4. For OLS in this case we have a classification rule of

$\hat{\beta}_0 + \hat{\beta}_1 x < C$ meaning our decision boundary is

$$x^* = \frac{c - \hat{\beta}_0}{\hat{\beta}_1}$$

So we will first use the normal equation to calculate $\hat{\beta}$

$$X^T X \beta = X^T y$$

$$\overbrace{\begin{bmatrix} 1 & \cdots & 1 \\ x_1 & \cdots & x_N \end{bmatrix} \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_N \end{bmatrix}}^{X^T X} = \begin{bmatrix} N & \Sigma x_i \\ \Sigma x_i & \Sigma x_i^2 \end{bmatrix} \qquad \overbrace{\begin{bmatrix} 1 & \cdots & 1 \\ x_1 & \cdots & x_N \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix}}^{X^T Y} = \begin{bmatrix} \Sigma y_i \\ \Sigma x_i y_i \end{bmatrix}$$

We ultimately want to show that the x^* for OLS is the same as the x^* for LDA so we will define for LDA

$N_1 = \Sigma y_i$

$N_0 = N - N_1$

$\hat{\mu}_0 = \sum_{i:g_0} x_i / N_0 =$

$\hat{\mu}_1 = \sum_{i:g_1} x_i / N_1 = \frac{\Sigma x_i y_i}{N_1}$

thus we can rewrite $X^T X = \begin{bmatrix} N & N_0\hat{\mu}_0 + N_1\hat{\mu}_1 \\ N_0\hat{\mu}_0 + N_1\hat{\mu}_1 & \Sigma x_i^2 \end{bmatrix}$

$X^T y = \begin{bmatrix} N_1 \\ N_1 \hat{\mu}_1 \end{bmatrix}$

Now we solve for $\hat{\sigma}^2 = \frac{1}{N-K} \sum_{k=1}^{K} \sum_{i:g_i = g_k} (x_i - \hat{\mu}_k)^2$

$$\hat{\sigma}^2 = \frac{1}{N-2}\left(\sum_{i:g_0} (x_i - \hat{\mu}_0)^2 + \sum_{i:g_1} (x_i - \hat{\mu}_1)^2 \right)$$

$$= \frac{1}{N-2}\left(\sum_{i:g_0} (x_i^2 - 2x_i\hat{\mu}_0 + \hat{\mu}_0^2) + \sum_{i:g_1} (x_i^2 - 2x_i\hat{\mu}_1 + \hat{\mu}_1^2) \right)$$

$$\hat{\sigma}^2(N-2) = \sum_{i:g_0} x_i^2 - 2\hat{\mu}_0 \sum_{i:g_0} x_i + \overset{\nearrow N_0}{(N-N_1)}\hat{\mu}_0^2 + \sum_{i:g_1} x_i^2 - 2\hat{\mu}_1 \sum_{i:g_1} x_i + N_1 \hat{\mu}_1^2$$

$$\hat{\sigma}^2(N-2) = \sum_i x_i^2 - 2\hat{\mu}_0^2 N_0 + \overset{\nearrow N_0\hat{\mu}_0^2}{\boxed{N\hat{\mu}_0^2 - N_1\hat{\mu}_0^2}} - 2\hat{\mu}_1^2 N_1 + N_1\hat{\mu}_1^2$$

$$\hat{\sigma}^2(N-2) = \sum_i x_i^2 - N_0\hat{\mu}_0^2 - N_1\hat{\mu}_1^2$$

$$\sum_i x_i^2 = \hat{\sigma}^2(N-2) + N_0\hat{\mu}_0^2 + N_1\hat{\mu}_1^2$$

Returning to solving for $\hat{\beta}$ we have

$$\begin{bmatrix} N & N_0\hat{\mu}_0 + N_1\hat{\mu}_1 \\ N_0\hat{\mu}_0 + N_1\hat{\mu}_1 & \Sigma x_i^2 \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = \begin{bmatrix} N_1 \\ N_1\hat{\mu}_1 \end{bmatrix}$$

$$N\hat{\beta}_0 + (N_0\hat{\mu}_0 + N_1\hat{\mu}_1)\hat{\beta}_1 = N_1$$
$$(N_0\hat{\mu}_0 + N_1\hat{\mu}_1)\hat{\beta}_0 + (\Sigma x_i^2)\hat{\beta}_1 = N_1\hat{\mu}_1$$

