VIT®

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

Essentials of Data Analytics - (CSE3506)

Faculty – Lakshmi Pathi Jakkamputi **Sir**

Lab-7

Harshanth k Prakash

19BCE1293

L21-22 Slot

Tasks for Week-7: Partitioning Based Clustering

Understand the following operations/functions on 'iris' data and perform similar operations on 'USArrests' dataset based on given instructions.

AIM

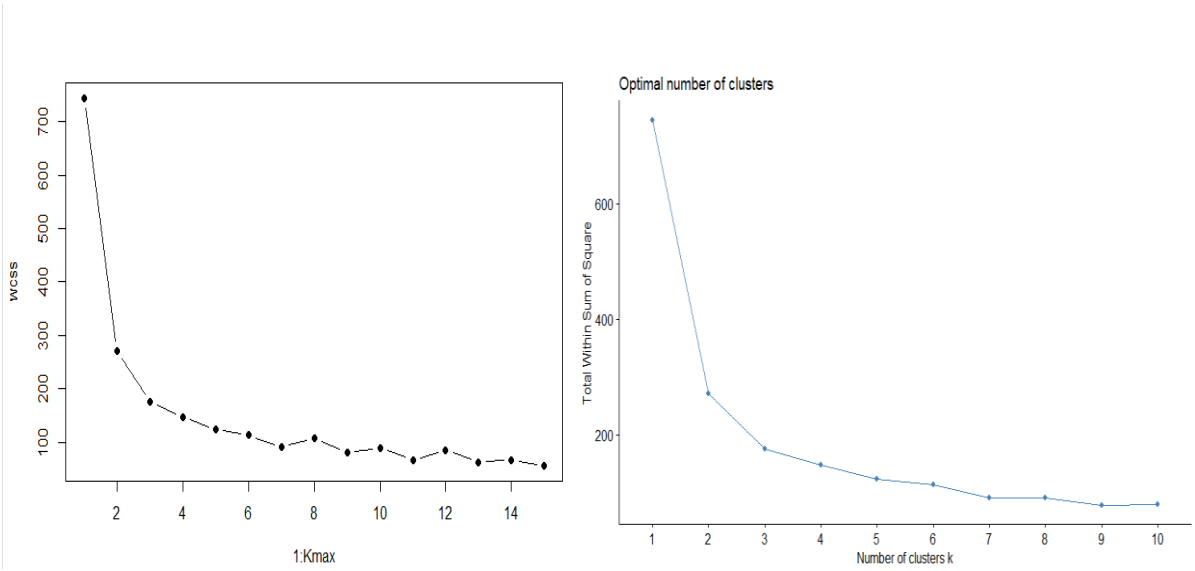
To Understand and perform operations/functions on 'iris' data and perform similar operations on 'USArrests' dataset based on given instructions.

