


Instituto Tecnológico de Tijuana
Ingeniería en Informática

Subject Name:

Data Mining

Exercise:

Evaluative Practice - Unit 4

Teacher:

Romero Hernandez Jose Christian

Students:
Castillo Ramirez Guadalupe 17213043
Ramos Rivera Manuel Isai 17212931

K-Means Clustering

Set our workspace

```
getwd()
setwd("F:/Data mining U1/PracticaEvaluatoriaU4")
getwd()
```

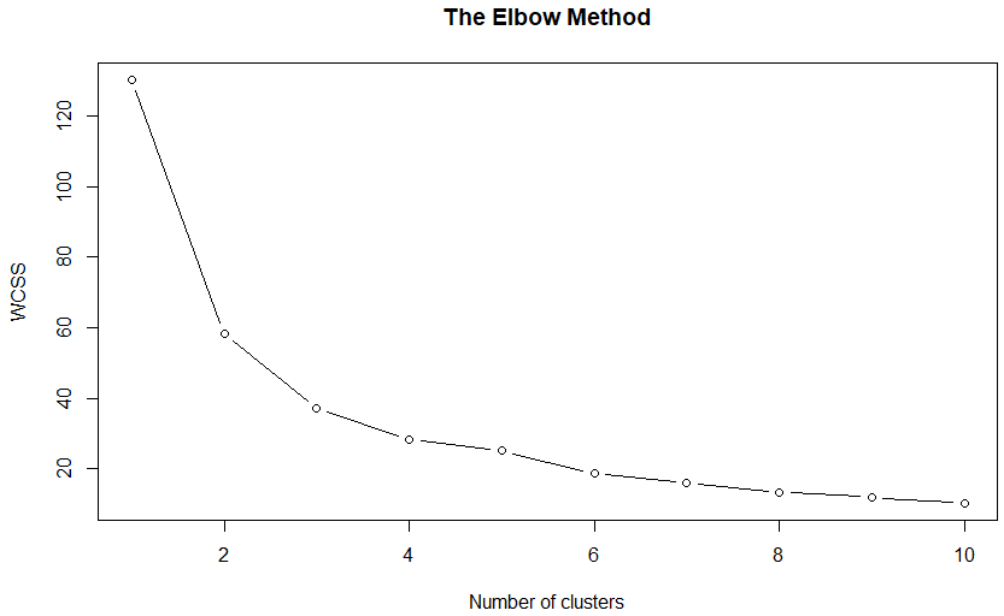
Importing the dataset

```
dataset = read.csv('iris.csv')
dt = dataset[1:2]
dt2 = dataset[3:4]
dt3 = dataset[c(1,4)]
```

