[[1]](#footnote-1)

**INFORME DE LABORATORIO PARCIAL PRIMER 35%**

J D Suárez Londoño  
juan.suarezl@usantoto.edu.co

***Abstract—*** ***The following report shows the elaboration and implementation of what was seen during the first cut pattern recognition subject by a series of exercises.***

***Index Terms—Bayesian, Sorter, Electronic, Recognition, Pattern.***

***Resumen— El siguiente informe muestra la elaboración y la implementación de lo visto durante el primer corte de la asignatura reconocimiento de patrones mediante una serie de ejercicios.***

***Palabras Claves—Bayesiano, Clasificador, Electrónica, Reconocimiento, Patrón.***

E

# INTRODUCCION

l reconocimiento de patrones es la ciencia que se ocupa de los procesos sobre ingeniería, computación y matemáticas relacionados con objetos físicos o abstractos, con el propósito de extraer información que permita establecer propiedades de entre conjuntos de dichos objetos.

El reconocimiento de patrones también llamado lectura de patrones, identificación de figuras y reconocimiento de formas consiste en el reconocimiento de patrones de señales. Los patrones se obtienen a partir de los procesos de segmentación, extracción de características y descripción donde cada objeto queda representado por una colección de descriptores. El sistema de reconocimiento debe asignar a cada objeto su categoría o clase (conjunto de entidades que comparten alguna característica que las diferencias del resto). Para poder reconocer los patrones se siguen los siguientes procesos:

* Adquisición de datos
* Extracción de características
* Toma de decisiones

El punto esencial del reconocimiento de patrones es la clasificación: se quiere clasificar una señal dependiendo de sus

características.

# OBJETIVOS

**GENERAL.**

Poner en practica la tematica vista en el primer corte de la materia reconocimiento de patrones

**ESPECIFICOS**

- Visualizar e interpretar datos.

- Diseñar e implementar clasificadores bayesianos.

- Utilizar el metodo de analisis de componentes principales.

# CONCEPTOS PREVIOS

### Clasificador Bayesiano

El enfoque probabilístico bayesiano provee un marco formal para construir clasificadores óptimos bajo ciertos criterios (como el minimizar el error clasificación o el costo de una mala clasificación).

Sin embargo, si aplicamos el enfoque bayesiano en una forma directa, la complejidad computacional (en memoria y tiempo) crece exponencialmente con el número de atributos de los objetos. Una alternativa para enfrentar este problema, es el clasificador bayesiano simple, que asume que todos los atributos de un objeto son estadísticamente independientes dada la clase. Esto hace que la complejidad crezca linealmente con el número de atributos, pero puede hacer que la efectividad del clasificador decrezca si los atributos no son realmente independientes dada la clase.

### PCA

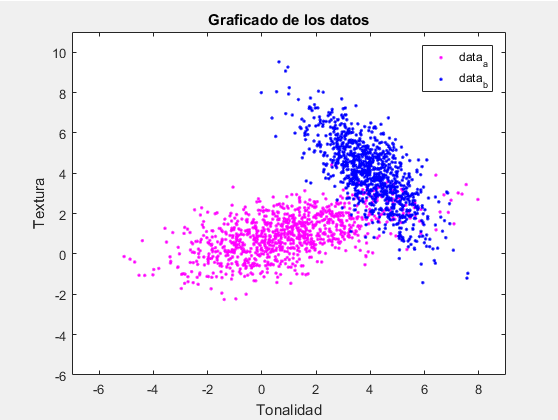
En estadística, el análisis de componentes principales (en español ACP, en inglés, PCA) es una técnica utilizada para reducir la dimensionalidad de un conjunto de datos. Técnicamente, el ACP busca la proyección según la cual los datos queden mejor representados en términos de mínimos cuadrados. Esta convierte un conjunto de observaciones de variables posiblemente correlacionadas en un conjunto de valores de variables sin correlación lineal llamadas componentes principales. El ACP se emplea sobre todo en análisis exploratorio de datos y para construir modelos predictivos. El ACP comporta el cálculo de la descomposición en autovalores de la matriz de covarianza, normalmente tras centrar los datos en la media de cada atributo. Debe diferenciarse del análisis factorial con el que tiene similaridades formales y en el cual puede ser utilizado como un método de aproximación para la extracción de factores.

# DESARROLLO DE LA PRÁCTICA

Nota: En esta parte no se presentará todo el código ya que se haría muy extenso y tedioso, por eso se anexa el código en Matlab con todas sus clases, subclases y ficheros, además de los comentarios/anotaciones para indicar donde están los ejercicios.

