$\Rightarrow$ 2 class classification problem, each class is gaussian

Let $\Theta = \{\alpha, \mu_1, \Sigma_1, \mu_2$

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

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

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

Class labels $= 1, 2$

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

$$L(\Theta) = \log P(\text{data}|\Theta)$$

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

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

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

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

Calculating max likelihood

for 1st term

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

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

2nd Term

$$L_2(\theta) = \sum_{y_i \in 1} \left[\log \frac{1}{(2\pi)^{D/2} \sqrt{|\Sigma_1|}} + \left(\frac{-1}{2}\right)(x_i - \mu_1^T)\Sigma_1^{-1}(x_i - \vec{\mu}_1)\right]$$

3rd Term

$$L_3(\theta) = \sum_{y_i \in 2} \left[\log \frac{1}{(2\pi)^{D/2} \sqrt{|\Sigma_2|}} + \left(\frac{-1}{2}\right)(x_i - \mu_2)^T \Sigma_2^{-1} (x_i - \mu_2)\right]$$

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

$$= \frac{\partial (L_1(\theta))}{\partial \alpha} + \frac{\partial (L_2(\theta))}{\partial \alpha} + \frac{\partial L_3(\theta)}{\partial \alpha}$$

(2)

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

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

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

Now, $\Sigma y_i = Z$

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

$$\alpha = Z/N$$

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

$$\mu_i^{\alpha} = \frac{\partial}{\partial \mu_i} L(\theta)$$

$$= \frac{\partial}{\partial \mu_i} L_2(\theta)$$

Now, $\frac{\partial \theta^T}{\partial \theta}\theta^T = 2\theta^T$

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

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

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

for μ_2^*

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

Using term 2

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

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

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

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

$$\therefore \frac{\partial LL(\theta)}{\partial A} = \frac{N_1}{2}(A^{-1})^T - \frac{1}{2}\sum_{i=1}^{N}\left[(\vec{x}_i - \vec{\mu}_1)(\vec{x}_i - \vec{\mu}_1)^T\right]$$

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

$$0 = \frac{N_1}{2}(\Sigma_1) - \frac{1}{2}\sum_{y_i \in 1}(\vec{x}_i - \vec{\mu}_1)(\vec{x}_i - \vec{\mu}_1)^T$$

$$\therefore \Sigma_1^* = \frac{1}{N_1}\sum_{y_i \in 1}(\vec{x}_i - \vec{\mu}_1)(\vec{x}_i - \vec{\mu}_1)^T$$

Similarly,

$$\Sigma_2^* = \frac{1}{N_2}\sum_{y_i \in 2}(\vec{x}_i - \vec{\mu}_2)(\vec{x}_i - \vec{\mu}_2)^T$$

8

$y = \arg\max_{\hat{y} = \{0,1\}} P(y^{\lambda} | 2$

Decision boundry

$$\sigma(x) = \frac{P(y=2 | x=2)}{P(y=1 | x=2)}$$

if $\sigma(x) > 1$, x is in 2^{nd} class

$\sigma(x) < 1$, x ~~belo~~ in 1^{st} class

taking log

$$\log(\sigma(x)) = \log P(y=2 | \cancel{X}=x) - \log P(y=1 | X=2)$$

Now,

$$P(y=j | X=x) = \frac{P(X=x | y=j)\, P(Y=j)}{P(X=x)}$$

and denotiing $\cancel{\lambda}_j = P(y=j)$

$$\log(\sigma(x)) = \log(P(\cancel{X}=x | y=2) + \log(\lambda_2)$$

$$-\log(P(X=\lambda) - \log P(X=x | y=1)$$

$$-\log(\lambda_1) + \log(P(X=x))$$

$= \log(P(X=2 \mid y=2)) + \log(\lambda_2) -$

$\log(P(X=2 \mid y=1)) - \log(\lambda_1)$

we know $P(X=x \mid y=y) =$

$$\frac{1}{\sqrt{(2\pi)^d |\Sigma_y|}} \exp\left(\frac{-(x-\mu_y)\Sigma_y^{-1}(x-\mu_y)'}{2}\right)$$

$$\log \delta(x) = \log \frac{P(X=x \mid y=2)}{P(X=x \mid y=1)} + \log \frac{\lambda_2}{\lambda_1}$$

$$= \log \sqrt{\frac{|\Sigma_1|}{|\Sigma_2|}} + \frac{-(x-\mu_2)\Sigma_2^{-1}(x-\mu_2)^T}{2}$$

$$+ \frac{(x-\mu_1)\Sigma_1^{-1}(x-\mu_1)^T}{2} + \log \frac{\lambda_2}{\lambda_1}$$

when $\Sigma_1 = \Sigma_2 = \Sigma$

$$\log \delta(x) = (x\Sigma^{-1}\mu_2' + \mu_2\Sigma^{-1}x' - \mu_1\Sigma^{-1}x' - x\Sigma^{-1}\mu_1)/2 + C$$

where $c = \log \frac{\lambda_2}{\lambda_1} + \mu_1 \Sigma^{-1} \mu_1 + \mu_2 \Sigma^{-1} \mu_2'$

c is a constant term

$$\log G(x) = \Sigma^{-1} x' (\mu_2 - \mu_1) + c$$

when $\Sigma_1 \neq \Sigma_2$

$$\log G(x) = \frac{x \Sigma_1^{-1} x' - x \Sigma_2^{-1} x' + \Sigma_2^{-1} x' \mu_2}{2}$$

$$- \Sigma_1^{-1} x' \mu_1 + c$$