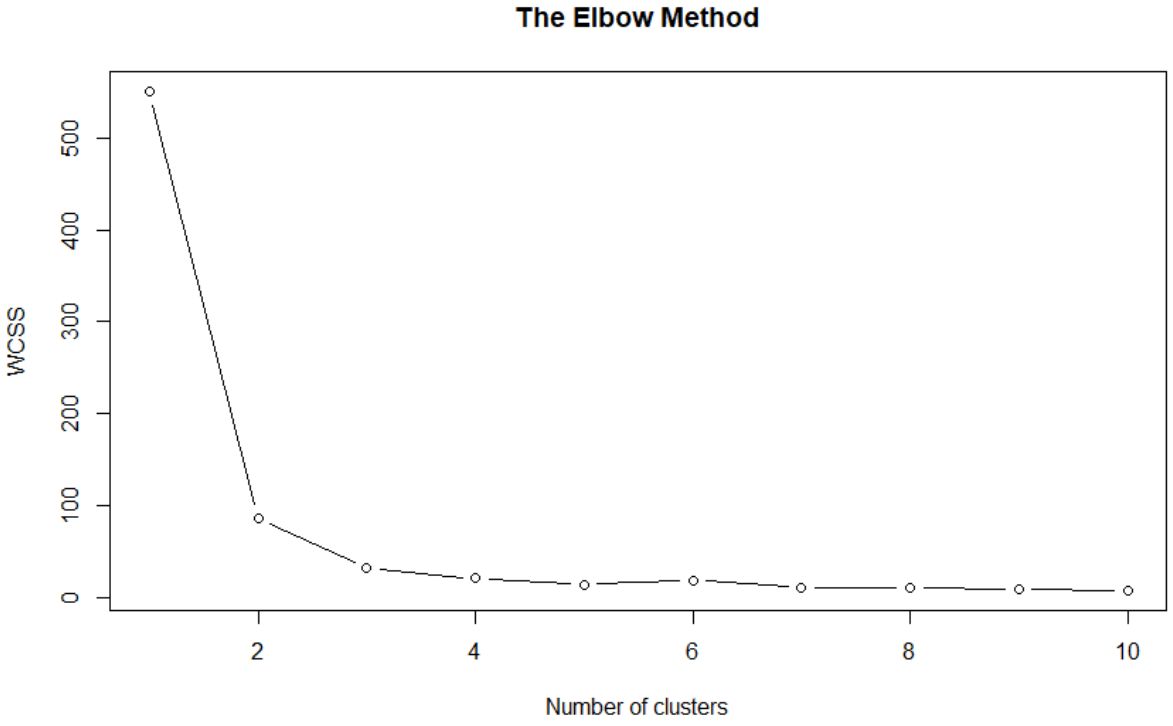
Using the elbow method to find the optimal number of clusters

```
TEM <- function(dataset){
  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')
}
```

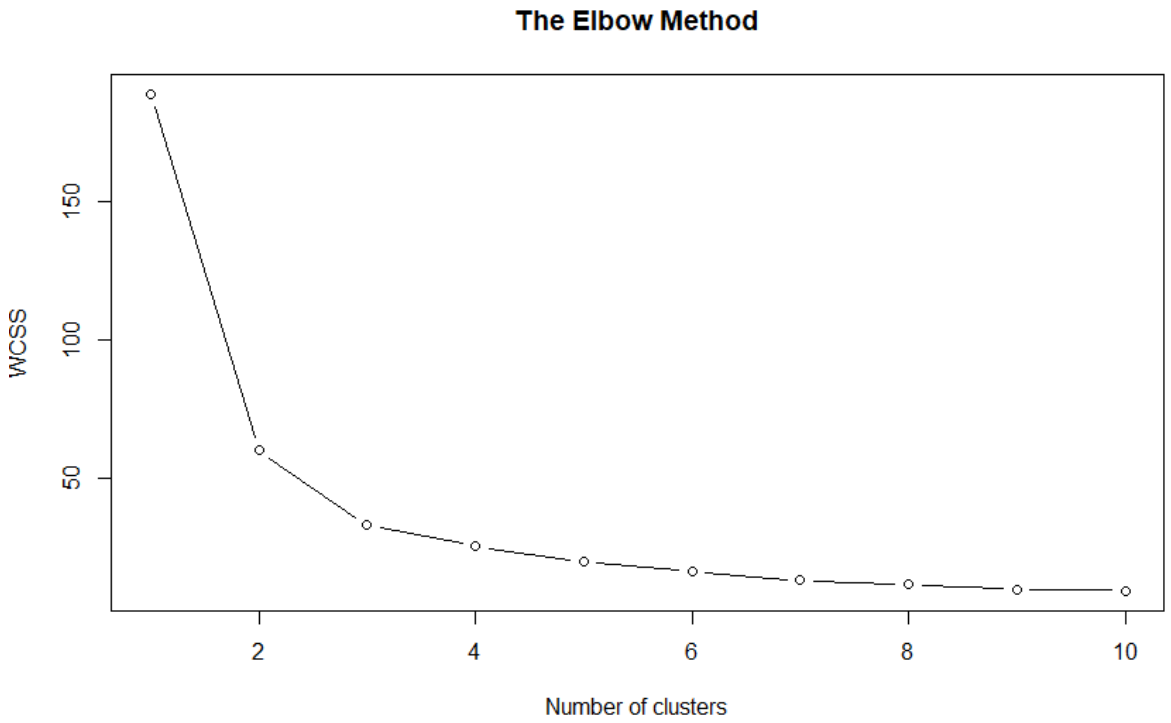
TEM(dt)