### El conjunto de datos data contiene datos pertenecientes a 2 clases {a, b}.

* Grafique los datos utilizando un color distintivo para cada clase.

Fig 1. Grafica datos de data\_a (magenta) y data\_b (azul).

* Determine de centro de cada clase {µa, µn}.

El centro o la media de cada clase se encuentra mediante el comando “mean”, usándolo de la siguiente manera:

meanA=(mean(data\_a))'; %media de la clase data\_a

meanB=(mean(data\_b))'; %media de la clase data\_b

* Determine las matrices de covarianza de cada clase {∑a, ∑b}. ¿Qué se puede concluir?

Las matrices de covarianza de cada clase se encuentran mediante el comando “cov”, usándolo de la siguiente manera:

kX=cov(data\_a); %matriz covarianza de la clase data\_a

kY=cov(data\_b); %matriz covarianza de la clase data\_b

* Determine los auto-vectores {v1, v2} y los auto-valores {λ1, λ2} de cada matriz de covarianza ∑.

Los auto-vectores y los auto-valores de cada clase se encuentran mediante el comando “eig”, usándolos de la siguiente manera:

autoValoresA=eig(kX); %auto-valores de kX

autoValoresB=eig(kY); %auto-valores de kY

[vectores\_x, valores\_x]=eig(kX); %auto-vectores de kX

[vectores\_y, valores\_y]=eig(kY); %auto-vectores de kY

* Determine y visualice el histograma de los datos.

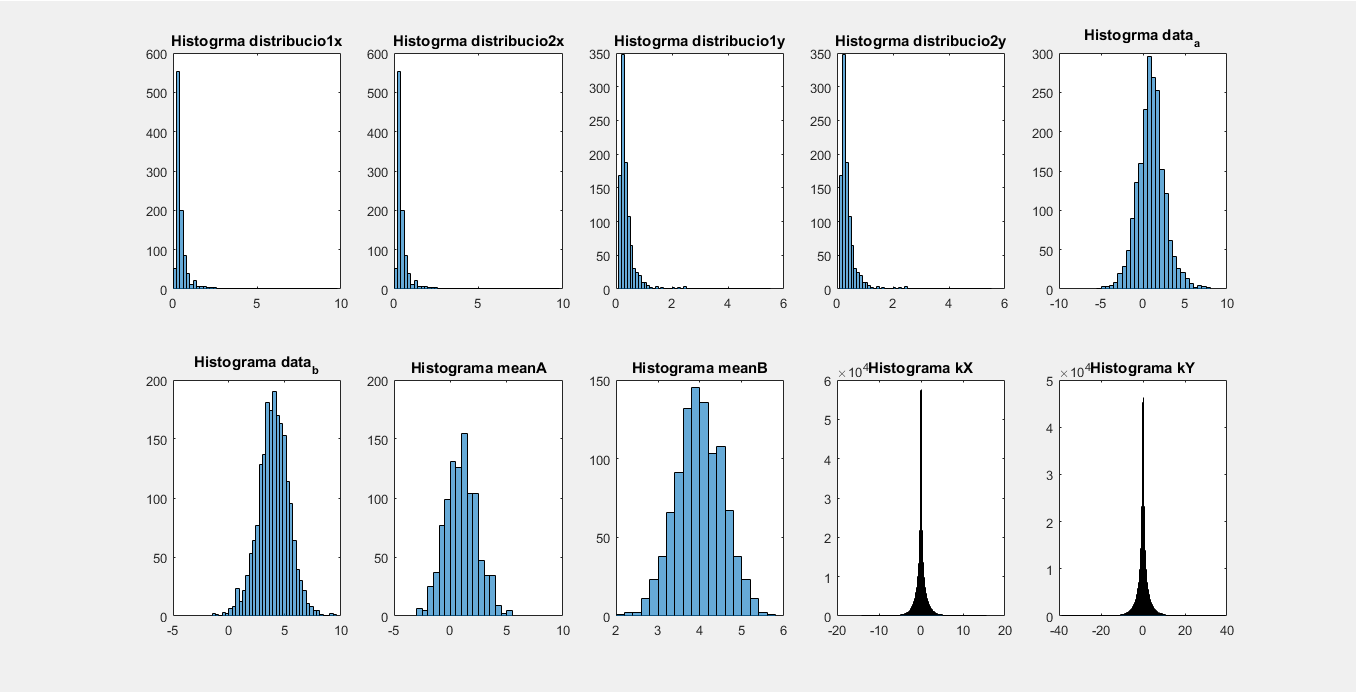


Fig 2. Graficas de los diferentes histogramas de datos, medias, matrices de covarianza y de distribución.