Algorithm

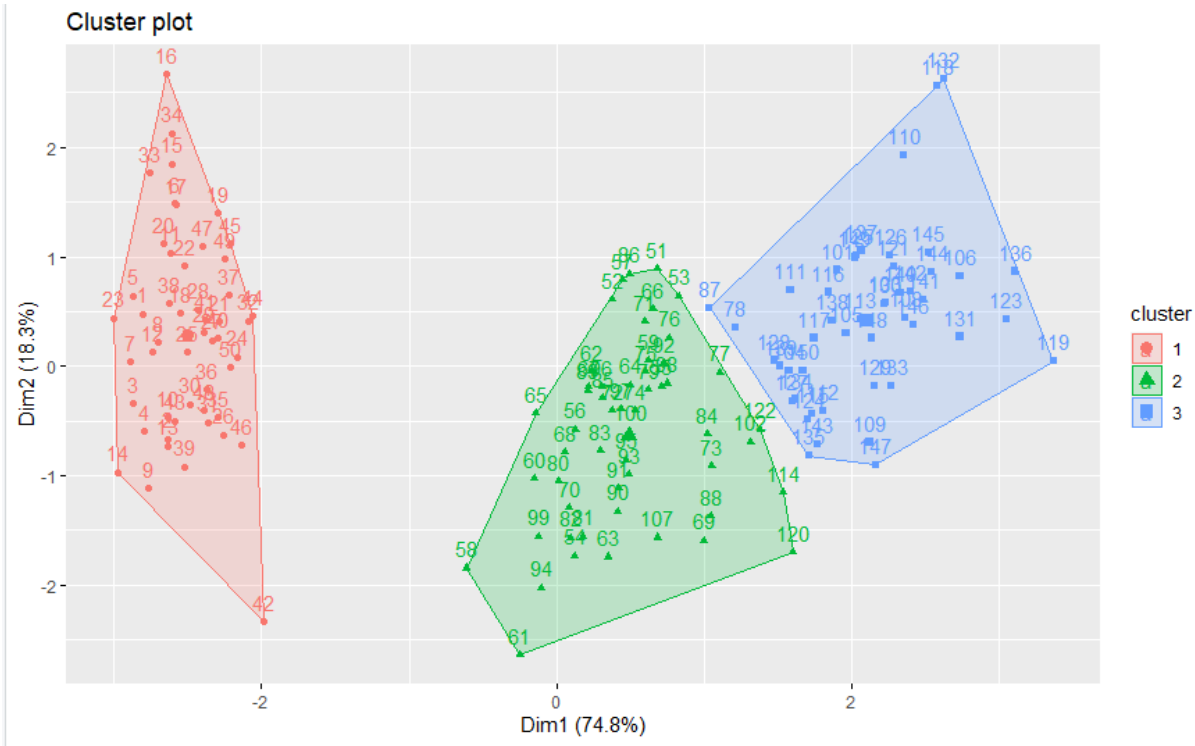
1. Start
2. Use rm(list=ls()) to clear environment variables.
3. Read the dataset from directory.
4. Make a scaled version of the original data.
5. Train the Kmeans model and go over all of its terms.
6. By using fit$cluster we can find the cluster values.
7. By using fit$size we can find the size of each cluster.
8. By using fit$withnss we can find with in cluster sum of squares for each cluster.
9. By using fit$tot.withnss we can find with in cluster sum of squares with respective to all clusters.
10. Find WCSS values and keep a list of them.
11. To find the elbow point, plot the graph.
12. Train the K-Medoids model and go over all of its terms.
13. By using the fitm$medoid we can find no of medoid
14. Using fviz_cluster function we can find the medoids
15. Stop.

