



UNIVERSIDAD AUTÓNOMA DE AGUASCALIENTES
CURSO INTERSEMESTRAL: MACHINE Y DEEP LEARNING

ACTIVIDAD 1_04. K-MEANS 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 K-Means, explique el funcionamiento del algoritmo de K-Means e implemente el algoritmo K-Means en R, el dataset que se utilizará, es el dataset visto en clase de notas escolares.

K-MEANS

Este método agrupa las observaciones en k clusters distintos, donde k lo determina el analista antes de ejecutar el algoritmo. K-means encuentra los k mejores clusters, es el mejor k el que minimiza la distancia intra cluster y maximiza la distancia inter cluster. El algoritmo empleado para ello es:

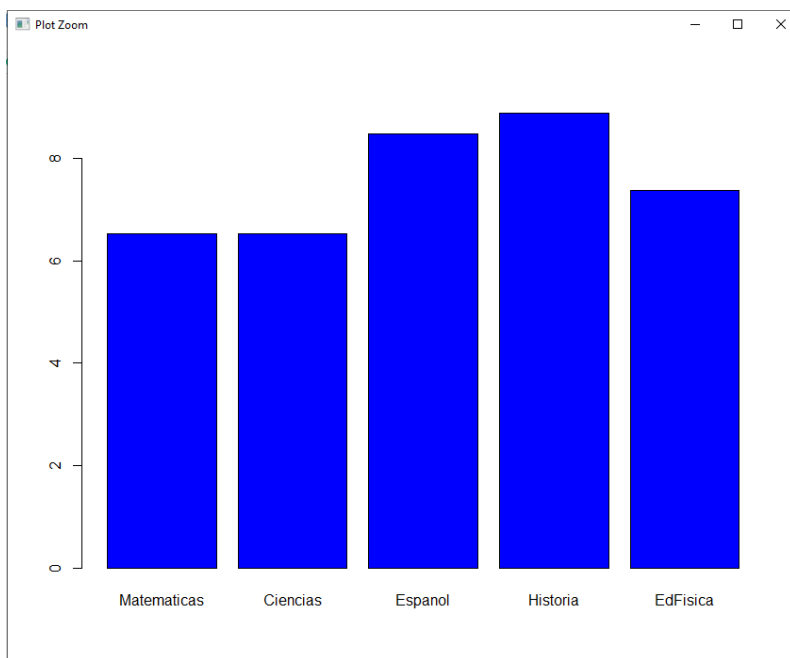
- (1). Especificar el numero k de clusters que se quieren crear.
- (2). Seleccionar de forma aleatoria k observaciones del set de datos como centroides iniciales.
- (3). Asignar cada una de las observaciones al centroide más cercano.
- (4). Para cada uno de los K clusters recalculan su centroide.
- (5). Repetir los pasos 3 y 4 hasta que las asignaciones no cambien o se alcance el numero maximo de iteraciones establecido.

