Extra exercises for STK4030

Geir Storvik

Fall 2013

Exercise 1 (Exercise 4.2)
From eq (4.11) we have that LDA for $K = 2$ corresponds to calculating the linear combination $x^T\widehat{\Sigma}^{-1}(\hat{\mu}_2 - \hat{\mu}_1)$. We will in this exercise show that the least squares criterion leads to the same linear combination.

(*a*) We have that with inserted estimates

$$\begin{aligned}\delta_k(x) =& \log\hat{\pi}_k - 0.5(x - \widehat{\mu}_k)^T\widehat{\Sigma}^{-1}(x - \hat{\mu}_k)\\ =& \log\hat{\pi}_k - 0.5x^T\widehat{\Sigma}^{-1}x + x^T\widehat{\Sigma}^{-1}\hat{\mu}_k - 0.5\hat{\mu}_k^T\widehat{\Sigma}^{-1}\hat{\mu}_k\\ =& \text{Const} + \log\hat{\pi}_k + \widehat{\mu}_k^T\widehat{\Sigma}^{-1}x - 0.5\hat{\mu}_k^T\widehat{\Sigma}^{-1}\hat{\mu}_k\end{aligned}$$

giving

$$\delta_2(x) > \delta_1(x)$$

being equivalent to

$$x^T\widehat{\Sigma}^{-1}(\hat{\mu}_2 - \hat{\mu}_1) > 0.5\hat{\mu}_2^T\widehat{\Sigma}^{-1}\hat{\mu}_2 - 0.5\hat{\mu}_1^T\widehat{\Sigma}^{-1}\hat{\mu}_1 + \log\frac{N_1}{N} - \log\frac{N_2}{N}$$

where we have inserted $\hat{\pi}_k = N_k/N$.

(*b*) We have that the least squares estimates are given by

$$\begin{aligned}\sum_{i=1}^N(y_i - \beta_0 - \beta^T x_i] =& 0\\ \sum_{i=1}^N(y_i - \beta_0 - \beta^T x_i]x_i^T =& 0\end{aligned}$$

or, by inserting $\hat{\beta}_0 = \bar{y} - \hat{\beta}^T\bar{x}$,

$$[\sum_{i=1}^N(x_i - \bar{x})(x_i - \bar{x})^T]\hat{\beta} = \sum_{i=1}^N(y_i - \bar{y})(x_i - \bar{x})$$

or

$$[\sum_{i=1}^{N} x_i x_i^T - N\bar{x}\bar{x}^T]\hat{\beta} = \sum_{i=1}^{N}(y_i - \bar{y})(x_i - \bar{x})$$

Defining

$$y_i = \begin{cases} -N/N_1 & \text{if class is equal to 1} \\ N/N_2 & \text{if class is equal to 1} \end{cases},$$

and assuming the first N_i's correspond to class i, we get that $\bar{y} = 0$ and

$$\begin{aligned}
\sum_{i=1}^{N}(y_i - \bar{y})(x_i - \bar{x})^T = & -\frac{N}{N_1}\sum_{i=1}^{N_1}(x_i - \bar{x})^T + \frac{N}{N_2}\sum_{i=N_1+1}^{N}(x_i - \bar{x})^T \\
= & N[\frac{1}{N_2}\sum_{i=N_1+1}^{N} x_i - \frac{1}{N_1}\sum_{i=1}^{N_1} x_i] \\
= & N[\hat{\mu}_2 - \hat{\mu}_1]
\end{aligned}$$

Further, using that

$$\begin{aligned}
\bar{x} = & \frac{N_1}{N}\hat{\mu}_1 + \frac{N_2}{N}\hat{\mu}_2, \\
\bar{x}\bar{x}^T = & [\frac{N_1}{N}\hat{\mu}_1 + \frac{N_2}{N}\hat{\mu}_2][\frac{N_1}{N}\hat{\mu}_1 + \frac{N_2}{N}\hat{\mu}_2]^T \\
= & \frac{1}{N^2}[N_1^2\hat{\mu}_1\hat{\mu}_1^T + N_2^2\hat{\mu}_2\hat{\mu}_2^T + N_1N_2\hat{\mu}_1\hat{\mu}_2^T + N_1N_2\hat{\mu}_2\hat{\mu}_1^T] \\
\widehat{\Sigma} = & \frac{1}{N-2}[(N_1 - 1)\hat{\Sigma}_1 + (N_2 - 1)\hat{\Sigma}_2] \\
= & \frac{1}{N-2}[\sum_{i=1}^{N_1}(x_i - \hat{\mu}_1)(x_i - \hat{\mu}_1)^T + \sum_{i=N_1+1}^{N}(x_i - \hat{\mu}_2)(x_i - \hat{\mu}_2)^T] \\
= & \frac{1}{N-2}[\sum_{i=1}^{N} x_i x_i^T - \sum_{i=1}^{N_1} x_i\hat{\mu}_1^T - \sum_{i=1}^{N_1}\hat{\mu}_1 x_i^T + \\
& N_1\hat{\mu}_1\hat{\mu}_1^T - \sum_{i=N_1+1}^{N} x_i\hat{\mu}_2^T - \sum_{i=N_1+1}^{N}\hat{\mu}_2 x_i^T + N_2\hat{\mu}_2\hat{\mu}_2^T] \\
= & \frac{1}{N-2}[\sum_{i=1}^{N} x_i x_i^T - N_1\hat{\mu}_1\hat{\mu}_1^T - N_2\hat{\mu}_2\hat{\mu}_2^T]
\end{aligned}$$

