Lecture 25

The EM algorithm in general

Let θ be an unknown parameter we want to estimate. Let Y be a set of observed data, and Z a set of unobserved latent/missing data.

$$L(\theta; Y) = P(Y \mid \theta) = \int P(Y \mid Z = z, \theta) P(Z = z \mid \theta) \, dz$$

Maximizing this likelihood is hard when we don't observe Z.

Idea: Instead of maximizing $\log L(\theta; Y)$

we will maximize $\mathbb{E}_{Z \mid Y, \theta} \left[\log L(\theta; Y, Z) \right]$

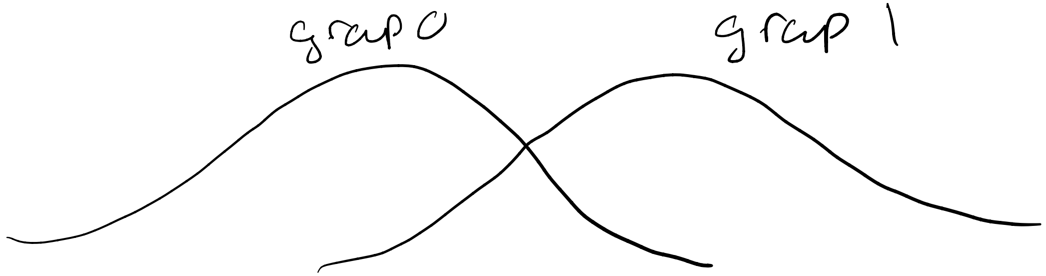
EM algorithm:

E-step: Let $\theta^{(k)}$ be the current estimate of θ

$$Q(\theta|\theta^{(k)}) = \mathbb{E}_{Z|Y,\theta^{(k)}}[\log L(\theta; Z, Y)]$$

M-step: $\theta^{(k+1)} = \arg\max_{\theta} Q(\theta|\theta^{(k)})$

Example: Gaussian mixtures



$Z \sim \text{Bernoulli}(\alpha)$

Z : group assignment

($Z=0 \Rightarrow$ group 0

$Z=1 \Rightarrow$ group 1)

$Y \mid (Z=j) \sim N(\mu_j, \sigma_j^2)$

$\theta = (\alpha, \mu_1, \mu_0, \sigma_1^2, \sigma_0^2)$

$$\Rightarrow \quad f(Y_i \mid Z_i = j, \theta) = \frac{1}{\sqrt{2\pi\sigma_j^2}} \exp\left\{-\frac{1}{2\sigma_j^2}(Y_i - \mu_j)^2\right\}$$

$$L(\theta; Y, Z) = \prod_{i=1}^{n} p(Y_i, Z_i \mid \theta) = \prod_{i=1}^{n} p(Z_i \mid \theta)\, p(Y_i \mid Z_i, \theta)$$

$$L(\theta; Y_i, Z_i) = p(Z_i \mid \theta)\, p(Y_i \mid Z_i, \theta)$$

$$= \alpha^{Z_i} (1-\alpha_i)^{1-Z_i} \frac{1}{\sqrt{2\pi\sigma_j^2}} \exp\left\{-\frac{1}{2\sigma_j^2}(Y_i - \mu_j)^2\right\}$$

$$Q(\theta \mid \theta^{(k)}) = \mathbb{E}_{Z \mid Y, \theta^{(k)}} \left[\log L(\theta; Y, Z) \right]$$

$$= \sum_{i=1}^{n} \mathbb{E}_{Z \mid Y, \theta^{(k)}} \left[\log L(\theta; Y_i, Z_i) \right]$$

$$\mathbb{E}_{Z \mid Y, \theta^{(k)}} \left[\log L(\theta; Y_i, Z_i) \right]$$

$$= \mathbb{E}_{Z \mid Y, \theta^{(k)}} \left[Z_i \log \alpha + (1 - Z_i) \log (1 - \alpha) - \frac{1}{2} \log (2\pi \sigma_{Z_i}^2) - \frac{1}{2\sigma_{Z_i}^2} (Y_i - \mu_{Z_i})^2 \right]$$

$$= \sum_{j=0}^{1} \left[\underset{\substack{\uparrow \\ P(Z=j)}}{\log \alpha_j} - \frac{1}{2} \log (2\pi \sigma_j^2) - \frac{1}{2\sigma_j^2} (Y_i - \mu_j)^2 \right] P(Z_i = j \mid Y_i, \theta^{(k)})$$

$$\Rightarrow Q(\theta \mid \theta^{(k)}) = \sum_{i=1}^{n} \sum_{j=0}^{1} \left[\log \alpha_j - \frac{1}{2} \log (2\pi \sigma_j^2) - \frac{1}{2\sigma_j^2} (Y_i - \mu_j)^2 \right] P(Z_i = j \mid Y_i, \theta^{(k)})$$

$$\mu_j^{(k+1)} = \underset{\mu_j}{\operatorname{argmax}} \, Q(\theta \mid \theta^{(k)}) \quad \left(\frac{\partial Q}{\partial \mu_j} \overset{\text{set}}{=} 0 \right) = \frac{\sum_{i=1}^{n} Y_i \, P(Z_i = j \mid Y_i, \theta^{(k)})}{\sum_i P(Z_i = j \mid Y_i, \theta^{(k)})}$$

Class activity

https://sta712-f23.github.io/class_activities/ca_lecture_25.html