Result

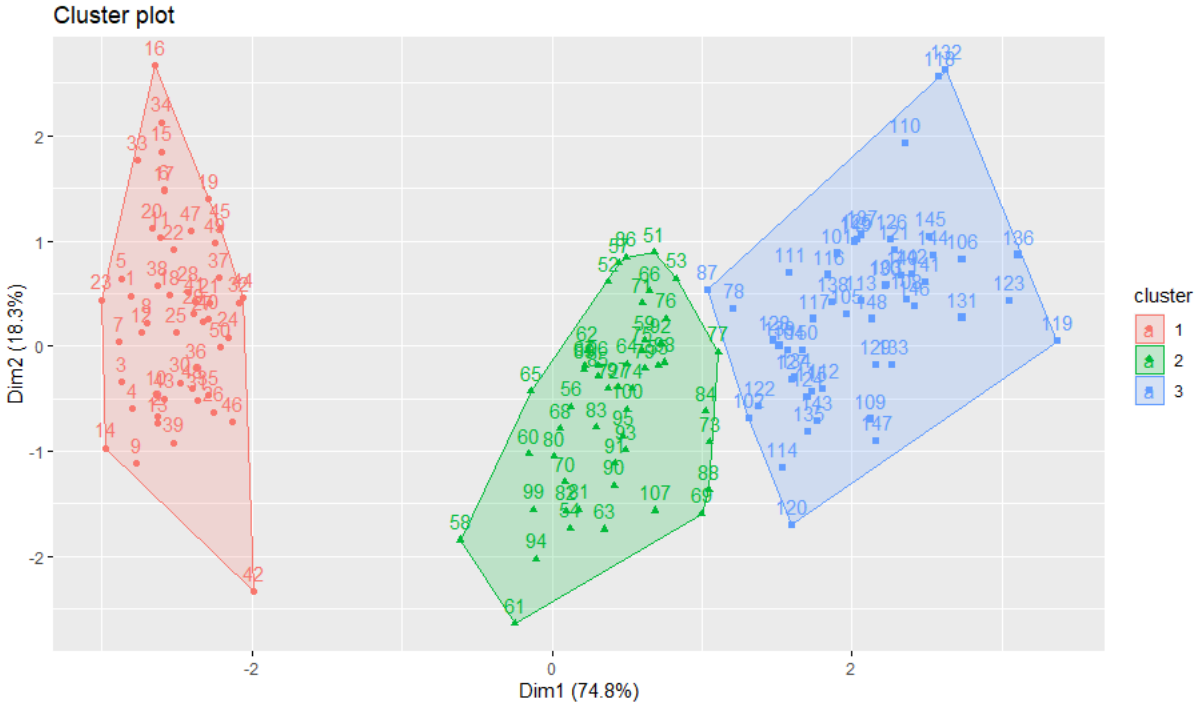
Case 1: Iris dataset



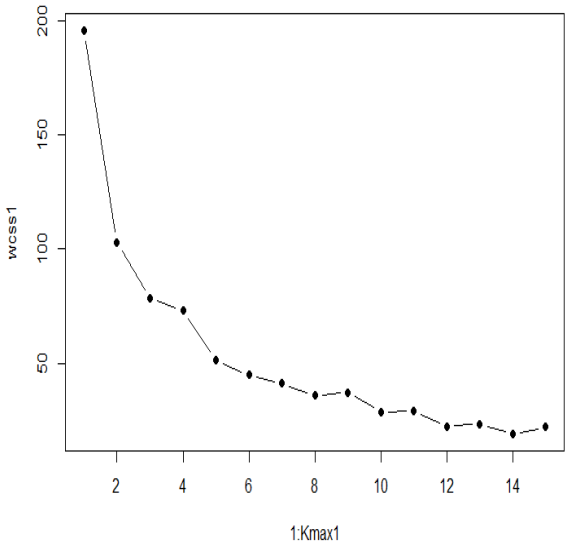
K-means centers:

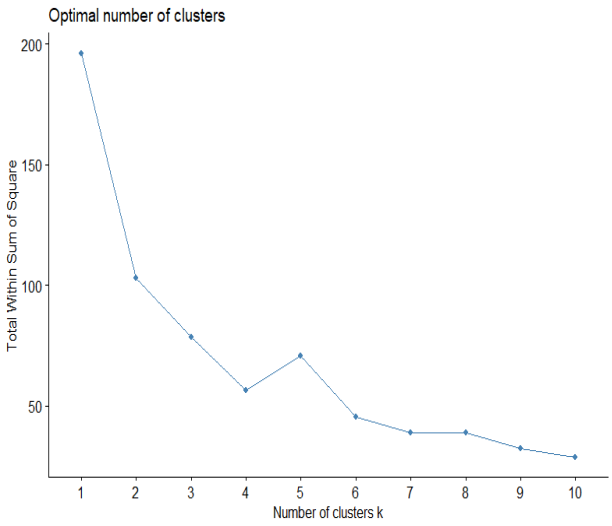


K-medoid centers:

