Ex. 4.2 (d)

Here we need to show that the results of Part (b) hold in the more general case where we have coded the two classes as $a, b \in \mathbb{R}$ s.t $a \neq b$. In this case $X^{*T} X^* \beta^*$ is unchanged but $X^{*T} y$ becomes:

$$\Rightarrow \begin{bmatrix} 1^T \\ x_1^T \\ \vdots \\ x_N^T \end{bmatrix} \begin{bmatrix} a \\ \vdots \\ a \\ b \\ \vdots \\ b \end{bmatrix} \begin{matrix} \left.\vphantom{\begin{matrix} a \\ \vdots \\ a \end{matrix}}\right\} N_1 \text{ times} \\ \left.\vphantom{\begin{matrix} b \\ \vdots \\ b \end{matrix}}\right\} N_2 \text{ times} \end{matrix} = \begin{bmatrix} N_1 a + N_2 b \\ a \sum_{i=1}^{N_1} x_i + b \sum_{i=N_1+1}^{N} x_i \end{bmatrix}$$

Now for solving for $\hat{\beta}_0$ and $\hat{\beta}$ we get:

1st eqn. $= N\hat{\beta}_0 + (N_1 \hat{\mu}_1^T + N_2 \hat{\mu}_2^T)\hat{\beta} = N_1 a + N_2 b$

$$\Rightarrow \hat{\beta}_0 = \frac{1}{N}\left[N_1 a + N_2 b - (N_1 \hat{\mu}_1^T + N_2 \hat{\mu}_2^T)\hat{\beta}\right]$$

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

$$\Rightarrow \frac{1}{N}(N_1\hat{\mu}_1 + N_2\hat{\mu}_2)(N_1 a + N_2 b) + \left[(N-2)\hat{\Sigma} + N\hat{\Sigma}_B\right]\hat{\beta}$$
$$= a\sum_{i=1}^{N_1} x_i + \sum_{i=N_1+1}^{N} x_i \qquad \left(\text{as Proved in Part (b)}\right)$$

e.g. we wish to solve:

$$a N_1 \hat{\mu}_1 + b N_2 \hat{\mu}_2 - \frac{1}{N}(N_1\hat{\mu}_1 + N_2\hat{\mu}_2)(N_1 a + N_2 b) \qquad (*)$$

grouping the 'a' terms gives:

$$a N_1 \hat{\mu}_1 - \frac{N_1^2}{N}\hat{\mu}_1 a - \frac{N_1 N_2}{N}\hat{\mu}_2 a$$

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

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

Notice!

$$N_1 - \frac{N_1^2}{N} = \frac{N N_1 - N_1^2}{N}$$

$$= \frac{N_1^2 - N_1^2 + N_1 N_2}{N} = \frac{N_1 N_2}{N}$$

and by grouping the 'b' terms is a similar way, (*) becomes:

$$\frac{N_1 N_2}{N} a(\hat{\mu}_1 - \hat{\mu}_2) + \frac{N_1 N_2}{N} b(\hat{\mu}_2 - \hat{\mu}_1) = \frac{N_1 N_2}{N}(b-a)(\hat{\mu}_2 - \hat{\mu}_1)$$

Therefore, in this more general case eqn. 4.56 becomes:

$$\left[(N-2)\hat{\Sigma} + N\hat{\Sigma}_B\right]\hat{\beta} = \frac{N_1 N_2}{N}(b-a)(\hat{\mu}_2 - \hat{\mu}_1)$$

It is easy to verify that in the case where $a = -N/N_1$ and $b = N/N_2$, this reduces back to eqn. 4.56.

Finally, it is clear that the results from Part (c) also hold for this eqn. and thus the result holds for any distinct coding of the two classes.