4.2 (b)

Setting $X^* = [1, X]$ and $\beta^* = \begin{bmatrix} \beta_0 \\ \beta \end{bmatrix}$

using block notation we get:

$$\sum_{i=1}^{N} (y_i - \beta_0 - x_i^T \beta)^2 = (y - X^*\beta^*)^T (y - X^*\beta^*)$$

Then, as usual, we obtain:

$$X^{*T} X^* \beta^* = X^{*T} y$$

Solving RHS:

$$X^{*T} y = \begin{bmatrix} 1^T \\ x_i^T \\ \vdots \\ x_N^T \end{bmatrix} \begin{bmatrix} -N/N_1 \\ \vdots \\ -N/N_1 \\ N/N_2 \\ \vdots \\ N/N_2 \end{bmatrix} \begin{matrix} \left.\vphantom{\begin{matrix} a \\ b \\ c \end{matrix}}\right\} N_1 \text{ of class } 1 \\ \left.\vphantom{\begin{matrix} a \\ b \\ c \end{matrix}}\right\} N_2 \text{ of class } 2 \end{matrix} = \begin{bmatrix} 0 \\ N(\hat{\mu}_2 - \hat{\mu}_1) \end{bmatrix} \begin{matrix} \} \text{ 1st row} \\ \} \text{ remaining } p\text{-rows} \end{matrix}$$

Because,

(a) $1^T y = N_1(-N/N_1) + N_2(N/N_2) = -N + N = 0$

(b) $X^T y = -N/N_1 \sum_{i=1}^{N_1} x_i + N/N_2 \sum_{i=N_1+1}^{N_2} x_i = N(\hat{\mu}_2 - \hat{\mu}_1)$

And solving LHS:

$$\begin{bmatrix} 1^T \\ X^T \end{bmatrix} [1, X] \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta} \end{bmatrix} \quad \text{where } X = \begin{bmatrix} x_1^T \\ \vdots \\ x_n^T \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} N & \sum_i x_i^T \\ \sum_i x_i & \sum_i x_i x_i^T \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta} \end{bmatrix}, \text{ with dimensions} \begin{bmatrix} (1\times1) & (1\times p) \\ (p\times1) & (p\times p) \end{bmatrix} \begin{bmatrix} (1\times1) \\ (p\times1) \end{bmatrix}$$

$\Rightarrow$ Now $\hat{\Sigma} = \sum_k^K \sum_{g_i \in k} (x_i - \hat{\mu}_k)(x_i - \hat{\mu}_k)^T / N-K$ (P109)

So we have:

$$\hat{\Sigma} = \frac{1}{N-2}\left( \sum_{i=1}^{N_1} (x_i - \hat{\mu}_1)(x_i - \hat{\mu}_1)^T + \sum_{i=N_1}^{N_1+N_2} (x_i - \hat{\mu}_2)(x_i - \hat{\mu}_2)^T \right)$$

$$= \frac{1}{N-2}\left[\left(\sum_{i=1}^{N_1} x_i x_i^T + N_1 \hat{\mu}_1 \hat{\mu}_1^T - \sum_{i=1}^{N_1} x_i \hat{\mu}_1^T - \sum_{i=1}^{N_1} \hat{\mu}_1 x_i^T\right) + \left(\sum_{i=N_1} x_i x_i^T + N_2 \hat{\mu}_2 \hat{\mu}_2^T - \sum_{i=N_1} (x_i \hat{\mu}_2^T) - \sum_{i=N_1} \hat{\mu}_2 x_i^T\right)\right] \quad (*)$$

Notice $\sum_{i=1}^{K} x_i \hat{\mu}_K^T = \frac{N_K}{N_K} \sum x_i \hat{\mu}_K^T = N_K \hat{\mu}_K \hat{\mu}_K^T$

