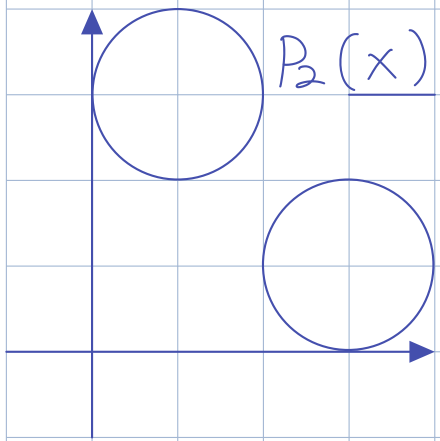
Gaussian Mixture Model And EM Algorithm

$D = \{x_1, \dots, x_N\}$ output $\hat{p}(x)$

$\hat{p}(x) =$ Gaussian Mixture Model



$p_2(x) = N(x, \mu_2, \Sigma_2)$ $\quad d = 2$

$p_1(x) = N(x, \mu_1, \Sigma_1)$

Given D and K (# of Gaussian)

Find $\hat{p}(x) = \sum_{i=1}^{K} \omega_i p_i(x)$ that is the best fit for D

Maximum Likelihood Criterion

$$\Omega = \{\omega_i, \mu_i, \Sigma_i\}_{i=1,2,\dots,K}$$

Find Ω that minimizes

$$\frac{1}{N} \sum_{n=1}^{N} -\log \hat{p}_{\Omega}(x_n)$$

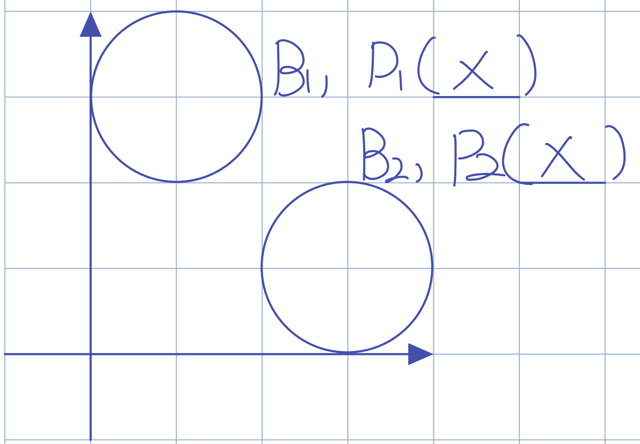
EM Algorithm

Auxiliary Variables

$D = B_1 \cup B_2 \cup \dots \cup B_k$

all B_i are mutually disjoint

$B_i \equiv$ set of points in D generated by $p_i(\underline{x})$



Subproblem 1:

Given $B_1, B_2, \dots, B_k$, find best choice of $\Omega\{w_i, \underline{\mu_i}, \Sigma_i\}_{i=1,2,\dots,K}$

It is sufficient to take emperical estimates

$N_i \equiv$ # of points in B_i

$$w_i = \frac{N_i}{N} \qquad \mu_i = \frac{1}{N_i}\sum_{x_n \in B_i} \underline{x_n}$$

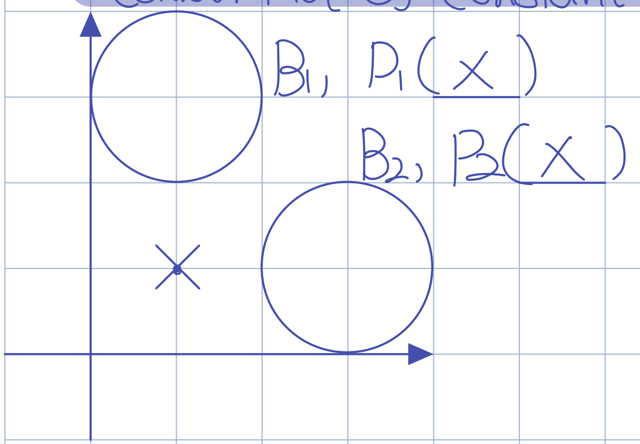
$$\Sigma_i = \frac{1}{N_i}\sum_{x_n \in B_i} \underline{(x_n - \mu_i)(x_n - \mu_i)^T}$$

Subproblem 2,

Given $\Omega\{w_i, \underline{\mu_i}, \underline{\Sigma_i}\}$
$i = 1, 2, \ldots, K$

Find $B_1, B_2, \ldots, B_K$

Contour Plot Of Constant Probability



Equivalently for each $\underline{x_n} \in D$
Find the density $p_j(\underline{x}) = N(x, \underline{\mu_j}, \underline{\Sigma_j})$
that is most likely to have generated it