### El conjunto de datos dato2 contiene datos pertenecientes a 2 clases {a, b}.

* Divida los datos aleatoriamente en conjunto de **prueba** (20%) y de **entrenamiento (80%).**

En este código se usó la función “randi” donde se le pone un rango de 1 a 4000, el cual es el rango de datos de cada característica. El código usado es el siguiente:

for i=1:1:3200

numRandom=randi([1 4000],1,1);

if i<=800

for j=1:1:3

pruebaA(i,j)=a(numRandom,j);

numRandom=randi([1 4000],1,1);

pruebaB(i,j)=b(numRandom,j);

numRandom=randi([1 4000],1,1);

entrenamientoA(i,j)=a(numRandom,j);

numRandom=randi([1 4000],1,1);

entrenamientoB(i,j)=b(numRandom,j);

end

else

for j=1:1:3

entrenamientoA(i,j)=a(numRandom,j);

numRandom=randi([1 4000],1,1);

entrenamientoB(i,j)=b(numRandom,j);

end

end

end

* Visualice el conjunto de entrenamiento con un color para cada clase.

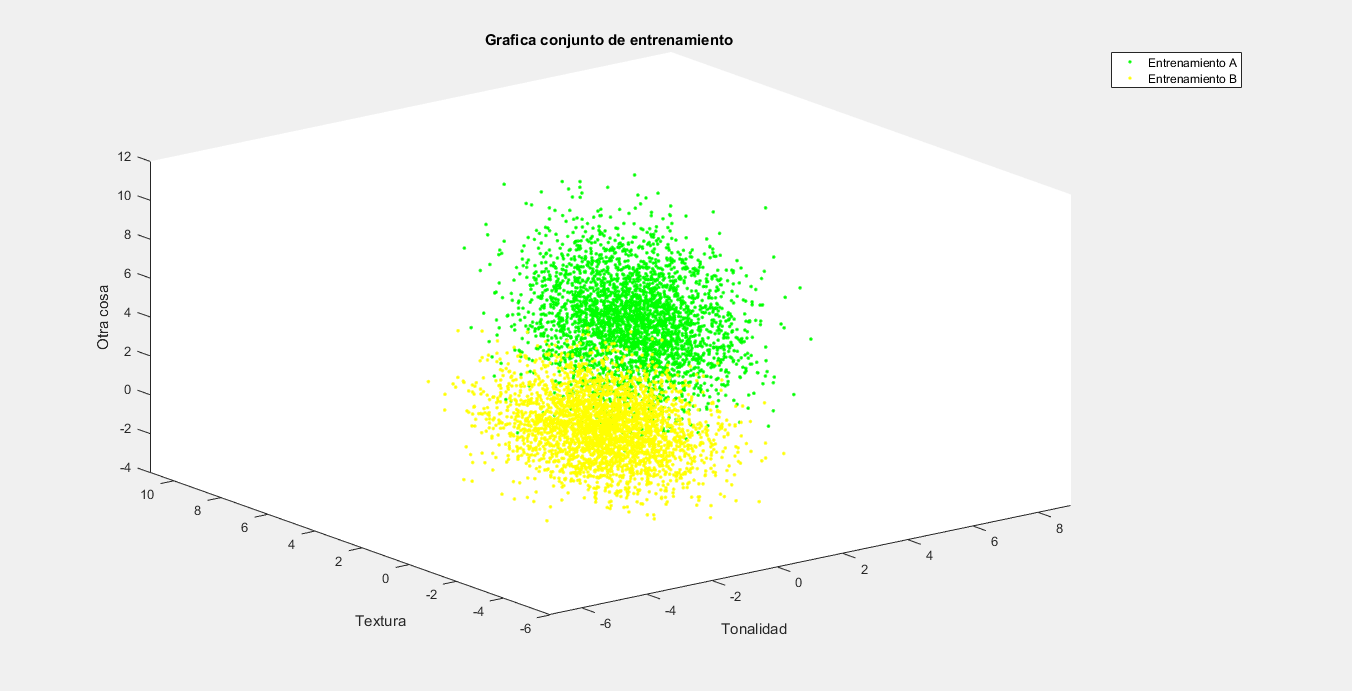


Fig 3. Grafica del conjunto de datos de entrenamientoA (verde) y entrenamientoB (Amarillo).