we have

$$
\begin{aligned}
&[(N-2)\widehat{\Sigma} + \frac{N_1N_2}{N}\widehat{\Sigma}_B] \\
&\quad = \sum_{i=1}^{N} x_i x_i^T - N_1\hat{\mu}_1\hat{\mu}_1^T - N_2\hat{\mu}_2\hat{\mu}_2^T + \frac{N_1N_2}{N}\hat{\mu}_2\hat{\mu}_2^T + \frac{N_1N_2}{N}\hat{\mu}_1\hat{\mu}_1^T - \\
&\quad \frac{N_1N_2}{N}\hat{\mu}_1\hat{\mu}_2^T - \frac{N_1N_2}{N}\hat{\mu}_2\hat{\mu}_1^T \\
&\quad = \sum_{i=1}^{N} x_i x_i^T - \frac{1}{N}[N_1^2\hat{\mu}_1\hat{\mu}_1^T + N_2^2\hat{\mu}_2\hat{\mu}_2^T + \frac{N_1N_2}{N}\hat{\mu}_1\hat{\mu}_2^T + \frac{N_1N_2}{N}\hat{\mu}_2\hat{\mu}_1^T] \\
&\quad = \sum_{i=1}^{N} x_i x_i^T - N\bar{x}\bar{x}^T
\end{aligned}
$$

we obtain the result.

(c) We have that

$$
\begin{aligned}
\widehat{\Sigma}_B\beta =&(\hat{\mu}_2 - \hat{\mu}_1)(\hat{\mu}_2 - \hat{\mu}_1)^T\beta \\
=&k(\hat{\mu}_2 - \hat{\mu}_1)
\end{aligned}
$$

where $k = (\hat{\mu}_2 - \hat{\mu}_1)^T\beta$ is a scalar. Thereby

$$
\begin{aligned}
&[(N-2)\widehat{\Sigma} + \frac{N_1N_2}{N}\widehat{\Sigma}_B]\hat{\beta} \\
&\qquad =[(N-2)\widehat{\Sigma}\beta + k\frac{N_1N_2}{N}(\hat{\mu}_2 - \mu_1)
\end{aligned}
$$

showing that

$$
\widehat{\Sigma}\hat{\beta} \propto (\hat{\mu}_2 - \mu_1)
$$

and

$$
\hat{\beta} \propto \widehat{\Sigma}^{-1}(\hat{\mu}_2 - \mu_1)
$$

(d) Assume now that $y_i = a + b\tilde{y}_i$ where

$$
\tilde{y}_i = \begin{cases} -N/N_1 & \text{if class is equal to 1} \\ N/N_2 & \text{if class is equal to 1} \end{cases},
$$

Define further $\tilde{x}_i$ such that $x_i = b\tilde{x}_i$. Then

$$
\sum_i (y_i - \beta_0 - \beta^T x_i)^2 = \sum_i (a + b\tilde{y}_i - \beta_0 - \beta^T b\tilde{x}_i)^2 = \quad b^2 \sum_i (\tilde{y}_i - (\beta_0 - a) - \beta^T \tilde{x}_i)^2
$$

Then the only change in the least squares solutions will be a shift in β_0 and a scaling of β which means that the direction will remain unchanged.

(e) We have that

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}^T \bar{x}$$

giving that

$$\begin{aligned}\hat{f} =& \bar{y} - \hat{\beta}^T \bar{x} + \hat{\beta}^T x \\ =& \bar{y} + \hat{\beta}^T (x - \bar{x}) \\ =& k(\hat{\mu}_2 - \hat{\mu}_1)^T \widehat{\Sigma}^{-1} (x - \bar{x}) \\ =& k[(\hat{\mu}_2 - \hat{\mu}_1)^T \widehat{\Sigma}^{-1} x - (\hat{\mu}_2 - \hat{\mu}_1)^T \widehat{\Sigma}^{-1} (\frac{N_1}{N}\hat{\mu}_1 + \frac{N_2}{N}\hat{\mu}_2)] \\ =& k[(\hat{\mu}_2 - \hat{\mu}_1)^T \widehat{\Sigma}^{-1} x - [\frac{N_1}{N} - \frac{N_2}{N}]\hat{\mu}_2 \widehat{\Sigma}^{-1} \hat{\mu}_1 - \frac{N_2}{N}\hat{\mu}_2 \widehat{\Sigma}^{-1} \hat{\mu}_2 + \frac{N_1}{N}\hat{\mu}_1 \widehat{\Sigma}^{-1} \hat{\mu}_1]\end{aligned}$$

and $\hat{f} > 0$ is equivalent to

$$(\hat{\mu}_2 - \hat{\mu}_1)^T \widehat{\Sigma}^{-1} x > \frac{N_2}{N}\hat{\mu}_2 \widehat{\Sigma}^{-1} \hat{\mu}_2 - \frac{N_1}{N}\hat{\mu}_1 \widehat{\Sigma}^{-1} \hat{\mu}_1 + \frac{N_1}{N} - \frac{N_2}{N}]\hat{\mu}_2 \widehat{\Sigma}^{-1} \hat{\mu}_1$$

If $N_1 = N_2$, then this reduces to

$$(\hat{\mu}_2 - \hat{\mu}_1)^T \widehat{\Sigma}^{-1} x > \frac{1}{2}\hat{\mu}_2 \widehat{\Sigma}^{-1} \hat{\mu}_2 - \frac{1}{2}\hat{\mu}_1 \widehat{\Sigma}^{-1} \hat{\mu}_1$$

which is the same as LDA!