HW2 write up

Q1. $\max a^T B a$

s.t. $a^T W a = 1$ by transforming to a standard eigenvalue problem

Lagrange function:

$\mathcal{L}(x, \lambda) = f(x) + \lambda g(x)$

in order to find the stationary points of of a function $f(x)$

subject to the equality constraint $g(x) = 0$,

Since $a^T W a = 1$. let $g(x) = a^T W a - 1 = 0$.

$\mathcal{L}(\lambda) = a^T B a + \lambda(a^T W a - 1)$

$$\Rightarrow \frac{d\mathcal{L}(\lambda)}{da} = (B + B^T)a - \lambda(W + W^T)a = 0$$

$$(B + B^T)a = \lambda(W + W^T)a$$

$$(W + W^T)^{-1}(B + B^T)a = \lambda a$$

$\Rightarrow$ the problem is transfered to a standard eigenvalue problem

Q2. $x \in \mathbb{R}^p$, two-class response, class sizes $N_1 . N_2$

targeted coded as $-\frac{N}{N_1}$, $\frac{N}{N_2}$

(a) $\pi_1 = -\frac{N}{N_1}$ $\pi_2 = \frac{N}{N_2}$

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

class 1: $\log \frac{Pr(G=1 \mid X=x)}{Pr(G=2 \mid X=x)} = \log \frac{f_1(x)}{f_2(x)} + \log \frac{\pi_1}{\pi_2}$

class 2: $\log \frac{Pr(G=2 \mid X=x)}{Pr(G=1 \mid X=x)} = -\log \frac{f_1(x)}{f_2(x)} - \log \frac{\pi_1}{\pi_2}$

$f_k(x) = \frac{1}{(2\pi)^{p/2} |\Sigma_k|^{1/2}} e^{-\frac{1}{2}(x-\mu_k)^T \Sigma_k^{-1} (x-\mu_k)}$

plug in

for class 2: If LDA rules classfies to class 2

$$\log \frac{Pr(G=2 \mid X=x)}{Pr(G=1 \mid X=x)} = -\log \frac{\pi_1}{\pi_2} - \frac{1}{2}(\mu_1+\mu_2)^T \Sigma^{-1} (\mu_2-\mu_1) + x^T \Sigma^{-1} (\mu_2-\mu_1)$$

should be greater than 0

$$\Rightarrow x^T \Sigma^{-1} (\mu_2 - \mu_1) > \log \frac{\pi_1}{\pi_2} + \frac{1}{2}(\mu_1+\mu_2)^T \Sigma^{-1} (\mu_2 - \mu_1)$$

$$= -\log \frac{N_2}{N_1} + \frac{1}{2}(\mu_1+\mu_2)^T \Sigma^{-1} (\mu_2 - \mu_1)$$

Same for class 1:

$$x^T \Sigma^{-1} (\mu_2 - \mu_1) < -\log \frac{N_2}{N_1} + \frac{1}{2}(\mu_1+\mu_2)^T \Sigma^{-1} (\mu_2 - \mu_1)$$

(b) minimization of the least square criterion

$$RSS(\beta') = \sum_{i=1}^{N} (y_i - \beta_0 - x_i^T \beta)^2$$

Let $U_i \in \mathbb{R}^n$ be the class indicator vector of class i

if $x_j \in$ class i. $j \in [0, N]$, $\Rightarrow U_{ij} = 1$ $U_{ij} = 0$ o.w.

$U = U_1 + U_2$ will be a vector of ones.

