LDA: $f_k(x) = \sim e^{-\frac{1}{2}(x-\mu_k)^T\Sigma^{-1}(x-\mu_k)}$

4.2(a) target coded as $-\frac{N}{N_1}, \frac{N}{N_2} \Rightarrow \pi_1 = \frac{N_1}{N}, \pi_2 = \frac{N_2}{N}$

$$\lg\frac{f_1(x)}{f_2(x)} = \lg\frac{\sim e^{-\frac{1}{2}(x-\mu_1)^T\Sigma^{-1}(x-\mu_1)}\frac{N_1}{N}}{\sim e^{-\frac{1}{2}(x-\mu_2)^T\Sigma^{-1}(x-\mu_2)}\frac{N_2}{N}} = \lg\frac{N_1}{N_2} - \frac{1}{2}\left[(x-\mu_1)^T\Sigma^{-1}(x-\mu_1) - (x-\mu_2)^T\Sigma^{-1}(x-\mu_2)\right]$$

$$= \lg\frac{N_1}{N_2} - \frac{1}{2}\left[-2\mu_1^T\Sigma^{-1}x + \mu_1^T\Sigma^{-1}\mu_1 + 2\mu_2^T\Sigma^{-1}x - \mu_2^T\Sigma^{-1}\mu_2\right]$$

$$= \lg\frac{N_1}{N_2} + \mu_1^T\Sigma^{-1}x - \mu_2^T\Sigma^{-1}x - \frac{1}{2}\mu_1^T\Sigma^{-1}\mu_1 + \frac{1}{2}\mu_2^T\Sigma^{-1}\mu_2$$

To classify to class 2, $\lg\frac{f_1(x)}{f_2(x)} < 0 \Rightarrow \lg\frac{N_1}{N_2} + \mu_1^T\Sigma^{-1}x - \mu_2^T\Sigma^{-1}x + \frac{1}{2}\mu_2^T\Sigma^{-1}\mu_2 - \frac{1}{2}\mu_1^T\Sigma^{-1}\mu_1 < 0$

$$\Rightarrow x^T\Sigma^{-1}(\mu_2-\mu_1) > \frac{1}{2}\mu_2^T\Sigma^{-1}\mu_2 - \frac{1}{2}\mu_1^T\Sigma^{-1}\mu_1 + \lg\frac{N_1}{N} - \lg\frac{N_2}{N}$$

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

$$= (Y - \beta_0\mathbb{1} - X\beta)^T(Y - \beta_0\mathbb{1} - X\beta)$$

$$\frac{\partial Loss}{\partial\beta_0} = -\mathbb{1}^T(Y-\beta_0\mathbb{1}-X\beta) = 0 \Rightarrow \beta_0\cdot\mathbb{1}^T\mathbb{1} = \mathbb{1}^T(Y-X\beta) \Rightarrow \beta_0 = \frac{\mathbb{1}^T(Y-X\beta)}{N} \cdots$$

$$\frac{\partial Loss}{\partial\beta} = 2X^T(Y-\beta_0\mathbb{1}-X\beta) = 0 \Rightarrow X^TX\beta - X^TY + X^T\beta_0\mathbb{1} = 0 \quad \cdots ②$$

$$X^TX\beta - X^TY + X^T\mathbb{1}\cdot\frac{\mathbb{1}^T(Y-X\beta)}{N} = 0$$

$① \to ②$: $X^TX\beta - X^TY + \frac{X^T\mathbb{1}\mathbb{1}^TY}{N} - \frac{X^T\mathbb{1}\mathbb{1}^TX\beta}{N} = 0$

$$\Rightarrow X^TX\beta - \frac{X^T\mathbb{1}\mathbb{1}^TX}{N}\beta = X^TY - \frac{X^T\mathbb{1}\mathbb{1}^TY}{N}$$

$$\Rightarrow \left(X^TX - \frac{X^T\mathbb{1}\mathbb{1}^TX}{N}\right)\beta = X^TY - \frac{X^T\mathbb{1}\mathbb{1}^TY}{N}$$

$$(N-2)\hat{\Sigma} = \sum_{i=1}^{N} x_i x_i^T - N_1\hat{\mu}_1\hat{\mu}_1^T - N_2\hat{\mu}_2\hat{\mu}_2^T = X^TX - N_1\hat{\mu}_1\hat{\mu}_1^T - N_2\hat{\mu}_2\hat{\mu}_2^T$$

