Problem 6.4

Chen Bo Calvin Zhang

09/11/2020

Import and preprocess the data.

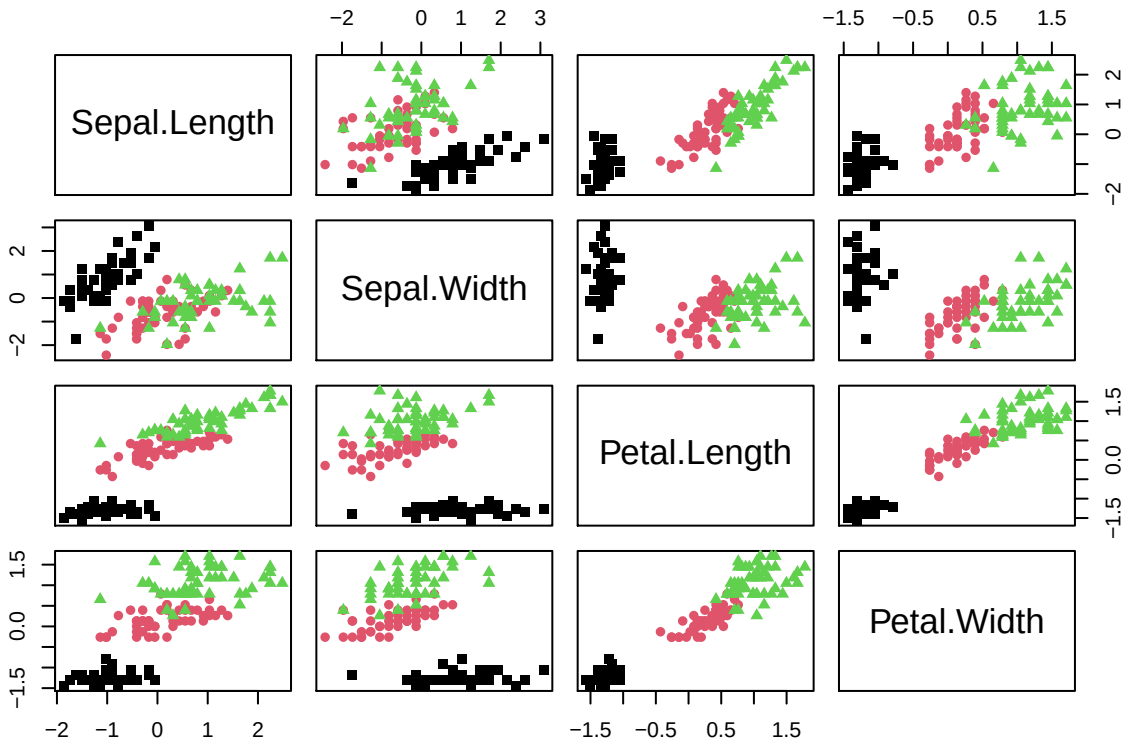
```
data(iris)

# preprocess
X.iris = scale((iris[, 1:4]), scale=TRUE) # center and standardise
L.iris = iris[, 5]

table(L.iris)
```

```
## L.iris
##     setosa versicolor  virginica
##         50         50         50
```

```
pairs(X.iris, col=as.integer(L.iris), pch=as.integer(L.iris)+14)
```



Now, apply a Gaussian Mixture Model (GMM) with G = 3.

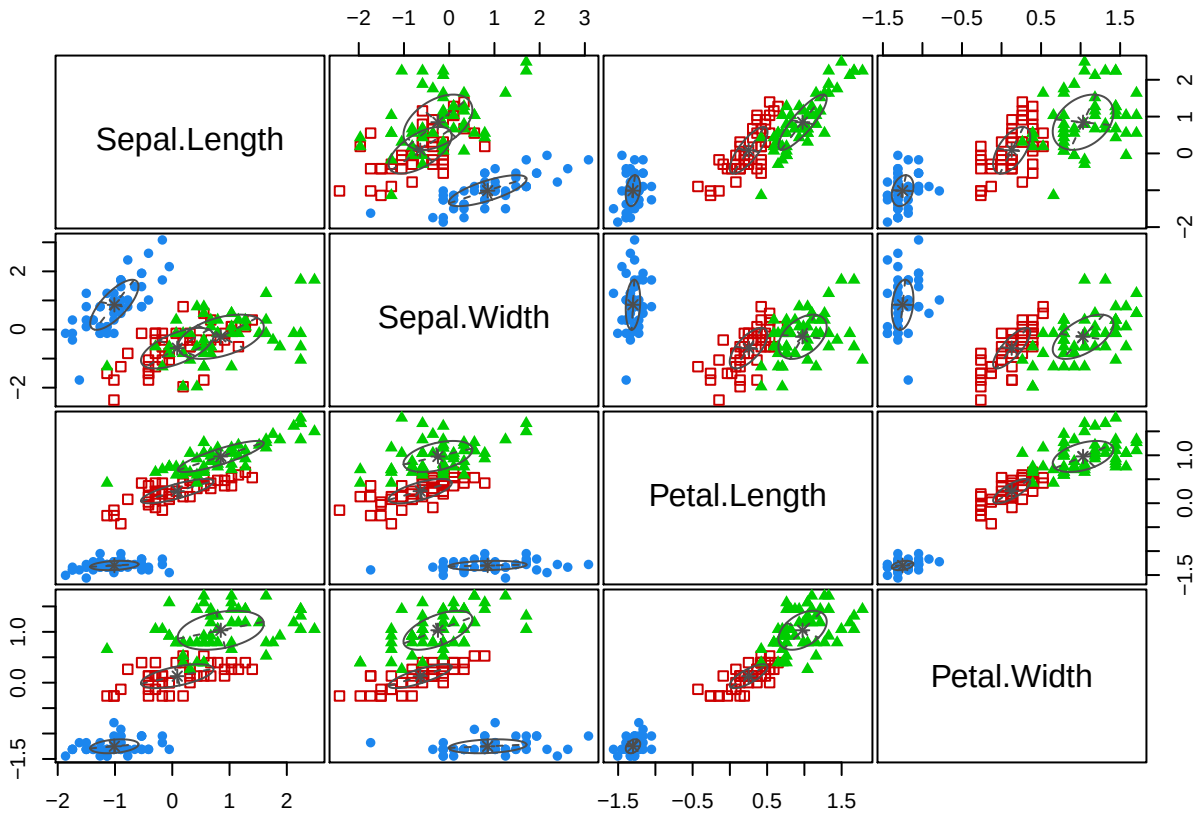
```
library("mclust")
```

```
## Package 'mclust' version 5.4.6
## Type 'citation("mclust")' for citing this R package in publications.
```

```
gmm3 = Mclust(X.iris, G=3)
```

