Ex 4.3

$$\delta_k(x) = \overset{(1^*)}{x^T \Sigma^{-1} \mu_k} - \overset{(2^*)}{\tfrac{1}{2}\mu_k^T \Sigma^{-1} \mu_k} + \overset{(3^*)}{\log \pi_k} \qquad (*)$$

Now transforming X to $\hat{Y}$ (and thus $\Sigma \to \hat{\Sigma}$, $\mu \to \hat{\mu}$) via linear regression, we wish to show that LDA using $\hat{Y}$ is identical to LDA in the original space e.g.

$$\overset{(1)}{y^T \hat{\Sigma}^{-1} \hat{\mu}_k} - \overset{(2)}{\tfrac{1}{2}\hat{\mu}_k^T \hat{\Sigma}^{-1} \hat{\mu}_k} + \overset{(3)}{\log \pi_k} = (*)$$

We will begin with a few observations. Firstly noticing the change in dimensionality:

(a) $x^T \to \cancel{(P \times N)}\ (1 \times P)$

$\hat{y}^T = x^T \hat{\beta} \to (1 \times P)(P \times K) = (1 \times K)$

$\Rightarrow$ We expect Σ^{-1} $(P \times P)$ to change such that we have $\hat{\Sigma}^{-1}$ $(K \times K)$

(b) $$\hat{\mu}_k = \sum_{g_i \in k} \hat{y}_i / N_k = \sum_{g_i \in k} \frac{\hat{\beta}^T x_i}{N_k}$$

Now we will use the fact that Y acts as an indicator matrix where each column corresponds to a class $k \in K$, and acts as an indicator vector taking the value 1 if row i is of class k and 0 otherwise. Thus we may transform $\hat{\mu}_k$ from a ~~(1×K)~~ $(K \times 1)$ mean vector for ~~observa~~ ~~class $k \in K$~~ ~~each class~~ ~~to~~ each ~~column~~ observation of class $g_i \in K$ to a more general $\hat{\mu}$ matrix $(K \times K)$ with each of the K ~~rows~~ columns corresponding to the mean vector $\hat{\mu}_k$ of that class.

$\Rightarrow \hat{\mu} = \hat{\beta}^T X^T Y D^{-1}$ where D is a $(K \times K)$ diagonal matrix with each entry corresponding to the number of observations in class k, e.g.

$$D = \begin{pmatrix} N_1 & & \cdots & 0 \\ \vdots & N_2 & & \vdots \\ & & \ddots & \\ 0 & \cdots & & N_K \end{pmatrix}$$

(c) Part (b) can also be demonstrated by noticing that $\hat{\beta}^T X^T = \hat{Y}^T$, thus making it clear that $\hat{\mu}_K = \frac{\hat{Y}^T y_K}{N_K}$ is simply the mean vector of $g_i = K$.

Similarly, we notice that $\frac{X^T y_K}{N_K} = \mu_K$.

(d) $$\hat{\Sigma} = \frac{1}{N-K} \sum_{K=1} \sum_{g_i \in K} (\hat{y}_i - \hat{\mu}_K)(\hat{y}_i - \hat{\mu}_K)^T$$

$$= \frac{1}{N-K} \sum_K \sum_{g_i} (\beta^T x_i - \hat{\mu}_K)(\beta^T x_i - \hat{\mu}_K)^T$$

$$= \frac{1}{N-K} \sum_K \sum_{g_i} (\beta^T x_i - \beta^T \mu_K)(\beta^T x_i - \beta^T \mu_K)^T \quad \text{(using (c))}$$

$$= \beta^T \left[\frac{1}{N-K} \sum_K \sum_{g_i} (x_i - \mu_K)(x_i - \mu_K)^T \right] \beta$$

$$= \beta^T \Sigma \beta \quad \text{(by definition)}$$

Now, returning to the original problem we will attempt to prove that 1, 2, 3 = 1*, 2*, 3* respectively.

① In fact, we will attempt to show $\hat{Y} \hat{\Sigma}^{-1} \hat{\mu} = X \Sigma^{-1} \mu$ as discussed in (b).

Now,

$$\hat{Y} \hat{\Sigma}^{-1} \hat{\mu} = (X\beta)(\beta^T \Sigma \beta)^{-1} (\beta^T X^T Y D^{-1}) \quad \text{(using (b) and (d))}$$

Now assuming β is full rank, then an inverse exists such that $\beta \beta^{-1} = \mathbb{I}_p$

$$= X \beta \beta^{-1} \Sigma^{-1} (\beta \beta^{-1})^T X^T Y D^{-1}$$

$$= X \Sigma^{-1} X^T Y D^{-1}$$

$$= X \Sigma^{-1} \mu \quad \text{(using (c))}$$

as required.

② W.T.S. $\hat{\mu}_k^T \hat{\Sigma}^{-1} \hat{\mu}_k = \mu_k^T \Sigma^{-1} \mu_k$

$\Rightarrow \hat{\mu}_k^T \hat{\Sigma}^{-1} \hat{\mu}_k$

$= (\beta^T \mu_k)^T (\beta^T \Sigma \beta)^{-1} (\beta^T \mu_k)$

$= \mu_k^T \Sigma^{-1} \mu_k$ (assuming β full rank as in ①)

③ $\log \pi_k = \log \pi_k$

Combining ①, ② and ③ we have proved that transforming X to $\hat{Y}$ gives identical results in LDA.