**PRACTICAL 10**

**Aim:** Implementation and analysis of Clustering algorithms like K-Means, Agglomerative Assessment.

**Theory:**

**K-MEANS CLUSTERING:**

. K-Means Clustering is an Unsupervised Learning algorithm, which groups the unlabelled dataset into different clusters. Here K defines the number of predefined clusters that need to be created in the process, as if K=2, there will be two clusters, and for K=3, there will be three clusters, and so on.

. It allows us to cluster the data into different groups and a convenient way to discover the categories of groups in the unlabelled dataset on its own without the need for any training.

. It is a centroid-based algorithm, where each cluster is associated with a centroid. The main aim of this algorithm is to minimize the sum of distances between the data point and their corresponding clusters.

. The algorithm takes the unlabelled dataset as input, divides the dataset into k number of clusters, and repeats the process until it does not find the best clusters. The value of k should be predetermined in this algorithm.

**.** The k-means clustering algorithm mainly performs two tasks:

a. Determines the best value for K center points or centroids by an iterative process.

b. Assigns each data point to its closest k-center. Those data points which are near to the particular k-center, create a cluster.

Hence each cluster has data points with some commonalities, and it is away from other clusters.

The working of the K-Means algorithm is explained in the below steps:

○ Step-1: Select the number K to decide the number of clusters.

○ Step-2: Select random K points or centroids. (It can be other from the input dataset).

○ Step-3: Assign each data point to their closest centroid, which will form the predefined K clusters.

○ Step-4: Calculate the variance and place a new centroid of each cluster.

○ Step-5: Repeat the third steps, which means assign each datapoint to the new closest centroid of

each cluster.

○ Step-6: If any reassignment occurs, then go to step-4 else go to FINISH.

○ Step-7: The model is ready.

**AGGLOMERATIVE CLUSTERING:**

. The agglomerative clustering is the most common type of hierarchical clustering used to group

objects in clusters based on their similarity. It’s also known as AGNES (Agglomerative Nesting).

The algorithm starts by treating each object as a singleton cluster. Next, pairs of clusters are

successively merged until all clusters have been merged into one big cluster containing all objects. The result is a tree-based representation of the objects, named dendrogram.

. Agglomerative clustering works in a “bottom-up” manner. That is, each object is initially considered as a single-element cluster (leaf).

. At each step of the algorithm, the two clusters that are the most similar are combined into a new bigger cluster (nodes).

. This procedure is iterated until all points are member of just one single big cluster (root)

Steps to agglomerative hierarchical clustering:

We’ll follow the steps below to perform agglomerative hierarchical clustering using R software:

○ Preparing the data

○ Computing (dis)similarity information between every pair of objects in the data set.

○ Using linkage function to group objects into hierarchical cluster tree, based on the distance

information generated at step 1. Objects/clusters that are in close proximity are linked together

using the linkage function.

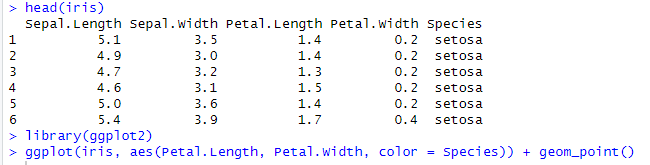
○ Determining where to cut the hierarchical tree into clusters. This creates a partition of the data.

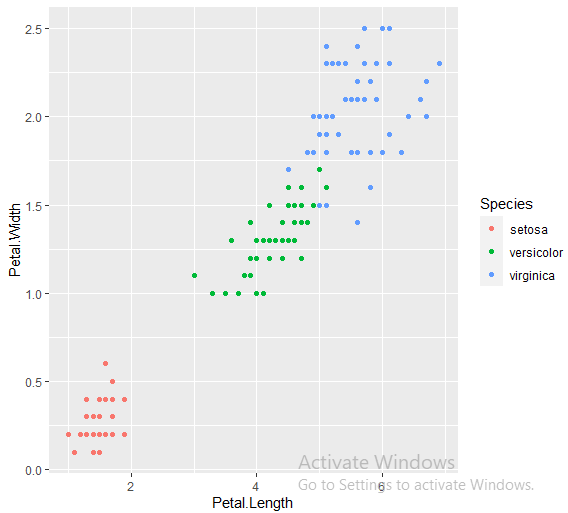
Implementation and analysis of clustering algorithms like

