

**Instituto Tecnológico de Tijuana**

**Nombre de Facultad**

**Ingeniería Informática**



**Proyecto / Tarea / Practica:**

Práctica Evaluatoria Unidad 4

**Materia:**

**Minería de datos**

**Facilitador:**

JOSE CHRISTIAN ROMERO HERNANDEZ

**Alumnos:**

**Erik Saul Rivera Reyes**

**Fecha:**

Tijuana Baja California a 13 de 12 2021

## Código

```
library(readxl)

dataset<-read.csv("iris.csv")
dataset = dataset[1:4]

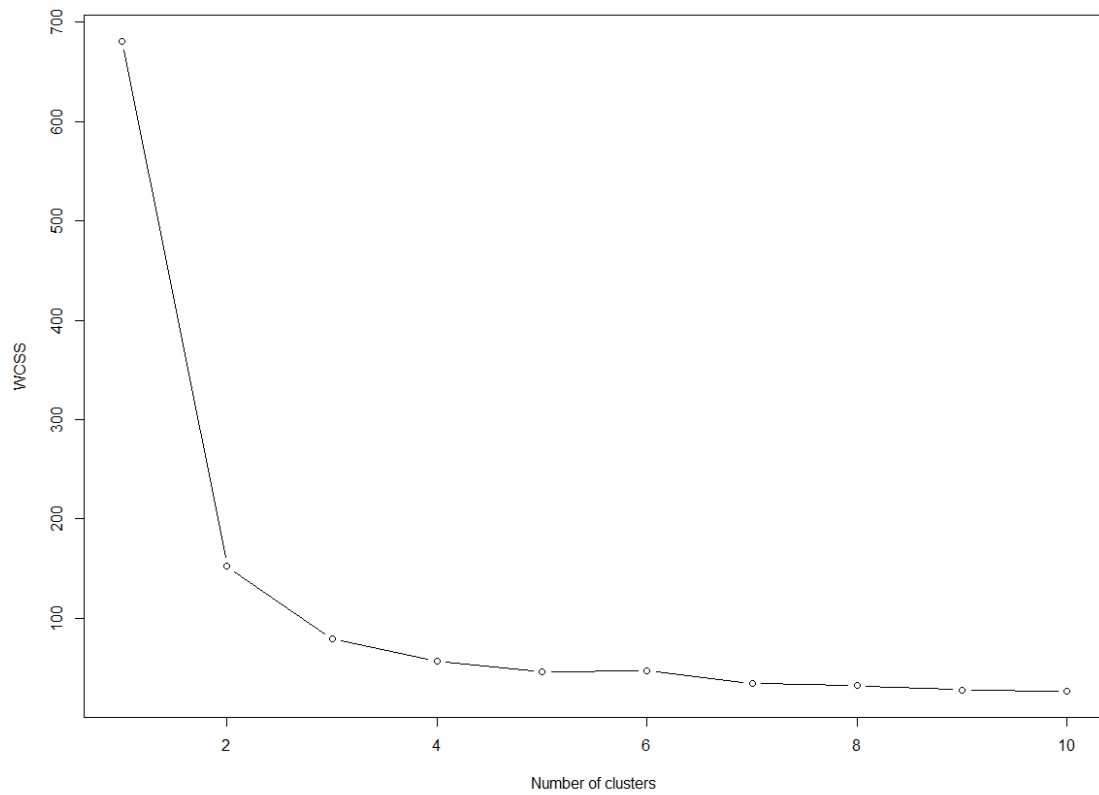
# Using the elbow method to find the optimal number of clusters
set.seed(6)
wcss = vector()
for (i in 1:10) wcss[i] = sum(kmeans(dataset, i)$withinss)
plot(1:10,
     wcss,
     type = 'b',
     main = paste('The Elbow Method'),
     xlab = 'Number of clusters',
     ylab = 'WCSS')

# Fitting K-Means to the dataset
set.seed(29)
kmeans = kmeans(x = dataset, centers = 5)
y_kmeans = kmeans$cluster

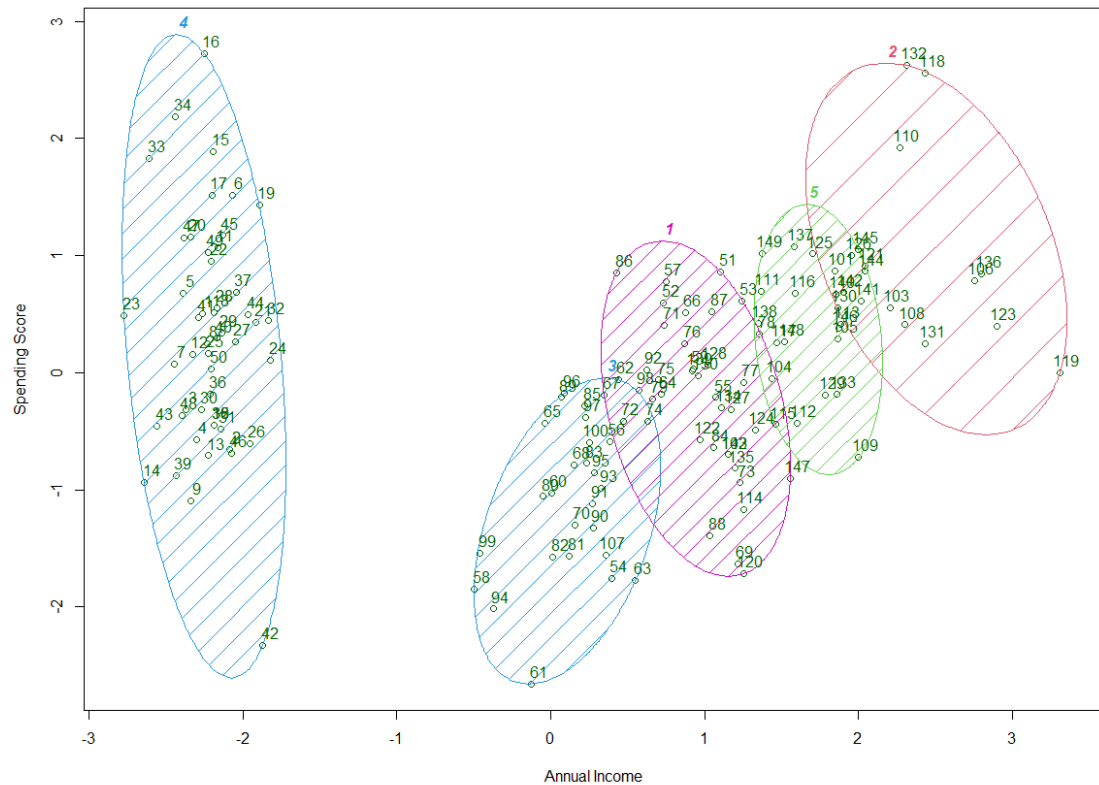
# Visualising the clusters
# install.packages('cluster')
library(cluster)
clusplot(dataset,
          y_kmeans,
          lines = 0,
          shade = TRUE,
          color = TRUE,
          labels = 2,
          plotchar = FALSE,
          span = TRUE,
          main = paste('Clusters of customers'),
          xlab = 'Annual Income',
          ylab = 'Spending Score')
```

# Graficas

The Elbow Method EXAM U4



K-MEANS EXAM U4



## ***Explicación***

El método K-means es un método de agrupamiento, que tiene como objetivo la partición de un conjunto de  $n$  observaciones en  $k$  grupos en el que cada observación pertenece al grupo cuyo valor medio es más cercano. Es un método utilizado en minería de datos. La agrupación del conjunto de datos puede ilustrarse en una partición del espacio de datos en celdas de Voronoi.

El problema es computacionalmente difícil (NP-hard). Sin embargo, hay eficientes heurísticas que se emplean comúnmente y convergen rápidamente a un óptimo local. Estos suelen ser similares a los algoritmos expectation-maximization de mezclas de distribuciones gaussianas por medio de un enfoque de refinamiento iterativo empleado por ambos algoritmos. Además, los dos algoritmos usan los centros que los grupos utilizan para modelar los datos, sin embargo k-means tiende a encontrar grupos de extensión espacial comparable, mientras que el mecanismo expectation-maximization permite que los grupos tengan formas diferentes.

## ***Defensa de mi examen***

**<https://youtu.be/fnQxCcvqvsI>**