$a_1 = -\frac{N}{N_1}$ $a_2 = \frac{N}{N_2}$ $Y = a_1 U_1 + a_2 U_2$ vector of labels

$$RSS(\beta, \beta_0) = \sum_{i=1}^{N} (y_i - \beta_0 - x_i^T \beta)^2$$

$$= (Y - \beta_0 U - X\beta)^T (Y - \beta_0 U - X\beta)$$

$$\frac{\partial RSS}{\partial \beta} = 2X^T X\beta - 2X^T Y + 2\beta_0 X^T U = 0 \quad ①$$

$$\frac{\partial RSS}{\partial \beta_0} = 2U^T U \beta_0 - 2U^T (Y - X\beta) = 2N\beta_0 - 2U^T(Y - X\beta) = 0 \quad ②$$

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

Plug in β_0. to ①

$$X^T X\beta - X^T Y + \frac{1}{N} U^T U (Y - X\beta) X^T = 0$$

$$X^T X\beta - X^T Y + \frac{1}{N} U^T U X^T Y - \frac{1}{N} U^T U X X^T \beta = 0$$

$$\boxed{(X^T X - \frac{1}{N} X^T X U^T U)\beta} = X^T Y - \frac{1}{N} X^T U^T U Y \quad ③$$

LHS

- $X^T U = X^T(U_1 + U_2) = N_1 \hat{\mu}_1 + N_2 \hat{\mu}_2$

$$(X^T X - \frac{1}{N}(N_1^2 \hat{\mu}_1 \hat{\mu}_1^T + N_2^2 \hat{\mu}_2 \hat{\mu}_2^T + N_1 N_2 \hat{\mu}_1 \hat{\mu}_2^T + N_1 N_2 \hat{\mu}_2 \hat{\mu}_1^T))\beta$$

Def of LDA

$$\hat{\Sigma} = \frac{1}{N-K} \sum_{k=1}^{K} \sum_{g_i = k} (x_i - \hat{\mu}_k)(x_i - \hat{\mu}_k)^T$$

$$\sum_{g_i = k} x_i = N_k \hat{\mu}_k$$

Sum of all variables ← count × mean

$$\sum_{k=1}^{2} \sum_{g_i = k} (x_i^2) = X^T X$$

$$(N-2)\hat{\Sigma} = \cancel{(N-2)} \frac{1}{\cancel{N-2}} \sum_{k=1}^{2} \sum_{g_i = k} (x_i - \hat{\mu}_k)(x_i - \hat{\mu}_k)^T$$

$$= \sum_{k=1}^{2} \sum_{g_i=k} (x_i^2 - 2x_i\mu_k^T + \mu_k\mu_k^T)$$

$$= XX^T + \sum_{i:g_i=a_1} (-2x_i\hat{\mu}_1^T + \hat{\mu}_1\hat{\mu}_1^T)$$

add

$$+ \sum_{i:g_i=a_2} (-2x_i\hat{\mu}_2^T + \hat{\mu}_2\hat{\mu}_2^T)$$

$$= XX^T - 2N_1\hat{\mu}_1\hat{\mu}_1^T + N_1\hat{\mu}_1\hat{\mu}_1^T$$

$$- 2N_2\hat{\mu}_2\hat{\mu}_2^T + N_2\hat{\mu}_2\hat{\mu}_2^T$$

$$= XX^T - N_1\hat{\mu}_1\hat{\mu}_1^T - N_2\mu_2\hat{\mu}_2^T$$

$$\hat{\Sigma}_B = \frac{N_1N_2}{N^2}(\hat{\mu}_2 - \hat{\mu}_1)(\hat{\mu}_2 - \hat{\mu}_1)^T$$

$$N\hat{\Sigma}_B = \frac{N_1N_2}{N}(\hat{\mu}_2 - \hat{\mu}_1)(\hat{\mu}_2 - \hat{\mu}_1)^T$$

$$\beta((N-2)\hat{\Sigma} + N\hat{\Sigma}_B) = \beta\Big(X^TX + \left(\frac{N_1N_2}{N} - N_1\right)\hat{\mu}_1\hat{\mu}_1^T +$$

$$\left(\frac{N_1N_2}{N} - N_2\right)\hat{\mu}_2\hat{\mu}_2^T - \frac{N_1N_2}{N}(\hat{\mu}_2\hat{\mu}_1^T + \hat{\mu}_1\hat{\mu}_2^T)\Big)$$

