media no se incluyen debido a que, la diferencia respecto a la selección de las líneas de características mas largas no es significativa, comparando los intervalos de confianza para el error promedio. Para la selección de las líneas mas cortas, el desempeño del clasificador decae de manera considerable como se puede notar en la gráfica 6-2 y en las tablas 6-25 y 6-26, comparando los desempeños y los intervalos de confianza. Esto evidencia que la selección de las líneas mas cortas, no permite realizar una caracterización adecuada del conjunto de datos provocando que el desempeño del clasificador no sea el mejor. Además, los resultados para este método de selección, de acuerdo a la longitud de los intervalos de confianza, son variables y no permite realizar una buena comparación de los resultados obtenidos.

De igual manera como sucedió con el conjunto de datos laryngeal 1, sucede con el conjunto laryngeal 2 (ver gráfica 6-14), aunque en este caso el desempeño de los clasificadores *svc*, *ldc* y *qdc*, a medida de que el número de objetos de representación aumenta, mejora notablemente, como se puede notar en los intervalos de confianza para el error promedio de la tabla 6-27, pero no es el mejor desempeño en clasificación. En cuanto al tiempo de cómputo, el costo del clasificador sigue siendo el peor, y de acuerdo al desempeño en clasificación el clasificador propuesto, no es adecuado para este conjunto de datos.

Para este conjunto de datos, los resultados para la selección de líneas de longitud media y

longitud corta, se incluyen dado que la diferencia de los resultados obtenidos, comparando los intervalos de confianza para el error promedio, es significativa como se puede observar en las graficas 6-15, 6-16 y en las tablas 6-29 a 6-32. La variabilidad de los resultados obtenidos es muy alta de acuerdo a la longitud de los intervalos de confianza, lo que permite concluir que no se realiza una adecuada caracterización del conjunto.

Ahora, al igual que para el conjunto, de problema de clasificación simple de dos clases gaussianas, para el conjunto laryngeal 3 (ver gráfica 6-17), el desempeño del clasificador de Parzen para el espacio generalizado de disimilitudes, como el de los demás clasificadores, no es bueno, aunque en este caso es sensiblemente mejor, como se puede verificar en los intervalos de confianza para el error promedio de la tabla 6-33 y los errores promedio de la tabla 6-34. Una vez más el tiempo de cómputo del clasificador propuesto es el peor y definitivamente no logra compensar con un buen desempeño en cuanto a el error de clasificación, porque, aunque es mejor que el de los demás, el error cometido es bastante alto al igual que el de los demás clasificadores, permitiendo afirmar que, para este conjunto de datos el clasificador de Parzen para el espacio generalizado de disimilitudes no es el mejor.

Para este conjunto de datos, los resultados para la selección de líneas de longitud media no se incluyen para este conjunto de datos dado que, la diferencia en cuanto al desempeño respecto a las líneas de características mas largas no es significativa. Esta conclusión se obtuvo comparando los intervalos de confianza para el error promedio, y teniendo en cuenta que las líneas de características más largas abarcan mayor cantidad de objetos, se obtiene una mejor caracterización del conjunto de datos. Ahora, la selección de las líneas mas cortas provoca que el desempeño del clasificador decaiga de manera considerable como se puede notar en la grafica 6-18 y en las tablas 6-35 y 6-36, comparando los desempeños y los intervalos de confianza, esto evidencia que la selección de las líneas mas cortas, no permite realizar una caracterización adecuada del conjunto de datos provocando, que el desempeño del clasificador no sea el mejor.

Para el conjunto iris, el desempeño de los 5 clasificadores es muy parejo (ver gráfica 6-20), excepto para el clasificador *qdc*, como se puede verificar en los intervalos de confianza para el error promedio de la tabla 6-38 y los errores promedio de la tabla 6-39, podemos afirmar entonces de esta manera que, en cuanto a el error de clasificación, el clasificador de Parzen para el espacio generalizado de disimilitudes es competitivo, aunque como ha pasado con todos los conjuntos de datos anteriores, el tiempo de cómputo de este clasificador es el peor, lo que definitivamente causa que no sea el más favorable, dado que se tiene un buen desempeño, pero un costo computacional alto respecto a los demás clasificadores.

Variando la metodología de selección de las líneas, para este conjunto de datos, los resultados para la selección de líneas de longitud media no se incluyen debido a que la diferencia respecto a la selección de las líneas de características mas largas no es significativa comparando

los intervalos de confianza para el error promedio.

