3) Given,

Parameter, $\theta = \{\alpha, \mu_1, \Sigma_1, \mu_2, \Sigma_2\}$

$$P(y|\theta) = \alpha^{y}(1-\alpha)^{1-y}.$$

$$P(x,y|\theta) = N(x|\mu_y, \Sigma_y)$$

Using the Parameter θ, we can generate the iid data from :-

$$P(x|y,\theta) = P(y|\theta)\, P(x|y,\theta)$$

We will use maximum likelihood, to recover the parameters from data.

$$L(\theta) = \log P(\text{Real data} \mid \theta)$$

$$= \sum_{i=1}^{N} \log P(x_i y_i|\theta)$$

According to the question there are two classes:-

Class 1 :- $y_i \in 0$

Class 2 :- $y_i \in 1$

2

$$\therefore L(\theta) = \sum_{i=1}^{N} \log\left(P(y_i|\theta) \cdot P(x_i|y_i, \theta)\right)$$

$$= \sum_{i=1}^{N} \log\left(P(y_i|\theta)\right) + \sum_{i=1}^{N} \log\left(P(x_i|y_i,\theta)\right)$$

$$= \sum_{i=1}^{N} \log P(y_i|\alpha) + \sum_{y \in 0} \log p(x_i/\mu_0 \epsilon_0) + \sum_{y \in 1} \log p(x_i|\mu_1 \epsilon_1) \rightarrow \text{I}$$

Estimating the likelihood for the about three terms would derive:-

$$P(y_i|\alpha) = \alpha^{y_i}(1-\alpha)^{1-y_i} \quad — (1)$$

$$P(x_i|\mu_0, \Sigma_0) = \frac{1}{2\pi^{D/2}(\sqrt{|\epsilon_0|})} \exp\left(-\frac{1}{2}(\vec{x}-\vec{\mu_0})^T \cdot \Sigma_0^{-1}(\vec{x}-\vec{\mu_0})\right) \rightarrow (2)$$

$$P(x_i|\mu_1, \Sigma_1) = \frac{1}{(2\pi)^{D/2}\sqrt{|\epsilon_1|}} \exp\left(-\frac{1}{2}(\vec{x}-\vec{\mu_1})^T \Sigma_1^{-1}(\vec{x}-\vec{\mu_1})\right) \rightarrow (3)$$

Differentiating eq (I) w.r.t α & equating to zero

$$\frac{\partial}{\partial}(L(\theta)) = \frac{\partial}{\partial\alpha}\sum_{i=1}^{N}\log\left(\alpha^{y_i}(1-\alpha)^{1-y_i}\right) + \frac{\partial}{\partial\alpha}\text{const} = 0$$

$$\frac{\partial}{\partial \alpha} \sum_{i=1}^{N} \left(y_i \log \alpha + (1-y_i) \log(1-\alpha) \right) = 0$$

$$\frac{\partial}{\partial \alpha} \left(\sum_{i \in class_1} \log \alpha + \sum_{i \in class_0} \log(1-\alpha) \right) = 0 \qquad \left[\begin{array}{l} \because y_i = 1 \text{ for class}_1 \\ 1-y_i = 1 \text{ for class } 0. \end{array} \right.$$

Solving for derivate

$$\sum_{i \in class_0} \frac{1}{\alpha} - \sum_{i \in class_1}$$

$$\frac{N_1}{\alpha} - \frac{N_0}{1-\alpha} = 0$$

$$\frac{N_1}{\alpha} = \frac{N_0}{1-\alpha}.$$

$$N_1(1-\alpha) = N_0(\alpha)$$

$$N_1 - N_1(\alpha) = N_0(\alpha)$$

$$N_1 = N_0(\alpha) + N_1(\alpha)$$

$$N_1 = \alpha(N_0 + N_1)$$

$$\boxed{\frac{N_1}{N_0 + N_1} = \alpha.}$$

Differentiating the equation (I) w.r.t to the given parameter μ_0 & equating to zero