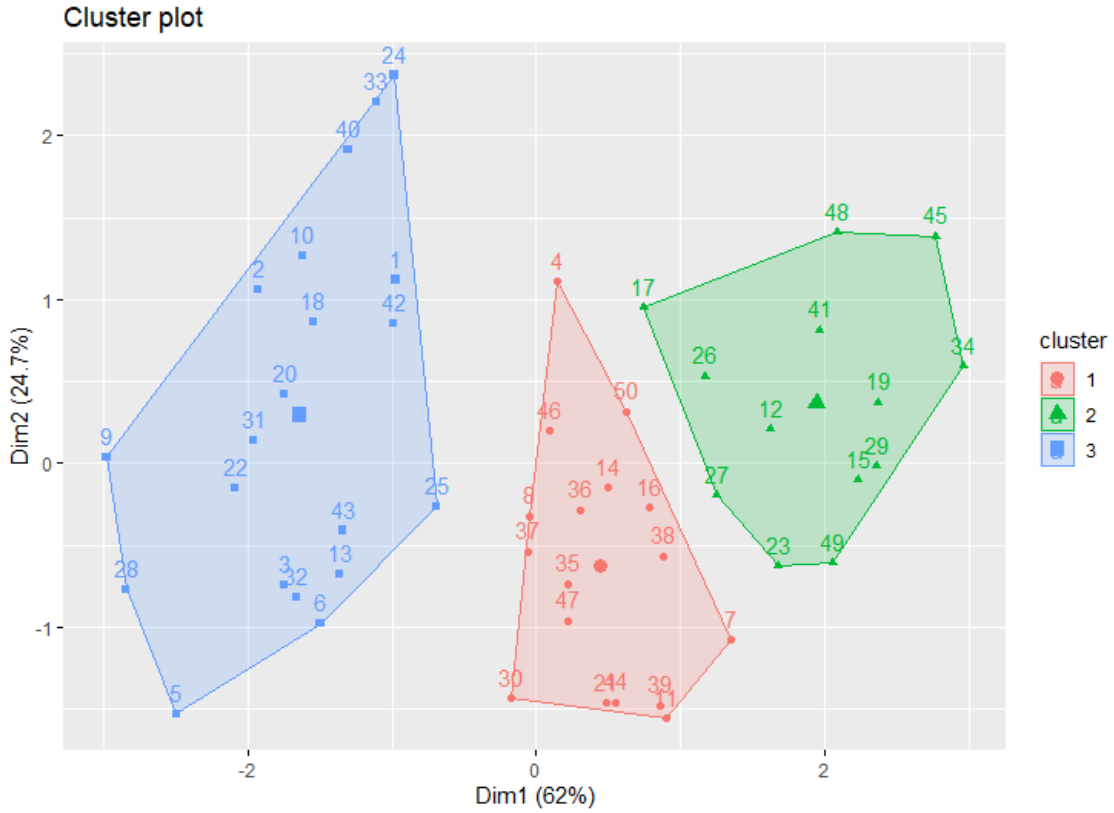


Case II: USArrests Dataset.

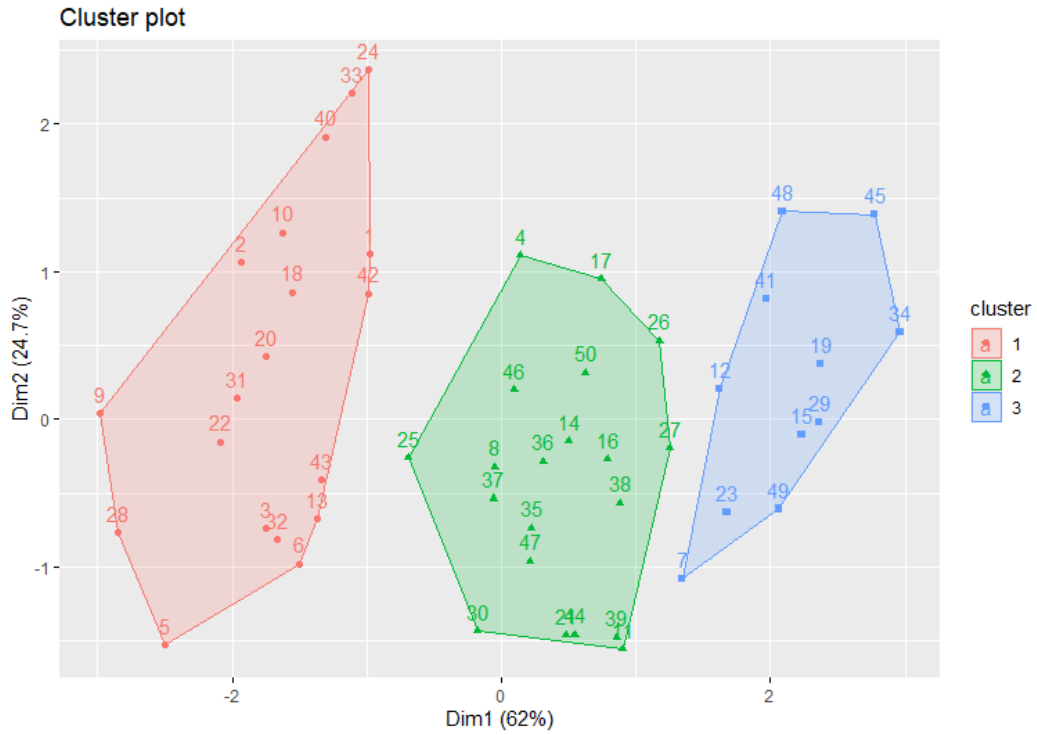




K-means centers:



