

## **Implement clustering techniques – Hierarchical and K-Means**

### **AIM:**

To implement Hierarchical and K-Means clustering techniques using R

### **Hierarchical Clustering**

#### **PROGRAM:**

```
# 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)
```

```
# 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
```

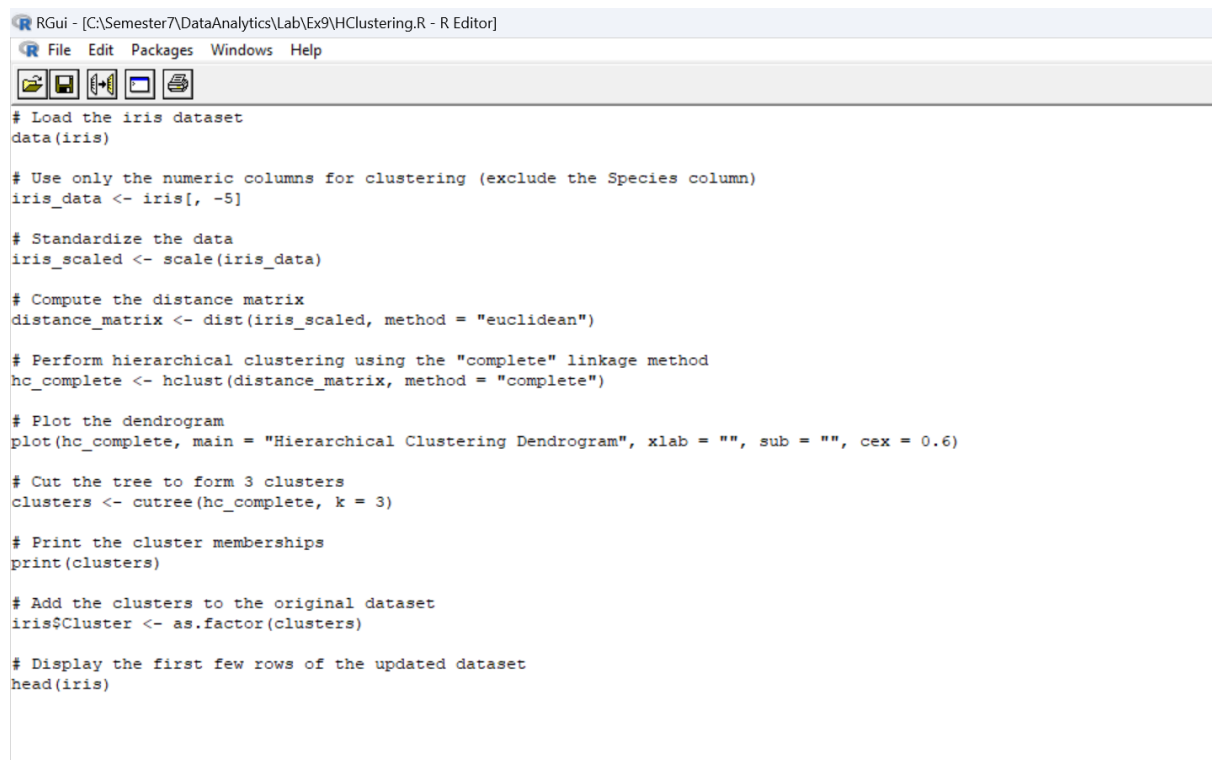
```
print(clusters)
```

```
# Add the clusters to the original dataset
```

```
iris$Cluster <- as.factor(clusters)
```

```
# Display the first few rows of the updated dataset
```

```
head(iris)
```



The screenshot shows an RGui window titled "RGui - [C:\Semester7\DataAnalytics\Lab\Ex9\HClustering.R - R Editor]". The window has a menu bar with "File", "Edit", "Packages", "Windows", and "Help". Below the menu bar is a toolbar with icons for file operations and execution. The main text area contains the following R code:

```
# 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)

# 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")

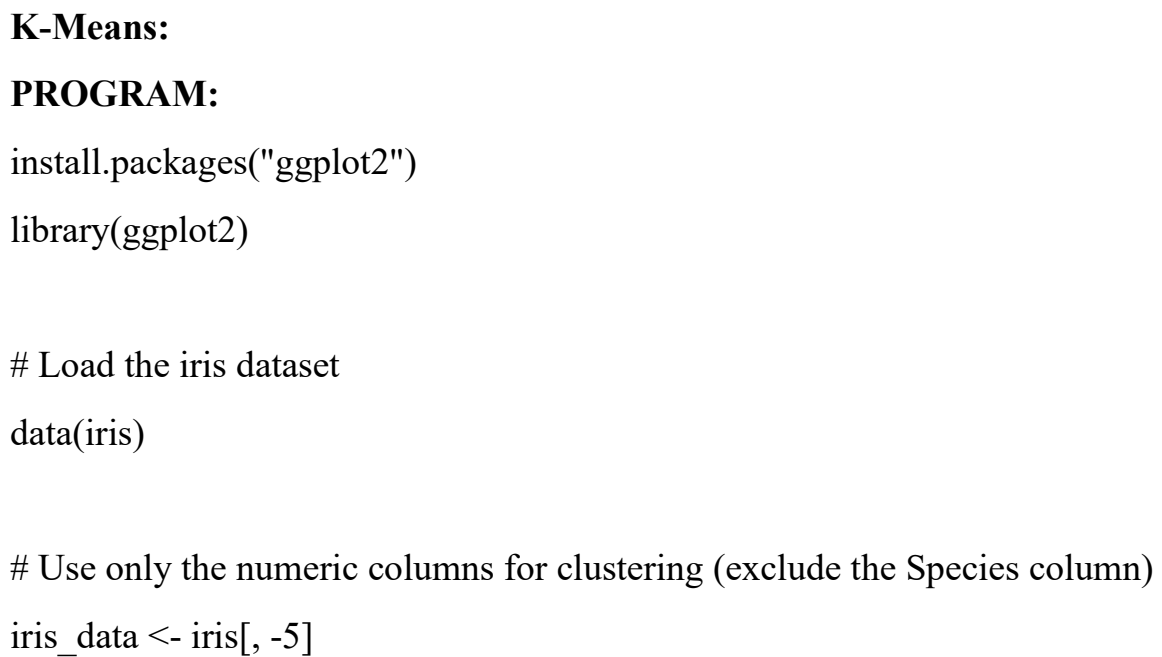
# Plot the dendrogram
plot(hc_complete, main = "Hierarchical Clustering Dendrogram", xlab = "", sub = "", cex = 0.6)

# 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)

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

[illegible]

