LDA

$X = \{x_1, \ldots, x_d\}$

$y = \{0, 1\}$

① $P[y=0] = P[y=1] = \frac{1}{2}$

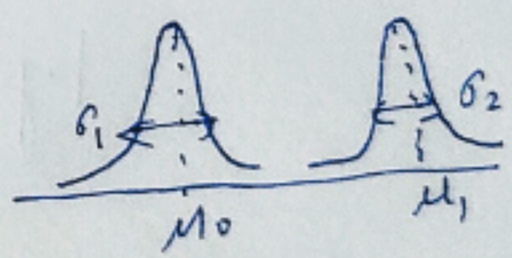
② $P[X = x \mid Y = y]$ is a Gaussian Distribution

$$P[X = x \mid Y = y] = \frac{1}{(2\pi)^{d/2} |\Sigma_y|^{1/2}} e^{-\frac{1}{2}(x-\mu_y)^T \Sigma_y^{-1} (x - \mu_y)}$$

$y = 0, 1$

③ The covariance matrixes of the Gaussian distributions are the same

$\Sigma_0 = \Sigma_1$



$\sigma_1 = \sigma_2$

$P[Y=0 \mid X=x]$ $\quad$ $P[Y=1 \mid X=x]$

$$h_{Bayes}(x) = \underset{y \in \{0,1\}}{\operatorname{argmax}} \; P[Y=y \mid X=x]$$

$$= \underset{y \in \{0,1\}}{\operatorname{argmax}} \; \frac{P[Y=y]\, P[X=x \mid Y=y]}{P[X=x]}$$

0.2 $\quad$ 0.8

0.2 $\quad$ 8

$$= \underset{y \in \{0,1\}}{\operatorname{arg\,max}} \; P[X=x \mid Y=y]$$

$$\max_{y \in \{0,1\}} \quad P[X = x \mid Y = y]$$

$y = 1$ if

$$P[X = x \mid Y = 1] > P[X = x \mid Y = 0]$$

$y = 0$ if $P[X = x \mid Y = 1] < P[X = x \mid Y = 0]$

$$P[X = x \mid Y = 1] > P[X = x \mid Y = 0]$$

$$\frac{P[X = x \mid Y = 1]}{P[X = x \mid Y = 0]} > 1$$

$$\log\left[\frac{P[X = x \mid Y = 1]}{P[X = x \mid Y = 0]}\right] > 0$$

$$\log\left[\frac{\frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}(x-\mu_1)^T\Sigma^{-1}(x-\mu_1)}}{\frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} e^{-1/2(x-\mu_0)^T\Sigma^{-1}(x-\mu_0)}}\right] > 0$$

$$\log\left(e^{-1/2(x-\mu_1)^T\Sigma^{-1}(x-\mu_1)}\right) - \log\left[e^{-\frac{1}{2}(x-\mu_0)^T\Sigma^{-1}(x-\mu_0)}\right] > 0$$

$$-(x-\mu_1)^T\Sigma^{-1}(x-\mu_1) + (x-\mu_0)^T\Sigma^{-1}(x-\mu_0) > 0$$

$$(x-\mu_0)^T\Sigma^{-1}(x-\mu_0) - (x-\mu_1)^T\Sigma^{-1}(x-\mu_1) > 0$$

$$(x^T-\mu_0^T)\Sigma^{-1}(x-\mu_0) - (x^T-\mu_1^T)\Sigma^{-1}(x-\mu_1) > 0$$

$$\cancel{x^T\Sigma^{-1}x} - x^T\Sigma^{-1}\mu_0 - \mu_0^T\Sigma^{-1}x + \mu_0^T\Sigma^{-1}\mu_0$$

$$- \cancel{x^T\Sigma^{-1}x} + x^T\Sigma^{-1}\mu_1 + \mu_1^T\Sigma x - \mu_1^T\Sigma^{-1}\mu_1 \quad > 0$$

$$\cancel{-2x^T} \quad -2\mu_0^T\Sigma^{-1}x + 2\mu_1^T\Sigma^{-1}x + \mu_0^T\Sigma^{-1}\mu_0 - \mu_1^T\Sigma^{-1}\mu_1 > 0$$

$$2(\mu_1^T-\mu_0^T)\Sigma^{-1}x + \mu_0^T\Sigma^{-1}\mu_0 - \mu_1^T\Sigma^{-1}\mu_1 > 0$$