K-medoid centers:



Statistics

Case 1: Iris dataset

K-means centers:

```
> fit$center
           X Sepal.Length Sepal.Width Petal.Length Petal.Width
1 -1.15087068  -1.01119138  0.85041372   -1.3006301  -1.2507035
2  0.07534946   0.03881135 -0.73324663    0.3059615   0.2137533
3  1.13936197   1.03196952 -0.07784286    1.0386287   1.0894947
```

K-medoid centers:

```
              X Sepal.Length Sepal.Width Petal.Length
[1,] -1.07030973   -0.7769106   0.7861738   -1.3357516
[2,] -0.08056095    0.3099591  -0.5903951    0.1370873
[3,]  0.95522266    0.7930124  -0.1315388    0.9868021
     Petal.Width
[1,]  -1.3110521
[2,]   0.1320673
[3,]   0.7880307
```

Case II: USArrests Dataset.

K-means centers:

```
> fit1$center
      Murder    Assault   UrbanPop       Rape
1 -0.4469795 -0.3465138  0.4788049 -0.2571398
2 -0.9615407 -1.1066010 -0.9301069 -0.9667633
3  1.0049340  1.0138274  0.1975853  0.8469650
```

K-medoid centers:

```
         Murder    Assault   UrbanPop       Rape
[1,]  0.8292944  1.3708088  0.3081225  1.1603196
[2,] -0.2727580 -0.2371077  0.1699510 -0.1315342
[3,] -1.2829727 -1.3770485 -0.5899924 -1.0603878
```

Program

Case 1: Iris dataset

```
rm(list=ls())
setwd("D:/6th Sem Works/A2- EDA/LAB/Lab7")
data1<-read.csv("D:/6th Sem Works/A2- EDA/LAB/Lab7/iris.csv")
View(data1)
df<-scale(data1)
fit<-kmeans(df,centers=2)
fit$cluster
fit$size
fit$withinss
fit$tot.withinss
Kmax<-15
wcss<-rep(NA,Kmax)
nClust<- list()
for(i in 1:Kmax){
  fit<-kmeans(df,i)
  wcss[i]<-fit$tot.withinss
  nClust[[i]]<-fit$size
}
plot(1:Kmax,wcss,type="b",pch=19)
fit<-kmeans(df,centers=3)
fit$cluster
fit$size
fit$center
library(factoextra)
fviz_nbclust(df, kmeans, method = "wss")
fviz_cluster(fit, data1)
library(cluster)
fitm <- pam(df, 3, metric = "manhattan")
fitm
fitm$medoids
fviz_cluster(fitm, data1)
```

Case II: USArrests Dataset.

```
rm(list=ls())
setwd("D:/6th Sem Works/A2- EDA/LAB/Lab7")
data2<-read.csv("D:/6th Sem Works/A2- EDA/LAB/Lab7/USArrests.csv")
view(data2)
data2<-data2[,-1]
df1<-scale(data2)
fit1<-kmeans(df1,centers=2)
fit1$cluster
fit1$size
fit1$withinss
fit1$tot.withinss
Kmax1<-15
wcss1<-rep(NA,Kmax1)
nClust1<- list()
for(i in 1:Kmax1){
  fit1<-kmeans(df1,i)
  wcss1[i]<-fit1$tot.withinss
  nClust1[[i]]<-fit1$size
}
plot(1:Kmax1,wcss1,type="b",pch=19)
fit1<-kmeans(df1,centers=3)
fit1$cluster
fit1$size
fit1$center
library(factoextra)
fviz_nbclust(df1, kmeans, method = "wss")
fviz_cluster(fit1, data2)
library(cluster)
fitm1 <- pam(df1, 3, metric = "manhattan")
fitm1
fitm1$medoids
fviz_cluster(fitm1, data2)
```