

Assignment 1.

PART 1.

Variation without scaling

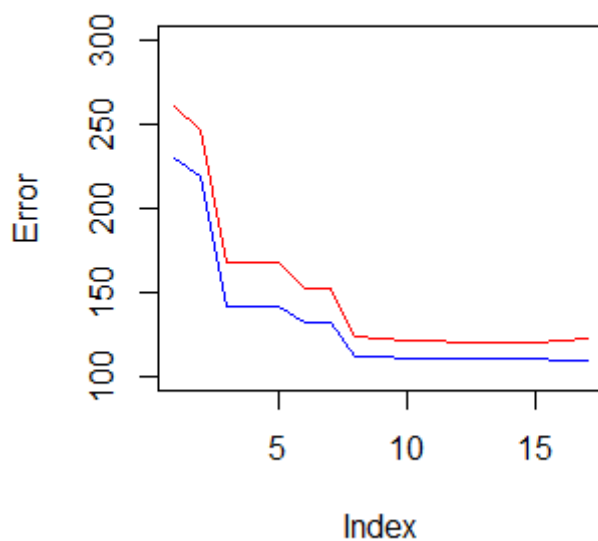
```
[1] "99.723" "99.935" "99.989" "100.000" "100.000" "100.000" "100.000" "100.000" "100.000" "100.000"
[10] "100.000" "100.000" "100.000" "100.000" "100.000" "100.000" "100.000" "100.000" "100.000" "100.000"
```

Variation with 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 is that the original data is on a very different scale → variation in one feature dominates variation in the other features.

PART 2.



When the number of components increases, the model becomes more complex and the bias goes down while variance goes up. The optimal model should have the lowest test error, in this case $M=14$. However, there are much simpler models with similar test errors, for ex. $M=8$.

PART 3.

```
pcrF=pcr(utime~., 8, data=train, validation="none", scale=T)
Yloadings(pcrF)
```

Loadings:

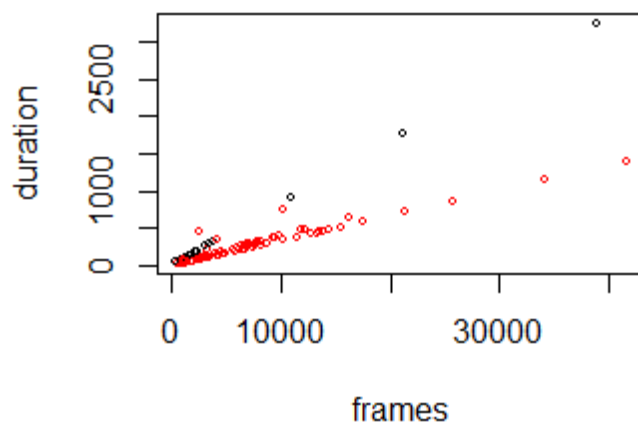
```
Comp 1 Comp 2 Comp 3 Comp 4 Comp 5 Comp 6 Comp 7 Comp 8
utime -1.196 1.966 6.048 0.156 0.195 -3.136 4.988
> mean(residuals(pcrF)^2)
```

[1] 156.092

Equation:

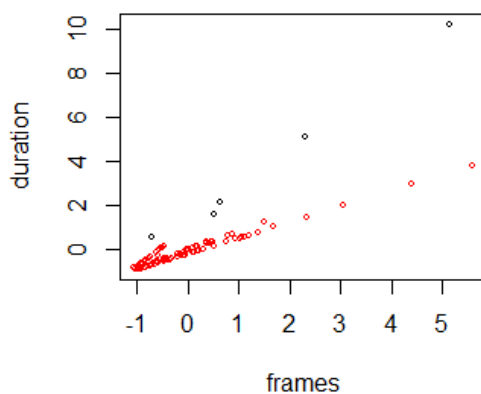
$$Utime_{scaled} \sim N(-1.196 \cdot \text{Comp1} + 1.966 \cdot \text{Comp2} + 6.048 \cdot \text{Comp3} + 0.156 \cdot \text{Comp4} + 0.195 \cdot \text{Comp5} - 3.136 \cdot \text{Comp6} + 4.988 \cdot \text{Comp8}, \quad \sigma^2 = 156)$$

PART 4.



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

PART 5.

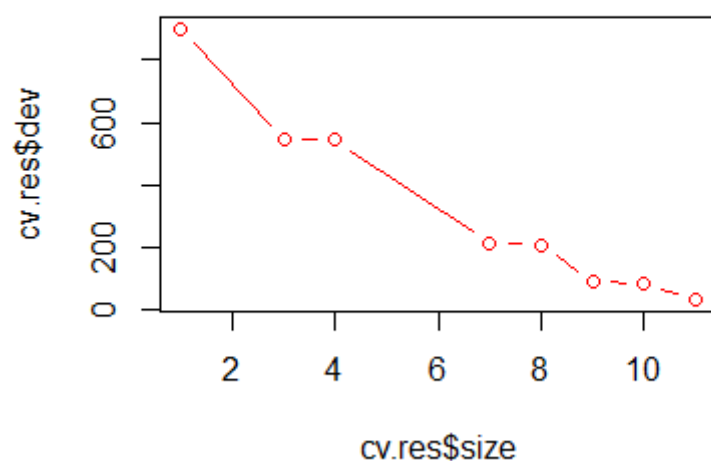


Misclassification error

[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 those that are 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.