



UNIVERSIDAD AUTÓNOMA DE AGUASCALIENTES  
CURSO INTERSEMESTRAL: MACHINE Y DEEP LEARNING

ACTIVIDAD 1 03. PCA EN R(PRÁCTICA).

DOCENTE:

FRANCISCO JAVIER LUNA ROSAS

ALUMNO:

EMILIO LUNA PÉREZ, ID:244182

JUNIO-AGOSTO DE 2021

## Evidencias de la Practica

Defina el algoritmo PCA, explique el funcionamiento del algoritmo PCA e implemente el algoritmo PCA en R, el dataset que se utilizará, es el dataset visto en clase de notas escolares.

### PCA

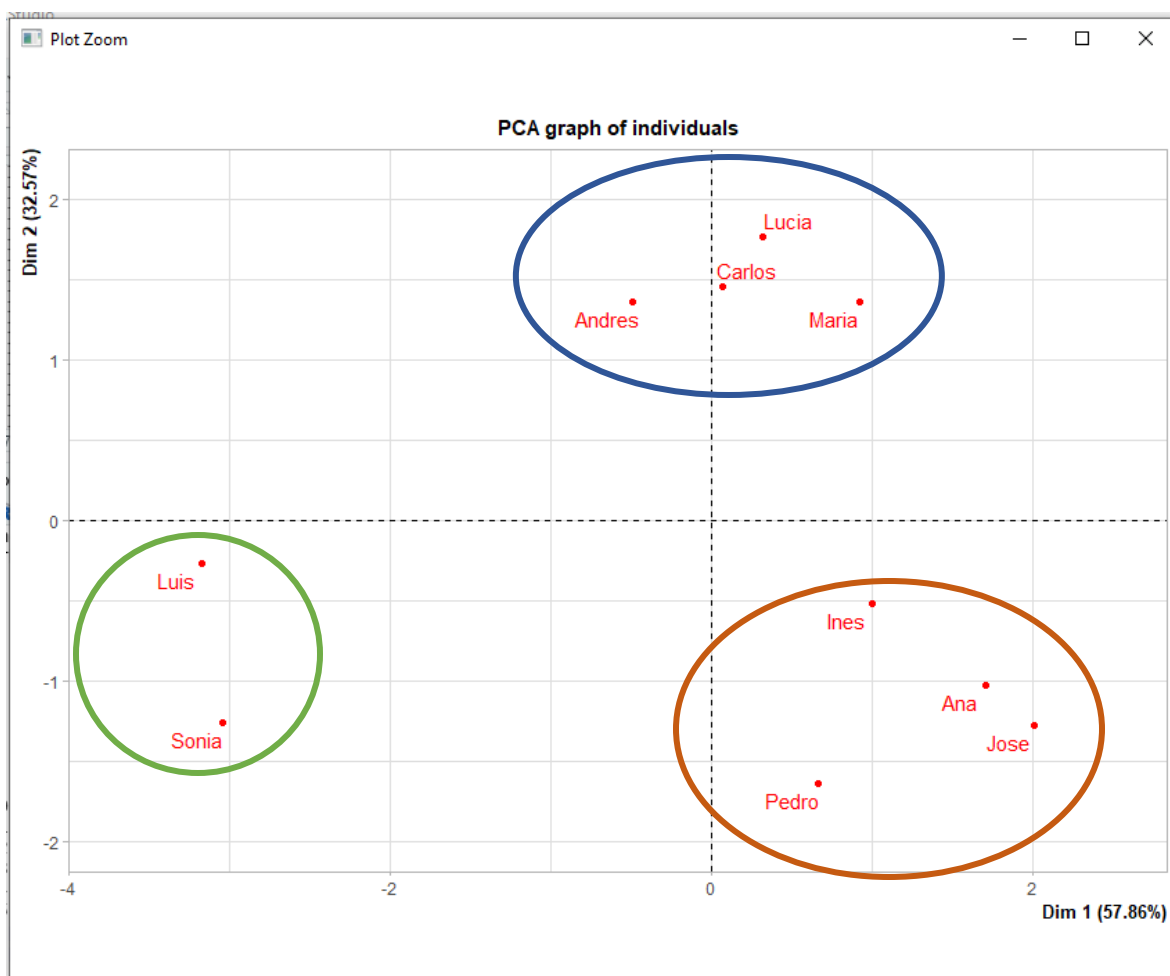
El Análisis de Componentes Principales (PCA) es una técnica del clustering (aprendizaje no supervisado) que suele utilizarse como parte del análisis exploratorio de los datos; lo que busca PCA es buscar el mejor ángulo de los datos, con el objetivo de bajar la dimensión; El objetivo es encontrar el mejor plano (subespacio para ver la nube de puntos), en otras palabras: construir un pequeño número de nuevas variables(componentes) en los cuales se concentre la mayor cantidad de información.

El Clustering (aprendizaje no supervisado) que se similar a la clasificación cuyo objetivo es particionar o segmentar un conjunto de datos o individuos en grupos que pueden ser disjuntos o no. Los grupos se forman basados en lo similar de los datos o individuos en ciertas variables.