Por otra parte, la selección de las líneas mas cortas provoca que el desempeño del clasificador decaiga de manera considerable como se puede notar en la gráfica 6-21 y en las tablas 6-40 y 6-41, comparando los desempeños y los intervalos de confianza. Esto evidencia que la selección de las líneas mas cortas, no permite realizar una caracterización adecuada del conjunto de datos, provocando que el desempeño del clasificador no sea el mejor. Además, los resultados para este método de selección, de acuerdo a la longitud de los intervalos de confianza, son variables y no permite realizar una buena comparación de los resultados obtenidos.

Para el conjunto wine (ver gráfica 6-23), el desempeño del clasificador de Parzen para el espacio generalizado de disimilitudes, aunque al igual que el de los demás clasificadores no es bueno, es el peor, no es competitivo dado que la diferencia, comparando los intervalos de confianza para el error promedio de la tabla 6-43 y los errores promedio de la tabla 6-44, son claramente mayores y en cuanto al tiempo de cómputo, como en los casos anteriores no es el mejor.

Para el conjunto de datos Malaysia (ver gráfica 6-24), la prueba vario un poco, debido al número de objetos por clase, debido a que algunas de las clases solo tenían solo 11 objetos, luego se modifico el número de objetos de representación de manera tal que la prueba realizada funcionara. Los resultados, en este caso, son absolutamente malos, el desempeño de cada clasificador es excesivamente bajo, como se puede verificar en los intervalos de confianza para el error promedio de la tabla 6-45 y los errores promedio de la tabla 6-46, además, los tiempos de cómputo son muy elevados (ver tabla 6-46), lo que permite afirmar que para este conjunto de datos ninguno de los clasificadores seleccionados resulta ser adecuado.

Para el conjunto diabetes (ver gráfica 6-25), el desempeño del clasificador de Parzen para el espacio generalizado de disimilitudes, al igual que el de los demás clasificadores, no es bueno y aunque no es el mejor tampoco es el peor, como se puede verificar en los intervalos de confianza para el error promedio de la tabla 6-47 y los errores promedio de la tabla 6-48. Aunque en términos generales, comparando los resultados de estas tablas se puede notar que, los desempeños de los clasificadores son muy parecidos, pero no son buenos. Como ha sucedido en ocasiones anteriores el tiempo de cómputo del clasificador *parzenc\_dis* es el peor de los 5 clasificadores y de acuerdo al desempeño obtenido no es, en términos generales, el mejor de los clasificadores, por el costo computacional y por el desempeño.

Para el conjunto biomed (ver gráfica 6-26), el desempeño del clasificador de Parzen para el espacio generalizado de disimilitudes, no es el mejor, pero tampoco es el peor, como se puede verificar en los intervalos de confianza para el error promedio de la tabla 6-47 y los errores promedio de la tabla 6-48, se podría llegar a considera competitivo y además, en términos generales, comparando los resultados de estas tablas, se puede notar que los desempeños

de los clasificadores son muy parecidos, aunque como ha sucedido en ocasiones anteriores el tiempo de cómputo del clasificador *parzenc\_dis* es el peor de los 5 clasificadores y de esta manera, de acuerdo al desempeño obtenido, no es el mejor de los clasificadores por costo computacional.

Para el conjunto heart (ver gráfica 6-27), el desempeño del clasificador de Parzen para el espacio generalizado de disimilitudes, al igual que el de los demás clasificadores, no es bueno, aunque en este caso es sensiblemente mejor, como se puede verificar en los intervalos de confianza para el error promedio de la tabla 6-51 y los errores promedio de la tabla 6-52. En términos generales, comparando los resultados de estas tablas se puede notar que, los desempeños de los clasificadores son muy parecidos, pero no son buenos. Estos resultados se deben indudablemente a que los objetos de las clases 1 y 2 se encuentra casi que en su totalidad traslapados, lo que inevitablemente complica la tarea de clasificación y provoca que los desempeños de los clasificadores no sean buenos.

Como ha sucedido en ocasiones anteriores el tiempo de cómputo del clasificador *parzenc\_dis* es el peor de los 5 clasificadores y de acuerdo al desempeño obtenido no es, en términos generales el mejor de los clasificadores por costo computacional y por desempeño.

Por otra parte, variando la metodología de selección de las líneas, para este conjunto de datos, los resultados para la selección de líneas de longitud media no se incluyen debido a que la diferencia respecto a la selección de las líneas de características mas largas no es significativa comparando los intervalos de confianza para el error promedio.

Por otra parte, la selección de las líneas mas cortas provoca que el desempeño del clasificador decaiga de manera considerable como se puede notar en la gráfica 6-28 y en las tablas 6-53 y 6-54, comparando los desempeños y los intervalos de confianza. Esto evidencia que la selección de las líneas mas cortas, no permite realizar una caracterización adecuada del conjunto de datos, provocando que el desempeño del clasificador no sea el mejor. Además, los resultados para este método de selección, de acuerdo a la longitud de los intervalos de confianza, son variables y no permite realizar una buena comparación de los resultados obtenidos.