We first use elimination to find β_1 and plug in for Σx_i^2

$$\cancel{N(N_0\hat{\mu}_0 + N_1\hat{\mu}_1)\hat{\beta}_0} + (N_0\hat{\mu}_0 + N_1\hat{\mu}_1)^2\hat{\beta}_1 = N_1(N_0\hat{\mu}_0 + N_1\hat{\mu}_1)$$
$$\cancel{-N(N_0\hat{\mu}_0 + N_1\hat{\mu}_1)\hat{\beta}_0} - N(\Sigma x_i^2)\hat{\beta}_1 = -NN_1\hat{\mu}_1$$

$$\left((N_0\hat{\mu}_0 + N_1\hat{\mu}_1)^2 - N(\hat{\sigma}^2(N-2) + N_0\hat{\mu}_0^2 + N_1\hat{\mu}_1^2)\right)\hat{\beta}_1 = N_1N_0\hat{\mu}_0 + N_1^2\hat{\mu}_1 - NN_1\hat{\mu}_1$$

$$\hat{\beta}_1 = \frac{N_1N_0\hat{\mu}_0 + N_1\hat{\mu}_1(N-N_0) - NN_1\hat{\mu}_1}{N_0^2\hat{\mu}_0^2 + 2N_0N_1\hat{\mu}_0\hat{\mu}_1 + N_1^2\hat{\mu}_1^2 - N\hat{\sigma}^2(N-2) - NN_0\mu_0^2 - NN_1\hat{\mu}_1^2}$$

$$\hat{\beta}_1 = \frac{N_1N_0\hat{\mu}_0 - N_1N_0\hat{\mu}_1}{N_0\hat{\mu}_0^2(N-N_1) + 2N_0N_1\hat{\mu}_0\hat{\mu}_1 + N_1\hat{\mu}_1^2(N-N_0) - N\hat{\sigma}^2(N-2) - NN_0\mu_0^2 - NN_1\hat{\mu}_1^2}$$

$$\hat{\beta}_1 = \frac{N_1N_0\hat{\mu}_0 - N_1N_0\hat{\mu}_1}{-N_1N_0\hat{\mu}_0^2 + 2N_0N_1\hat{\mu}_0\hat{\mu}_1 - N_1N_0\hat{\mu}_1^2 - N\hat{\sigma}^2(N-2)}$$

$$\hat{\beta}_1 = \frac{N_1N_0(\hat{\mu}_1 - \hat{\mu}_0)}{N\hat{\sigma}^2(N-2) + (N_1N_0(\hat{\mu}_0^2 - 2\hat{\mu}_0\hat{\mu}_1 + \hat{\mu}_1^2))} = \frac{N_1N_0(\hat{\mu}_1 - \hat{\mu}_0)}{N\hat{\sigma}^2(N-2) + N_1N_0(\hat{\mu}_0 - \hat{\mu}_1)^2}$$

Plugging back in for $\hat{\beta}_0$ we get

$$N\hat{\beta}_0 + \frac{(N_0\hat{\mu}_0 + N_1\hat{\mu}_1)(N_1N_0\hat{\mu}_1 - N_1N_0\hat{\mu}_0)}{N\hat{\sigma}^2(N-2) + (N_1N_0(\hat{\mu}_0 - \hat{\mu}_1)^2)} = N_1$$

$$N\hat{\beta}_0 + \frac{(N_1N_0^2\hat{\mu}_0\hat{\mu}_1 - N_1N_0^2\hat{\mu}_0^2 + N_1^2N_0\hat{\mu}_1^2 - N_1^2N_0\mu_0\mu_1)}{N\hat{\sigma}^2(N-2) + N_1N_0(\hat{\mu}_0 - \hat{\mu}_1)^2} = N_1$$