Let us plot the clusters obtained with GMM and compare the resutls to the true labels.

```
plot(gmm3, what="classification")
```




```
print(table(L.iris, gmm3$classification))
```

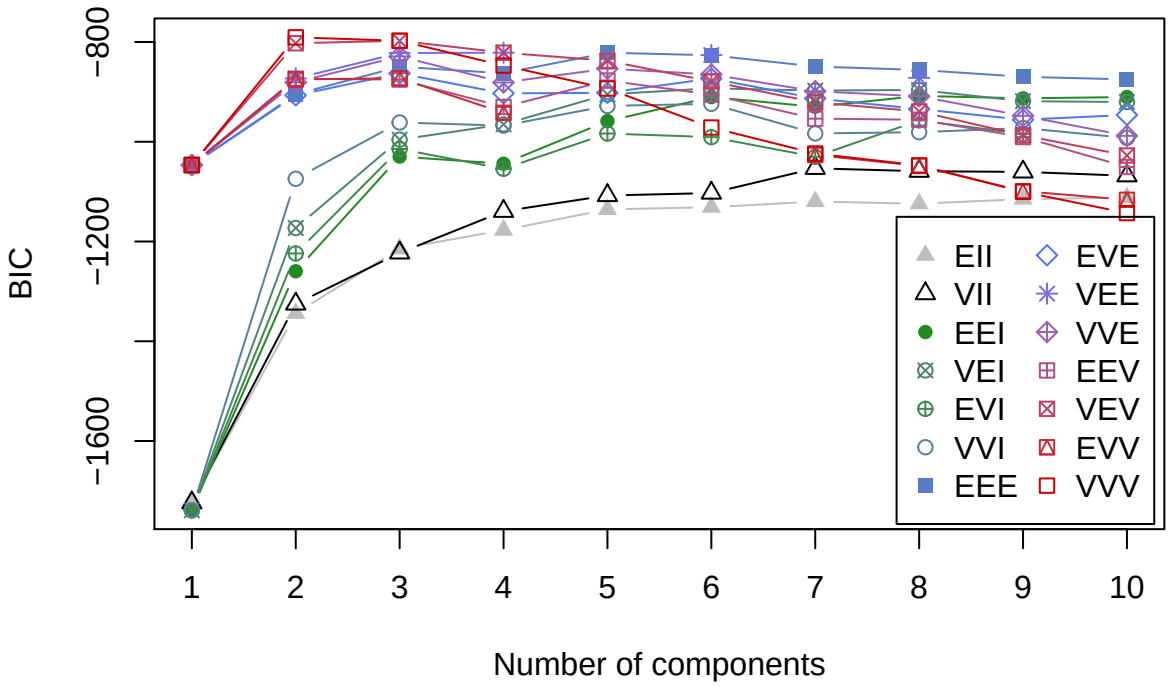
```
##
## L.iris         1  2  3
##    setosa     50  0  0
##    versicolor  0 45  5
##    virginica   0  0 50
```

Let us apply a GMM with varying values of G and check how the BIC varies.

```
gmm = Mclust(X.iris, G=1:10)
print(gmm$G) # the optimal number of G
```

```
## [1] 2
```

```
plot(gmm, what="BIC")
```


BIC
−800
−1200
−1600
1
2
3
4
5
6
7
8
9
10
Number of components
EII
VII
EEI
VEI
EVI
VVI
EEE
EVE
VEE
VVE
EEV
VEV
EVV
VVV