$$\frac{\partial (l(\theta))}{\partial \mu_0} = \frac{\partial}{\partial \mu_0}\left(\sum_{y_i \in 0} \log\left(\frac{1}{(2\pi)^{D/2}\sqrt{|\Sigma_0|}} - \exp\left(\frac{-1}{2}(\vec{x}_i - \vec{\mu}_0)^T \Sigma_{10}^{-1} (\vec{x}_i - \vec{\mu}_0)\right.\right.\right.$$

$$= \frac{\partial}{\partial \mu_0}\left(\sum_{y_i \in 0}\left(-\frac{D}{2}\log(2\pi) - \frac{1}{2}\log(\Sigma_{10}^{-1})\right) - \frac{1}{2}(\vec{x}_i - \vec{\mu}_0)^T \Sigma_{10}^{-1} (\vec{x}_i - \vec{\mu}_0)\right.$$

Since $-\frac{D}{2}\log(2\pi)$ & $-\frac{1}{2}\log \Sigma_{10}^{-1}$ will be const their derivative is zero.

$$\frac{\partial}{\partial \mu_0}\sum_{y_i \in 0}\left[\frac{-1}{2}(\vec{x}_i - \vec{\mu}_0)^T \Sigma_{10}^{-1}(\vec{x}_i - \vec{\mu}_0)\right] = 0$$

$$\sum_{y_i \in 0}\left[\frac{-1}{2} \times 2 \times \Sigma_{10}^{-1}(\vec{x}_i - \vec{\mu}_0)\right] = 0 \quad \left\{ \because \frac{\partial x^T B x}{\partial x} = 2Bx \right\}$$

$$\sum_{y_i \in 0}(\vec{x}_i - \vec{\mu}_0)^T \Sigma_{10}^{-1} = 0.$$

The value of covariance (i.e) Σ_{10}^{-1} cannot be zero

$$\sum_{y_i \in 0} (\vec{x} - \vec{\mu}_0)^T = 0$$

$$\sum_{y_i \in 0} \vec{x} - N_0 \mu_0 = 0$$

$$N_0 \mu_0 = \sum_{y_i \in 0} \vec{x}$$

$$\boxed{\mu_0 = \frac{\sum_{y_i \in 0} \vec{x}}{N_0}}$$

$||ly$

$$\boxed{\mu_1 = \frac{\sum_{y_i \in 1} \vec{x}}{N_1}}$$

Now, we will differentiate w.r.t Σ_0^{-1}.

$$\frac{\partial (L(\theta))}{\partial \Sigma_0^{-1}} = \frac{\partial}{\partial \Sigma_0^{-1}} \sum_{y_i \in 0} \log P(x_i | \mu_0, \Sigma_0) + \frac{\partial}{\partial \Sigma_0^{-1}} (\text{const})$$

$$= \frac{\partial}{\partial \Sigma_0^{-1}} \sum_{y_i \in 0} \log P(x_i | \mu_0, \Sigma_0)$$

$$= \frac{\partial}{\partial \Sigma_0^{-1}} \left(\sum_{y_i \in 0} \log \left(\frac{1}{(2\pi)^{D/2} \sqrt{|\Sigma_0|}} \exp\left(\frac{-1}{2} (\vec{x}_i - \vec{\mu}_0)^T \Sigma_0^{-1} (\vec{x}_i - \vec{\mu}_0) \right) \right) \right) = 0$$

$$= \frac{\partial}{\partial \epsilon_0^{-1}} \left(\frac{N_0}{2} \log |\Sigma_0^{-1}| - \frac{1}{2} \sum_{y_i \in 0} - \frac{1}{2} (\vec{x_i} - \vec{\mu_0})^T \Sigma_{10}^{-1} (\vec{x_i} - \vec{\mu_0}) \right)$$

$$= \frac{\partial}{\partial \epsilon_0^{-1}} \left(\frac{N_0}{2} \log |\Sigma_0^{-1}| - \frac{1}{2} \sum_{y_i \in 0} tr\left[(\vec{x_i} - \vec{\mu_0})^T \Sigma_{10}^{-1} (\vec{x_i} - \vec{\mu_0}) \right] \right) = 0$$

$\hookrightarrow$ since $\{ x^T B x = tr(x^T B x). \}$

let us assume $\Sigma_{10}^{-1} = A$.