$$N = N_1 + N_2 \rightarrow = \Big(XX^T - \frac{1}{N}(N_1^2\hat{\mu}_1\hat{\mu}_1^T + N_2\hat{\mu}_2\hat{\mu}_2^T$$

$$+ N_1N_2\hat{\mu}_2\hat{\mu}_1^T + N_1N_2\hat{\mu}_1\hat{\mu}_2^T)\Big)$$

- RMS $\quad U^TY = a_1N_1 + a_2N_2$

$$X^TU = N_1\hat{\mu}_1 + N_2\hat{\mu}_2$$

$$X^TY - \frac{1}{N}(U^TYX^TU) = X^T(a_1U_1 + a_2U_2) - \frac{1}{N}(a_1N_1 + a_2N_2)(N_1\hat{\mu}_1 + N_2\hat{\mu}_2)$$

$$= a_1N_1\hat{\mu}_1 + a_2N_2\hat{\mu}_2 - \frac{\hat{\mu}_1(a_1N_1^2 + a_2N_1N_2)}{N} - \frac{\hat{\mu}_2(a_1N_1N_2 + a_2N_2^2)}{N}$$

$$a_1N_1^2 = a_1N_1(N - N_2) = a_1N_1N - a_1N_1N_2$$

$$a_2N_2^2 = a_2N_2N - a_2N_2N_1$$

$$a_1 N_1 \hat{\mu}_1 + a_2 N_2 \hat{\mu}_2 - \frac{\hat{\mu}_1 (a_1 N_1^2 + a_2 N_1 N_2)}{N} - \frac{\hat{\mu}_2 (a_1 N_1 N_2 + a_2 N_2^2)}{N}$$

$$= \frac{\hat{\mu}_1 (a_1 N_1 N - a_1 N_1 N + a_1 N_1 N_2 - a_2 N_1 N_2)}{N}$$

$$+ \frac{\hat{\mu}_2 (a_2 N_2 N - a_1 N_1 N - a_2 N_2 N + a_2 N_1 N_2)}{N}$$

$$= \frac{N_1 N_2}{N} (\hat{\mu}_1 - \hat{\mu}_2)(a_1 - a_2)$$

$$a_1 = -\frac{N}{N_1} \quad a_2 = \frac{N}{N_2} \quad a_1 - a_2 = \frac{-N^2}{N_1 N_2}$$

$$= \frac{N_1 N_2}{N} (\hat{\mu}_1 - \hat{\mu}_2)\left(-\frac{N^2}{N_1 N_2}\right)$$

$$= -N(\hat{\mu}_1 - \hat{\mu}_2)$$

Thus. $(N-2)\hat{\Sigma} + N\hat{\Sigma}_B)\hat{\beta} = X^T Y - \frac{1}{N}(U^T U X^T Y)$

(c) $\hat{\Sigma}_B \beta \quad (\hat{\mu}_2 - \hat{\mu}_1) \Rightarrow \hat{\beta} \propto \hat{\Sigma}^{-1}(\hat{\mu}_2 - \hat{\mu}_1)$

$$\hat{\Sigma}_B \beta = (\hat{\mu}_2 - \hat{\mu}_1)(\hat{\mu}_2 - \hat{\mu}_1)^T \beta$$

$$= c(\hat{\mu}_2 - \hat{\mu}_1)$$

$\Rightarrow \hat{\Sigma}_B \beta$ is in the direction of $(\hat{\mu}_2 - \hat{\mu}_1)$

$\Rightarrow \hat{\beta} \propto^{-1} (\hat{\mu}_2 - \hat{\mu}_1)$

(d) $\frac{N_1 N_2}{N}(\hat{\mu}_1 - \hat{\mu}_2)(a_1 - a_2)$ hold for any encoding a_1, a_2, this result in c) hold for any coding of the two classes

