Writing Exercise

Bochen Wang
Tianyuan Liu

September 29, 2015

1 Exercise 4.1

1.1

$$\max_a a^T B a$$
$$\text{s.b.t } a^T W a = 1$$

corresponding to the lagrangian

$$\begin{aligned} L_P &= a^T Ba - \lambda(a^T Wa - 1) \\ \text{Let } \frac{\partial L_P}{\partial a} &= 2Ba - 2\lambda Wa = 0 \\ W^{-1}Ba &= \lambda a \end{aligned}$$

$\therefore$ a is a eigenvector of $W^{-1}B$

$$\begin{aligned} \because Ba &= \lambda Wa \\ \therefore a^T Ba &= \lambda a^T Wa \\ \because a^T Wa &= 1 \\ \therefore a^T Ba &= \lambda \end{aligned}$$

$\therefore$ a is the eigenvector corresponding to the biggest eigenvalue of $W^{-1}B$

2 Exercise 4.2

2.1

From equation 4.10 we know the linear discriminant function is:

$$\delta_k(x) = x^T \Sigma^{-1} \mu_k - \frac{1}{2}\mu_k^T \Sigma^{-1} \mu_k + \log \pi_k$$

so LDA classifies to class 2 if

$$
\begin{aligned}
\delta_2(x) &> \delta_1(x) \\
x^T\Sigma^{-1}\mu_2 - \frac{1}{2}\mu_2^T\Sigma^{-1}\mu_2 + \log\pi_2 &> x^T\Sigma^{-1}\mu_1 - \frac{1}{2}\mu_1^T\Sigma^{-1}\mu_1 + \log\pi_1 \\
x^T\Sigma^{-1}(\mu_2-\mu_1) &> \frac{1}{2}\mu_2^T\Sigma^{-1}\mu_2 - \frac{1}{2}\mu_1^T\Sigma^{-1}\mu_1 + \frac{1}{2}\mu_1^T\Sigma^{-1}\mu_2 \\
&- (\frac{1}{2}\mu_1^T\Sigma^{-1}\mu_2)^T + \log(N_1) - \log(N_2) \\
x^T\Sigma^{-1}(\mu_2-\mu_1) &> \frac{1}{2}(\mu_2+\mu_1)^T\Sigma^{-1}(\mu_2-\mu_1) - \log(\frac{N_2}{N_1})
\end{aligned}
$$

2.2

Let U_i be the class indicator vector of class i, so $U = U_1 + U_2$ is a vector with all entries 1. When we encode the target as $a1 = -\frac{N}{N_1}$ and $a2 = \frac{N}{N2}$, we have $Y = a_1U_1 + a_2U_2$.
Then, $RSS = \sum_{i=1}^{N}(y_i - \beta_0 - \beta^T x_i)^2 = (Y - \beta_0 U - X\beta)^T(Y - \beta_0 U - X\beta)$.
To minimize the RSS.

$$
\begin{aligned}
\frac{\partial RSS}{\partial \beta} &= 2X^TX\beta - 2X^TY + 2\beta_0 X^TU = 0 \\
\frac{\partial RSS}{\partial \beta_0} = 2U^TU\beta_0 - 2U^T(Y - X\beta) &= 2N\beta_0 - 2U^T(Y - X\beta) = 0
\end{aligned}
$$

From the second equation, we get:

$$\beta_0 = \frac{U^T(Y - X\beta)}{N} \tag{1}$$

Plug it back to first equation, we get:

$$(X^TX - \frac{X^TUU^TX}{N})\beta = X^TY - \frac{X^TUU^TY}{N} \tag{2}$$

For l.h.s:
Notice that $X^TU = X^T(U_1 + U_2) = N_1\mu_1 + N_2\mu_2$. plug it to l.h.s. We get:

$$l.h.s = X^TX - \frac{1}{N}(N_1^2\mu_1\mu_1^T N_2^2\mu_2\mu_2^2 + N_1N_2\mu_1\mu_2^T + N_1N_2\mu_2\mu_1^T) \tag{3}$$