* Implemente un clasificador bayesiano gaussiano.

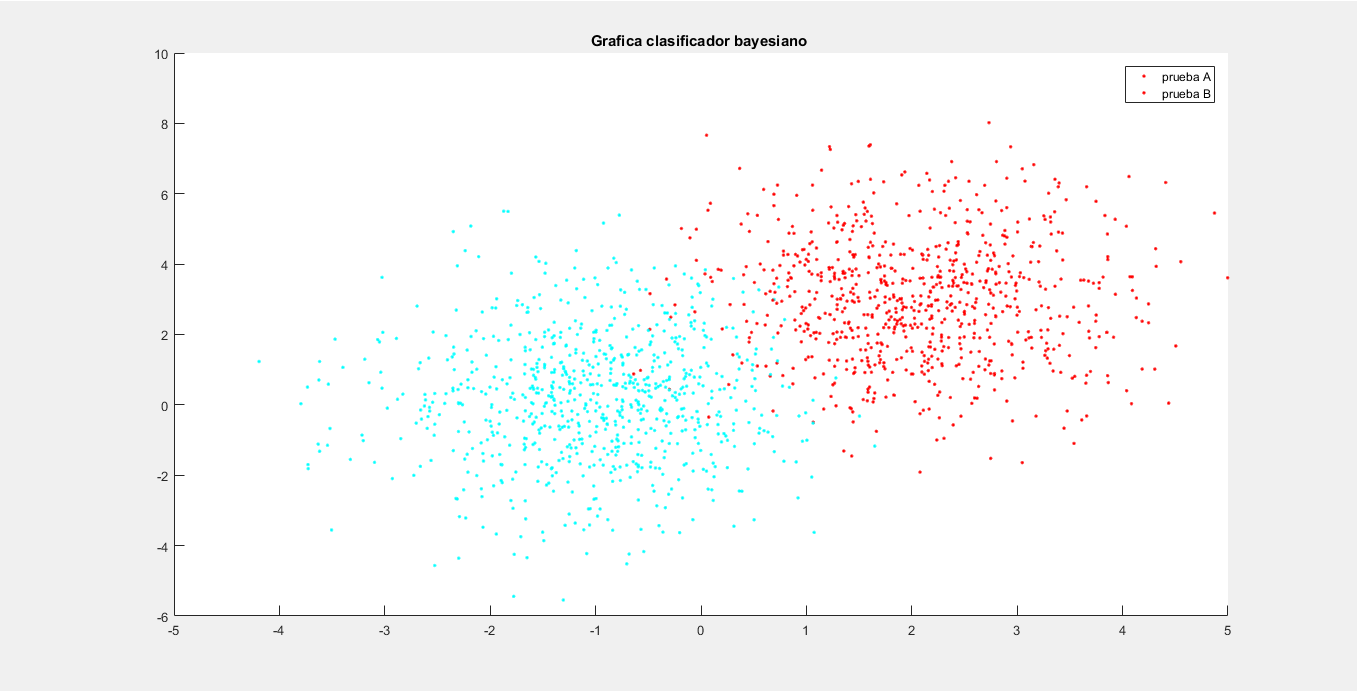


Fig 4. Grafica del conjunto de datos del Clasificador Bayesiano Gaussiano de pruebaA (rojo) y pruebaB (cian).

* Implemente un clasificador bayesiano gaussiano naive.