Therefore we will differentiate the above equation. w.r.t A.

$$\frac{\partial}{\partial A} \left(\frac{N_0}{2} \log |A| - \frac{1}{2} \sum_{y_i \in 0} tr\left[(\vec{x_i} - \vec{\mu_0}) A (\vec{x_i} - \vec{\mu_0}) \right] \right) = 0$$

$$\frac{N_0}{2} (A^{-1})^T - \frac{1}{2} \sum_{y_i \in 0} \left[(\vec{x_i} - \vec{\mu_0}) (\vec{x_i} - \vec{\mu_0})^T \right]^T = 0$$

$$\frac{N_0}{2} A^{-1} - \frac{1}{2} \sum_{y_i \in 0} (\vec{x_i} - \vec{\mu_0}) (\vec{x_i} - \vec{\mu_0})^T = 0$$

$$\frac{N_0}{2}\Sigma_0 - \frac{1}{2}\sum_{y_i \in G_0}(\vec{x_i}-\vec{\mu_0})(\vec{x_i}-\vec{\mu_0})^T = 0$$

$$\frac{N_0}{\cancel{2}}(\Sigma_0) = \frac{1}{\cancel{2}}\sum_{y_i \in G_0}(\vec{x_i}-\vec{\mu_0})(\vec{x_i}-\vec{\mu_0})^T$$

$$\boxed{\Sigma_0 = \frac{1}{N_0}\sum_{y_i \in G_0}(\vec{x_i}-\vec{\mu_0})(\vec{x_i}-\vec{\mu_0})^T}$$

Therefore when we differentiate w.r.t Σ_1 we get:-

$$\Sigma_1 = \frac{1}{N_1}\sum_{y_i \in G_1}(\vec{x_i}-\vec{\mu_1})(\vec{x_i}-\vec{\mu_1})^T \quad \{\text{from symmetry}\}$$

There we now have all the parameters values for given $\theta = \{\alpha, \mu_1, \Sigma_1, \mu_2, \Sigma_2\}$

The given Bayes optimal decision is:-

$$y = \operatorname{argmax}_{\hat{y} = \{0,1\}} P(\hat{y} \mid x)$$

For linear decision Boundary of a 2-class classification problems,

$$P(y=1 \mid x) = P(y=0 \mid x) = 0.5$$

Using the conditional parameters we get:-

$$P(y=1 \mid x) = \frac{P(x, y=1)}{P(x)}$$

$$= \frac{P(x, y=1) + P(y=1)}{\sum_y P(x,y)}$$

$$= \frac{P(x \mid y=1)\, P(y=1)}{P(x, y=1) + P(x, y=0)}$$

$$= \frac{P(x \mid y=1)\, P(y=1)}{P(x \mid y=1)\, P(y=1) + P(x \mid y=0)\, P(y=0)}$$

$P(y=1) = \alpha^1 (1-\alpha)^0$
$= \alpha.$

$P(y=0) = \alpha^0 (1-\alpha)^{1-0}$
$= (1-\alpha)$

Therefore,

$P(x|y=1) = N(x|\mu, \Sigma_1)$

$$\therefore P(y=1|x) = \frac{\alpha \cdot N(x|\mu_1 \Sigma_1)}{\alpha(N(x|\mu_1 \Sigma_1) + (1-\alpha) N(x|\mu_0, \Sigma_0))}$$

$||^{ly}$.