Scanned with CamScanner

$$N\hat{\beta}_0 + \frac{N_1N_0\hat{\mu}_0\hat{\mu}_1(N-N_1) - N_1N_0\hat{\mu}_0^2(N-N_1) + N_1^2N_0\hat{\mu}_1^2 - N_1^2N_0\hat{\mu}_0\hat{\mu}_1}{N\hat{\sigma}^2(N-2) + N_1N_0(\hat{\mu}_0-\hat{\mu}_1)^2} = N_1$$

$$N\hat{\beta}_0 + \frac{NN_1N_0\hat{\mu}_0\hat{\mu}_1 - 2N_1^2N_0\hat{\mu}_0\hat{\mu}_1 - NN_1N_0\hat{\mu}_0^2 + N_1^2N_0\hat{\mu}_0^2 + N_1^2N_0\hat{\mu}_1^2}{N\hat{\sigma}^2(N-2) + N_1N_0(\hat{\mu}_0-\hat{\mu}_1)^2} = N_1$$

$$N\hat{\beta}_0 + \frac{NN_1N_0\hat{\mu}_0(\hat{\mu}_1-\hat{\mu}_0) + N_1^2N_0(\hat{\mu}_0-\mu_1)^2}{N\hat{\sigma}^2(N-2) + N_1N_0(\hat{\mu}_0-\hat{\mu}_1)^2} = N_1$$

$$\cancel{N}\hat{\beta}_0 = \frac{\cancel{N}N_1\hat{\sigma}^2(N-2) + \cancel{N_1^2N_0(\hat{\mu}_0-\hat{\mu}_1)^2} - \cancel{N}N_1N_0\hat{\mu}_0(\hat{\mu}_1-\hat{\mu}_0) - \cancel{N_1^2N_0(\hat{\mu}_0-\hat{\mu}_1)^2}}{N\hat{\sigma}^2(N-2) + N_1N_0(\hat{\mu}_0-\hat{\mu}_1)^2}$$

$$\hat{\beta}_0 = \frac{N_1\hat{\sigma}^2(N-2) + N_1N_0\hat{\mu}_0(\hat{\mu}_0-\hat{\mu}_1)}{N\hat{\sigma}^2(N-2) + N_1N_0(\hat{\mu}_0-\hat{\mu}_1)^2}$$

For LDA, we know $x^* = \frac{\sigma^2 \log\frac{\pi_m}{\pi_k} + \frac{\mu_k^2-\mu_m^2}{2}}{\mu_k-\mu_m}$ from Lec 12 pg 68

so setting the boundary of LDA and OLS to each other we can solve for c and show the equivalence of this situation

$$\frac{\hat{\sigma}^2\log\overset{\frac{N_0}{N_1}}{\frac{\pi_0}{\pi_1}} + \frac{\hat{\mu}_1^2-\hat{\mu}_0^2}{2}}{\hat{\mu}_1-\hat{\mu}_0} = \frac{\dfrac{c(N\hat{\sigma}^2(N-2)+N_1N_0(\hat{\mu}_0-\hat{\mu}_1)^2) - N_1\hat{\sigma}^2(N-2) - N_1N_0\hat{\mu}_0(\hat{\mu}_0-\hat{\mu}_1)}{N\hat{\sigma}^2(N-2)+N_1N_0(\hat{\mu}_0-\hat{\mu}_1)^2}}{\dfrac{N_1N_0(\hat{\mu}_1-\hat{\mu}_0)}{N\hat{\sigma}^2(N-2)+N_1N_0(\hat{\mu}_0-\hat{\mu}_1)^2}}$$