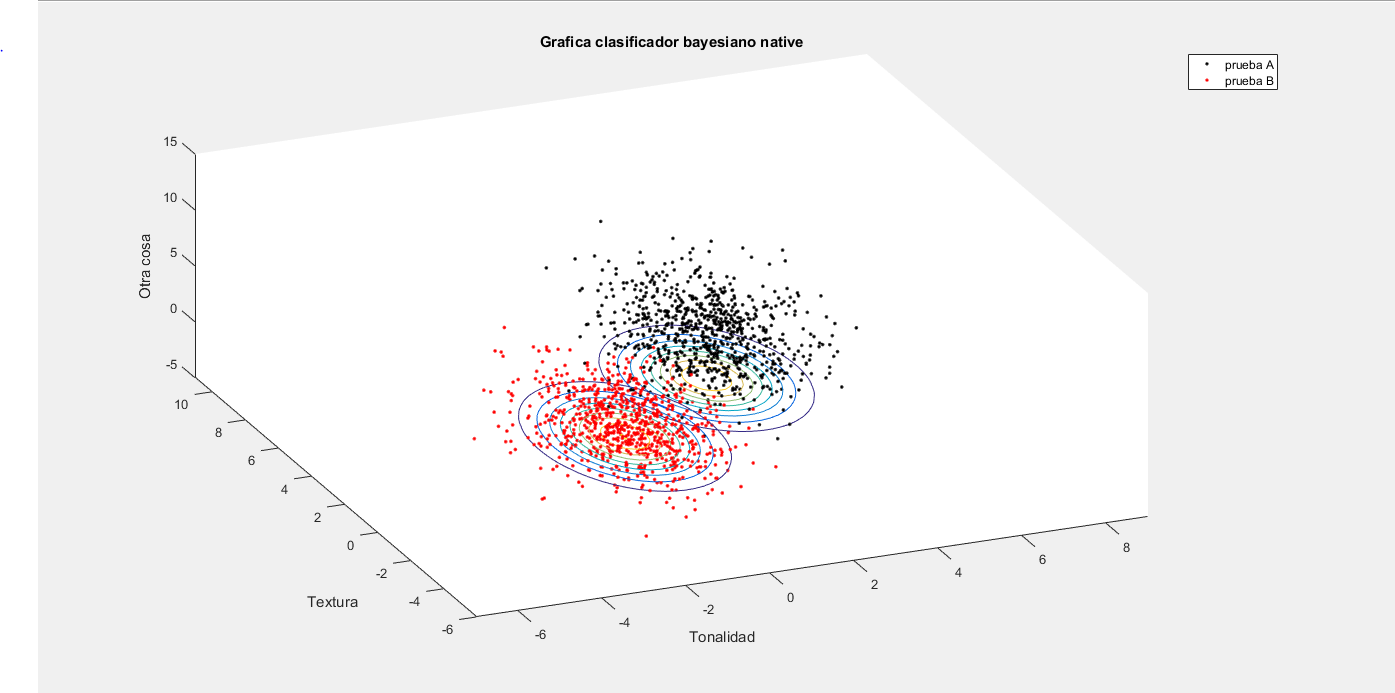


Fig 5. Grafica del Clasificador Bayesiano Gaussiano Naive de los datos prueba A (negro) y prueba B (rojo)

### Utilizando el conjunto de datos del Punto 2.

* Determine la transformación de a utilizando PCA (sobre el conjunto de entrenamiento).

La reducción dimensional de los datos de entrenamiento que eran de a un espacio dimensión o mejor dicha como la proyección de los datos aplicando el análisis de componentes principales se realizo mediante el comando “pca”, el siguiente código es el usado para llegar a proyectar en

[WA pcA] = pca(entrenamientoA');

[WB pcB] = pca(entrenamientoB');

pcA=pcA';

pcB=pcB';

WA=WA';

WB=WB';

plot(pcA(1,:),pcA(2,:),'.r'); % Teniendo las principales

plot(pcB(1,:),pcB(2,:),'.b'); % caracteristicas solo se

title('PCA'); % grafica en 2D

legend('Entrenamiento A','Entrenamiento B')

xlabel('PC 1'); ylabel('PC 2')

* Implemente un clasificador bayesiano gaussiano sobre los datos transformados.