The estimate of the covariance matrix used in LDA is:

$$
\begin{aligned}
(N-2)\Sigma &= \sum_{i:y_i=a_1}(x_i-\mu_1)(x_i-\mu_1)^T + \sum_{i:y_i=a_2}(x_i-\mu_2)(x_i-\mu_2)^T \\
&= X^TX - N_1\mu_1\mu_1^T - N_2\mu_2\mu_2^T
\end{aligned}
$$

Also $\Sigma_B = (\mu_1 - \mu_2)(\mu_1 - \mu_2)^T$. So

$$
\begin{aligned}
(N-2)\Sigma + \frac{N_1N_2}{N}\Sigma_B &= X^TX + (\frac{N_1N_2}{N} - N_1)\mu_1\mu_1^T + (\frac{N_1N_2}{N} - N_2)\mu_2\mu_2^T \\
&- \frac{N_1N_2}{N}\mu_1\mu_2^T - \frac{N_1N_2}{N}\mu_1\mu_2^T \\
&= X^TX - \frac{N_1^2}{N}\mu_1\mu_1^T - \frac{N_2^2}{N}\mu_2\mu_2^T - \frac{N_1N_2}{N}\mu_1\mu_2^T - \frac{N_1N_2}{N}\mu_2\mu_1^T
\end{aligned}
$$

Which is the same as equation(3).
For r.h.s: The first term is equal to :

$$X^TY = X^T(a_1U_1 + a_2U_2) = a1N_1\mu_1 + a2N_2\mu_2$$

The second term is equal to:

$$\begin{aligned}\frac{1}{N}X^TUU^TY &= \frac{1}{N}(N_1\mu_1 + N_2\mu_2)\\ &= \frac{a_1N_1^2 + a_2N_1N_2}{N}\mu_1 + \frac{a2N_2^2 + a_1N_1N_2}{N}\mu_2\end{aligned}$$

Combing two terms together, and with l.h.s we get:

$$[(N-2)\Sigma + \frac{N_1N_2}{N}\Sigma_B]\beta = \frac{N_1N_2}{N}(a_1 - a_2)(\mu_1 - \mu_2) \tag{4}$$

plug $a_1 = -\frac{N}{N_1}$ and $a_2 = \frac{N}{N_2}$ back to the equation:

$$r.h.s = \frac{N_1N_2}{N}(-\frac{N^2}{N_1N_2})(\mu_1 - \mu_2) = N(\mu_2 - \mu_1) \tag{5}$$

So we get:

$$[(N-2)\Sigma + \frac{N_1N_2}{N}\Sigma_B]\beta = N(\mu_2 - \mu_1) \tag{6}$$

2.3

Let c = $(\mu_2 - \mu_1)^T\beta$

$$\Sigma_B\beta = (\mu_2 - \mu_1)(\mu_2 - \mu_1)\beta = c(\mu_2 - \mu_1)$$

$\Sigma_B\beta$ is int the direction of $(\mu_2 - \mu_1)$. Combined with result of last question, $\beta \propto \Sigma^{-1}(\mu_2 - \mu_1)$. Therefore the least-squares regression coefficient is identical to the LDA coefficient, up to a scalar multiple.

2.4

Since (5) holds for any encoding a1 and a2, the result also holds for any encoding.

2.5

Notice that $U^TY = a_1N_1 + a_2N_2 = -N + N = 0$. So from equation(1), we get: $\beta_0 = -\frac{1}{N}(U^TX\beta) = -(\frac{N_1}{N}\mu_1^T + \frac{N_2}{N}\mu_2^T)\beta$. so the decision function is:

$$f(x) = x^T\beta + \beta_0 = (X^T - \frac{N_1}{N}\mu_1^T - \frac{N_2}{N}\mu_2^T)\beta$$

Notice from the result of 1.3. We have $\beta = c\Sigma^{-1}(\mu_2 - \mu_1)$ for some value c. The decision is:

$$\begin{aligned}f(x) &> 0\\ x^T\Sigma^{-1}(\mu_2 - \mu_1) &> (\frac{N_1}{N}\mu_1^T + \frac{N_2}{N}\mu_2^T)\Sigma^{-1}(\mu_2 - \mu_1)\end{aligned}$$

So this is not the same as the LDA rule unless the classes have equal numbers of observations.

