10.R

AVShrez

2024-05-31

```r
myData <- iris[,-5]


# K-means clusters with k=2 and k=3

# # with K = 3
km.outTwo <- kmeans(myData, 2, nstart = 20)

# # with K = 3
km.outThree <- kmeans(myData, 3, nstart = 20)

km.outThree
```

```
## K-means clustering with 3 clusters of sizes 62, 38, 50
##
## Cluster means:
##   Sepal.Length Sepal.Width Petal.Length Petal.Width
## 1     5.901613    2.748387     4.393548    1.433871
## 2     6.850000    3.073684     5.742105    2.071053
## 3     5.006000    3.428000     1.462000    0.246000
##
## Clustering vector:
##   [1] 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3
##  [38] 3 3 3 3 3 3 3 3 3 3 3 3 3 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
##  [75] 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 2 2 2 2 1 2 2 2 2
## [112] 2 2 1 1 2 2 2 2 1 2 1 2 1 2 2 1 1 2 2 2 2 2 1 2 2 2 2 1 2 2 2 1 2 2 2 1 2
## [149] 2 1
##
## Within cluster sum of squares by cluster:
## [1] 39.82097 23.87947 15.15100
##  (between_SS / total_SS =  88.4 %)
##
## Available components:
##
## [1] "cluster"      "centers"      "totss"        "withinss"     "tot.withinss"
## [6] "betweenss"    "size"         "iter"         "ifault"
```

```r
# Plotting the clusters with k =3
plot(myData[c("Sepal.Length","Sepal.Width")],col = km.outThree$cluster, main = "K-means with 3 clusters"

# # Interpretration
# After clustering by taking centers, we've used two variables here (Sepal Length and Sepal Width) to
```

```
# plot the graph of the data myData which is the data taking the first four variables of the iris data.
# We can clearly see three different clusters being formed here:
# The first cluster has sepal width on the higher end of the spectrum of the data and sepal length on t
# The second cluster has sepal width on the mid to the lower end of the spectrum while the sepal length
# The thrid cluster has sepal width on the middle of the spectrum while the sepal length is clearly on

# Cluster centers for plot of k = 3
km.outThree$centers
```

```
##    Sepal.Length Sepal.Width Petal.Length Petal.Width
## 1      5.901613    2.748387     4.393548    1.433871
## 2      6.850000    3.073684     5.742105    2.071053
## 3      5.006000    3.428000     1.462000    0.246000
```

```
points(km.outThree$centers[,c("Sepal.Length","Sepal.Width")],col = 1:3, pch = 8, cex = 3)
```

K−means with 3 clusters
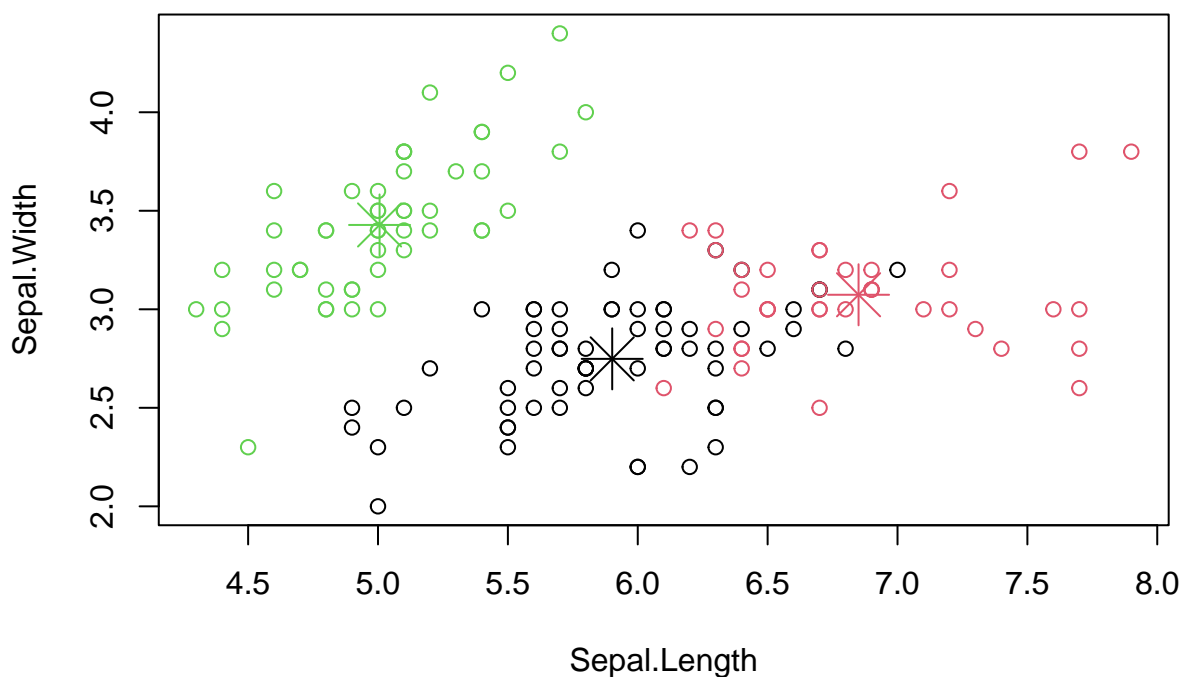

```
# # Interpretation
# We can see that the first cluster has the center (5.06, 3.42)
# We can see that the second cluster has the center (5.9, 2.7)
# We can see that the third cluster has the center (6.8, 2.07)


# Confusion Matrix to compare with Species variable
cmThree <- table(iris$Species, km.outThree$cluster)
cmThree
```

```
##
##               1  2  3
##   setosa      0  0 50
##   versicolor 48  2  0
##   virginica  14 36  0
```

```r
#Accuracy
(accuracy <- sum(diag(cmThree))/sum(cmThree))
```

```
## [1] 0.01333333
```

```r
(mce <- 1 - accuracy)
```

```
## [1] 0.9866667
```