Extra exercises STK4030-f13

Extra exercise 4.1
Extend the R-program in exercise 4.9 for quadratic discriminant analysis to cover also

a) Linear regression of an indicator matrix

b) Linear discriminant analysis.

Apply the program on the vowel data, and compare the result with Table 4.1 in the textbook.

Extra exercise 5.1

a) Show that the smoother matrix S_λ and the penalty matrix K have the same eigenvectors when $\lambda > 0$.

b) Use the property of `trace` to find an expression of df_λ in terms of the penalty λ and the eigenvalues of the penalty matrix, K.

c) Argue that df_λ is monotone in λ.

d) Explain what happens when $\lambda = 0$ and $\lambda \to \infty$. Is this reasonable?

Extra exercise 6.1
Consider the data set `phoneme` available on the web-page of *ElemStatLearn*. It consists of 4509 pronunciations of the phonemes "sh" "iy" "dcl" "aa" or "ao", together with log periodograms at 256 frequencies. We consider smoothing the response, coded as 1 or 0, separately for each group against frequency 128.

a) Make five data sets, one for each phoneme with occurrence of the phoneme in that class coded as 1.

b) Smooth each data set using the Nadaraya-Watson estimator. As shown in Exercise 6.5 this is the same as fitting a locally constant logit model.

[R-hints: If e.g the phoneme "sh" is placed in a 4509×2matrix where sh[,1] is the column consisting of 0 or 1's and sh[,2] contains the values of the log periodogram at frequency 128, the commands

```
plot(sh[,2],sh[,1])
lines(ksmooth(sh[,2],sh[,1],kernel="normal",bandwidth=1.0))
```

will plot the observations and the smoothed values. Use the command `help(ksmooth)` for more information.