(e) $a_1 = -\frac{N}{N_1} \quad a_2 = \frac{N}{N_2} \quad Y = a_1 U_1 + a_2 U_2$

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

$$= \frac{1}{N} U^T (a_1 U_1 + a_2 U_2 - X\beta)$$

plug in a_1, a_2 $= \frac{1}{N} U^T \left(-\frac{N}{N_1} U_1 + \frac{N}{N_2} U_2 - X\beta\right)$

$$= -\frac{U^T U_1}{N_1} + \frac{1}{N_2} U^T U_2 - \frac{1}{N} U^T X\beta$$

$$= -\frac{N_1}{N_1} + \frac{N_2}{N_2} - \frac{1}{N} U^T X \beta$$
$$= -\frac{1}{N} U^T X \beta$$
$$= -\frac{1}{N} (N_1 \hat{\mu}_1^T + N_2 \hat{\mu}_2^T) \beta$$

$$\hat{f}(x) = \hat{\beta}_0 + x^T \hat{\beta} = -\frac{1}{N}(N_1 \hat{\mu}_1^T + N_2 \hat{\mu}_2^T)\hat{\beta} + x^T \hat{\beta}$$
$$\hat{\beta} = c \hat{\Sigma}^{-1} (\hat{\mu}_2 - \hat{\mu}_1)$$
$$\hat{f}(x) = x^T c \Sigma^{-1} (\hat{\mu}_2 - \hat{\mu}_1) - \frac{N_1 \hat{\mu}_1^T}{N} c \Sigma^{-1} (\hat{\mu}_2 - \hat{\mu}_1) - \frac{N_2 \hat{\mu}_2^T}{N} c \Sigma^{-1} (\hat{\mu}_2 - \hat{\mu}_1)$$
$$\hat{f}(x) > 0$$
$$x^T c \Sigma^{-1} (\hat{\mu}_2 - \hat{\mu}_1) - \frac{N_1 \hat{\mu}_1^T}{N} c \Sigma^{-1} (\hat{\mu}_2 - \hat{\mu}_1) - \frac{N_2 \hat{\mu}_2^T}{N} c \Sigma^{-1} (\hat{\mu}_2 - \hat{\mu}_1) > 0$$
$$x^T \Sigma^{-1} (\hat{\mu}_2 - \hat{\mu}_1) - \frac{N_1 \hat{\mu}_1^T}{N} \Sigma^{-1} (\hat{\mu}_2 - \hat{\mu}_1) - \frac{N_2 \hat{\mu}_2^T}{N} \Sigma^{-1} (\hat{\mu}_2 - \hat{\mu}_1) > 0$$

$$\Longrightarrow$$

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

if $\hat{f}(x) > 0$ classify to class 2 or class 1 o.w.

when $N_1 = N_2$.

this simplifies to the LDA decision f in (a)

Q3. (a)

$$M^TM = \begin{bmatrix} 1 & 3 & 2 & 0 & 5 \\ 0 & 7 & -2 & -1 & 8 \\ 3 & 2 & 8 & 1 & 7 \end{bmatrix} \begin{bmatrix} 1 & 0 & 3 \\ 3 & 7 & 2 \\ 2 & -2 & 8 \\ 0 & -1 & 1 \\ 5 & 8 & 7 \end{bmatrix}$$

$$= \begin{bmatrix} 39 & 57 & 60 \\ 57 & 118 & 53 \\ 60 & 53 & 127 \end{bmatrix}$$

$$MM^T = \begin{bmatrix} 1 & 0 & 3 \\ 3 & 7 & 2 \\ 2 & -2 & 8 \\ 0 & -1 & 1 \\ 5 & 8 & 7 \end{bmatrix} \begin{bmatrix} 1 & 3 & 2 & 0 & 5 \\ 0 & 7 & -2 & -1 & 8 \\ 3 & 2 & 8 & 1 & 7 \end{bmatrix}$$

$$= \begin{bmatrix} 10 & 9 & 26 & 3 & 26 \\ 9 & 62 & 8 & -5 & 85 \\ 26 & 8 & 72 & 10 & 50 \\ 3 & -5 & 10 & 2 & -1 \\ 26 & 85 & 50 & -1 & 138 \end{bmatrix}$$