Computing, $Pr(\text{cluster} = j \mid \underline{x_n})$ Postern Probability

$$j^* = \underset{j \in \{1, 2, \ldots, K\}}{\operatorname{argmax}} Pr(\text{cluster} = j \mid \underline{x_n})$$

$$\underline{x_n} \in B_{j^*}$$

Bayes Rule's

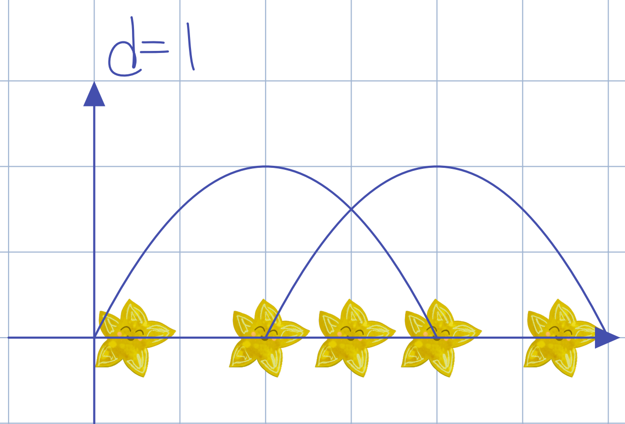
$$Pr(\text{cluster} = j \mid \underline{x_n}) = \frac{Pr(\underline{x_n} \mid \text{cluster} = j)\, Pr(\text{cluster} = j)}{Pr(\underline{x_n})}$$

$$= \frac{N(x_n, \mu_j, \Sigma_j) \cdot \omega_j}{\sum_{j=1}^{k} Pr(x_n | cluster = j)\, Pr(cluster = j)}$$

$$= \frac{N(x_n, \mu_j, \Sigma_j) \cdot \omega_j}{\sum_{i=1}^{k} N(x_n, \mu_i, \Sigma_i) \cdot \omega_i}$$ does not depend on j

$$\boxed{j^* = \underset{1 \le j \le k}{\operatorname{argmax}} \left(N(x_n, \mu_j, \Sigma_j) \cdot \omega_j \right.}$$

Density Function



Hard Decision EM Algorithm

Start with arbitrary $B_1, B_2, \ldots, B_k$

Step 1, Given $B_1, B_2, \ldots, B_k$

Find Ω via problem 1

Step 2, Given Ω

Find $B_1, B_2, \ldots, B_k$ via problem 2

Alternate between step 1 and step 2 until convergence

Soft Decision EM Algorithm

At each iterate t

$\gamma_{n,j}(t) \equiv$ fraction of $\underline{x_n}$ belonging to cluster B_j

$\gamma_{n,j}(t) \geq 0$

Soft Decisions

$$\sum_{j=1}^{K} \gamma_{n,j}(t) = 1$$

Problem 1

Given $\{\gamma_{n,j}(t)\}$ for $1 \leq n \leq N$ and $1 \leq j \leq K$

Find $\Omega = \{\omega_j(t), \mu_j(t), \Sigma_j(t)\}_{1 \leq j \leq K}$

$N_j(t) \equiv$ Effective # of points in cluste B_j

$$= \sum_{n=1}^{N} \gamma_{n,j}(t) = \gamma_{1,j}(t) + \dots + \gamma_{N,j}(t)$$

Then compute emperical means

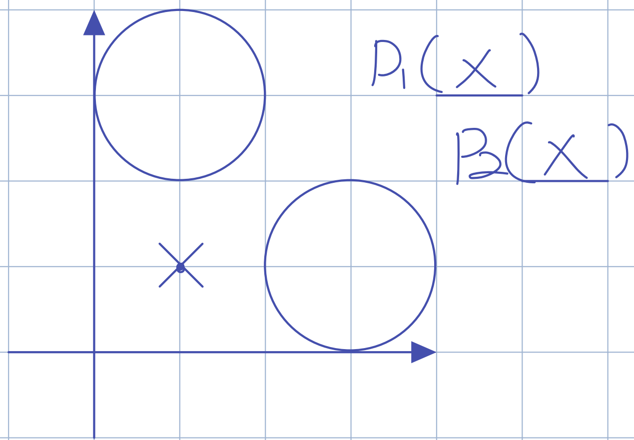
$$\mu_j(t) = \frac{1}{N_j(t)} \sum_{n=1}^{N} \gamma_{n,j}(t) \underline{x_n}$$

$$\Sigma_j(t) = \frac{1}{N_j(t)} \sum_{n=1}^{N} \gamma_{n,j}(t) \underline{(x_n - \mu_j(t))(x_n - \mu_j(t))^T}$$

$$\omega_j(t) = \frac{N_j(t)}{N}$$

Problem 2

Given $\Omega(t)$, Compute $\{\gamma_{n,j}(t)\}_{1\le n\le N}^{1\le j\le K}$



$$\gamma_{n,j}(t) = \Pr(\text{cluster} = j \mid x_n)$$

Using Bayes Rule

Alternate between problem 1 and 2 until convergence

In practic, soft decision EM perform better