Para el conjunto ionosphere (ver gráfica 6-30), el desempeño del clasificador de Parzen para el espacio generalizado de disimilitudes es el peor, no es competitivo dado que la diferencia, comparando los intervalos de confianza para el error promedio de la tabla 6-56 y los errores promedio de la tabla 6-57, son claramente mayores y en cuanto al tiempo de cómputo, como en los casos anteriores no es el mejor.

Para el conjunto IMOX el desempeño de los 5 clasificadores es muy parejo (ver gráfica 6-20), como se puede verificar en los intervalos de confianza para el error promedio de la tabla 6-38 y los errores promedio de la tabla 6-39, podemos afirmar de esta manera que en cuanto a el

error de clasificación, el clasificador de Parzen para el espacio generalizado de disimilitudes es competitivo, aunque como ha pasado con todos los conjuntos de datos anteriores el tiempo de cómputo de este clasificador es el peor, incluso aumentado unos segundos más respecto a los anteriores conjuntos, lo que definitivamente causa que no sea el más favorable, dado que se tiene un buen desempeño, pero un costo computacional alto respecto a los demás clasificadores y los desempeños de estos son muy parecidos al del clasificador propuesto. Además, respecto a los intervalos de confianza, se puede notar que el error es muy variable, dado que el rango de la mayoría de intervalos de confianza es grande.

Para el conjunto glass (ver gráfica 6-32), el desempeño del clasificador de Parzen para el espacio generalizado de disimilitudes, aunque al igual que el de los demás clasificadores no es bueno, es el peor, no es competitivo dado que la diferencia, comparando los intervalos de confianza para el error promedio de la tabla 6-64 y los errores promedio de la tabla 6-65, son claramente mayores y en cuanto al tiempo de cómputo, como en los casos anteriores no es el mejor.

Para el conjunto ORL (ver gráfica 6-33 a 6-35), el desempeño de los clasificadores es muy parejo, exceptuando el clasificador *qdc*, como se puede verificar en los intervalos de confianza para el error promedio de la tabla 6-51 y los errores promedio de la tabla 6-52, a partir de estos resultados podemos afirmar que el clasificador de Parzen para el espacio generalizado de disimilitudes es competitivo, aunque como ha sucedido en ocasiones anteriores el tiempo de cómputo del clasificador *parzenc\_dis* es el peor de los 5 clasificadores, aunque es muy parecido al tiempo del clasificador *svc*.

Variando la metodología de selección de las líneas, para este conjunto de datos, los resultados para la selección de líneas de longitud media no se incluyen debido a que la diferencia respecto a la selección de las líneas de características mas largas no es significativa comparando los intervalos de confianza para el error promedio.

Por otra parte, la selección de las líneas mas cortas provoca que el desempeño del clasificador decaiga de manera considerable como se puede notar en la gráfica 6-36 y en las tablas 6-53 y 6-54, comparando los desempeños y los intervalos de confianza. Esto evidencia que la selección de las líneas mas cortas, no permite realizar una caracterización adecuada del conjunto de datos, provocando que el desempeño del clasificador no sea el mejor. Además, los resultados para este método de selección, de acuerdo a la longitud de los intervalos de confianza, son variables y no permite realizar una buena comparación de los resultados obtenidos.

De otro lado, en las cuatro pruebas realizadas para el método de selección basado en algoritmo genético (GA)[10], el tiempo de computo fue excesivo, por eso solo se presentaron estos pocos resultados, que además fueron muy pobres desde el punto de vista estadístico,

dado que tan solo se repito una vez cada prueba, por el tiempo de computo, el error es muy variable como se puede evidenciar en las graficas 6-19, 6-22, 6-29 y 6-37 y en las tablas 6-37, 6-42, 6-55 y 6-66 y el tiempo de computo es alto para el clasificador *parzenc\_dis*. En general, el clasificador de Parzen para el espacio generalizado de disimilitudes no tiene el mejor desempeño, sin importar el método de selección utilizado (longitud de las líneas o algoritmo genético). En la mayoría de los conjuntos de alta dimensión no logra ser competitivo y en aquellos donde las clases están muy traslapadas empeora su desempeño. La estructura de los conjuntos de datos es vital y juega un papel relevante en el desempeño del clasificador, por ejemplo, para el caso del conjunto círculos concéntricos, el conjunto banana, lituanas, de clases distribuidas de manera gaussiana, el desempeño fue el peor y a medida que aumentaba el número de objetos de representación, empeoraba. Ahora, para conjuntos de estructura lineal (caso bidimensional), como es el caso del conjunto de clases difíciles distribuidas normalmente, el conjunto problema de truncamiento, e incluso iris (caso tridimensional), el desempeño es bueno, es competitivo respecto a los demás clasificadores y en algunos casos es el mejor. En los restantes conjuntos de datos, como se menciono anteriormente, al existir un traslape entre clases, el clasificador empeora notablemente su desempeño y en el caso del conjunto Malaysia, el desempeño fue pésimo aunque paso lo mismo con los demás clasificadores.