$$\frac{\hat{\sigma}^2\log\frac{N_0}{N_1} + \frac{\hat{\mu}_1^2-\hat{\mu}_0^2}{2}}{\hat{\mu}_1-\hat{\mu}_0} = \frac{c(N\hat{\sigma}^2(N-2)+N_1N_0(\hat{\mu}_0-\hat{\mu}_1)^2) - N_1\hat{\sigma}^2(N-2) - N_1N_0\hat{\mu}_0(\hat{\mu}_0-\hat{\mu}_1)}{N_1N_0(\hat{\mu}_1-\hat{\mu}_0)}$$

$$\frac{N_1N_0\sigma^2\log\frac{N_0}{N_1} + N_1N_0\left(\frac{\hat{\mu}_1^2-\hat{\mu}_0^2}{2}\right)}{N_1N_0(\hat{\mu}_1-\hat{\mu}_0)} = \quad ||$$

$$C = \frac{N_1 N_0 \hat{\sigma}^2 \log \frac{N_0}{N_1} + N_1 N_0 \left(\frac{\hat{\mu}_1^2 - \hat{\mu}_0^2}{2}\right) + N_1 \hat{\sigma}^2 (N-2) + N_1 N_0 \hat{\mu}_0 (\hat{\mu}_0 - \hat{\mu}_1)}{N \hat{\sigma}^2 (N-2) + N_1 N_0 (\hat{\mu}_0 - \hat{\mu}_1)^2}$$

$$C = \frac{N_1 \hat{\sigma}^2 \left(N_0 \log \frac{N_0}{N_1} + N - 2\right) + N_1 N_0 \left(\frac{\hat{\mu}_1^2 - \hat{\mu}_0^2}{2}\right) + N_1 N_0 \hat{\mu}_0 (\hat{\mu}_0 - \hat{\mu}_1)}{N \hat{\sigma}^2 (N-2) + N_1 N_0 (\hat{\mu}_0 - \hat{\mu}_1)^2}$$

$$C = \frac{2 N_1 \hat{\sigma}^2 \left(N_0 \log \frac{N_0}{N_1} + N - 2\right) + N_1 N_0 (\hat{\mu}_1^2 - \hat{\mu}_0^2) + 2 N_1 N_0 \hat{\mu}_0 (\hat{\mu}_0 - \hat{\mu}_1)}{2 N \hat{\sigma}^2 (N-2) + 2 N_1 N_0 (\hat{\mu}_0 - \hat{\mu}_1)^2}$$

$$C = \frac{2 N_1 \hat{\sigma}^2 \left(N_0 \log \frac{N_0}{N_1} + N - 2\right) + N_1 N_0 \hat{\mu}_1^2 + N_1 N_0 \hat{\mu}_0^2 - 2 N_1 N_0 \mu_1 \mu_0}{2 N \hat{\sigma}^2 (N-2) + 2 N_1 N_0 (\hat{\mu}_0 - \hat{\mu}_1)^2}$$

$$C = \frac{2 N_1 \hat{\sigma}^2 \left(N_0 \log \frac{N_0}{N_1} + N - 2\right) + N_1 N_0 (\hat{\mu}_0 - \hat{\mu}_1)^2}{2 N \hat{\sigma}^2 (N-2) + 2 N_1 N_0 (\hat{\mu}_0 - \hat{\mu}_1)^2}$$

If $N_1 = N_0 = \frac{N}{2}$

$$C = \frac{N \hat{\sigma}^2 (N-2) + \frac{N^2}{4} (\hat{\mu}_0 - \hat{\mu}_1)^2}{2 N \hat{\sigma}^2 (N-2) + \frac{N^2}{2} (\hat{\mu}_0 - \hat{\mu}_1)^2} = \frac{4 N \hat{\sigma}^2 (N-2) + N^2 (\hat{\mu}_0 - \hat{\mu}_1)^2}{8 N \hat{\sigma}^2 (N-2) + 2 N^2 (\hat{\mu}_0 - \hat{\mu}_1)^2}$$