## **Exp:9**

## <u>Implement clustering techniques – Hierarchical and K-Means</u>

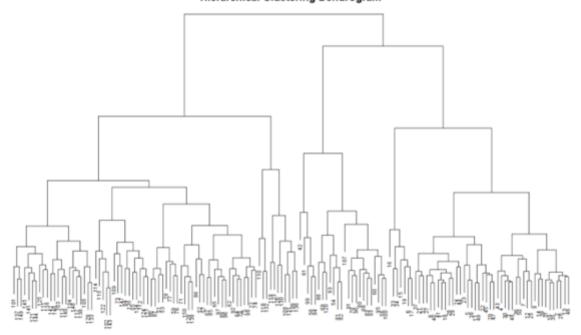
# a) HIERARCHIAL CLUSTERING

```
# Load the iris dataset
data(iris)
# Use only the numeric columns for clustering (exclude the Species column)
iris data <- iris[, -5]
# Standardize the data
iris_scaled <- scale(iris_data)</pre>
# Compute the distance matrix
distance matrix <- dist(iris scaled, method = "euclidean")
# Perform hierarchical clustering using the "complete" linkage method
hc complete <- hclust(distance matrix, method = "complete")
# Plot the dendrogram
plot(hc_complete, main = "Hierarchical Clustering Dendrogram", xlab = "", sub = "", cex =
0.6)
# Cut the tree to form 3 clusters
clusters <- cutree(hc_complete, k = 3)
# Print the cluster memberships
print(clusters)
# Add the clusters to the original dataset
iris$Cluster <- as.factor(clusters)</pre>
# Display the first few rows of the updated dataset
head(iris)
```

```
# Perform nierarchical clustering using the "complete" linkage method
hc_complete <- hclust(distance_matrix, method = "complete")
# Plot the dendrogram
plot(hc_complete, main = "Hierarchical Clustering Dendrogram", xlab = "", sub = "", cex =</pre>
               0.6)
 # Cut the tree to form 3 clusters
 clusters <- cutree(hc_complete, k = 3)
# Print the cluster memberships
 [51]
# Add the clusters to the original dataset iris$Cluster <- as.factor(clusters)
# Display the first few rows of the updated dataset head(iris)
 Sepal.Length Sepal.Width Petal.Length Petal.Width Species Cluster
                                                                    1.4
                                                                                             0.2
                  5.1
4.9
4.7
4.6
                                   3.5
                                                                                                      setosa
                                                                                                       setosa
                                                                                                                                 1
                                                                                                       setosa
                                                                    1.5
                                           3.1
                                                                                             0.2 setosa
                                                                                                                                 1
                   5.0
                                                                                                       setosa
                   5.4
                                          3.9
                                                                                             0.4 setosa
                                                                                                                                 1
                                                          7 obs. of 2 variables
List of 7
150 obs. of 6 variables
150 obs. of 4 variables
num [1:150, 1:4] -0.898 -1.139 -1.381 -1.501 -1.018 ...
List of 12
List of 30
32 obs. of 11 variables
List of 31
45 obs. of 5 variables
105 obs. of 5 variables
List of 14
O data
O hc_complete
O iris
                                                                                                                                                                                                                 o o
O iris data
iris_scaled
Olinear_model
Ologistic_model
 mtcars
                                                                                                                                                                                                                  Q,
 osvm_model
test_data
train_data
tree_model
Values
    accuracy
clusters
                                                           0.9777777777778
int [1:150] 1 1 1 1 1 1 1 1 1 1 ...
'table' int [1:3, 1:3] 14 0 0 0 18 0 0 1 12
'dist' num [1:11175] 1.172 0.843 1.1 0.259 1.035 ...
num [1:7] 150 160 165 170 175 180 185
Named num [1:32] 0.461 0.461 0.598 0.492 0.297 ...
Factor w/ 3 levels "setosa", "versicolor", ... 1 1 1 1 1 1 1 1 1 ...
int [1:105] 14 50 118 43 150 148 90 91 143 92 ...
num [1:7] 55 60 62 68 70 75 80
```

#### **Hierarchical Clustering Dendrogram**

confusion\_matrix distance\_matrix distance\_matrix heights predicted\_probs predictions sample\_indices weights



