

Problem 6.2

Chen Bo Calvin Zhang

09/11/2020

First generate the data.

```
# generate the "mouse" data
library("mnormt")

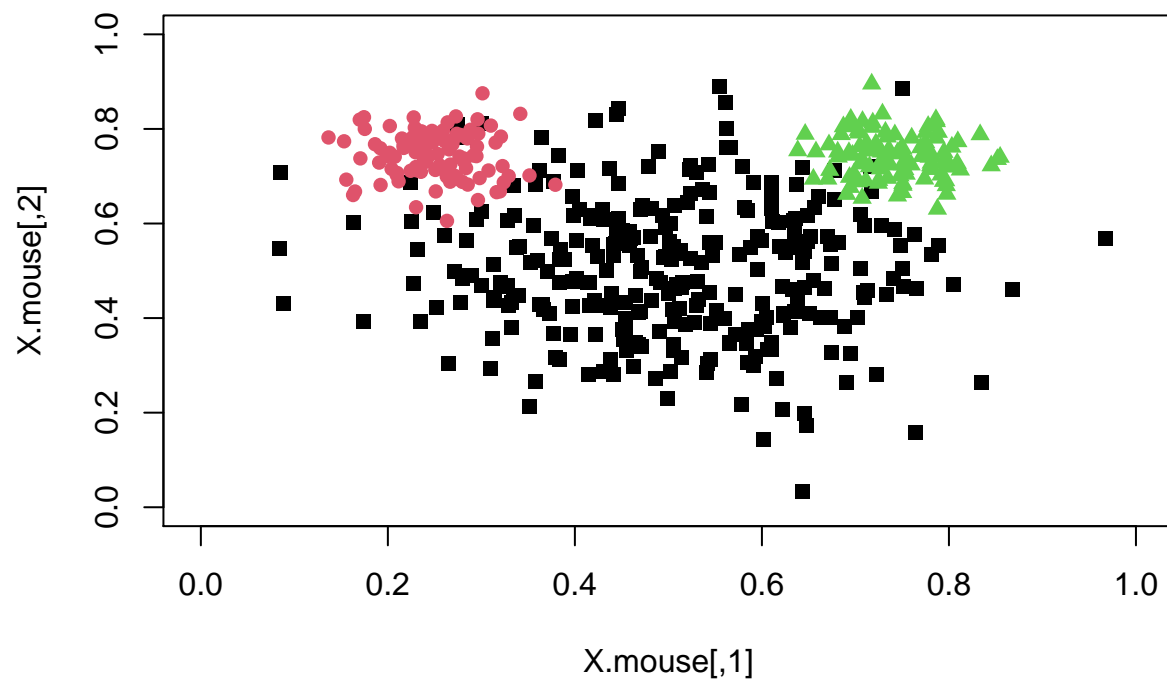
# class one (head)
n1 = 300
mu1 = c(0.5,0.5)
Sigma1 = matrix(c(0.15^2, 0, 0, 0.15^2), 2)
x1 = rmnorm(n1,mean=mu1,var=Sigma1)

# class two (left ear)
n2=100
mu2 = c(0.25,0.75)
Sigma2 = matrix(c(0.05^2, 0, 0, 0.05^2), 2)
x2 = rmnorm(n2,mean=mu2,var=Sigma2)

# class three (right ear)
n3 = 100
mu3 = c(0.75,0.75)
Sigma3 = matrix(c(0.05^2, 0, 0, 0.05^2), 2)
x3 = rmnorm(n3,mean=mu3,var=Sigma3)

# put all data together in one matrix
X.mouse = rbind(x1,x2,x3)
L.mouse = factor(c( rep("Head", n1), rep("Left Ear", n2), rep("Right Ear", n3) ))

plot(X.mouse, col=as.integer(L.mouse),
      pch=as.integer(L.mouse)+14, ylim=c(0,1), xlim=c(0,1))
```



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

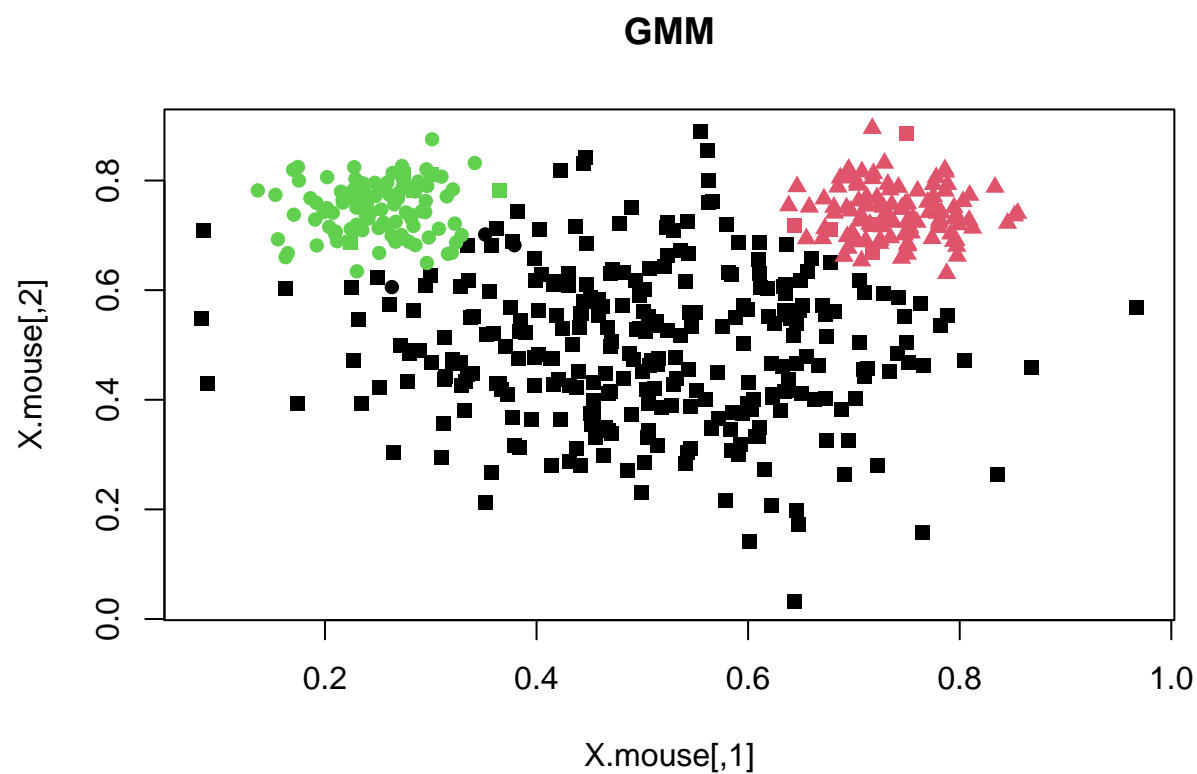
```
library("mclust")
```

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

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

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

```
plot(X.mouse, col = gmm3$classification, pch = as.integer(L.mouse) + 14, main = "GMM")
```



```
table(L.mouse, gmm3$classification)
```

```
##
## L.mouse      1   2   3
## Head        290  5   5
## Left Ear     3   0  97
## Right Ear    0 100   0
```

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

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

```
## [1] 3
```

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

