

Experiment No. 10

Title: To Implement K means clustering algorithm in R

Problem:

The iris dataset contains data about sepal length, sepal width, petal length, and petal width of flowers of different species. K means algorithm is applied to this dataset in order to classify the flowers into 3 species.

Following are the steps to be followed:

1. Load and view dataset

```
require("datasets")  
data("iris") # load Iris Dataset  
str(iris) #view structure of dataset  
  
summary(iris) #view statistical summary of the dataset  
  
head(iris) #view top rows of the dataset
```

2. Preprocess the dataset: Since clustering is a type of Unsupervised Learning, we would not require Class Label(output) during the execution of our algorithm. We will, therefore, remove Class Attribute "Species" and store it in another variable.

```
iris.new<- iris[,c(1,2,3,4)]  
iris.class<- iris[, "Species"]  
head(iris.new)  
  
head(iris.class)  
  
normalize <- function(x){  
  return ((x-min(x))/(max(x)-min(x)))
```

```
}
```

```
iris.new$Sepal.Length<- normalize(iris.new$Sepal.Length)
iris.new$Sepal.Width<- normalize(iris.new$Sepal.Width)
iris.new$Petal.Length<- normalize(iris.new$Petal.Length)
iris.new$Petal.Width<- normalize(iris.new$Petal.Width)
head(iris.new)
```

3. Apply the k-means clustering algorithm.

```
result<- kmeans(iris.new,3) #apply k-means algorithm with no. of
centroids(k)=3
result$size # gives no. of records in each cluster

result$centers # gives the value of cluster center data point value(3
centers for k=3)
```

```
result$cluster #gives cluster vector showing the cluster where each record
falls
```

4. Verify the results of clustering

```
par(mfrow=c(2,2), mar=c(5,4,2,2))
plot(iris.new[c(1,2)], col=result$cluster)# Plot to see how Sepal.Length and
Sepal.Width data points have been distributed in clusters
plot(iris.new[c(1,2)], col=iris.class)# Plot to see how Sepal.Length and
Sepal.Width data points have been distributed originally as per "class"
attribute in dataset
plot(iris.new[c(3,4)], col=result$cluster)# Plot to see how Petal.Length and
Petal.Width data points have been distributed in clusters
plot(iris.new[c(3,4)], col=iris.class)

table(result$cluster,iris.class)
```



