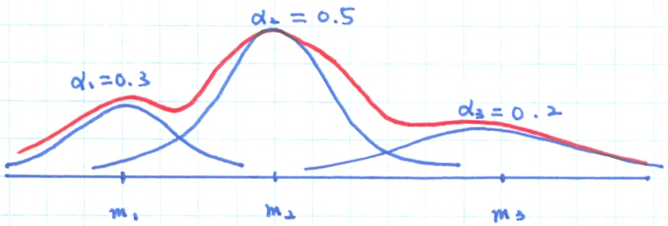
EM algorithm for parameter estimation of Gaussian mixture models

- **Gaussian mixture**

Example : a mixture of three Gaussian pdf



Component i is selected with prob. α_i

and then x is generated $\sim N(m_i, b_i^2)$

The mixture is parameterized by

$$\theta = (\alpha_1, \alpha_2, \alpha_3, m_1, m_2, m_3, b_1^2, b_2^2, b_3^2)$$

$$p(x \mid \theta) = \sum_{i=1}^{3} \alpha_i \, p_{N(m_i, b_i^2)}(x)$$

$$= \sum_{i=1}^{3} \alpha_i \frac{1}{\sqrt{2\pi}\, b_i} \exp\left[- \frac{(x_i - m_i)^2}{2 b_i^2} \right]$$

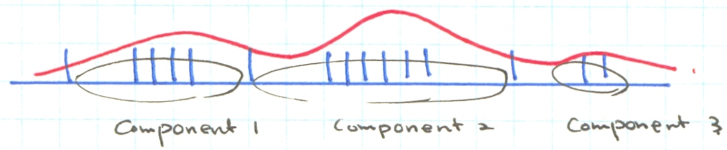
- **Objective :**

Given observed data $x_1, x_2, \cdots, x_N$,

Find θ which maximizes

$$\prod_{i=1}^{N} p(x_i \mid \theta) \quad \text{// likelihood of } \theta.$$

Such θ is called the ML estimate.



- **Missing data** y_i : (this is the segmentation result)

$y_i = k$ means that x_i comes from component k.

- **ML estimate of θ when y_i's are known.**

$S_1 = \{ x_i \mid y_i = 1 \}$ $\quad S_2 = \{ x_i \mid y_i = 2 \}$ $\quad S_3 = \{ x_i \mid y_i = 3 \}$

$$\alpha_k = \frac{|S_k|}{N}$$

$$m_k = \frac{1}{|S_k|} \sum_{x_i \in S_k} x_i$$

$$b_k^2 = \frac{1}{|S_k|} \sum_{x_i \in S_k} (x_i - m_k)^2$$

- EM algorithm

  (It attempts to obtain the ML estimate of θ when y_i's are unknown, and reaches a local maximum of the likelihood)

  - It starts with an estimated θ and updates it to θ'

$$\boxed{\alpha_k' = \frac{1}{N}\sum_{i=1}^{N} P(y_i = k \mid x_i, \theta)} \quad \cdots ①$$

Meaning of $p(y_i = k \mid x_i, \theta)$

= the probability that x_i comes from component k when θ is given.

Computation of $p(y_i = k \mid x_i, \theta)$

$$= \frac{p(y_i = k, x_i \mid \theta)}{p(x_i \mid \theta)} = \frac{p(y_i = k, x_i \mid \theta)}{\sum_{j=1}^{3} p(y_i = j, x_i \mid \theta)}$$

$$= \frac{\alpha_k \, P_{N(m_k, b_k^2)}(x_i)}{\sum_{j=1}^{3} \alpha_j \, P_{N(m_j, b_j^2)}(x_i)}$$

	x_1	x_2	x_3	x_4	x_5
$p(y_i = 1 \mid x_i, \theta)$	0.9	0.8	0.7	0.2	0.1
$p(2 \mid ,)$	0.05	0.15	0.2	0.7	0.1
$p(3 \mid ,)$	0.05	0.05	0.1	0.1	0.8

$$\alpha_1' = \frac{0.9 + 0.8 + 0.7 + 0.2 + 0.1}{5}$$

$$\alpha_2' = \frac{0.05 + 0.15 + 0.2 + 0.7 + 0.1}{5}$$

$$\alpha_3' = \frac{0.05 + 0.05 + 0.1 + 0.1 + 0.8}{5}$$

$$\boxed{\begin{aligned} m_k' &= \frac{\sum_{i=1}^{N} x_i \, p(y_i = k \mid x_i, \theta)}{\sum_{i=1}^{N} p(y_i = k \mid x_i, \theta)} \\ (b_k^2)' &= \frac{\sum_{i=1}^{N} (x_i - m_k')^2 \, p(y_i = k \mid x_i, \theta)}{\sum_{i=1}^{N} p(y_i = k \mid x_i, \theta)} \end{aligned}} \quad \cdots ②$$

After the estimation of θ,

x_i is declared to come from component k

if $p(y_i = k \mid x_i, \theta) \geq p(y_i = l \mid x_i, \theta)$ for all l.

- Generalization of ① and ② to d-dimensional space

$$\alpha_k' = \frac{1}{N}\sum_{i=1}^{N} p(y_i = k \mid x_i, \theta)$$

d-dimensional vector

$$\mu_k' = \frac{\sum_{i=1}^{N} x_i \, p(y_i = k \mid x_i, \theta)}{\sum_{i=1}^{N} p(y_i = k \mid x_i, \theta)}$$

mean vector

$$\Sigma_k' = \frac{\sum_{i=1}^{N} (x_i - \mu_k')(x_i - \mu_k')^T \, p(y_i = k \mid x_i, \theta)}{\sum_{i=1}^{N} p(y_i = k \mid x_i, \theta)}$$

Covariance matrix

Note.

$$N(\mu, \Sigma) = \frac{1}{(2\pi)^{\frac{d}{2}} |\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x - \mu)^T \Sigma^{-1} (x - \mu)\right).$$