LDA for the case of 2 features :-

$$\frac{P\left(C=1 \,\middle|\, X=\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right)}{P\left(C=0 \,\middle|\, X=\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right)} = \frac{\dfrac{1}{(\sqrt{2\pi})^2\sqrt{|\hat{\Sigma}_1|}}\, e^{-\frac{1}{2}\begin{bmatrix} x_1-\hat{\mu}_1^{x_1} \\ x_2-\hat{\mu}_1^{x_2} \end{bmatrix}^T \hat{\Sigma}_1^{-1} \begin{bmatrix} x_1-\hat{\mu}_1^{x_1} \\ x_2-\hat{\mu}_1^{x_2} \end{bmatrix}}}{\dfrac{1}{(\sqrt{2\pi})^2\sqrt{|\hat{\Sigma}_0|}}\, e^{-\frac{1}{2}\begin{bmatrix} x_1-\hat{\mu}_0^{x_1} \\ x_2-\hat{\mu}_0^{x_2} \end{bmatrix}^T \hat{\Sigma}_0^{-1} \begin{bmatrix} x_1-\hat{\mu}_0^{x_1} \\ x_2-\hat{\mu}_0^{x_2} \end{bmatrix}}}$$

But for LDA, $\hat{\Sigma}_0 = \hat{\Sigma}_1 = \dfrac{N_0 \cdot \hat{\Sigma}_0 + N_1 \cdot \hat{\Sigma}_1}{(N_0+N_1)} = \underline{\underline{\Sigma_p}}$

So,

$$\frac{P\left(C=1 \mid X=\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right)}{P\left(C=0 \mid X=\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right)} = \frac{\cancel{\frac{1}{(\sqrt{2\pi})^2 \sqrt{|\Sigma_p|}}} e^{-\frac{1}{2}\begin{bmatrix} x_1-\hat{\mu}_1^{x_1} \\ x_2-\hat{\mu}_1^{x_2} \end{bmatrix}^T \Sigma_p^{-1} \begin{bmatrix} x_1-\hat{\mu}_1^{x_1} \\ x_2-\hat{\mu}_1^{x_2} \end{bmatrix}}}{\cancel{\frac{1}{(\sqrt{2\pi})^2 \sqrt{|\Sigma_p|}}} e^{-\frac{1}{2}\begin{bmatrix} x_1-\hat{\mu}_0^{x_1} \\ x_2-\hat{\mu}_0^{x_2} \end{bmatrix}^T \Sigma_p^{-1} \begin{bmatrix} x_1-\hat{\mu}_0^{x_1} \\ x_2-\hat{\mu}_0^{x_2} \end{bmatrix}}}$$

$$= e^{\frac{1}{2}\begin{bmatrix} x_1-\hat{\mu}_0^{x_1} \\ x_2-\hat{\mu}_0^{x_2} \end{bmatrix}^T \Sigma_p^{-1} \begin{bmatrix} x_1-\hat{\mu}_0^{x_1} \\ x_2-\hat{\mu}_0^{x_2} \end{bmatrix} - \frac{1}{2}\begin{bmatrix} x_1-\hat{\mu}_1^{x_1} \\ x_2-\hat{\mu}_1^{x_2} \end{bmatrix}^T \Sigma_p^{-1} \begin{bmatrix} x_1-\hat{\mu}_1^{x_1} \\ x_2-\hat{\mu}_1^{x_2} \end{bmatrix}}$$

Let's Imagine the Σ_p is non singular with the following structure –

$$\Sigma_p = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Therefore, now –

$$= \frac{1}{2}\begin{bmatrix} x_1 - \hat{\mu}_0^{x_1} \\ x_2 - \hat{\mu}_0^{x_2} \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}^{-1}\begin{bmatrix} x_1 - \hat{\mu}_0^{x_1} \\ x_2 - \hat{\mu}_1^{x_1} \end{bmatrix} -$$

$$e$$

$$\frac{1}{2}\begin{bmatrix} x_1 - \hat{\mu}_1^{x_1} \\ x_2 - \hat{\mu}_1^{x_2} \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}^{-1}\begin{bmatrix} x_1 - \hat{\mu}_1^{x_1} \\ x_2 - \hat{\mu}_1^{x_2} \end{bmatrix}$$

$$= e^{\frac{1}{2}\left[(x_1-\hat{\mu}_0^{x_1})^2+(x_2-\hat{\mu}_0^{x_2})^2-(x_1-\hat{\mu}_1^{x_1})^2-(x_2-\hat{\mu}_1^{x_2})^2\right]}$$