TEM(dt2)



```
TEM(dt3)
```



Fitting K-Means to the dataset

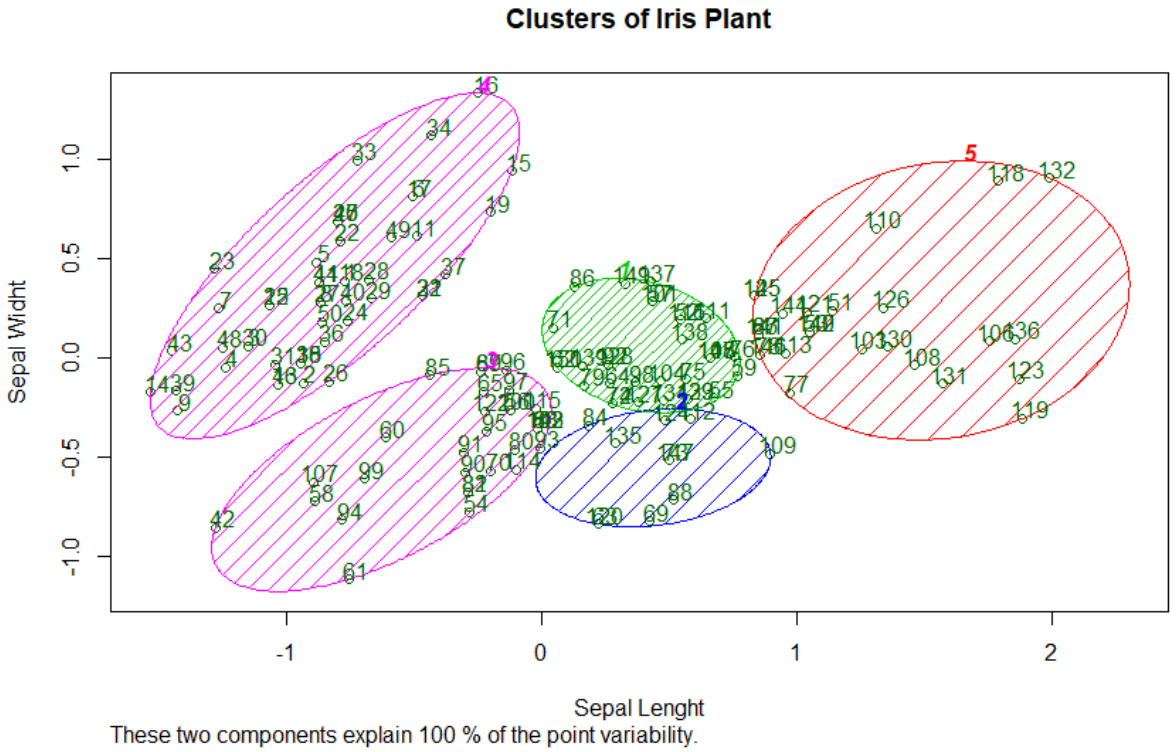
```
Clusters <- function(dataset, cnt){
  set.seed(29)
  kmeans = kmeans(x = dataset, centers = cnt)
  y_kmeans = kmeans$cluster
}
```

```
ykmeans <- Clusters(dt, 5)
ykmeans2 <- Clusters(dt2, 2)
ykmeans3 <- Clusters(dt3, 5)
```