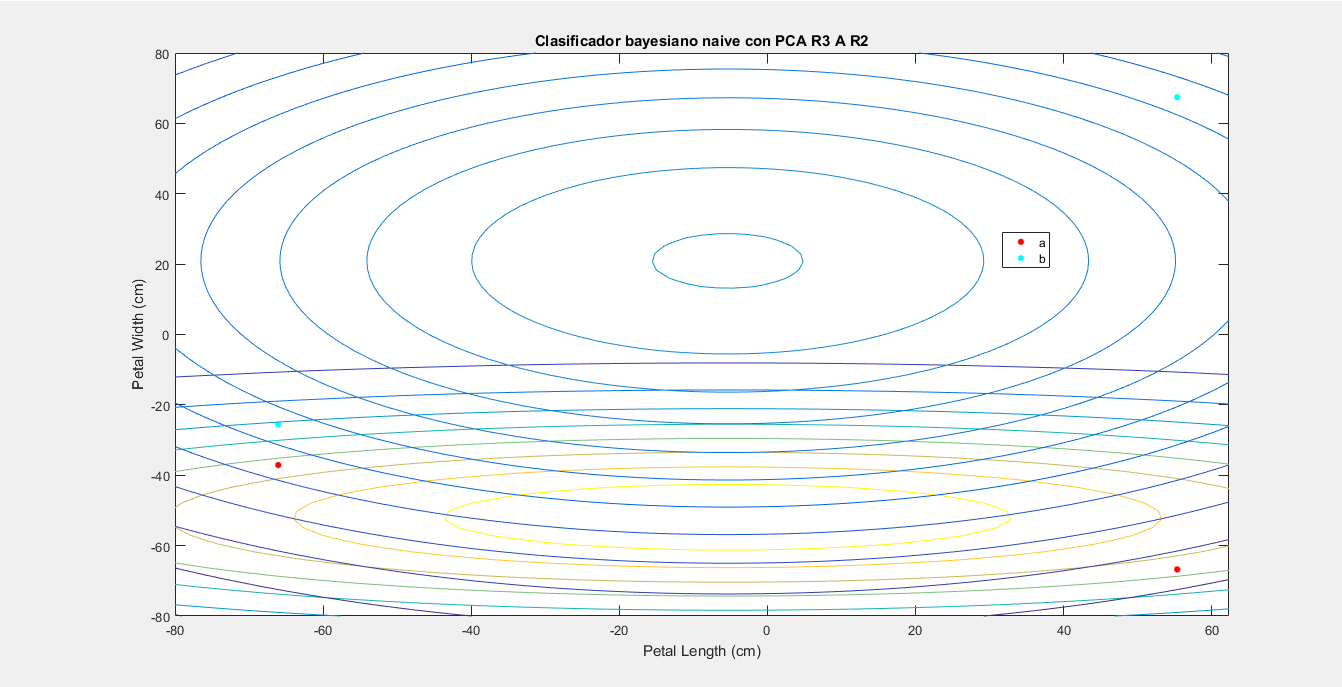


Fig 6. Grafica del Clasificador Bayesiano Gaussiano Naive de los datos de entrenamientoA (rojo) y entrenamientoB (cian) luego de aplicar PCA y proyectarlos en 2D.

### Concluya sobre los resultados obtenidos de los 3 clasificadores construidos anteriormente.

# CONCLUSIONES

* Para el primer clasificador se puede ver que su implementacion fue la mas sensilla ya que contaba con dos caracteristicas y dos clases, por lo que siempre se manejo un espacio de

Los datos de este clasificador tienen un punto de convergencia y esto se puede ver en el histograma de las medias, donde si se superpusieran estarían tocándose en valores alrededor de 4. Esto ultimo se corrobora en el histograma de los datos

* Para el segundo clasificador, al tener 3 características llevo el espacio en donde se manejan los datos a donde se puede ver que los datos que hay sufren de un comportamiento similar al del punto uno, habiendo un punto convergente para los datos corroborado por las probabilidades vistas en el clasificador Bayesiano Gaussiano.

El clasificador Bayesiano Gaussiano Naive en este grupo de datos fue en especial complejo de aplicar, ya que tenia que llevarse la densidad de probabilidades a 3 dimensiones, ya que había 3 características por clase y no 2 como en el punto uno. Pero igualmente al aplicar uno de 2 dimensiones a 2 de las 3 características, se puede apreciar que si da los resultados esperados.

* Para el tercer clasificador, este fue por el método PCA que es el análisis de los componentes principales, la cual es una técnica de visualización para datos de altas dimensiones y para pre-procesar datos. Como podemos ver en los resultados, el PCA reduce la dimensionalidad (número de variables) de nuestros datos manteniendo la mayor varianza posible, permitiendo así la proyección o transformación de a siendo menor o igual a

# REFERENCIAS

[1]http://www.maginvent.org/articles/pidht/pidtoot/Reconocimiento\_Patrones.html

**[2]** Sucar, L. E., & Tonantzintla, M. (2006). Redes Bayesianas. BS Araujo, Aprendizaje Automático: conceptos básicos y avanzados, 77-100.

# BIOGRAFÍA

******

***Suarez L. Juan Diego.*** Nació el 27 de enero del año 1994 en la ciudad de Tunja, hijo único de Beatriz Elvira Londoño Ruiz y Hugo Ernesto Suarez Leandro. Estudio y se graduó en el Gimnasio Campestre del Norte, Promoción 2010. Actualmente studia Ingeniería Electrónica en la Universidad Santo Tomas seccional Tunja.

1. Manuscrito del 23 de Marzo de 2018

   Juan Diego Suarez Londoño, Universidad Santo Tomás Tunja, juan.suarezl@usantoto.edu.co, Reconocimiento de patrones [↑](#footnote-ref-1)