If you solve this then you will observe that the x_1^2, x_2^2 terms will cancel <u>out</u>.

$$= e^{\left[\overbrace{\frac{\hat{\mu}_0^{x_1^2} + \hat{\mu}_0^{x_2^2} + \hat{\mu}_1^{x_1^2} + \hat{\mu}_1^{x_2^2}}{2}}^{\hat{\theta}_0} - \cancel{2} x_1 \cdot \hat{\mu}_0^{x_1} - \cancel{2} x_2 \cdot \hat{\mu}_0^{x_2} + \cancel{2} x_1 \cdot \hat{\mu}_1^{x_1} + \cancel{2} x_2 \cdot \hat{\mu}_1^{x_2}\right]}$$

$$= e^{\left[\hat{\theta}_0 + \underbrace{\left(\hat{\mu}_1^{x_1} - \hat{\mu}_0^{x_1}\right)}_{\hat{\theta}_1} \cdot x_1 + \underbrace{\left(\hat{\mu}_1^{x_2} - \hat{\mu}_0^{x_2}\right)}_{\hat{\theta}_2} \cdot x_2\right]}$$

Therefore, $\hat{\theta}_1 = \frac{1}{2}\left[\hat{\mu}_0^{x_1^2} + \hat{\mu}_1^{x_1^2} + \hat{\mu}_0^{x_2^2} + \hat{\mu}_1^{x_2^2}\right]$

$\hat{\theta}_1 = \left(\hat{\mu}_1^{x_1} - \hat{\mu}_0^{x_1}\right)$

Plug our values

$\hat{\theta}_2 = \left(\hat{\mu}_1^{x_2} - \hat{\mu}_0^{x_2}\right)$

QED

$$\frac{P\left(C=1 \mid X=\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right)}{P\left(C=0 \mid X=\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right)} = e^{\left(\hat{\theta}_0 + \hat{\theta}_1 x_1 + \hat{\theta}_2 x_2\right)}$$

$$\frac{P\left(C=1 \mid X=\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right)}{1-P\left(C=1 \mid X=\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right)} = e^{\left(\hat{\theta}_0+\hat{\theta}_1 x_1+\hat{\theta}_2 x_2\right)}$$

$$\frac{1-P\left(C=1 \mid X=\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right)}{P\left(C=1 \mid X=\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right)} = e^{-\left(\hat{\theta}_0+\hat{\theta}_1 x_1+\hat{\theta}_2 x_2\right)}$$

$$\frac{1}{P\left(C=1 \middle| X=\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right)} - 1 = e^{-\left(\hat{\theta}_0 + \hat{\theta}_1 x_1 + \hat{\theta}_2 x_2\right)}$$

$$\frac{1}{P\left(C=1 \middle| X=\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right)} = 1 + e^{-\left(\hat{\theta}_0 + \hat{\theta}_1 x_1 + \hat{\theta}_2 x_2\right)}$$

$$\boxed{P\left(c=1 \,\middle|\, X=\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \frac{1}{1+e^{-(\hat{\theta}_0+\hat{\theta}_1 x_1+\hat{\theta}_2 x_2)}}}$$

$$\log_e \frac{p\left(C=1 \mid X=\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right)}{p\left(C=0 \mid X=\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right)} = \left(\hat{\theta}_0 + \hat{\theta}_1 x_1 + \hat{\theta}_2 x_2\right)$$

where $\hat{\theta} = \begin{bmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{bmatrix}$, $X = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

$$\log_e \frac{P\left(C=1 \,\middle|\, X=\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right)}{P\left(C=0 \,\middle|\, X=\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right)} = y = f(x_1, x_2) = f(X) =$$

$$(\hat{\theta}_0 + \hat{\theta}_1 x_1 + \hat{\theta}_2 x_2) = (\hat{\theta}_0 + \hat{\theta}^T X) = \text{linear}$$

discriminant function