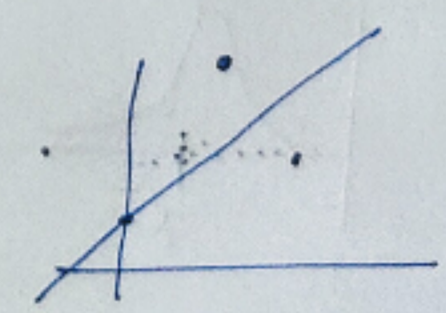
$$w^Tx + b > 0$$

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

$$b = \mu_0^T\Sigma\mu_0 - \mu_1^T\Sigma^{-1}\mu_1$$

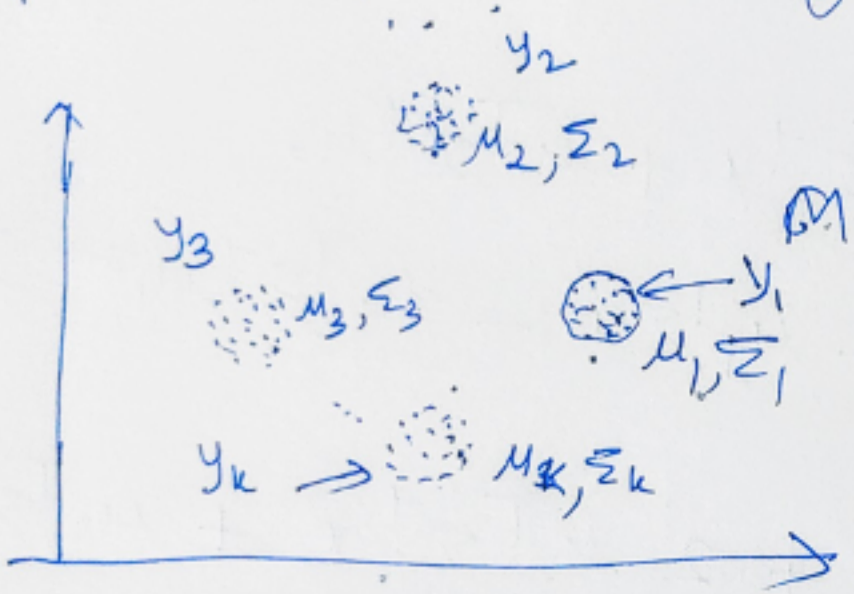
$$\checkmark\ w^Tx + b > 0 \Rightarrow y = 1$$

$$w^Tx + b < 0 \Rightarrow y = 0$$



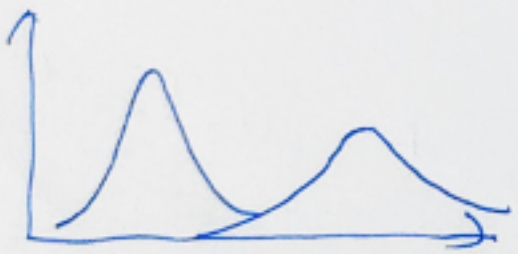
Expectation Maximization Algorithm.