Ahora bien, para el conjunto de imágenes ORL, el desempeño del clasificador *parzenc\_dis* fue bastante competitivo, haciendo uso de la selección de las líneas de características de mayor longitud, permitiendo afirmar que para conjuntos de alta dimensión el método es adecuado. Además, para este conjunto en particular, dado que cada clase tiene pocos objetos (10), el método de estimación de Parzen en el espacio generalizado de disimilitudes logra enriquecer y lograr una adecuada caracterización del conjunto, de tal manera que se logra una buena clasificación.

También es importante notar que el desempeño del clasificador no fue el mejor, de acuerdo a la longitud de las líneas de características, dado que la dimensión de la mayoría de conjuntos de datos es "pequeña", ocasionando que la probabilidad de que existiera cruce entre líneas de características fuera alta, afectando directamente el desempeño, mientras que para conjuntos de datos como ORL, al tener una dimensión alta, la probabilidad de cruce entre las líneas es casi nula, favoreciendo el desempeño del clasificador.

Por otra parte, como se puede verificar en las tablas 6-67 y 6-68, los desempeños tanto como en el espacio generalizado de disimilitudes, como en el espacio de características son parecidos, aunque, para algunos conjuntos de datos el desempeño de los clasificadores en el espacio de características mejora, como por ejemplo el conjunto Malaysia, conjunto de clases difíciles distribuidas normalmente de dimensión 20, conjunto banana, conjunto de círculos concéntricos. En general, los desempeños en el conjunto de características mejoran

aunque, en aquellos conjuntos en donde el clasificador *parzen\_dis* tuvo un buen desempeño, los clasificadores en el espacio de características tuvieron desempeños muy parecidos. Esto permite afirmar que el trabajo de clasificación sobre el espacio generalizado de disimilitudes es válido. Ahora respecto al costo computacional, los desempeños para todos los conjuntos de datos fueron muy parecidos, por consiguiente en este sentido el costo computacional es básicamente el mismo.



## 7. Conclusiones

En general, como se menciona en el capítulo anterior, el clasificador de Parzen para el espacio generalizado de disimilitudes no tiene el mejor desempeño, en la mayoría de los conjuntos de alta dimensión no logra ser competitivo y en aquellos donde las clases están muy traslapadas empeora su desempeño. En el caso de los conjuntos de datos artificiales, de conjuntos bidimensionales o de dimensión pequeña, la estructura de los conjuntos de datos es vital y juega un papel relevante en el desempeño del clasificador, por ejemplo, para el caso del conjunto círculos concéntricos, el conjunto banana, lituanas, de clases distribuidas de manera gaussiana, el desempeño fue el peor y a medida que aumentaba el número de objetos de representación empeoraba. Ahora, para conjuntos de estructura lineal, como es el caso del conjunto de clases difíciles distribuidas normalmente, el conjunto problema de truncamiento, e incluso iris, el desempeño es bueno, es competitivo respecto a los demás clasificadores y en algunos casos es el mejor. Para los conjuntos de datos de la vida real, al no poder establecer una estructura en particular para cada uno de ellos, el desempeño del clasificador es variable, aunque, para el conjunto de datos ORL, al ser un conjunto de alta dimensión, el desempeño del clasificador fue competitivo, tomando como base la selección de las líneas de características de mayor longitud. Resulta importante tener en cuenta el método de estimación de densidades de probabilidad de Parzen, inicialmente planteado sobre el espacio de características, para el espacio generalizado de disimilitudes (generalizado mediante líneas de características). Es importante por las posibilidades que podría brindar este método de estimación en clasificación automática de patrones. El planteamiento inicial, es lograr caracterizar este método en el espacio generalizado de disimilitudes, particularmente, que tipo de estructura de datos resulta conveniente en clasificación de manera que, el método propuesto, con las modificaciones hechas sobre el algoritmo de clasificación, sea optimo en cuanto al desempeño de clasificación, medido en el porcentaje de error, y el costo computacional que trae consigo este método alternativo de estimación y clasificación en el espacio generalizado de disimilitudes.

