P3

$$P(x \mid \mu, \Sigma) = \frac{1}{(2\pi)^{D/2}\sqrt{|\Sigma|}} \exp\left(\frac{1}{2}(\vec{x}-\mu)^T \Sigma^{-1} (\vec{x}-\vec{\mu})\right)$$

$$\cancel{P(x \mid \theta) =} \quad P(y \mid \theta) = \alpha^{y}(1-\alpha)^{1-y}$$

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

Class labels $= 1, 2$

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

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

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

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

$$= \sum_{i=1}^{N} \log P(y_i \mid \theta) + \sum_{y_i \in 1} \log P(x_i, \mu_1 \mid \Sigma_1)$$

$$+ \sum_{y_i \in 2} \log P(x_i, \mu_2 \mid \Sigma_2)$$

calculating max likelihood seperately.

Term 1

$$L_1(\theta) = \sum_{i=1}^{N} \left(\log \alpha^{y_i} + \log(1-\alpha)^{y_i}\right)$$

$$= \sum_{i=1}^{N} y_i \log \alpha + \sum_{i=1}^{N} (1-y_i) \log(1-\alpha)$$

Term 2

$$L_1(\theta) = \sum_{y_i \in 1} \frac{1}{(2\pi)^{D/2}\sqrt{|\Sigma_1|}} \exp\left(-\frac{1}{2}(x_i-\mu_1)^T \Sigma_1^{-1} (\vec{x}-\vec{\mu_1})\right)$$

$$LL_2(\theta) = \sum_{y_i \in 1} \left[\log \frac{1}{(2\pi)^{D/2}\sqrt{|\Sigma_1|}} + \frac{-1}{2}(x_i-\mu_1^T)\Sigma_1^{-1}(x_i-\vec{\mu_1}) \right]$$

Term 3

$$LL_3(\theta) = \sum_{y_i \in 2} \left[\log \frac{1}{(2\pi)^{D/2}\sqrt{|\Sigma_2|}} + \frac{-1}{2}(x_i-\mu_2)^T \Sigma_2^{-1}(x_i-\mu_2) \right]$$

$$\therefore \alpha^* = \frac{\partial LL(\theta)}{\partial \alpha}$$

$$= \frac{\partial (LL_1(\theta) + LL_2(\theta) + LL_3(\theta))}{\partial \alpha}$$

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

$$\frac{1}{1-\alpha}\sum(1-y_i) = \frac{1}{\alpha}\sum y_i$$

$$\frac{\alpha}{1-\alpha} = \frac{\sum y_i}{\sum(1-y_i)}$$

Let $\sum y_i = z$

$$\frac{\alpha}{1-\alpha} = \frac{z}{N-z}$$

$$\alpha N - \alpha z = z - \alpha z$$

$$\alpha = z/N$$

$$\alpha = \frac{1}{N} \sum_{i=1}^{N} y_i$$

$$\mu_1^* = \frac{\partial}{\partial \mu_1} LL(\theta)$$

$$= \frac{\partial}{\partial \mu_1} LL_2(\theta)$$

$$= \cancel{0 - \frac{\partial}{\partial \mu_1} \frac{1}{2}} ($$

By using $\frac{\partial \theta^T \theta^T}{\partial \theta} = 2\theta^T$

$$\mu_1^* = \sum -\frac{1}{2} 2(x_i - \mu_1)^T (-1) [\Sigma_1^{-1}]$$

$$0 = \sum_{y_i \in 1} (x_i - \mu_1)^T \Sigma^{-1}$$

$$\boxed{\mu_1^* = \frac{1}{N_1} \sum_{y_i \in 1} x_i}$$

Similarly for μ_2^*

$$\boxed{\mu_2^* = \frac{1}{N_2} \sum_{y_i \in 2} x_i}$$

Using Term 2

$$\Sigma_1^* = \frac{\partial LL(\theta)}{\partial \Sigma_1}$$

$$= \frac{\partial}{\partial \Sigma_1} \left[\sum_{y_i \in 1} \log \frac{1}{(2\pi)^{D/2}} + \log \frac{1}{\sqrt{|\Sigma|}} + \log\left(-\frac{1}{2}(x_i - \mu_1)^T \Sigma^{-1} (x_i - \mu_1)\right) \right]$$

$$= \left[0 + \sum_{y_i \in 1} \log\sqrt{|\Sigma^{-1}|} + \sum \log\left(-\frac{1}{2}(x_i-\mu_1)^T \Sigma^{-1} (x_i-\mu_1)\right)\right]$$

Let $A = \Sigma^{-1}$

$$\therefore \frac{\partial LL(\theta)}{\partial A} = \frac{N_1}{2}(A^{-1})^T - \frac{1}{2}\sum_{i=1}^{N}\left[(\vec{x_i}-\vec{\mu_1})(\vec{x_i}-\vec{\mu_1})^T\right]^T$$

By $\frac{\partial \log |A|}{\partial A} = (A^{-1})^T$ & $\frac{\partial \operatorname{tr}[BA]}{\partial A} = B^T$

$$0 = \frac{N_1}{2}(\Sigma_1) - \frac{1}{2}\sum_{\substack{i=1 \\ y_i \in 1}}^{N}(\vec{x_i}-\vec{\mu_1})(\vec{x_i}-\vec{\mu_1})^T$$

$$\boxed{\therefore \Sigma_1^{*} = \frac{1}{N_1}\sum_{y_i \in 1}(\vec{x_i}-\vec{\mu_1})(\vec{x_i}-\vec{\mu_1})^T}$$

Similarly

$$\boxed{\Sigma_2^{*} = \frac{1}{N_2}\sum_{y_i \in 2}(\vec{x_i}-\vec{\mu_2})(\vec{x_i}-\vec{\mu_2})^T}$$