## **b) K-MEANS CLUSTERING**

```
# Load the iris dataset
data(iris)
# Use only the numeric columns for clustering (exclude the Species column)
iris_data <- iris[, -5]
# Standardize the data
iris_scaled <- scale(iris_data)</pre>
# Set the number of clusters
set.seed(123) #For reproducibility
k <- 3 # Number of clusters
# Perform K-Means clustering
kmeans result <- kmeans(iris scaled, centers = k, nstart = 25)
# Print the K-Means result
print(kmeans_result)
# Print the cluster centers
print(kmeans_result$centers)
# Add the cluster assignments to the original dataset
iris$Cluster <- as.factor(kmeans_result$cluster)</pre>
# Display the first few rows of the updated dataset
head(iris)
# Plot the clusters
library(ggplot2)
ggplot(iris, aes(x = Sepal.Length, y = Sepal.Width, color = Cluster)) +
 geom_point(size = 3) +
 labs(title = "K-Means Clustering of Iris Dataset", x = "Sepal Length", y = "Sepal Width")
```

### Output

```
install.packages("ggplot2")

# Load the iris dataset

data(iris)

# Use only the numeric columns for clustering (exclude the Species column)

iris_data <- iris[, -5]

# Standardize the data

iris_scaled <- scale(iris_data)

# Set the number of clusters

set.seed(123) # For reproducibility

k <- 3 # Number of clusters

set.seed(123) # For reproducibility

k <- 3 * Number of clusters

print k-Means clustering

kmeans_result <- kmeans(iris_scaled, centers = k, nstart = 25)

# Print the K-Means result

print(kmeans_result)

# Print the cluster centers

print(kmeans_resultScenters)

# Add the cluster assignments to the original dataset in iris$Cluster <- as.factor(kmeans_resultScluster)

# Display the first few rows of the updated dataset head(iris)

# Plot the clusters

ibrary(ggplot2)

ggplot(iris, aes(x = Sepal.Length, y = Sepal.Width, color = Cluster)) +

geom_point(size = 3) +

labs(title = "K-Means Clustering of Iris Dataset", x = "Sepal Length", y = "Sepal Width"
```

```
R 4.4.1 · ~/ AVAITABLE COMPONENTS.
[1] "cluster"
                    "centers"
                                    "totss"
                                                     "withinss"
                                                                     "tot.withinss"
[6] "betweenss"
                   "size"
                                                    "ifault"
                                    "iter"
> # Print the cluster centers
> print(kmeans_result$centers)
  Sepal.Length Sepal.Width Petal.Length Petal.Width
1 -1.01119138 0.85041372 -1.3006301 -1.2507035
  -0.05005221 -0.88042696
1.13217737 0.08812645
                             0.3465767 0.2805873
0.9928284 1.0141287
                                             0.2805873
> # Add the cluster assignments to the original dataset
> iris$Cluster <- as.factor(kmeans_result$cluster)</pre>
> # Display the first few rows of the updated dataset > head(iris)
  Sepal.Length Sepal.Width Petal.Length Petal.Width Species Cluster
                                             0.2 setosa
1
           5.1
                        3.5
                                      1.4
                                                                      1
2
           4.9
                        3.0
                                      1.4
                                                   0.2 setosa
3
            4.7
                         3.2
                                      1.3
                                                   0.2
                                                        setosa
4
            4.6
                         3.1
                                      1.5
                                                   0.2 setosa
           5.0
                         3.6
                                      1.4
                                                   0.2
                                                       setosa
                                                                       1
6
           5.4
                        3.9
                                      1.7
                                                   0.4 setosa
                                                                       1
> # Plot the clusters
> library(ggplot2)
> ggplot(iris, aes(x = Sepal.Length, y = Sepal.Width, color = Cluster)) +
    geom_point(size = 3) +
    labs(title = "K-Means Clustering of Iris Dataset", x = "Sepal Length", y = "Sepal Width")
  Files Plots Packages Help Viewer Presentation
                                                                                              -0
  🧢 🗼 🔑 Zoom 😕 Export 🕶 🝳
                                                                                      🗫 Publish 💌 🎯
       K-Means Clustering of Iris Dataset
    4.5 -1
    4.0 -
    3.5
                                                                                            Cluster
  Sepal Width
                                                                                             • 1
                                                                                             • 2
    3.0
    2.5 -
```

Sepal Length

2.0