Este trabajo muestra un nuevo camino, un clasificador basado en la estimación de densidades de probabilidad de Parzen. Para llegar a este cometido la tarea inicio con las modificaciones directamente en el algoritmo de clasificación. El método de estimación de Parzen está planteado en principio para el espacio de características al igual que el algoritmo de clasificación. Teóricamente el método de estimación no tiene inconvenientes sobre el espacio generalizado de disimilitudes, dado que, la naturaleza de los objetos en este espacio son fundamentalmente

iguales a la del espacio de características, el cambio principal es este espacio alternativo de trabajo es la interpretación de las componentes de cada uno de los vectores que representan los objetos. Es así que hechas las modificaciones sobre el algoritmo de clasificación, se realizan las pruebas planteadas y las comparaciones de desempeño de los clasificadores seleccionados en esta tesis contra el clasificador implementado producto de las modificaciones realizadas. La experiencia de estas pruebas permite hacer las siguientes anotaciones, respecto al objetivo principal de caracterizar el método de estimación de Parzen para el espacio generalizado de disimilitudes:

- En todos los experimentos, la selección de objetos de acuerdo al algoritmo genético (*GA*), provoca que el desempeño del clasificador decaiga de manera notable, esto permite afirmar que no es un método adecuado de selección. Además el tiempo de cómputo para cada uno de los experimentos es muy alto, tomando como punto de partida que se realizaron de manera sencilla, con el fin de estimar un tiempo aproximado de las pruebas.
- La selección de líneas de características de longitud media o corta, mostraron no ser métodos adecuados de selección, a comparación de las líneas de longitud mayor, debido a la variabilidad de los resultados, al desempeño bajo del clasificador y además por la longitud de los intervalos de confianza para el error promedio. En las siguientes conclusiones, nos referimos en particular a la selección de líneas de acuerdo a la mayor longitud de las mismas.
- Para conjunto de datos artificiales o de dimensión pequeña, el clasificador *parzen\_dis* correspondiente a las modificaciones hechas en el algoritmo de clasificación de Parzen y de acuerdo a la longitud de las líneas de características, funciona bien si el conjunto de datos tiene una estructura lineal tales como el conjunto de clases difíciles, el conjunto problema de truncamiento, el conjunto iris. El desempeño es bueno, es competitivo respecto a los demás clasificadores, teniendo en cuenta que es competitivo si el error de clasificación es bajo y además el tiempo de cómputo es bueno. Utilizando *GA* (algoritmo genético), el desempeño del clasificador no es bueno, debido a la aleatoriedad de este método.
- Para el conjunto de clases de highleyman el desempeño del clasificador es bueno e incluso mejor a los demás clasificadores y este conjunto según la grafica de dispersión 5-5, tiene una estructura cuadrática, tomando como referencia las fronteras de decisión.
- Si el conjunto de datos no es lineal, el desempeño del clasificador decae notablemente, por ejemplo para el conjunto banana, o de círculos concéntricos, fue el desempeño fue el peor de los 5 clasificadores y además el tiempo de cómputo es también el peor. La Razón principal es el cruce de las líneas en regiones de objetos pertenecientes a una clase diferente a la de las líneas de características.

- 
- Ahora si existe un traslape notorio entre los objetos de cada clases de los conjuntos, el desempeño también decae de manera notable y el tiempo de computo aumenta también, por ejemplo para el conjunto de clases lituanas, de clases distribuidas de manera gaussiana, el desempeño fue de los peores y a medida que aumentaba el número de objetos de representación empeoraba. La razón principal es por el cruce entre líneas de características perteneciente a clases diferentes.
  - En todas las pruebas el desempeño en cuanto costo computacional del clasificador *parzen\_dis* fue siempre el peor, aunque en la mayoría de los casos la diferencia es solo de unos pocos segundos, lo que lleva a la pregunta si sacrificar tiempo de cómputo por mejor desempeño o simplemente tomar totalmente en cuenta el tiempo de cómputo.
  - Para la base de datos Malaysia el desempeño de los 5 clasificadores fue claramente pésimo.
  - En términos generales la estructura de los datos debería ser lineal para lograr un buen desempeño del clasificador de Parzen para el espacio generalizado de disimilitudes. Realizando una comparación con el desempeño de los clasificadores *ldc*, *qdc*, *svc*, *parzenc* junto con el clasificador *knnnc*, se concluye que la diferencia con el desempeño en el espacio generalizado de disimilitudes no es significativa tomando en cuenta los errores promedio y los intervalos de confianza para estos errores, permitiendo afirmar que el trabajo en este espacio alternativo de trabajo junto con la estimación de probabilidad de Parzen es adecuado, y eficiente en algunos casos de acuerdo al desempeño.
  - Para el conjunto de datos ORL, el desempeño del clasificador *parzenc\_dis*, es competitivo. Es importante resaltar esta base de datos dado que la dimensión del conjunto es alta, lo cual permite asumir que para conjuntos cuya dimensión es alta el desempeño del clasificador mejora, debido a que la posibilidad de que las líneas de características se crucen es muy baja, favoreciendo el desempeño del clasificador.