```
# Standardize the data
iris_scaled <- scale(iris_data)

# 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)

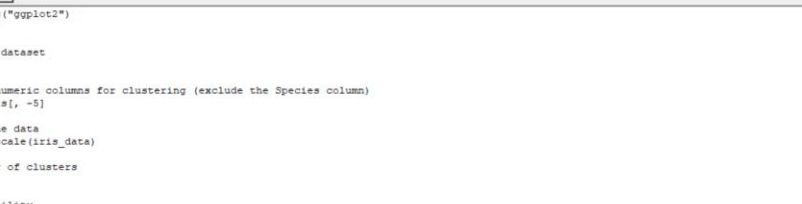
# Print the cluster centers
print(kmeans_result$centers)

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

# 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")
```



The screenshot shows the RStudio interface with the following components:

- Top Panel:** The title bar reads "RGui - [C:\Semester7\DataAnalytics\Lab\Ex9\KMeans.R - R Editor]". The menu bar includes "File", "Edit", "Packages", "Windows", and "Help". Below the menu is a toolbar with icons for file operations and running code.
- Source Panel:** Contains the R script code for K-Means clustering and visualization.
 

```
install.packages("ggplot2")
library(ggplot2)

# Load the iris dataset
data(iris)

# Use only the numeric columns for clustering (exclude the Species column)
iris_data <- iris[, 1:5]

# Standardize the data
iris_scaled <- scale(iris_data)

# Set the number of clusters
set.seed(123)

# For reproducibility
c <- 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)

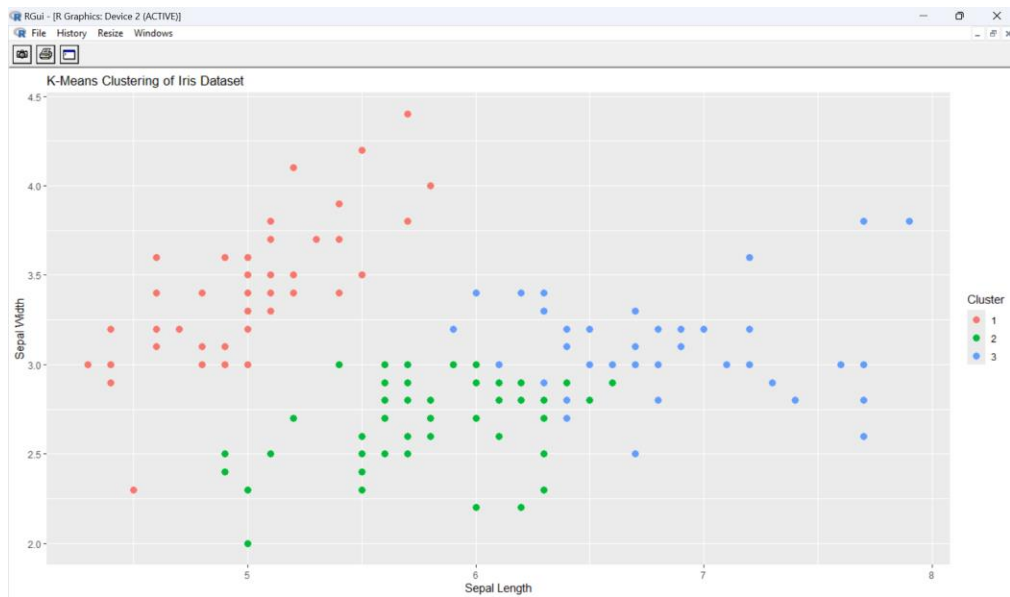
# Add the cluster assignments to the original dataset
iris$Cluster <- as.factor(kmeans_result$cluster)

# 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:**

[illegible]



## RESULT:

Thus, the hierarchical clustering and K-Means algorithm is implemented successfully using R.