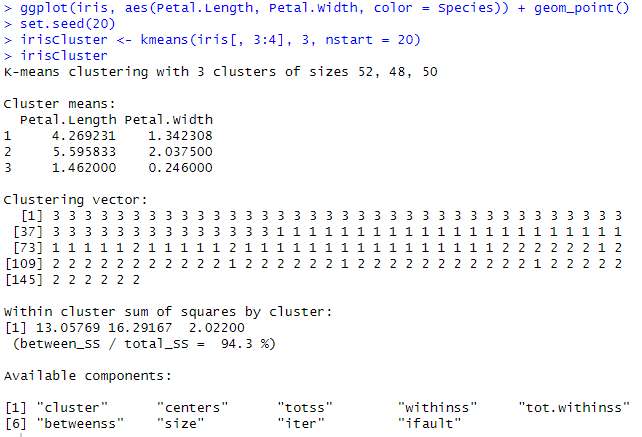
1. K-Means

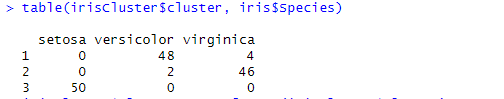
2. Agglomerative

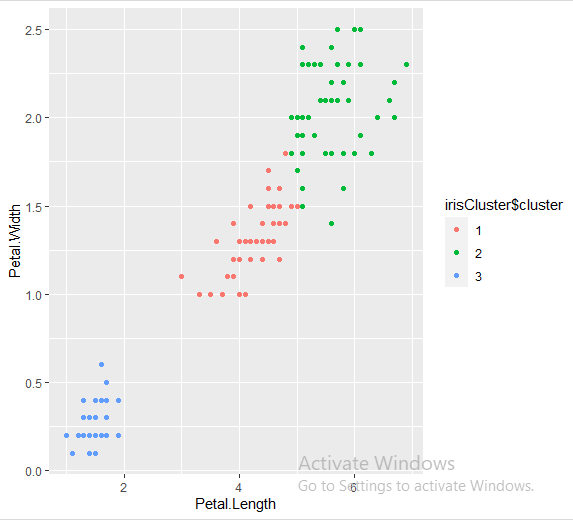
1. K-Means



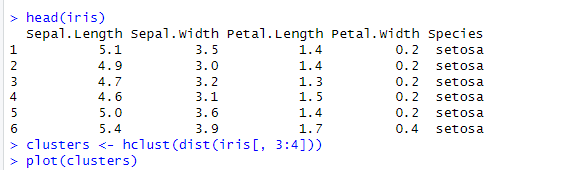


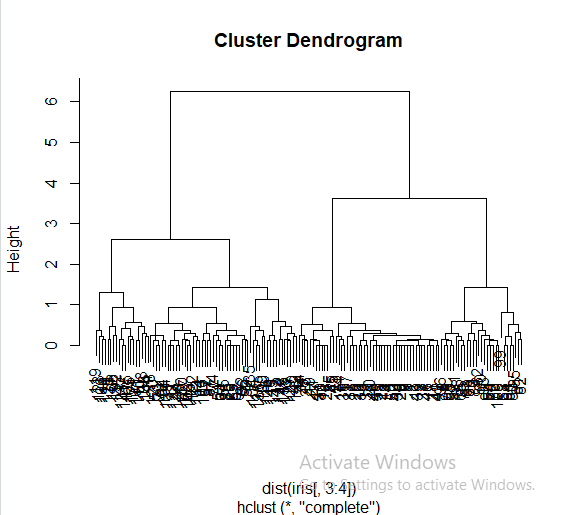






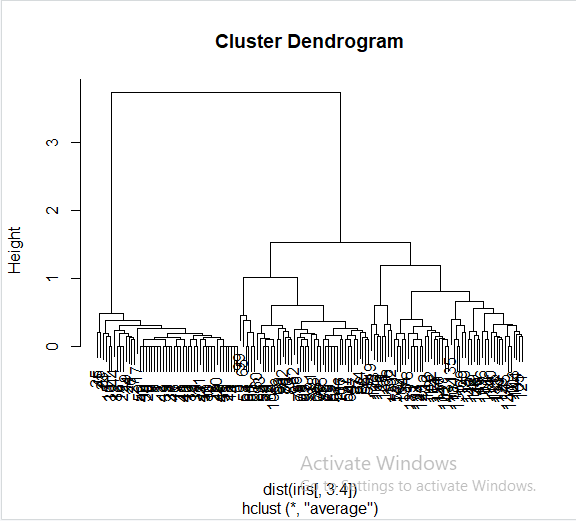
**#Agglomerative Clustering**

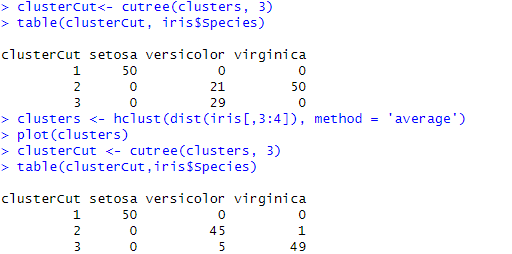




Text

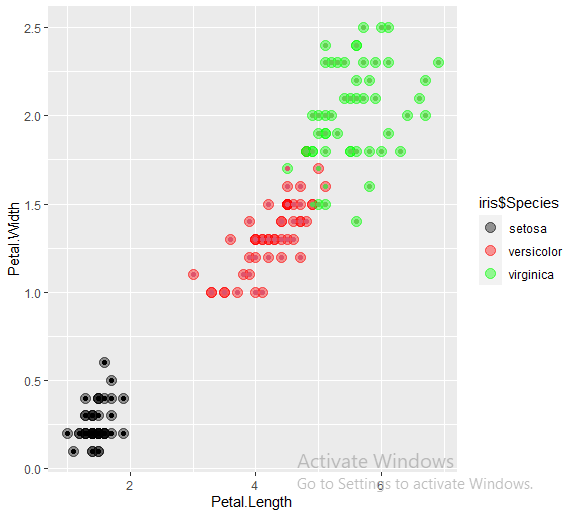
Description automatically generated





Icon

Description automatically generated



**Conclusion:** Successfully implemented k-means clustering and agglomerative clustering algorithms.