Teóricamente, se ha logrado dar una primera caracterización del espacio generalizado de disimilitudes al lograr establecer una isometría entre este espacio y el conjunto  $\mathbb{R}^m$  mediante la métrica  $\ell_\infty$ , permitiendo establecer una relación directa entre los elementos del conjunto  $\mathbb{R}^m$  y el del espacio de generalizado de disimilitudes, en particular poder llevar las propiedades de  $\mathbb{R}^m$  a este de acuerdo a la isometría hallada, este es un avance desde el punto de vista teórico dado que se da un primer paso en la formalización del trabajo realizado y en poder lograr caracterizar totalmente a este espacio alternativo de trabajo.

## A. Anexo: PRTools

Ahora, como se mencionó anteriormente el uso de herramientas computacionales es fundamental en reconocimiento automático de patrones, una de las herramienta de mayor desarrollo y uso es el software **Matlab** y en nuestro caso del toolbox **PRTools**, desarrollado para brindar eficacia y comodidad en reconocimiento de patrones. Éste contiene un número importante de funciones, asignaciones y clasificadores, que facilitan el desarrollo e implementación de métodos que permiten dar solución a problemas de clasificación.

En reconocimiento de patrones estadístico, se estudian técnicas para la generalización de las reglas de decisión que serán utilizadas para la detección y reconocimiento de patrones en datos experimentales. En este toolbox se definen:

- Dataset (conjunto de datos): Consiste de un conjunto de objetos representados por una matriz de vectores de medida. Se adjunta a esta matriz de etiquetas de clase, una para cada objeto y un conjunto de nombres de las características. Por otra parte se almacena un conjunto de probabilidades a priori, uno para cada clase. En Matlab son de tipo struct.
- Datafiles (Archivos de datos): estos se refieren al conjunto de datos que se crean a partir de directorios de archivos, básicamente se construyen para resolver el problema de la memoria relacionada con los datasets. Con estos se tiene la posibilidad de ser utilizados cuando se desee, se pueden manipular y modelar como un dataset.
- Clasificadores y asignaciones: Operan transformaciones sobre un conjunto de datos, se consideran de tipos formadas o entrenadas y sin formación o no entrenada, esto si se operan sobre un conjunto de evaluación o entrenamiento respectivamente. Son básicamente funciones o aplicaciones, las cuales son las herramientas base para la tarea de clasificación.

Cada uno de las anteriores herramientas de este toolbox se pueden manipular según el caso, existen, como se mencionó anteriormente, un número importante de funciones propias del toolbox que claramente utilizan el lenguaje de Matlab aparte de funciones naturales del mismo.

## B. Anexo: Valores del parametro de suavizado para cada uno de los conjuntos de datos

**Tabla B-1.:** Valor promedio del parametro de suavizado

Valor promedio del parametro de suavizado

Objetos de representaci3n	Clases Dificiles	Clases Dificiles	Clases de Highleyman	Conjunto banana	C3rculos conc3ntricos
4	18.291	37.912	0.8565	3.3946	20.550
5	18.194	32.486	0.8558	2.8097	21.010
6	17.851	28.435	0.9418	2.8183	19.566
7	17.070	26.698	0.973	2.6312	19.016
8	16.864	24.630	0.9877	2.5111	18.516
9	16.851	23.327	10.256	2.4924	18.148
10	16.524	22.472	0.9912	2.4192	17.302
11	16.567	21.415	10.140	2.3539	17.594
12	16.677	20.579	10.065	2.3123	16.834
13	16.102	20.172	10.281	2.3807	16.891
14	16.395	19.366	10.300	2.2533	16.704
15	16.232	18.788	10.344	2.3810	16.569
16	15.797	18.529	10.004	2.1551	16.263
17	15.901	17.830	10.135	2.1901	15.871
18	16.153	17.773	10.193	2.2927	16.008
19	16.074	17.246	10.108	2.2342	15.685
20	16.124	17.075	10.125	2.1232	15.886

