Mixture Models : Multinomials + Gaussians

2 clusters

mixture components

$$\underbrace{K \; ; \; \mathcal{N}(x; \mu^{(j)}, \sigma_j^2) \; j=1,\dots,K}$$

mixture weights $\quad p_1, p_2, \dots p_K \;, \; \sum p_i = 1$

$$j \sim \text{Multinomial}(p_1, \dots, p_K)$$

$$x \sim \mathcal{N}(\mu^{(j)}, \sigma_j^2)$$ Generative!

$$\Theta : \; p_1, \dots p_K, \; \mu^{(1)} \dots \mu^{(K)}, \; \sigma_1^2, \dots \sigma_K^2$$

$$P(x|\Theta) = \sum \left(p_i \frac{1}{\sqrt{2\pi}\,\sigma_i} e^{-\frac{(x-\mu^{(i)})^2}{2\sigma_i^2}} \right)$$

$$P(S_n|\Theta) = \prod_{j=}^{n} \underbrace{\sum_{i=1}^{K} p_i \frac{1}{\sqrt{2\pi}\,\sigma_i} e^{-\frac{(x^{(j)}-\mu^{(i)})^2}{2\sigma_i^2}}}$$

Set of data

Using observed case!

$$\delta(j|i) = \begin{cases} 1, & x^{(i)} \text{ is assigned to } j \end{cases}$$

$\log P(S_n | \Theta)$ $\quad$ 0, otherwise

$$= \sum_{i=1}^{n} \left[\sum_{j=1}^{k} \delta(j|i) \log P_j N(x^{(i)}; \mu^{(j)}, \sigma_j^2) \right]$$

$$= \sum_{j=1}^{k} \left[\sum_{i=1}^{n} \delta(j|i) \log P_j N(x^{(i)}; \mu^{(j)}, \sigma_j^2) \right]$$

$$\# x \in \text{Model } j. = \hat{n}_j = \sum_{i=1}^{n} \delta(j|i)$$

$$\hat{P}_j \Leftarrow \frac{\partial \log P(S_n | \Theta)}{\partial P_j} = 0 \rightarrow \hat{P}_j = \frac{n_j}{n}$$

$$\hat{\mu}^{(j)} = \frac{1}{\hat{n}_j} \sum_{i=1}^{n} \delta(j|i) x^{(i)} \quad (\text{MLE})$$

$$\hat{\sigma}_j^2 = \frac{1}{\hat{n}_j} \sum_{i=1}^{n} \delta(j|i) \cdot \| x^{(i)} - \mu^{(j)} \|^2$$

$$\partial \sum_{j=1}^{k} \left[\sum_{i=1}^{n} \delta(j|i) \cdot \log P_j \frac{1}{\sqrt{2\pi \sigma_j^2}} e^{-\frac{\| x^{(i)} - \mu^{(j)} \|^2}{2\sigma_j^2}} \right]$$

$$\frac{\qquad}{\partial \mu^{(j)}}$$

$$= \frac{\partial \sum_{j=1}^{k} \sum_{i=1}^{n} \delta(j|i) \left[C + - \frac{\|x^{(i)} - \mu^{(j)}\|^2}{2\sigma_j^2} \right]}{\partial \mu^{(j)}}$$

$$= \sum_{j=1}^{k} \sum_{i=1}^{n} \delta(j|i) \frac{\cancel{2}(\mu^{(j)} - x^{(i)})\cancel{(-1)}}{\cancel{2}\sigma_j^2}$$

$$= 0$$

$$\Rightarrow \sum_{i=1}^{n} \delta(j|i) \left(\frac{\mu^{(j)} - x^{(i)}}{\cancel{\sigma_j^2}} \right) = 0$$

$$\Rightarrow \mu^{(j)} = \frac{\sum_{i=1}^{n} \delta(j|i)\, x^{(i)}}{n^{(j)}}$$

E-M Algo

1. Randomly initialized all $\mu^{(1)}, \cdots \mu^{(k)}$

$\sigma_1^2, \cdots, \sigma_k^2$

$P_1, \cdots, P_k$

① Step E-step

$x^{(i)}$

$$P(j|i) = \frac{P_j N(\mu^{(j)}, \sigma_j^2)}{\sum_{j=1}^{k} \left(P(x_i; \mu^{(j)}, \sigma_j^2) \right)}$$