ALGORITMO DEL K-MEANS EN R

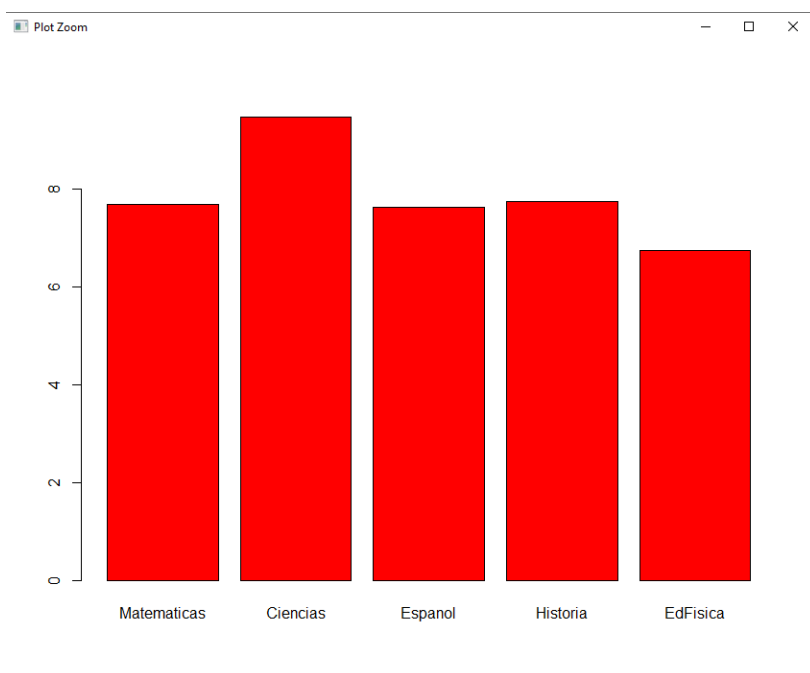
```
#
library(FactoMineR)
#
setwd("C:/Users/emili/Desktop/VACACIONES 2021/MACHINE LEARNING")
Datos <- read.table("EjemploEstudiantes.csv", header=TRUE, sep=";", dec=".", row.names=1)
Datos
#
#entrena el modelo
grupos = kmeans (Datos, 3, iter.max = 1500)
#grafica los clusters
barplot (grupos$center[1,], col= "blue") #centro de gravedad del cluster1
barplot (grupos$center [2,], col= "red") #centro de gravedad del cluster2
barplot (grupos$center [3,], col = "green") #centro de gravedad del cluster3
#grafica los tres clusters completos
rownames (grupos$centers)= c("cluster 1", "cluster 2", "cluster 3")
barplot(t(grupos$centers), beside = TRUE, col = heat.colors (5))

# En grupos verifico a que cluster pertenece cada fila de la tabla de datos
NDatos=cbind (Datos, Grupo=grupos$cluster)
NDatos
setwd("C:/Users/emili/Desktop/VACACIONES 2021/MACHINE LEARNING")
write.csv (NDatos, "NDatos.csv")
```

