



### Experiment 2.3

**Student Name:** Atul Yadav

**UID:** 20BCS9425

**Branch:** CSE

**Section/Group:** 20BCS\_DM\_720/A

**Semester:** 6th

**Date of Performance:** 24/04/2023

**Subject Name:** Data Mining Lab

**Subject Code:** 20CSP-376

#### 1. Aim:

To perform the cluster analysis by the k-means method using R

#### 2. Objective:

Performing the cluster analysis by the k-means method using R.

#### 3. Script and Output:

```
# Loading data  
data(iris)
```

```
# Structure  
str(iris)  
# Installing Packages  
#install.packages("ClusterR")  
#install.packages("cluster")
```

```
# Loading package  
library(ClusterR)  
library(cluster)
```

```
# Removing initial label of  
# Species from original dataset  
iris_1 <- iris[, -5]
```

```
# Fitting K-Means clustering Model  
# to training dataset  
set.seed(240) # Setting seed
```

```
kmeans.re <- kmeans(iris_1, centers = 3, nstart = 20)
```

```
kmeans.re
#Cluster identification for
# each observation
kmeans.re$cluster

# Confusion Matrix
cm <- table(iris$Species, kmeans.re$cluster)
cm

# Model Evaluation and visualization
plot(iris_1[c("Sepal.Length", "Sepal.Width")])
plot(iris_1[c("Sepal.Length", "Sepal.Width")],
col = kmeans.re$cluster)
plot(iris_1[c("Sepal.Length", "Sepal.Width")],
col = kmeans.re$cluster,
main = "K-means with 3 clusters")

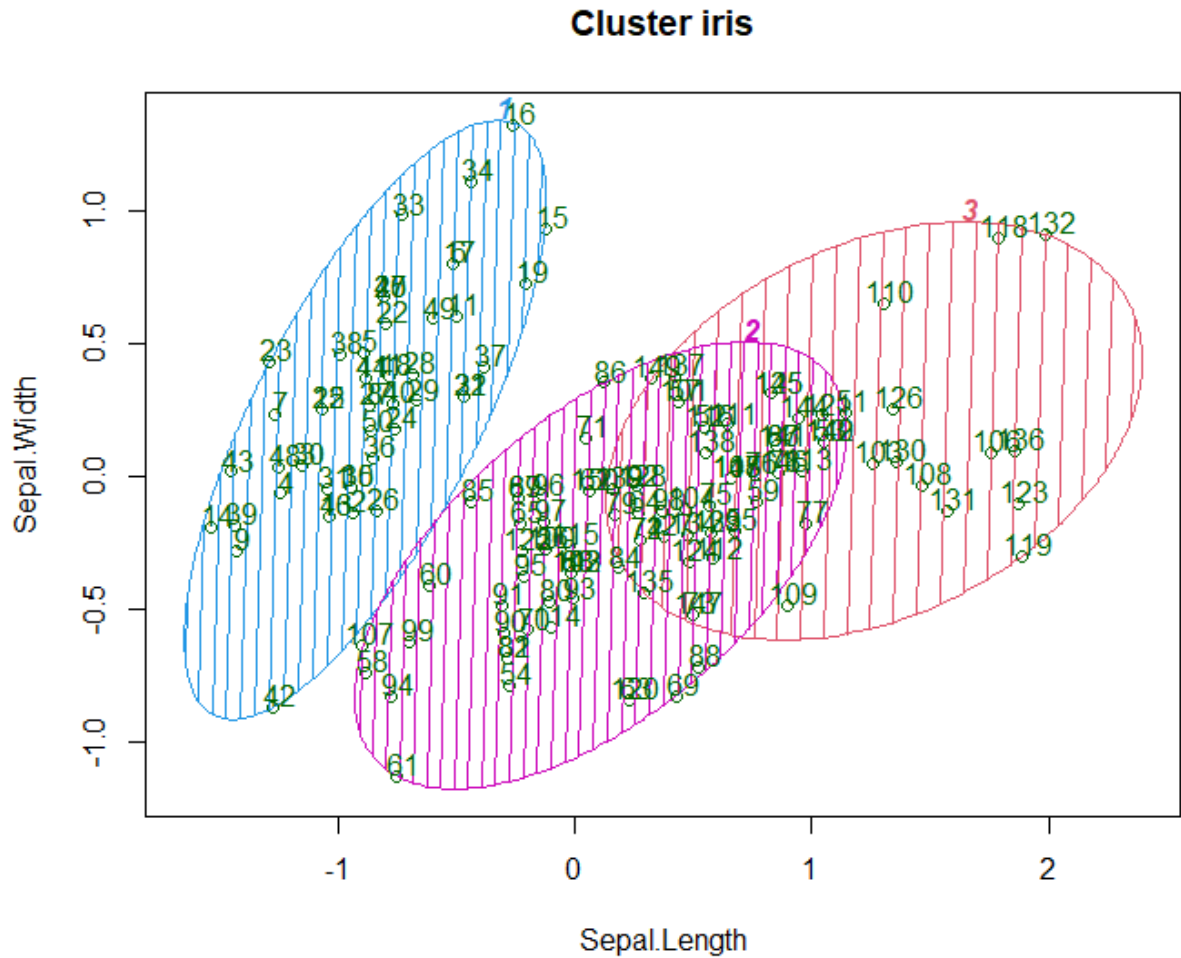
## Plotting cluster centers
kmeans.re$centers
kmeans.re$centers[, c("Sepal.Length", "Sepal.Width")]

# cex is font size, pch is symbol
points(kmeans.re$centers[, c("Sepal.Length", "Sepal.Width")],
col = 1:3, pch = 8, cex = 3)
## Visualizing clusters
y_kmeans <- kmeans.re$cluster
clusplot(iris_1[, c("Sepal.Length", "Sepal.Width")],
y_kmeans,
lines = 0,
shade = TRUE,
color = TRUE,
labels = 2,
plotchar = FALSE,
span = TRUE,
main = paste("Cluster iris"),
xlab = 'Sepal.Length',
ylab = 'Sepal.Width')
```