(b) Eigenvalue of M^TM

$$\det(M - \lambda I) = \det \begin{bmatrix} 39-\lambda & 57 & 60 \\ 57 & 118-\lambda & 53 \\ 60 & 53 & 127-\lambda \end{bmatrix} = 0$$

$$(39-\lambda)\left[(118-\lambda)(127-\lambda) - 53^2\right] - 57\left((57\times(127-\lambda)) - 60\times 53\right)$$
$$+ 60\left[57\times 53 - 60\times(118-\lambda)\right] = 0$$

$(39-\lambda)(14986-245\lambda+\lambda^2-2809)-57(7239-57\lambda-3180)$
$+60(3021-7080+60\lambda)=0$
$\lambda = 214.6705,\ 69.3295$

Eigenvalue of MM^T

$$\det(M-\lambda I) = \det\begin{bmatrix} 10-\lambda & 9 & 26 & 3 & 26 \\ 9 & 62-\lambda & 8 & -5 & 85 \\ 26 & 8 & 72-\lambda & 10 & 50 \\ 3 & -5 & 10 & 2-\lambda & -1 \\ 26 & 85 & 50 & -1 & 138-\lambda \end{bmatrix} = 0$$

$\lambda = 214.6705,\ 69.3295$

(c) Eigenvector of M^TM:

$$\begin{bmatrix} 0.4261 \\ 0.6150 \\ 0.6634 \end{bmatrix} \quad \begin{bmatrix} -0.0146 \\ -0.7285 \\ 0.6847 \end{bmatrix}$$

Eigenvector of MM^T

$$\begin{bmatrix} -0.1649 \\ -0.4716 \\ -0.3364 \\ -0.0033 \\ -0.7982 \end{bmatrix} \quad \begin{bmatrix} 0.2449 \\ -0.4533 \\ 0.8294 \\ 0.1697 \\ -0.1331 \end{bmatrix}$$

(d) Find SVD $\quad A = U\Sigma V^T$

Singular values: $\sqrt{69.3295} = 8.3264 \quad \sqrt{214.6705} = 14.6516$

$$\Sigma = \begin{bmatrix} 14.6516 & 0 \\ 0 & 8.3264 \end{bmatrix}$$

$MM^T \Rightarrow$ eigenvector $\Rightarrow V$

$V = \begin{bmatrix} 0.4261 & -0.0146 \end{bmatrix}$

$$\begin{bmatrix} 0.6150 & -0.1285 \\ 0.6634 & 0.6847 \end{bmatrix}$$

$M^TM \Rightarrow$ eigenvector $\Rightarrow U$

$$U = \begin{bmatrix} -0.1649 & 0.2449 \\ -0.4716 & -0.4533 \\ -0.3364 & 0.8294 \\ -0.0033 & 0.1697 \\ -0.7982 & -0.1331 \end{bmatrix}$$

e) $\Sigma' = \begin{bmatrix} 14.6516 & 0 \\ 0 & 0 \end{bmatrix}$

$$U\Sigma' V^T = \begin{bmatrix} -0.1649 & 0.2449 \\ -0.4716 & -0.4533 \\ -0.3364 & 0.8294 \\ -0.0033 & 0.1697 \\ -0.7982 & -0.1331 \end{bmatrix} \begin{bmatrix} 14.6516 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0.4261 & 0.6150 & 0.6634 \\ -0.0146 & -0.7285 & 0.6847 \end{bmatrix}$$

$$= \begin{bmatrix} -1.02948 & -1.48587 & -1.60281 \\ -2.94422 & -4.24946 & -4.58389 \\ -2.10016 & -3.03121 & -3.26976 \\ -0.020600 & -0.029735 & -0.03207 \\ -4.9832 & -7.19337 & -77584 \end{bmatrix}$$