$S = \{x_1, \dots, x_n\}$ $\qquad \max_{\theta} \ell(S,\theta) = \log(P_\theta(X=x))$



Mixture of Gaussian Distributions

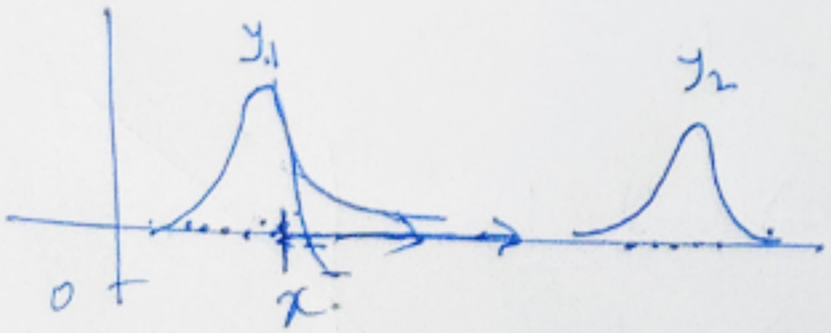
$$\theta = (\mu_1, \mu_2 \dots \mu_k, \Sigma_1, \Sigma_2, \dots \Sigma_k)$$



Consider each Gaussian distribution Separately. Let us consider k random variables $y_1, y_2 \dots y_k$ (Latent Variables)

$$P[Y = y] = c_y \qquad y \in \{1, 2, \dots, k\}$$

$$\sum_{y=1}^{k} c_y = 1$$



$P[X=x \mid y=1] >$
$P[X=x \mid y=2]$

We choose X on the basis of the value Y according to a Gaussian distribution

$$P[X = x \mid Y = y] = \left[\frac{1}{(2\pi)^{d/2} |\Sigma_y|^{1/2}}\right] e^{-\frac{1}{2}(x-\mu_y)^T \Sigma_y^{-1} (x-\mu_y)}$$

$$\max L(S;\theta) = \max_{\theta} \underline{P_{\theta}[x=x]}$$

Marginal Distribution

$A, B \quad P(A,B) \leftarrow$ Jonit distribution

$A \downarrow \{0,1,2\}$, $B \downarrow \{0,1,2\}$

$A = a$

$$P(A) = \sum_{B \in \{0,1,2\}} P(A,B) \quad , \quad P(B) = \sum_{A \in \{0,1,2\}} P(A,B)$$

$\uparrow$ Marginal distribution

$$P_{\theta}(x,y)$$

$$P_{\theta}(x) = \sum_{y=1}^{k} P_{\theta}(X=x, Y=y)$$

$$P_{\theta}[X=x] = \sum_{y=1}^{k} P_{\theta}[X=x, Y=y]$$

$$P[X=x, Y=y] = P[Y=y]\, P[X=x \mid Y=y]$$

$$P_{\theta}[X=x] = \sum_{y=1}^{k} \underline{P[Y=y]}\, P[X=x \mid Y=y] \quad \text{—— (1)}$$

$$P_\theta[X = x] = \sum_{y=1}^{k} c_y \frac{1}{(2\pi)^{d/2} |\Sigma_y|^{1/2}} e^{-\frac{1}{2}(x-\mu_y)^T \Sigma_y^{-1} (x-\mu_y)} \quad (2)$$

$$S = \{x_1, \ldots, x_m\}$$

$$\theta = \{\mu_1, \mu_2, \ldots, \mu_k, \Sigma_1, \Sigma_2, \ldots \Sigma_k\}$$

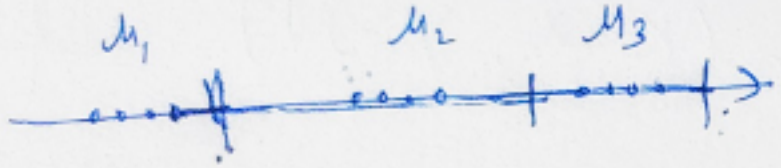
$$\max_\theta \log(P_\theta[X = x])$$

$$\max_\theta \log\left(\prod_{i=1}^{m} P_\theta[X = x_i]\right)$$

$$\max_\theta \sum_{i=1}^{m} \log(P_\theta[X = x_i])$$

from (1) $\Rightarrow$ $$\max_\theta \sum_{i=1}^{m} \log\left(\sum_{y=1}^{k} P[Y=y]\, P[X = x_i \mid Y = y]\right)$$

$$\max_\theta \sum_{i=1}^{m} \log\left[\sum_{y=1}^{k} P_\theta[X = x_i, Y = y]\right]$$