## 4. Dataset Used:

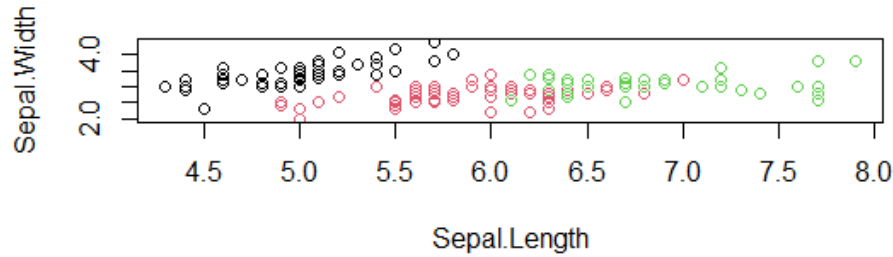
A1									
Id									
	A	B	C	D	E	F	G	H	I
1	Id	SepalLeng	SepalWidt	PetalLengt	PetalWidt	Species			
2	1	5.1	3.5	1.4	0.2	Iris-setosa			
3	2	4.9	3	1.4	0.2	Iris-setosa			
4	3	4.7	3.2	1.3	0.2	Iris-setosa			
5	4	4.6	3.1	1.5	0.2	Iris-setosa			
6	5	5	3.6	1.4	0.2	Iris-setosa			
7	6	5.4	3.9	1.7	0.4	Iris-setosa			
8	7	4.6	3.4	1.4	0.3	Iris-setosa			
9	8	5	3.4	1.5	0.2	Iris-setosa			
10	9	4.4	2.9	1.4	0.2	Iris-setosa			
11	10	4.9	3.1	1.5	0.1	Iris-setosa			
12	11	5.4	3.7	1.5	0.2	Iris-setosa			
13	12	4.8	3.4	1.6	0.2	Iris-setosa			
14	13	4.8	3	1.4	0.1	Iris-setosa			
15	14	4.3	3	1.1	0.1	Iris-setosa			
16	15	5.8	4	1.2	0.2	Iris-setosa			
17	16	5.7	4.4	1.5	0.4	Iris-setosa			
18	17	5.4	3.9	1.3	0.4	Iris-setosa			
19	18	5.1	3.5	1.4	0.3	Iris-setosa			
20	19	5.7	3.8	1.7	0.3	Iris-setosa			
21	20	5.1	3.8	1.5	0.3	Iris-setosa			
22	21	5.4	3.4	1.7	0.2	Iris-setosa			
23	22	5.1	3.7	1.5	0.4	Iris-setosa			
24	23	4.6	3.6	1	0.2	Iris-setosa			
25	24	5.1	3.3	1.7	0.5	Iris-setosa			

## 5. Output:



These two components explain 100 % of the point variability.

## K-means with 3 clusters



Project: (None)

Environment History Connections Tutorial

Import Dataset 279 MiB

List

R Global Environment

m1

List of 6

PlantGrowth

30 obs. of 2 variables

test\_set

100 obs. of 3 variables

training\_set

300 obs. of 3 variables

Values

a

2

accuracy\_test

0.88

accuracy\_Test

0.933333333333333

accuracy\_train

0.886666666666667

b

6

cm

'table' int [1:3, 1:3] 50 0 0 0 48 14 0 2 36

cm\_test

'table' int [1:2, 1:2] 60 8 4 28

cm\_train

'table' int [1:2, 1:2] 180 21 13 86

dataTestPred

Factor w/ 3 levels "setosa","versicolor",...: 1 1 1 1 1 1 ...

H

num [1:7] 7 15 25 35 5 45 20

M

chr [1:7] "Marc" "Apr" "May" "Jun" "Nov" "Oct" "Aug"

predictions

Factor w/ 3 levels "ctrl","trtl",...: 2 3 3 3 2 2 3 2 3 3 ...

s

1963.49540849362

sp1

logi [1:5] FALSE TRUE TRUE FALSE TRUE

split

logi [1:400] TRUE TRUE TRUE FALSE TRUE TRUE ...

table\_matrix

'table' int [1:3, 1:3] 18 0 0 2 18 0 0 2 20

v

56.5486677646163

X

int [1:10] 1 2 3 4 5 6 7 8 9 10

Y

num [1:10] 1 4 9 16 25 36 49 64 81 100

y\_kmeans

int [1:150] 1 1 1 1 1 1 1 1 1 1 ...

y\_pred\_test

Factor w/ 2 levels "0","1": 1 2 1 1 1 1 1 2 2 1 ...

y\_pred\_train

Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...

Functions

volcylinder

function (dia = 5, len = 8)

Files Plots Packages Help Viewer Presentation

Zoom Export

Publish

figure margins too large

ENG INTL

22:13 27-04-2023