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

$$= \frac{(1-\alpha) N(x|\mu_0, \Sigma_0)}{\alpha N(x|\mu_1, \Sigma_1) + (1-\alpha) N(x|\mu_0, \Sigma_0)}$$

In order to calculate the decision Boundary,

$P(y=1|x) = p(y=0|x)$

$\therefore (1-\alpha) N(x|\mu_0, \Sigma_0) = \alpha (N(x|\mu_0, \Sigma_0))$

Substituting the value of N we get,

$$\frac{\alpha}{(2\pi)^{D/2}\sqrt{|\Sigma_1|}} \exp\left(\frac{-1}{2}(\vec{x}-\vec{\mu_1})\Sigma_1^{-1}(\vec{x}-\vec{\mu_1})^T\right) = \frac{(1-\alpha)}{2\pi^{D/2}\sqrt{|\Sigma_0|}} \exp\left((\vec{x}-\vec{\mu_0})\Sigma_0^{-1}(\vec{x}-\vec{\mu_0})\right)$$

Simplifying and taking log we get,

$$\frac{-1}{2}(\vec{x}-\vec{\mu_1})\Sigma_{11}^{-1}(\vec{x}-\vec{\mu_1})^T + \frac{1}{2}(\vec{x}-\vec{\mu_0})\Sigma_{10}^{-1}(\vec{x}-\vec{\mu_0})^T = \log\left(\frac{1-\alpha}{\alpha}\sqrt{\frac{|\Sigma_1|}{|\Sigma_0|}}\right)$$

Now, for linear classification,

$$\Sigma_0 = \Sigma_1 = \Sigma_1 \quad \& \quad \alpha = \frac{N_1}{N_0+N_1}$$

$$\frac{-1}{2}(\vec{x}-\vec{\mu_1})\Sigma_1^{-1}(\vec{x}-\vec{\mu_1})^T + \frac{1}{2}(\vec{x}-\vec{\mu_0})\Sigma_1^{-1}(\vec{x}-\vec{\mu_0}) = \log\left(\frac{N_0}{N_1}\right)$$

$$\Sigma_1^{-1}(\mu_1-\mu_0)^T\vec{x} + \frac{1}{2}(\mu_0+\mu_1)^T\Sigma_1^{-1}(\mu_0-\mu_1) + \log\left(\frac{N_1}{N_0}\right) = 0$$

The above equation is of the form.

$w^T x + b = 0 \rightarrow$ linear equation

where $w = \Sigma^{-1}(\mu_1-\mu_0)$

and $b = \frac{1}{2}(\mu_0 + \mu_1)^T \Sigma_1^{-1} (\mu_0 - \mu_1) - \log \frac{N_1}{N_0}$

Therefore this can be written as:-

$$f(x) = \text{sign}\cdot(w^T x + b)$$

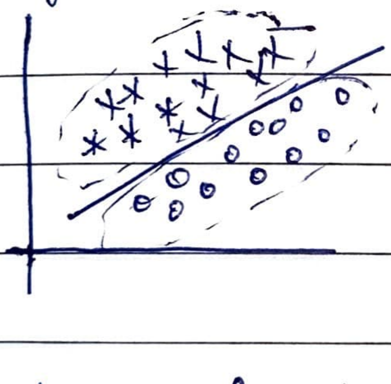
Therefore, we will get a linear decision boundary.

1^{ly} if $\Sigma_0 \neq \Sigma_1$ we will get, a quadratic function of x.

$$-\frac{1}{2}(\vec{x} - \vec{\mu_1})^T \Sigma_1^{-1} (\vec{x} - \vec{\mu_1}) + \frac{1}{2}(\vec{x} - \mu_0)^T \Sigma_0^{-1} (\vec{x} - \vec{\mu_0}) + \frac{1}{2}\log\frac{\Sigma_0}{\Sigma_1} + \log\frac{N_1}{N_0}$$

Therefore:-

→ linear function when, $\Sigma_0 = \Sigma_1$



→ quadratic function. when $\Sigma_0 \neq \Sigma_1$