Assignment 1.

PART 1.

Variation without scaling

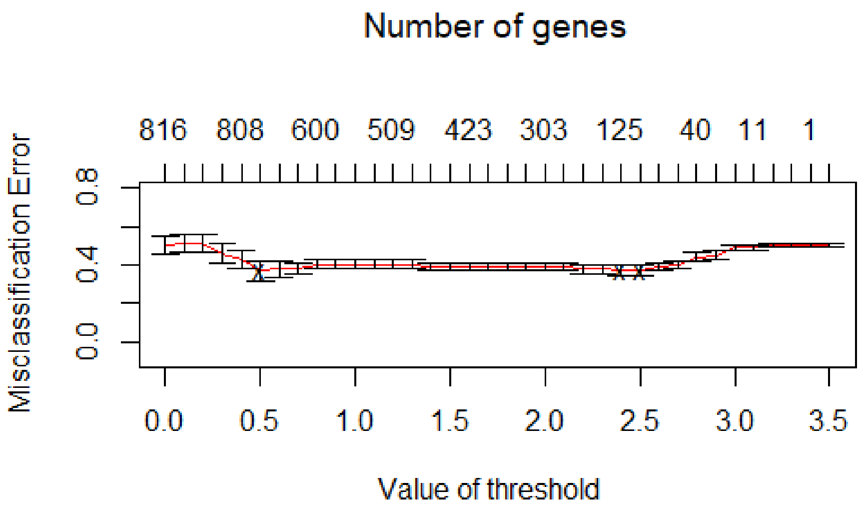
```
[1] "99.723"  "99.935"  "99.989"  "100.000" "100.000" "100.000" "100.000" "100.000" "1
[10] "100.000" "100.000" "100.000" "100.000" "100.000" "100.000" "100.000" "100.000"
```

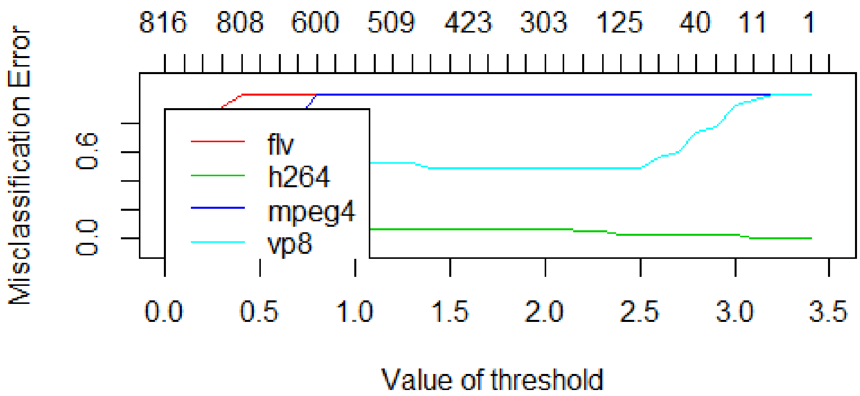
Variation witht scaling

```
 [1] "35.076"  "50.934"  "63.495"  "70.112"  "76.285"  "81.969"  "87.425"  "92.320"
  "95.623"
[10] "97.555"  "99.139"  "99.653"  "99.932"  "99.968"  "100.000" "100.000" "100.000"
```

1 component in unscaled data, 9 components in the scaled data are needed to explain 95% variation. The reason in that the original data is on a very different scale → variation in one feature dominates variation in the other features.

PART 2.





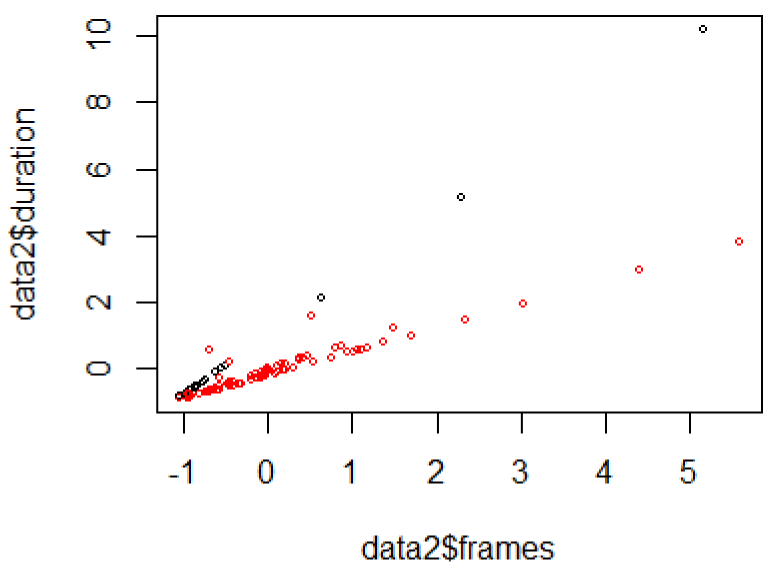
Higher threshold leads to less complex models, the higher the complexity the lower the bias and the higher is the variance

PART 3.

```
> cvmodel$threshold[which.max(cvmodel$loglik)]
[1] 2.5
```

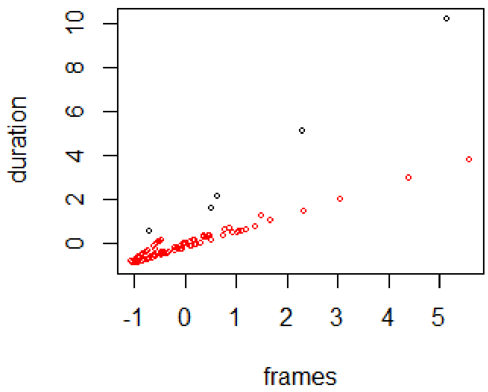
Multinomial likelihood is used because this is a classification problem.

PART 4.



Classes seen to be rather clearly linearly separable (with exception of a few cases near to the origin).

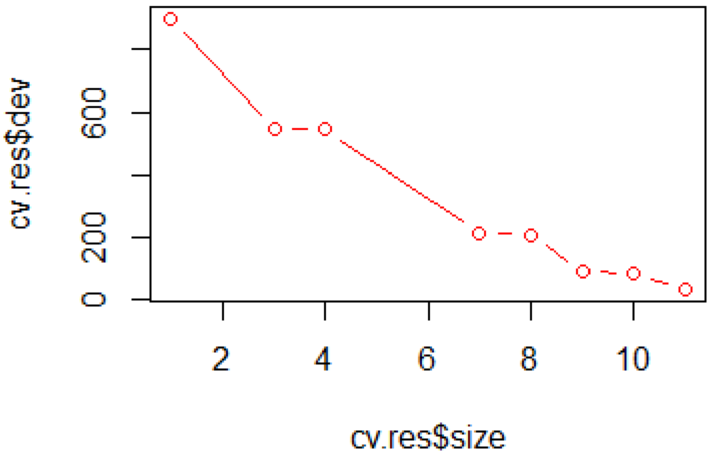
PART 5.




```
Misclassification error
[1] 0.172
```

The result of classification is rather bad. It is clear that covariance matrices per class are very different. In addition, class-conditional distributions do not look like multivariate normal.

PART 6.




```
Misclassification error
[1] 0.001
```

According to the cross-validation plot, the optimal tree is the largest one among the ones that were grown with default settings. The optimal tree among these has 11 leaves.

Such a complicated tree is needed because the optimal decision boundary is linear but not parallel to any of the coordinate axes. Accordingly, decision tree would need to make many splits and produce a stair-kind of decision boundary that would approximate this linear decision boundary.