Qno_10.R

arpan

2023-07-27

```r
data(iris)

# Select first four variables
iris_data = iris[,1:4]

head(iris_data)
```

```
##   Sepal.Length Sepal.Width Petal.Length Petal.Width
## 1          5.1         3.5          1.4         0.2
## 2          4.9         3.0          1.4         0.2
## 3          4.7         3.2          1.3         0.2
## 4          4.6         3.1          1.5         0.2
## 5          5.0         3.6          1.4         0.2
## 6          5.4         3.9          1.7         0.4
```
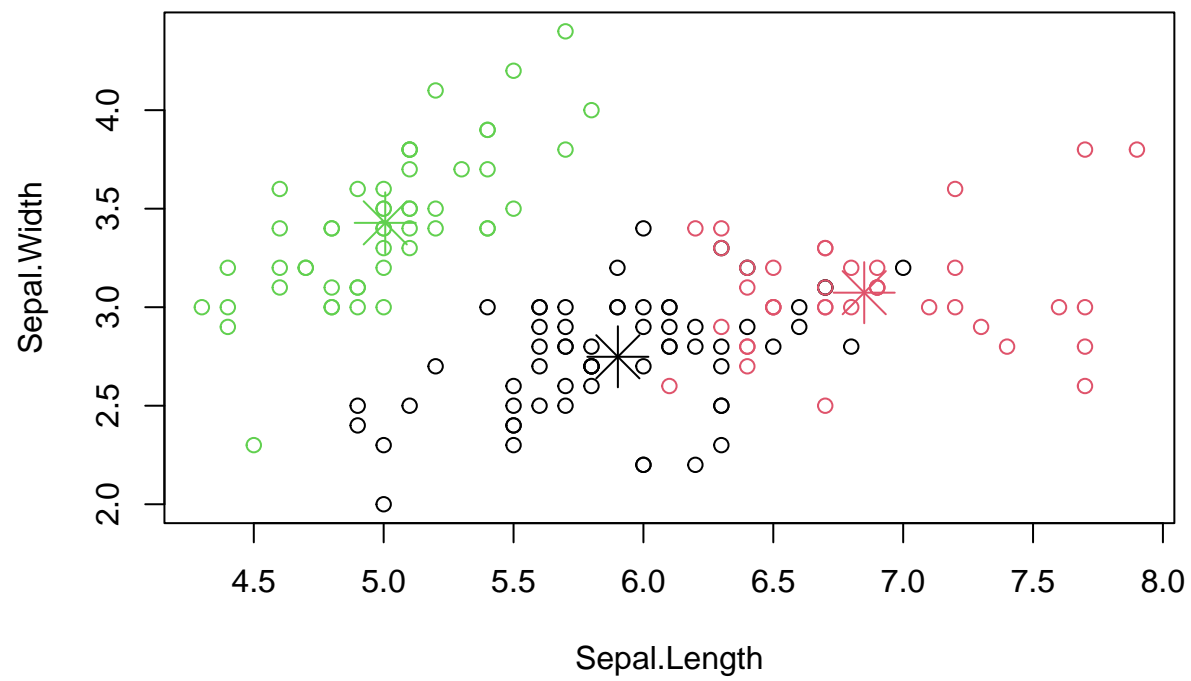
```r
#a
#  k=2
k2_model = kmeans(iris_data, centers = 2)

# k=3
k3_model = kmeans(iris_data, centers = 3)

#b
# Plot of clusters formed with k=3
plot(iris_data[,1:2], col = k3_model$cluster)

# The plot of iris data with first four variable shows that the are three clusters in
#  the data.


#c
# Add cluster centers
points(k3_model$centers, col = 1:3, pch = 8, cex = 3)
```
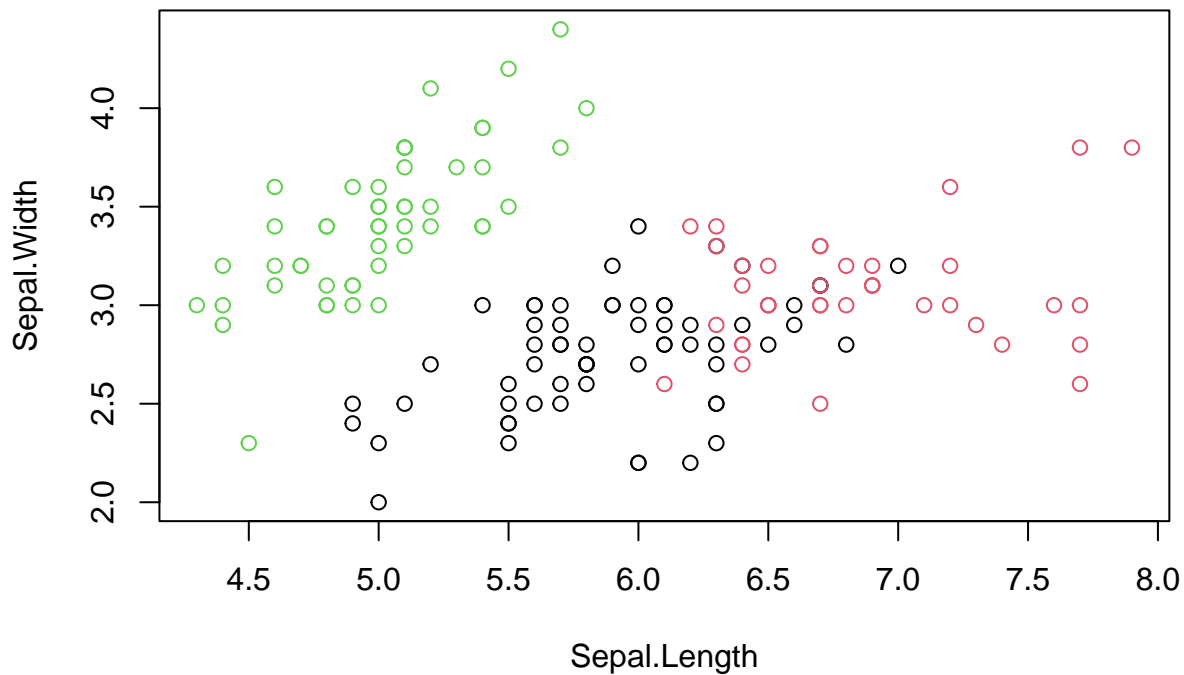
```
plot(iris_data[,1:2], col = k3_model$cluster)
```

```
# Here we add the centers for the plot of clusters formed with k=3 and plot the cluster again.



#d
cm = table(iris$Species, k3_model$cluster)

#confusion matrix
print(cm)
```

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

```
# Interpretation:
# The confusion matrix shows that the k=3 cluster variable is able to correctly classify 95% of the obs
# The misclassifications are all for some of the species, which are being classified as another species
# This is likely because the misclassified species is intermediate between the two species,
# and the k=3 cluster variable is not able to distinguish between them.
```