Expectation Maximization

Ex: Suppose we have data

$X_1, \dots, X_n \overset{ind}{\sim} \lambda N(\mu_1, \Sigma_1) + (1-\lambda) N(\mu_2, \Sigma_2)$

from a mixture model.

$f(x_i) = \lambda \phi(x_i; \mu_1, \Sigma_1) + (1-\lambda) \phi(x_i; \mu_2 \Sigma_2)$

We could define $\theta = (\lambda, \mu_1, \mu_2, \Sigma_1, \Sigma_2)$

and find $\hat{\theta}_{MLE}$.

We could also use latent variables

$z_i \in \{0, 1\}$ that indicate whether

the i-th data came form

component 1 or component 2

$X_i | Z_i \overset{ind}{\sim} z_i N(\mu_1, \Sigma_i) + (1 - z_i) N(\mu_2, \Sigma_2)$

$z_i \overset{iid}{\sim} Bern(\lambda)$

So if we knew then we can simplify using calculations.

Consider the z_i as latent or missing.

Thus

$$\mathcal{L}(\theta; x, z) = \prod_{i=1}^{n} \mathbb{P}(x_i \mid z_i)\,\mathbb{P}(z_i)$$

$$= \prod_{i=1}^{n} \phi(x_i; \mu_1, \Sigma_1)^{z_i}\, \phi(x_i; \mu_2, \Sigma_2)^{1-z_i}\, \lambda^{z_i} \lambda^{1-z_i}$$

$$\ell(\theta; x, z) = \sum \Big\{ z_i \big[\log \phi(x_i; \mu_1, \Sigma_1) + \log \lambda\big]$$

$$+ (1-z_i)\big[\log \phi(x_i; \mu_2, \Sigma_2) + \log(1-\lambda)\big]$$

To define Q we need $\mathbb{E}_{z \mid x, \theta^{(t)}}[z_i]$

$$\pi_i^{(t)} = \mathbb{E}_{z \mid x_i, \theta^{(t)}(z_i)}[z_i]$$

$$= \mathbb{P}(z_i = 1 \mid x ; \theta^{(t)})$$

$$= \frac{\sum_{\substack{z_j \in \{0,1\} \\ j \neq i}} \mathbb{P}(\{z_j\}_j, z_i = 1, x, \theta^{(t)})}{\sum_{z_i} \mathbb{P}(z, x, \theta^{(t)})}$$

$$= \frac{\sum_{z_j \in \{0,1\}} \prod_{\substack{k=1 \\ k \neq i}} \mathbb{P}(x_k \mid z_k ; \theta^{(t)}) \, \mathbb{P}(x_i \mid z_i = 1\, \theta) \times \mathbb{P}(z_i = 1 ; \theta^{(t)})}{\sum \prod_{k=1}^{n} \mathbb{P}(x_k \mid \tilde{z}_{k-1}, \theta^{(t)}) \, \mathbb{P}(\tilde{z}_k ; \theta^{(t)})}$$

$$= \frac{\mathbb{P}(x_i \mid z_i = 1) \, \mathbb{P}(z_i = 1)}{\sum_{\tilde{z}_i} \mathbb{P}(x_i \mid \tilde{z}_i) \mathbb{P}(\tilde{z}_i)}$$

z_i

So

$$\pi_i^{(t)} = \frac{\phi(x_i; \mu_1^{(t)}, \Sigma_1^{(t)})\lambda^{(t)}}{\phi(x_i; \mu_1^{(t)}, \Sigma_1^{(t)})\lambda^{(t)} + \phi(x_i; \mu_2^{(t)}, \Sigma_2^{(t)})(1-\lambda^{(t)})}$$

$\boxed{E}$ - Step

$$Q(\theta, \theta^{(t)}) = \mathbb{E}_{z|x, \theta^{(t)}}\left[\ell(\theta; x, z)\right]$$

$$= \sum_{i=1}^{n} \pi_i^{(t)}\left[\log \phi(x_i; \mu_1, \Sigma_1) + \log \lambda\right]$$

$$+ (1-\pi_i^{(t)})\left[\log \phi(x_i; \mu_2, \Sigma_2) + \log(1-\lambda)\right]$$

$\boxed{M}$ $$\frac{\partial Q}{\partial \lambda} = \sum_{i=1}^{n} \frac{\pi_i^{(t)}}{\lambda} - \frac{1-\pi_i^{(t)}}{1-\lambda}$$

$$\Rightarrow \boxed{\lambda^{(t+1)} = \frac{1}{n}\sum_{i=1}^{n} \pi_i^{(t)}}$$

$$\log \phi(X; \mu, \Sigma) = -\frac{1}{2}\log|\Sigma|$$
$$-\frac{1}{2}(X-\mu)^T \Sigma^{-1} (X-\mu)$$
$$-\frac{n}{2}\log 2\pi$$

So

$$\frac{dQ}{d\mu_i} = \sum_{i=1}^{n} \left\{ \pi_i^{(t)} \left(-\Sigma_i (x_i - \mu_i^{(t+1)}) \right) \right\} = 0$$

$$\Rightarrow \boxed{\mu_1^{(t+1)} = \frac{\sum_{i=1}^{n} \pi_i^{(t)} x_i}{\sum_{i=1}^{n} \pi_i^{(t)}}}$$

By symmetry

$$\boxed{\mu_2^{(t+1)} = \frac{\sum_{i=1}^{n} (1-\pi_i^{(t)}) x_i}{\sum_{i=1}^{n} (1-\pi_i^{(t)})}}$$

One can show

$$\boxed{\Sigma_1^{(t+1)} = \frac{\sum_{i=1}^{n} \pi_i^{(t)} \left(x_i - \mu_1^{(t+1)}\right)\left(x_i - \mu_1^{(t+1)}\right)^T}{\sum_{i=1}^{n} \pi_i^{(t)}}}$$

$$\boxed{\Sigma_2^{(t+1)} = \frac{\sum_{i=1}^{n} \left(1-\pi_i^{(t)}\right) \left(x_i - \mu_2^{(t+1)}\right)\left(x_i - \mu_2^{(t+1)}\right)^T}{\sum_{i=1}^{n} 1-\pi_i^{(t)}}}$$

Then lastly update the $\pi_i^{(t+1)}$