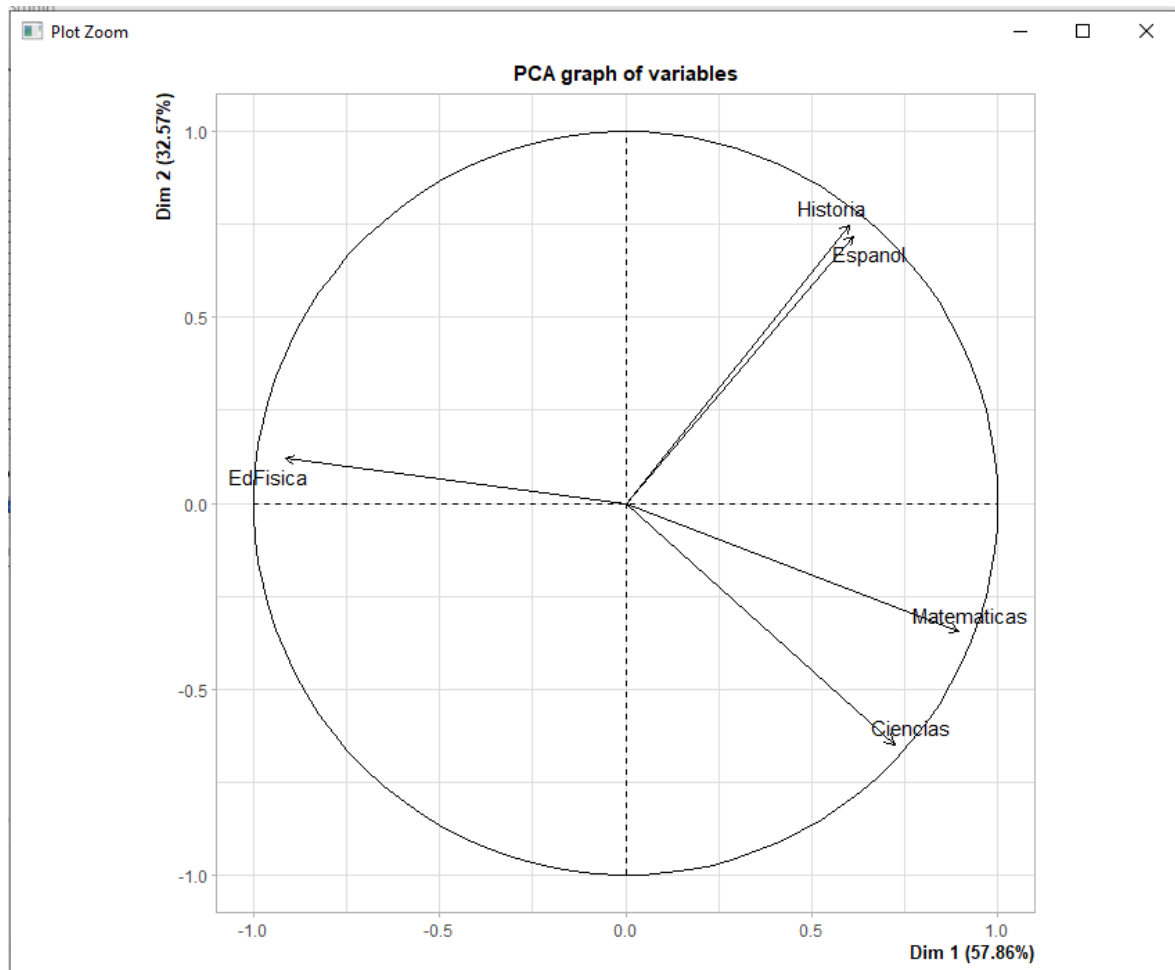
### Algoritmo del PCA en R

```
#  
library(FactoMineR)  
#  
setwd("C:/Users/emili/Desktop/VACACIONES 2021/MACHINE LEARNING")  
Datos <- read.table("EjemploEstudiantes.csv",header=TRUE,sep=";",dec="," ,row.names=1)  
#  
#PCA  
#  
res <- PCA(Datos,scale.unit = TRUE,ncp = 5,graph = FALSE)  
res  
#  
plot(res,axes = c(1,2),choix = "ind",col.ind = "red",new.plot = TRUE)  
#  
#  
plot(res,axes = c(1,2),choix = "var",col.ind = "blue",new.plot = TRUE)  
#circulo de correlaciones  
#
```

## PLANO PRINCIPAL



## CIRCULO DE CORRELACIÓN



### Clustering (Aprendizaje No Supervisado De Análisis De Conglomerados)

El clúster verde: Luis y Sonia: es representado por Luis y Sonia con buenas notas de educación física: son buenos deportistas. Podemos hacer notar que son malos en ciencias y matemáticas porque tienen un ángulo mayor a  $180^\circ$  con respecto a educación física. A su vez son malos en español e historia, pues se oponen negativamente en un Angulo mayor de  $45^\circ$ .

El clúster rojo: Ana, Inés, José y Pedro: parece ser el opuesto del clúster de educación física. Sus miembros son buenos en las ciencias y las matemáticas, pero son malos en deportes.

El clúster azul: Andrés, Carlos, Lucía y María: tiene a los estudiantes destacados en el área de español e historia. No son buenos en deportes, pero no son tan malos

## Conclusiones

Es la primera semana que nos adentramos de lleno en machine learning, pues estábamos descubriendo el lenguaje de R, que era indispensable para un buen desarrollo de este curso, en cuanto a la práctica de PCA, me pareció complejo el algoritmo del PCA, pero al menos no tuvimos que implementarlo en R, sino que utilizamos una librería que facilita el uso de funciones. Me parecen interesantes la elaboración de las graficas porque nos aportan más información de manera visual.

## Referencias

Download R-4.1.0 for Windows. (2021). The R-project for statistical computing. Consultado en julio 28, 2021, de R-project.org Sitio web: <https://cran.r-project.org/bin/windows/base/>

Download the RStudio IDE. (2021). Consultado en julio 28, 2021, de Rstudio.com Sitio web: <https://www.rstudio.com/products/rstudio/download/>

RPubs (2019) Análisis de componentes principales (PCA). Consultado en Agosto 1, 2021, de Rpubs.com Sitio web: [https://rpubs.com/Cristina\\_Gil/PCA](https://rpubs.com/Cristina_Gil/PCA)