② Step M-step

(i) $\hat{n}_j = \sum_{i=1}^{n} \delta(j|i)$ $\quad \sum_{i=1}^{n} P(j|i)$!!!

(ii) $\hat{P}_j = \frac{\hat{n}_j}{n}$

(iii) $\hat{\mu}_j = \frac{1}{\hat{n}_j} \sum_{i=1}^{n} \delta(j|i) x^{(i)}$

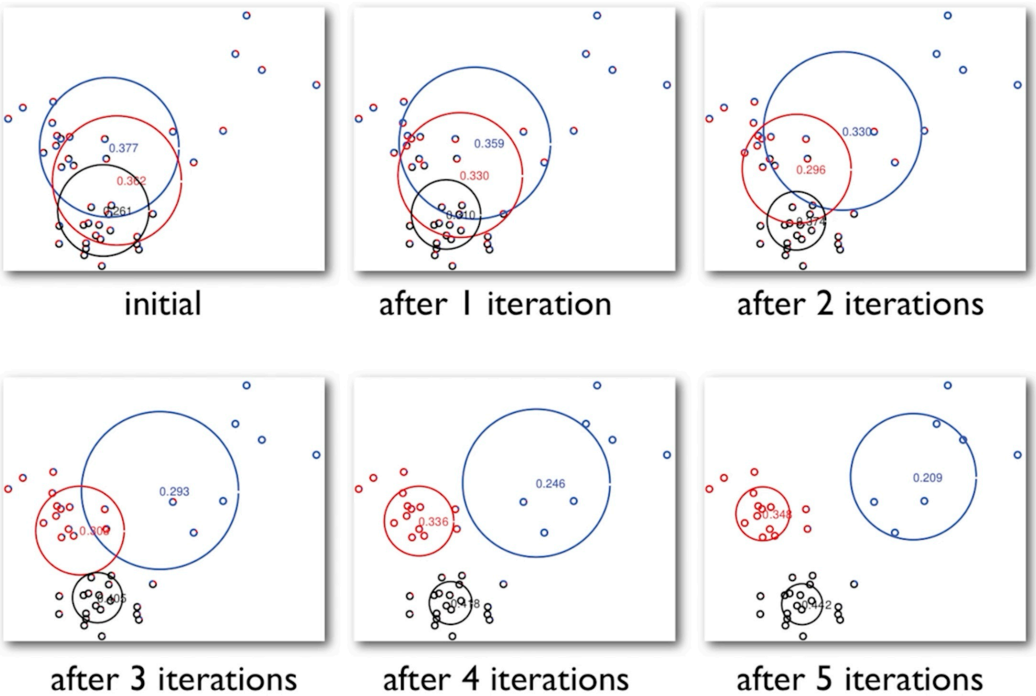
$P(j|i)$

(iv) $\hat{\sigma}^2 = \frac{1}{\ } \sum^{n} \delta(j|i) \| x^{(i)} - \hat{\mu}^{(j)} \|^2$

$$\hat{\sigma}_j^2 = \frac{1}{\hat{n}_j} \sum_{i=1} \delta(j|i)(\ldots$$

$p(j|i)$

Iteration !



initial | after 1 iteration | after 2 iterations

after 3 iterations | after 4 iterations | after 5 iterations

We can use k-means first for k, be careful about the initialization method ! (weak point !)

It is guaranteed to cvg locally !

M-step

$O(nkd)$ same as KNN