Reference:
http://cbio.ensmp.fr/ jvert/svn/tutorials/course/1102Senegal/solution.pdf

3 RLU Exercise 11.3.1

3.1

$$
\begin{aligned}
M^T M &= \begin{bmatrix} 1 & 3 & 5 & 0 & 1 \\ 2 & 4 & 4 & 2 & 3 \\ 3 & 5 & 3 & 4 & 5 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 3 & 4 & 5 \\ 5 & 4 & 3 \\ 0 & 2 & 4 \\ 1 & 3 & 5 \end{bmatrix} \\
&= \begin{bmatrix} 36 & 37 & 38 \\ 37 & 49 & 61 \\ 38 & 61 & 84 \end{bmatrix} \\
M M^T &= \begin{bmatrix} 1 & 2 & 3 \\ 3 & 4 & 5 \\ 5 & 4 & 3 \\ 0 & 2 & 4 \\ 1 & 3 & 5 \end{bmatrix} \begin{bmatrix} 1 & 3 & 5 & 0 & 1 \\ 2 & 4 & 4 & 2 & 3 \\ 3 & 5 & 3 & 4 & 5 \end{bmatrix} \\
&= \begin{bmatrix} 14 & 26 & 22 & 16 & 22 \\ 26 & 50 & 46 & 28 & 40 \\ 22 & 46 & 50 & 20 & 32 \\ 16 & 28 & 20 & 20 & 26 \\ 22 & 40 & 32 & 26 & 35 \end{bmatrix}
\end{aligned}
$$

3.2

Two eigenvalues for $M^T M$ and $M M^T$ are: **15.43** and **153.57**.

3.3

Two eigenvectors for $M^T M$ are:

$$
\begin{bmatrix} -0.816 & 0.409 \\ -0.126 & 0.564 \\ 0.564 & 0.718 \end{bmatrix}
$$

Two eigenvectors for $M M^T$ are:

$$
\begin{bmatrix} 0.159 & 0.298 \\ -0.033 & 0.571 \\ -0.736 & 0.521 \\ 0.510 & 0.323 \\ 0.414 & 0.459 \end{bmatrix}
$$

3.4

$$
\begin{aligned}
MM^T &= (UDV^T)(UDV^T)^T \\
&= UDV^TVD^TU^T \\
&= UD^2U^T
\end{aligned}
$$

This is the eigendecomposition of MM^T.
So U should be a N * N matrix where ith column is the ith eigenvector.
U is:

$$
\begin{bmatrix}
0.298 & 0.159 \\
0.571 & -0.033 \\
0.521 & -0.736 \\
0.323 & 0.510 \\
0.459 & 0.414
\end{bmatrix}
$$

Similarly, V is:

$$
\begin{bmatrix}
0.409 & -0.816 \\
0.564 & -0.126 \\
0.718 & 0.564
\end{bmatrix}
$$

Σ is the diagonal matrix which the ith diagonal value is the square of ith eigenvalue.

$$
\begin{bmatrix}
12.39 & 0 \\
0 & 3.93
\end{bmatrix}
$$

3.5

$$
\begin{bmatrix}
0.298 \\
0.571 \\
0.521 \\
0.323 \\
0.459
\end{bmatrix}
* 12.39 *
\begin{bmatrix}
0.409 \\
0.564 \\
0.718
\end{bmatrix}^T
=
\begin{bmatrix}
1.51 & 2.08 & 2.65 \\
2.89 & 3.98 & 5.07 \\
2.64 & 3.64 & 4.64 \\
1.64 & 2.25 & 2.87 \\
2.33 & 3.20 & 4.08
\end{bmatrix}
$$

3.6

The energy retained is $\frac{12.39^2}{12.39^2+3.93^2} = 0.91$.