**Tabla B-2.:** Valor promedio del parametro de suavizado

Valor promedio del parametro de suavizado

Objetos de representacion	Clases lituanas	Clases distribuidas de manera gaussiana	Truncamiento de 2 clases gaussianas	Laryngeal 1	Laryngeal 2
4	22.592	20.873	26.948	253.043	201.029
5	20.449	20.715	25.009	208.224	199.251
6	20.004	19.429	23.392	192.092	187.983
7	18.701	18.394	21.738	182.028	159.340
8	19.151	17.361	20.720	178.036	174.576
9	19.165	16.779	19.998	174.825	168.688
10	18.790	16.114	19.393	176.509	163.930
11	18.186	15.365	18.759	163.912	172.922
12	18.358	15.212	18.255	169.173	162.862
13	17.968	14.777	17.829	168.796	163.383
14	18.189	14.281	17.217	174.745	166.041
15	18.304	14.001	16.922	162.621	162.097
16	17.853	13.665	16.567	161.619	159.515
17	18.051	13.463	16.499	165.576	159.584
18	18.074	13.323	16.022	165.443	156.011
19	18.105	12.999	15.900	166.326	160.778
20	18.164	12.840	15.652	158.710	160.870

**Tabla B-3.:** Valor promedio del parametro de suavizado

Valor promedio del parametro de suavizado

Objetos de representacion	Laryngeal 3	Iris	Wine	Malaysia	Diabetes
4	233.986	146.013	255.136	9018.100	429.093
5	210.703	141.227	213.475	9213.232	402.856
6	194.863	140.491	191.928	9103.408	372.369
7	200.544	137.633	179.849	9205.967	335.303
8	186.456	137.201	174.214	9004.949	327.430
9	180.181	136.848	165.401	9156.521	307.455
10	186.104	136.319	159.796	9041.048	307.950
11	183.788	135.993	159.592	—	310.949
12	179.550	133.644	166.493	—	301.328
13	184.700	134.002	153.692	—	291.494
14	178.131	135.947	162.321	—	283.473
15	178.384	135.867	161.328	—	296.495
16	180.686	135.074	151.456	—	295.876
17	180.502	133.713	157.994	—	284.567
18	174.699	135.142	154.745	—	289.539
19	181.697	134.999	155.453	—	287.315
20	179.342	135.143	155.372	—	276.664

**Tabla B-4.:** Valor promedio del parametro de suavizado

Valor promedio del parametro de suavizado

Objetos de representacion	IMOX	Biomed	Heart	Glass	Ionosphere	ORL
4	42.504	1023.002	267.327	16.033	16.678	40.466
5	39.362	845.639	243.333	15.890	18.180	38.645
6	39.533	963.326	232.385	15.817	17.491	37.489
7	41.497	980.619	226.179	15.490	17.137	36.536
8	34.373	970.647	213.064	15.392	16.523	35.903
9	34.948	1032.947	204.362	15.015	16.114	35.457
10	36.694	1101.184	205.382	15.516	15.721	35.072
11	32.926	1047.146	195.864	15.460	15.536	—
12	36.512	1055.064	197.020	15.226	15.022	—
13	36.455	1023.845	191.307	15.251	14.854	—
14	35.635	1076.245	189.211	14.910	14.870	—
15	35.146	1061.703	190.218	15.219	14.636	—
16	33.956	1048.804	189.266	15.043	14.606	—
17	34.260	1119.696	182.902	—	14.173	—
18	35.312	1142.489	183.088	—	14.389	—
19	34.619	1161.305	180.988	—	14.239	—
20	33.800	1098.349	179.307	—	13.913	—



# Bibliografía

- [1] BICEGO, Manuele ; CRISTANI, Marco ; MURINO, Vittorio ; PEKALSKA, Elzbieta ; DUIN, Robert P. W.: Clustering-Based Construction of Hidden Markov Models for Generative Kernels. En: *EMMCVPR*, 2009, p. 466–479
- [2] BICEGO, Manuele ; MURINO, Vittorio ; FIGUEIREDO, Mario A.: Similarity-based classification of sequences using hidden Markov models. En: *Pattern Recognition* Vol. 37 (2004), p. 2281–2291
- [3] BICEGO, Manuele ; PEKALSKA, Elzbieta ; TAX, David M. J. ; DUIN, Robert P. W.: Component-based discriminative classification for hidden Markov models. En: *Pattern Recognition* Vol. 42 (2009), p. c2637–2648
- [4] DUDA, Richard O. ; HART, Peter E. ; STORK, David G. ; WILEY-INTERSCIENCE (Ed.): *Pattern Classification (2nd Edition)*. John-Wiley, 2000
- [5] DUIN, R. P. W. ; JUSZCZAK, P ; DE RIDDER, D ; PACLIK, P ; PEKALSKA, E ; TAX, D. M. J.: *PRTools4: a Matlab Toolbox for Pattern Recognition*. <http://www.prtools.org/> : The Netherlands Delft University of Technology, 2004
- [6] DUIN, Robert P. W.: On the Choice of Smoothing Parameters for Parzen Estimators of Probability Density Functions. En: *IEEE Transactions on Computers* Vol. 25 Issue 11 (November 1976), p. 1175–1179
- [7] DUIN, Robert P. W. ; JAIN, Anil K. ; MAO, Jianchang: Statistical Pattern Recognition: A Review. En: *IEEE Transactions on Pattern Analysis and Machine Intelligence* Vol. 22 (2000), p. 4–37
- [8] FRANK, A. ; ASUNCION, A.: *UCI Machine Learning Repository*. <http://archive.ics.uci.edu/ml>, 2010
- [9] FUKUNAGA, Keinosuke ; KOONTZ, Warren L. G.: Asymptotic Analysis of a Nonparametric Clustering Technique. En: *IEEE Transactions on Computers* Vol. c-21, No. 9 (September 1972), p. 967–974
- [10] GOLDBERG, D. E. ; CO., Addison-Wesley Longman P. (Ed.): *Genetic Algorithms in Search, Optimization and Machine Learning 1st*. Addison-Wesley Longman Publishing Co., 1989