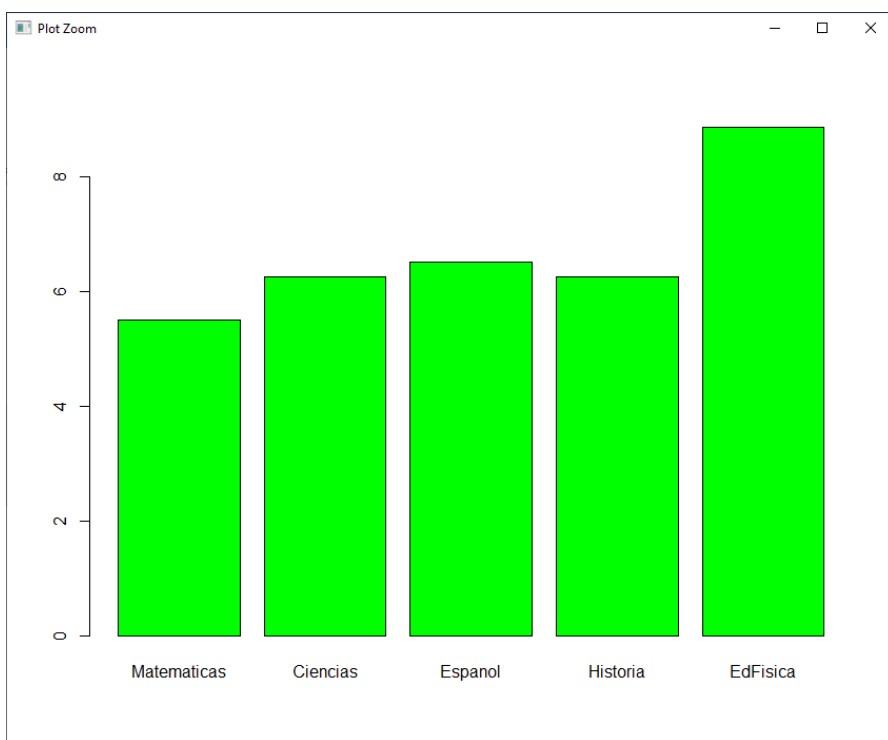
Centro de gravedad del cluster 1



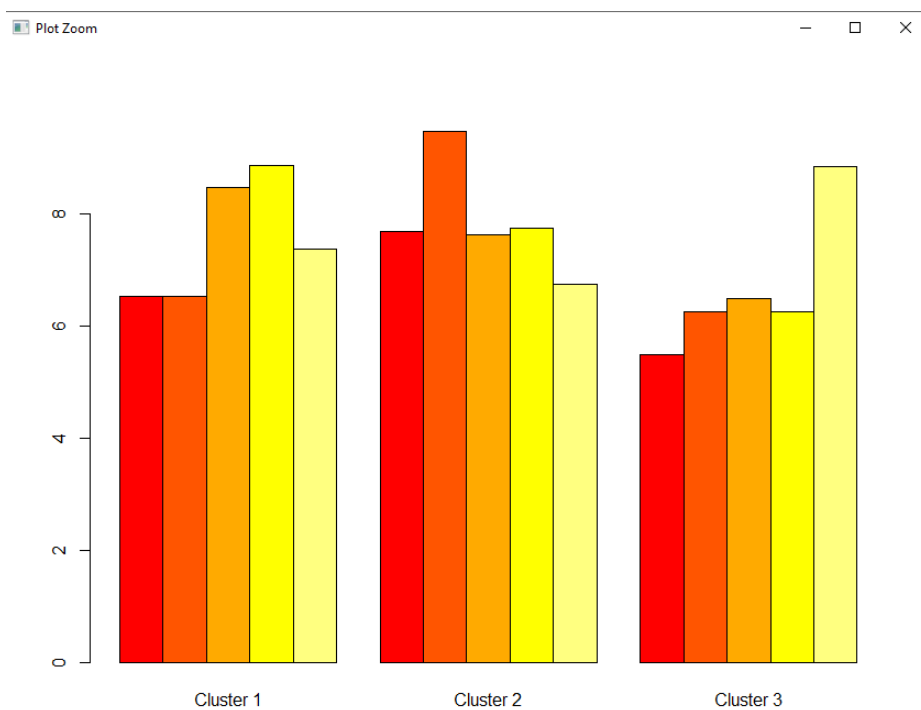
Centro de gravedad del cluster 2



Centro de gravedad del cluster 3



Centro de gravedad de los clusters



En grupos verifico a que cluster pertenece cada fila de la tabla de datos

```
# En grupos verifico a que cluster pertenece cada fila de la tabla de datos
NDatos=cbind (Datos, Grupo=grupos$cluster)
Datos
NDatos
setwd("C:/Users/emili/Desktop/VACACIONES 2021/MACHINE LEARNING")
write.csv (NDatos, "NDatos.csv")
```

	Matematicas	Ciencias	Espanol	Historia	EdFisica	Grupo
Lucia	7.0	6.5	9.2	8.6	8.0	1
Pedro	7.5	9.4	7.3	7.0	7.0	2
Ines	7.6	9.2	8.0	8.0	7.5	2
Luis	5.0	6.5	6.5	7.0	9.0	3
Andres	6.0	6.0	7.8	8.9	7.3	1
Ana	7.8	9.6	7.7	8.0	6.5	2
Carlos	6.3	6.4	8.2	9.0	7.2	1
Jose	7.9	9.7	7.5	8.0	6.0	2
Sonia	6.0	6.0	6.5	5.5	8.7	3
Maria	6.8	7.2	8.7	9.0	7.0	1

```
> setwd("C:/Users/emili/Desktop/VACACIONES 2021/MACHINE LEARNING")
> write.csv (NDatos, "NDatos.csv")
```

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° con respecto a educación física. A su vez son malos en español e historia, pues se oponen negativamente en un Angulo mayor de 45° .

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

En cuanto a aprendizaje no supervisado, este me pareció un método muy visual en cuanto al manejo de datos, el algoritmo de Kmeans con la librería FactoMiner vuelve el trabajo fácil, pero me pude dar cuenta que es algo complejo la implementación de este algoritmo pues necesitamos bases matemáticas en cuanto a distancias y otros relacionados con matemáticas.

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 (2020). Método Kmeans Consultado en agosto 1, 2021, de Rpubs.com Sitio web: <https://www.rpubs.com/Giorgia96/kmeans>