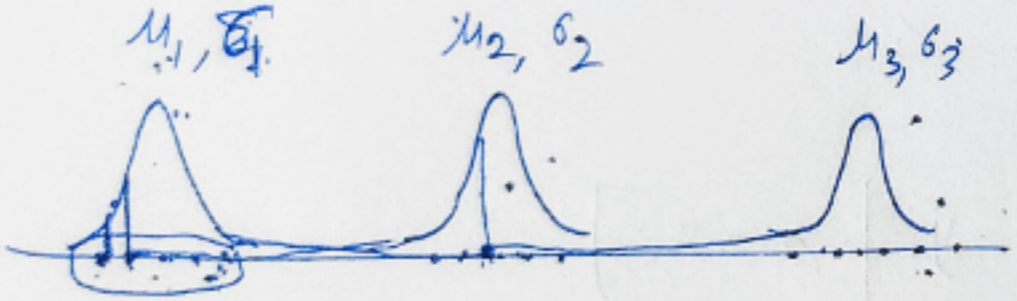


$$\sum_{i=1}^{K} \sum_{x \in C_i} \|x - \mu_i\|_2^2$$

$$\sum_{i=1}^{K} \sum_{j=1}^{m} C_{j,i} \|x - \mu_i\|_2^2$$

$C_{j,i} = 1$ if j^{th} point belongs to the i^{th} cluster

$= 0$



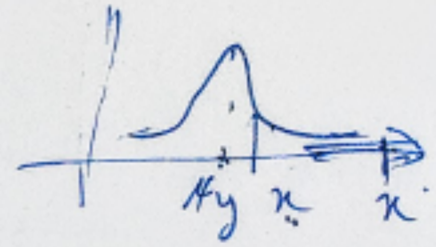
y_1 in $G(\mu_1, \sigma_1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(\mu - x)^2}{2}}$

- If we know the source distribution of all the points then we can easily determine the parameter

- If we know the parameters for all the distributions then we can easily determine the source distribution for each point

(1, 1) (8, 1) (5, 1)

$$F(Q,\theta) = \sum_{i=1}^{m}\sum_{y=1}^{K} Q_{i,y}\, \log\left(P_\theta\left(X=x_i;\ Y=y\right)\right)$$



$$Q_{i,y} = \cancel{P_\theta[X=x]} \quad P_\theta[Y=y \mid X=x]$$

$$\cancel{(\theta^{(t)},}$$

$$(\theta^{(0)},\ Q^{(0)}) \quad (\theta^{(1)},\ Q^{(1)}) \quad (\theta^{(2)},\ Q^{(2)}) \ \dots$$

$$Q^{(t+1)} = P_{\theta^{(t)}}[Y=y \mid X=x_i]$$

$$= \frac{P_{\theta^{(t)}}[X=x_i \mid Y=y]\ (P_{\theta^{(t)}}[y]}{P[X=x_i]}$$

$$= \frac{c_y^{(t)}\, e^{-\frac{1}{2}\|x_i-y\|^2}}{z_i} \qquad \Sigma = I$$

$$\theta^{(t+1)} = \max_{\theta} F(Q^{(t+1)}, \theta)$$

$$= \max_{\theta} \sum_{i=1}^{m} \sum_{y=1}^{k} P_{\theta^{(t)}}[Y=y \mid X=x_i] \left[\log(c_y) - \frac{1}{2}\|x_i - \mu_y\|^2\right]$$

$$y = 1, 2, \cdots, k$$

$$\frac{\partial}{\partial \mu_y} F(Q^{(t+1)}, \theta) = 0$$

$$\mu_y = \frac{\sum_{i=1}^{m} P_{\theta^{(t)}}[Y=y \mid X=x_i]\, x_i}{\sum_{i}^{m} P_{\theta^{(t)}}[Y=y \mid X=x_i]}$$