So $(*)$ cancels and becomes

$$\frac{1}{N-2}\left[\left(\sum_{i=1}^{N_1} x_i x_i^T - N_1 \hat{\mu}_1 \hat{\mu}_1^T\right) + \left(\sum_{i=N_1}^{N} x_i x_i^T - N_2 \hat{\mu}_2 \hat{\mu}_2^T\right)\right]$$

$$\Rightarrow \frac{1}{N-2}\left[\sum_{i=1}^{N} x_i x_i^T - N_1 \hat{\mu}_1 \hat{\mu}_1^T - N_2 \hat{\mu}_2 \hat{\mu}_2^T\right] = \hat{\Sigma}$$

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

Also Notice: $\sum_{i=1}^{N} x_i = \sum_{i=1}^{N_1} x_i + \sum_{i=N_1+1}^{N} x_i = N_1 \hat{\mu}_1 + N_2 \hat{\mu}_2$ (2)

Subbing (1) and (2) into the LHS we are left with two equations and two unknowns e.g.

$$\begin{bmatrix} N & , N_1 \hat{\mu}_1^T + N_2 \hat{\mu}_2^T \\ N_1 \hat{\mu}_1 + N_2 \hat{\mu}_2 & , (N-2)\hat{\Sigma} + N_1 \hat{\mu}_1 \hat{\mu}_1^T + N_2 \hat{\mu}_2 \hat{\mu}_2^T \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta} \end{bmatrix} = \begin{bmatrix} 0 \\ N(\hat{\mu}_2 - \hat{\mu}_1) \end{bmatrix}$$

Solving the first we get:

$$N\hat{\beta}_0 + (N_1 \hat{\mu}_1^T + N_2 \hat{\mu}_2^T)\hat{\beta} = 0$$

$$\Rightarrow \hat{\beta}_0 = -\frac{1}{N}(N_1 \hat{\mu}_1^T + N_2 \hat{\mu}_2^T)\hat{\beta}$$

and subbing into the second eqn. we get:

$$\hat{\beta}(N_1\hat{\mu}_1 + N_2\hat{\mu}_2)\left(-\frac{1}{N}(N_1\hat{\mu}_1^T + N_2\hat{\mu}_2^T)\right) + \hat{\beta}\left[(N-2)\hat{\Sigma} + N_1\hat{\mu}_1\hat{\mu}_1^T + N_2\hat{\mu}_2\hat{\mu}_2^T\right]$$

$$= N(\hat{\mu}_2 - \hat{\mu}_1) \quad ③$$

Now we just need to show that:

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

$$= N \cdot \frac{N_1 N_2}{N}(\hat{\mu}_2 - \hat{\mu}_1)(\hat{\mu}_2 - \hat{\mu}_1)^T = N\hat{\Sigma}_B$$

$$\text{LHS} = -\frac{N_1N_2}{N}\left[\left(\frac{\hat{\mu}_1}{N_2} + \frac{\hat{\mu}_2}{N_1}\right)(N_1\hat{\mu}_1^T + N_2\hat{\mu}_2^T) - \frac{N}{N_2}\hat{\mu}_1\hat{\mu}_1^T - \frac{N}{N_1}\hat{\mu}_2\hat{\mu}_2^T\right]$$

$$= -\frac{N_1N_2}{N}\left[\frac{N_1}{N_2}\hat{\mu}_1\hat{\mu}_1^T + \frac{N_2}{N_1}\hat{\mu}_2\hat{\mu}_2^T + \hat{\mu}_1\hat{\mu}_2^T + \hat{\mu}_2\hat{\mu}_1^T - \frac{N}{N_2}\hat{\mu}_1\hat{\mu}_1^T - \frac{N}{N_1}\hat{\mu}_2\hat{\mu}_2^T\right]$$

Now using $\frac{N_1 - N}{N_2} = -\frac{N_2}{N_2} = -1$

$$= -\frac{N_1N_2}{N}\left[-\hat{\mu}_1\hat{\mu}_1^T - \hat{\mu}_2\hat{\mu}_2^T + \hat{\mu}_1\hat{\mu}_2^T + \hat{\mu}_2\hat{\mu}_1^T\right]$$

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

finally, subbing this result into ③ we get:

$$\left[(N-2)\hat{\Sigma} + N\hat{\Sigma}_B\right]\hat{\beta} = N(\hat{\mu}_2 - \hat{\mu}_1) \quad \text{as required.}$$