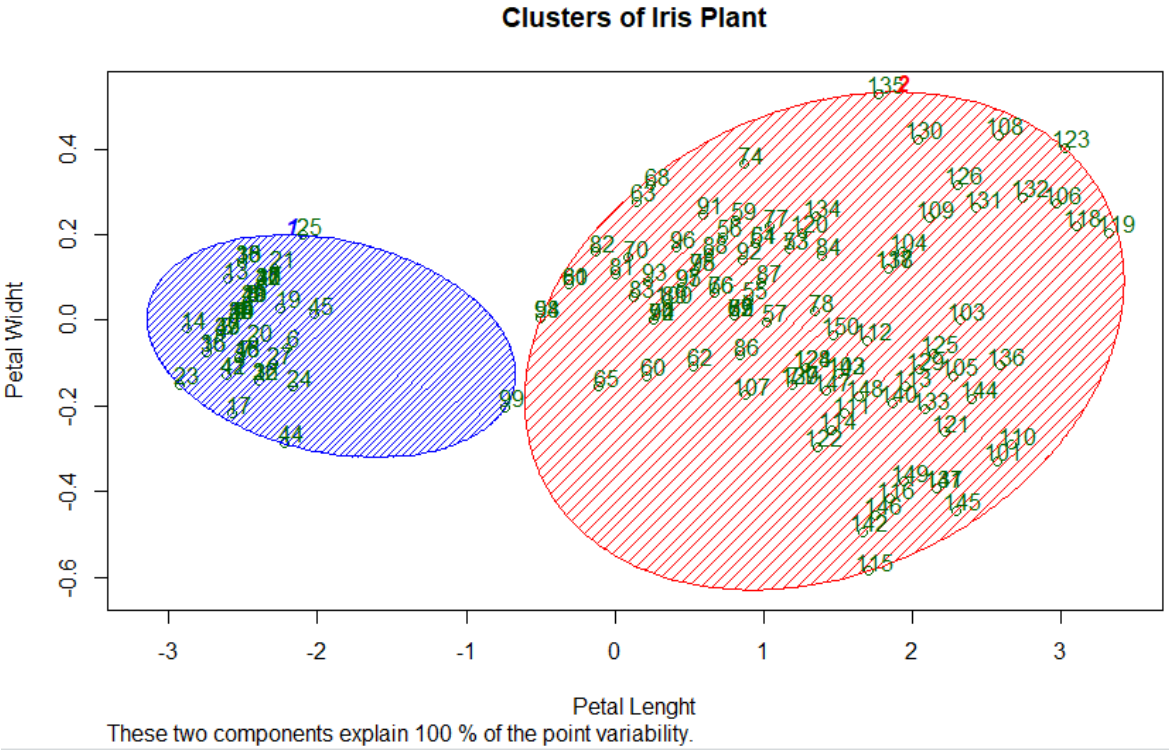
Visualising the clusters

```
install.packages('cluster')
library(cluster)

clusplot(dt,
         ykmeans,
         lines = 0,
         shade = TRUE,
         color = TRUE,
         labels = 2,
         plotchar = FALSE,
         span = TRUE,
         main = paste('Clusters of Iris Plant'),
         xlab = 'Sepal Lenght',
         ylab = 'Sepal Widht')
```



```
clusplot(dt2,
         ykmeans2,
         lines = 0,
         shade = TRUE,
         color = TRUE,
         labels = 2,
         plotchar = FALSE,
         span = TRUE,
         main = paste('Clusters of Iris Plant'),
         xlab = 'Petal Lenght',
         ylab = 'Petal Widht')
```



```
clusplot(dt3,
        ykmeans3,
        lines = 0,
        shade = TRUE,
        color = TRUE,
        labels = 2,
        plotchar = FALSE,
        span = TRUE,
        main = paste('Clusters of Iris Plant'),
        xlab = 'Sepal Lenght',
        ylab = 'Petal Widht')
```