Solución taller 2

Daniel Zambrano Acosta1

Pontificia universidad Javeriana, Ingeniería Electrónica, Bogotá D.C, Colombia

Objetivos:

* Utilizar el método de Análisis de Componentes Principales (PCA) y analizar su funcionalidad.

1. Teniendo en cuenta los datos data\_3D con dos clases{a,b}.
2. Transformación de

Para la trasformación de datos se utilizó el método PCA calculando el valor que aporta cada una de las características de cada clase, para ello se dividieron los datos en el 20% para la prueba y 80% para el entrenamiento y se trabajaron en los datos de entrenamiento.

A continuación, se muestra cada uno de los pasos que se realizó para identificar las características relevantes de los datos de entrenamiento.

En un primer paso se tiene la extracción y división de los datos.

A picture containing graphical user interface

Description automatically generated

Ilustración 1 Extracción y división de datos

Posterior a ello, se centraliza los datos obteniendo la media de cada una de las clases y restándolo al valor original, como se muestra a continuación.

A screenshot of a computer

Description automatically generated with medium confidence

Ilustración 2 centralización de datos

Posteriormente, se procede al cálculo de la matriz de covarianza, de los vectores y valores propios.

Text

Description automatically generated

Ilustración 3 Cálculo de matriz de covarianza, vectores y valores propios

Una vez hecho esto, se debe calcular u observar que características aportan al entrenamiento, por ello se realizó la siguiente grafica.

Chart, histogram

Description automatically generated

Ilustración 4 Características de clase a

Chart, histogram

Description automatically generated

Ilustración 5 Características de clase b

De las anteriores ilustraciones se puede observar que las características que mayor aportan son las dos ultimas para las dos clases.

Una vez identificadas las características importantes se construye la matriz de los datos no correlacionados como se muestra en la siguiente fracción de código.

Text

Description automatically generated

Ilustración 6 Matriz de datos no correlacionados

1. Clasificador bayesiano

Para implementar este clasificador, se debe centralizar los datos de la matriz de entrenamiento teniendo en cuenta el proceso realizado en la reducción de dimensión, como se muestra en las siguientes líneas de código.

Text

Description automatically generated

Ilustración 7 Matriz de entrenamiento.

Una vez construida la matriz de entrenamiento se procedió a la construcción de la función de verosimilitud para calcular la probabilidad y así poder clasificar los datos en las dos clases, como se muestra en la siguiente fracción de código.

Text

Description automatically generated

Ilustración 8 Clasificación bayesiana

En esta ilustración se calcula las medias muestreales de cada clase y las matrices de covarianza, para poder encontrar la probabilidad de los resultados. Los datos al estar centrados se tiene una media muestral de 0.

En la variable “Y\_bayes” se guardó la clasificación con un indicador de cero si es la clase a y de uno si es la clase b. sin embargo, para apreciar la clasificación de los datos según las características, se realizó el siguiente código.

Text

Description automatically generated

Ilustración 9 Características de cada clase

Chart, scatter chart

Description automatically generated

Ilustración 10 clasificación de los datos

Error de clasificación.

Para el cálculo del error se tuvo en cuenta un vector con los datos ideales como se muestra en el siguiente código.

Text

Description automatically generated

Al momento de compilar el código se encontró un error del 10.5%. con esto se puede concluir que a comparación de lo anterior