- [11] HART, P.: The Condensed Nearest Neighbor Rule. En: *IEEE Transactions on Information Theory* Vol. 14 (1968), p. 515 – 516
- [12] VAN DER HEIJDEN, F. ; DUIN, R. P. W. ; DE RIDDER, D. ; TAX, D. M. J. ; JOHN-WILEY (Ed.): *Classification, Parameter Estimation and State Estimation an Engineering Approach using MATLAB*. John-Wiley, 2004
- [13] HIGHLEYMAN, W. H. ; OF THE IRE, Proceedings (Ed.): *Linear decision functions, with application to pattern recognition*. Proceedings of the IRE, 1962. – 1501–1514 p.
- [14] KIM, Sang-Woon: On using a dissimilarity representation method to solve the small sample size problem for face recognition. En: *Advanced concepts for intelligent vision systems* Vol. 4179 (September 2006), p. 1174–1185
- [15] KIM, S.W. ; DUIN, R. P. W.: On Improving Dissimilarity-Based Classifications Using a Statistical Similarity Measure. En: *Lecture Notes in Computer Science* Vol. 6419 (2010), p. 418–425
- [16] KUNCHEVA, L.I.: *Real medical data sets*. <http://pages.bangor.ac.uk/~mas00a/>, 2005
- [17] LERNER, Boaz ; GUTERMAN, Hugo ; ALADJEM, Mayer ; DINSTEIN, Its'hak ; ROMEN, Yitzhak: On patterns clasification with Sammon's nonlinear mapping - an experiment study. En: *Pattern Recognition* 31 (1998), p. 371–381
- [18] LI, S. Z. ; LU, J.: Face recognition using the nearest feature line method. En: *IEEE Transactions Neural Networks* vol. 10, no. 2 (1999), p. 439–443
- [19] MUNKRES, James R. ; HALL, Prentice (Ed.): *Topology (second edition)*. Prentice Hall, 2000
- [20] OROZCO-ALZATE, Mauricio: *Generalized dissimilarity representations for pattern recognition*, Universidad Nacional de Colombia sede Manizales, Tesis de Grado, 2008
- [21] OROZCO-ALZATE, Mauricio ; DUIN, Robert P. W. ; CASTELLANOS-DOMÍNGUEZ, Germán: A generalization of dissimilarity representations using feature lines and feature planes. En: *Pattern Recognition Letters* Vol. 30 (2009), p. 242–254
- [22] PANG, Y. ; YUAN, Y. ; LI, X.: Generalised nearest feature line for subspace learning. En: *Electronics Letters* vol. 43 (2007), p. 1079 – 1080
- [23] PARZEN, Emanuel: On estimation of a probability density function and mode. En: *Stanford University* (1961)
- [24] PEKALSKA, Elzbieta ; DUIN, Robert P. W. ; SCIENTIFIC, World (Ed.): *The Dissimilarity Representations in Pattern Recognition Foundations and Applications*. World Scientific, 2005

- 
- [25] PEKALSKA, Elzbieta ; DUIN, Robert P. W.: The use of dissimilarities for object recognition. En: *EOS Conference on Industrial Image and Machine Vision*, 2005
  - [26] PEKALSKA, Elzbieta ; DUIN, Robert P. W.: Prototype selection for dissimilarity-based classifiers. En: *Pattern Recognition* Vol. 39 (2006), p. 189–208
  - [27] PEKALSKA, Elzbieta ; DUIN, Robert P.: Classification on dissimilarity data: a first look. En: *ASCI* (2000), p. 221–228
  - [28] PEÑA, Daniel ; MCGRAW-HILL (Ed.): *Análisis de datos multivariantes*. McGraw-Hill, 2002
  - [29] WANATABE, Satosi ; SONS, John Wiley (Ed.): *Pattern Recognition: Human and Mechanical*. John-Wiley, 1985