$$\hat{\Sigma}_B = \hat{\mu}_1\hat{\mu}_1^T - \hat{\mu}_1\hat{\mu}_2^T - \hat{\mu}_2\hat{\mu}_1^T + \hat{\mu}_2\hat{\mu}_2^T$$

$$(N-2)\hat{\Sigma} + \frac{N_1N_2}{N}\hat{\Sigma}_B = X^TX - \frac{N_1^2}{N}\hat{\mu}_1\hat{\mu}_1^T - \frac{N_2^2}{N}\hat{\mu}_2\hat{\mu}_2^T - \frac{N_1N_2}{N^2}\hat{\mu}_1\hat{\mu}_2 - \frac{N_1N_2}{N^2}\hat{\mu}_2$$

$$= X^TX - \frac{X^T\mathbb{1}\mathbb{1}^TX}{N}$$

右: $X^Ty = N(\hat{\mu}_2 - \hat{\mu}_1)$, $X^T\mathbb{1}\mathbb{1}^TY = (x_1, \cdots x_N)\begin{pmatrix}1\\ \vdots\\ 1\end{pmatrix}(1, \cdots 1)Y$

$$\left(-\frac{N}{N_1}\right)\cdot N_1 + \frac{N}{N_2}\cdot N_2 = 0.$$

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

(d) Let the coding be $\begin{cases} \text{class 1: } a \\ \text{class 2: } b. \end{cases}$

$\hat{\Sigma}_B\hat{\beta}$ doesn't change with coding, $\propto (\hat{\mu}_2-\hat{\mu}_1)$

$$X^T y - \frac{X^T \mathbb{1}\mathbb{1}^T Y}{N} = (N_1 a\hat{\mu}_1 + N_2 b\hat{\mu}_2) - \hat{\mu}(N_1 a + N_2 b)$$
$$= N_1 a(\hat{\mu}_1-\hat{\mu}) + N_2 b(\hat{\mu}_2-\hat{\mu})$$
$$= N_1 a \cdot \frac{N_2}{N}(\hat{\mu}_1-\hat{\mu}_2) + N_2 b\cdot\frac{N_1}{N}(\hat{\mu}_2-\hat{\mu}_1)$$
$$= \frac{N_1 N_2(\mu_2-\mu_1)(b-a)}{N} \propto (\hat{\mu}_2-\hat{\mu}_1)$$

(e) $\hat{f}(x) = \beta_0 + \hat{\beta}x^T$

$$= \frac{\mathbb{1}^T(Y-X\hat{\beta})}{N} + \hat{\beta}x^T$$
$$= \frac{\mathbb{1}^T Y}{N} + (x^T - \frac{\mathbb{1}^T X}{N})\hat{\beta}$$

$\because Y$ is coded as $-\frac{N}{N_1}$ and $\frac{N}{N_2}$, $\frac{\mathbb{1}^T Y}{N} = 0$

$\therefore \hat{f}(x) > 0 \Rightarrow (x^T - \frac{\mathbb{1}^T X}{N})\hat{\beta} > 0.$

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

$\therefore (x^T - \frac{\mathbb{1}^T X}{N})\Sigma^{-1}(\hat{\mu}_2-\hat{\mu}_1) > 0$

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

When $N_1 = N_2$, $\Rightarrow x^T\Sigma^{-1}(\hat{\mu}_2-\hat{\mu}_1) = \hat{\mu}_2^T\Sigma^{-1}\hat{\mu}_2 - \hat{\mu}_1^T\Sigma^{-1}\hat{\mu}_1$

The same as LDA decision boundary given $N_1 = N_2$.

Q1: $\max a^T B a$ s.t. $a^T W a = 1$

$$\Leftrightarrow \max \frac{a^T B a}{a^T W a}$$

$$\frac{d}{da} \frac{a^T B a}{a^T W a} = \frac{2Ba(a^T W a) - 2(a^T B a) W a}{(a^T W a)^2} = 0$$

$$\Rightarrow Ba(a^T W a) = (a^T B a) W a$$

$$\Rightarrow Ba = \frac{a^T B a}{a^T W a} W a$$

$$\Rightarrow Ba = \lambda(a) W a$$

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

$\Rightarrow a$ is eigen vector of $W^{-1}B$, eigen value is $\lambda(a)$