Aim: Perform the data clustering using clustering algorithm using R/Python.

Code:

> require(graphics)

> x<-rbind(matrix(rnorm(100,sd=0.3),ncol=2),

+ +

+ + + matrix(rnorm(100,mean=1,sd=0.3),ncol=2))

> colnames(x)<-c("x","y")

> (cl<-kmeans(x,2))

K-means clustering with 2 clusters of sizes 51, 49

Cluster means:

x y

1 -5.006193e-05 0.08726408

2 1.039479e+00 0.98298082

Clustering vector:

[1] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

[38] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2

[75] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2

Within cluster sum of squares by cluster:

[1] 8.797201 10.017265

(between\_SS / total\_SS = 71.4 %)

Available components:

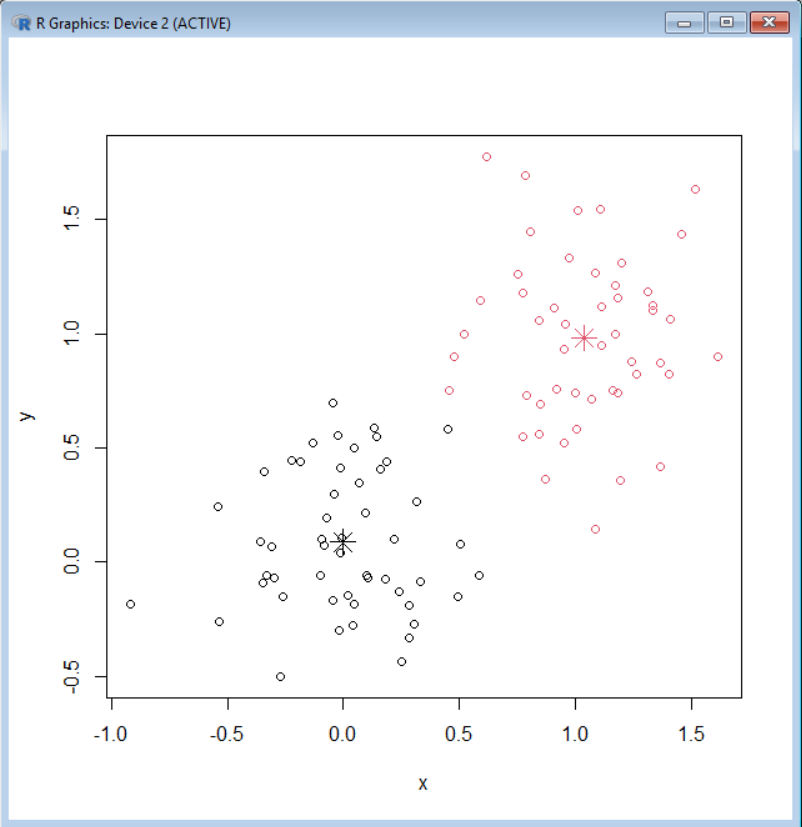
[1] "cluster" "centers" "totss" "withinss" "tot.withinss"

[6] "betweenss" "size" "iter" "ifault"

> plot(x, col = cl$cluster)

> points(cl$centers, col = 1:2,pch = 8,cex =2)

Output:



> kmeans(x,1)$withinss

[1] 56.29457

> (cl <- kmeans(x,6,nstart=29))

K-means clustering with 6 clusters of sizes 13, 13, 20, 17, 14, 23

Cluster means:

x y

1 -0.16338399 -0.1788409

2 0.68687428 1.2172188

3 0.85554268 0.6656313

4 1.16595029 1.0738302

5 -0.03960536 0.3035177

6 0.32897786 -0.1684106

Clustering vector:

[1] 6 1 1 6 6 6 1 1 1 5 6 5 6 5 6 6 1 6 1 6 1 6 6 1 6 6 5 6 6 5 5 5 6 6 5 6 1

[38] 1 6 6 5 5 5 1 6 1 5 6 5 5 2 2 3 2 3 2 3 4 3 2 2 4 4 3 3 3 3 4 4 4 2 2 3 2

[75] 3 4 4 2 2 4 3 3 4 3 3 4 2 4 3 3 4 3 2 4 3 4 4 3 3 4

Within cluster sum of squares by cluster:

[1] 0.6456872 0.6924334 1.1075181 1.0046304 0.9729954 1.4522806

(between\_SS / total\_SS = 89.6 %)

Available components:

[1] "cluster" "centers" "totss" "withinss" "tot.withinss"

[6] "betweenss" "size" "iter" "ifault"

> plot(x,col=cl$cluster)

> points(cl$centers,col = 1:5,pch=8)

Output:

