

From gaussian multi-variate normal distribution we get the following equation for each class:

$$p(x|y = 0) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} \exp \left( -\frac{1}{2} (x - \mu_0)^T \Sigma^{-1} (x - \mu_0) \right)$$

$$p(x|y = 1) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} \exp \left( -\frac{1}{2} (x - \mu_1)^T \Sigma^{-1} (x - \mu_1) \right)$$

We have taken a common covariance matrix for both the classes because we want a linear boundry, if not taken we will get a quadratic boundry (we can see that in the solution). The covariance matrix tells about the shape of the classes in the space and orientation.

We have to also take account of the prior probability, which can be obtained from the sample:

$$p(y) = \begin{cases} y = 1 & \varnothing \\ y = 0 & 1 - \varnothing \end{cases} \text{ where } \varnothing = \frac{1}{m} \sum_{i=1}^m 1\{y^i = 1\}$$

Then we can find the individual probability from the following formula:

$$p(y = 1|x) = \frac{p(y = 1)p(x|y = 1)}{p(y = 0)p(x|y = 0) + p(y = 1)p(x|y = 1)}$$

Then we can get the boundary (i.e. hyperplane) from equating the probability to half, as the given sample will have equal chance of going into both classes:

$$\frac{p(y = 1)p(x|y = 1)}{p(y = 0)p(x|y = 0) + p(y = 1)p(x|y = 1)} = \frac{1}{2}$$

$$\Rightarrow p(y = 0)p(x|y = 0) = p(y = 1)p(x|y = 1)$$

$$\Rightarrow (\varnothing) \exp \left( -\frac{1}{2} (x - \mu_1)^T \Sigma^{-1} (x - \mu_1) \right) = (1 - \varnothing) \exp \left( -\frac{1}{2} (x - \mu_0)^T \Sigma^{-1} (x - \mu_0) \right)$$

rest of the terms gets cancelled as they are same, and now taking log on both sides we get:

$$\log(\varnothing) - \frac{1}{2} (x - \mu_1)^T \Sigma^{-1} (x - \mu_1) = \log(1 - \varnothing) - \frac{1}{2} (x - \mu_0)^T \Sigma^{-1} (x - \mu_0)$$

distributing and rearranging the terms we get the below equation as  $x^T \Sigma^{-1} x$  gets cancelled making it linear in nature otherwise it would have been quadratic,

$$\begin{aligned}
& 2\log(\varnothing) - \left( x^T \sum^{-1} x - x^T \sum^{-1} \mu_1 - \mu_1^T \sum^{-1} x + \mu_1^T \sum^{-1} \mu_1 \right) = \\
& 2\log(1 - \varnothing) - \left( x^T \sum^{-1} x - x^T \sum^{-1} \mu_0 - \mu_0^T \sum^{-1} x + \mu_0^T \sum^{-1} \mu_0 \right) \\
\Rightarrow & x^T \sum^{-1} (\mu_1 - \mu_0) + (\mu_1 - \mu_0)^T \sum^{-1} x = \mu_1^T \sum^{-1} \mu_1 - \mu_0^T \sum^{-1} \mu_0 + 2K
\end{aligned}$$

$$\text{where } K = \log(1 - \varnothing) - \log(\varnothing)$$

The blue and red highlighted terms are transpose of each other and one cross one matrix hence both of them are equal. Then we get:

$$(\mu_1 - \mu_0)^T \sum^{-1} x = \frac{1}{2} \left( \mu_1^T \sum^{-1} \mu_1 - \mu_0^T \sum^{-1} \mu_0 \right) + K$$

$(\mu_1 - \mu_0)^T \sum^{-1}$  term is a row vector and  $x$  is column vector and the RHS is a constant hence it form a hyperplane of form as follows:

$$a_1 x_1 + a_2 x_2 + \dots = a_0$$

comparing the logistic regression model we get the  $W_0$  and  $W_1$  as following:

$$W^T X + W_0 = 0 \quad (\text{logistic model})$$

$$(\mu_1 - \mu_0)^T \sum^{-1} x + \frac{1}{2} \left( \mu_0^T \sum^{-1} \mu_0 - \mu_1^T \sum^{-1} \mu_1 \right) - K = 0 \quad (GDA)$$

Then we get the values as,

$$\begin{aligned}
W_0 &= \frac{1}{2} \left( \mu_0^T \sum^{-1} \mu_0 - \mu_1^T \sum^{-1} \mu_1 \right) + \log\left(\frac{\varnothing}{1 - \varnothing}\right) \\
W^T &= (\mu_1 - \mu_0)^T \sum^{-1}
\end{aligned}$$

**Hence it can be concluded that Gaussian dicriminant analysis forms hyperplane similar too Logistic Regression.**