# EM Algorithm Examples

Let $z_i = \begin{cases} 1 & \text{if Group 1} \\ 0 & \text{if Group 2} \end{cases}$

$$x_i \mid z_i = 1 \sim N(\mu_1, \sigma_1^2)$$

$$x_i \mid z_i = 2 \sim N(\mu_2, \sigma_2^2)$$

$Pr(z_i = 1) = \pi$
$Pr(z_i = 2) = 1 - \pi$

- Goal: Estimate $\mu_1, \sigma_1^2, \pi$

Problem: $z_i$ is not observed

- Likelihood for individual $i$

$$N(x_i \mid \mu_1, \sigma_1^2)\,\pi + N(x_i \mid \mu_2, \sigma_2^2)\,(1-\pi)$$

- Likelihood:

$$L = \prod_{i=1}^{n} \left[ N(x_i \mid \mu_1, \sigma_1^2)\,\pi + N(x_i \mid \mu_2, \sigma_2^2)\,(1-\pi) \right]$$

- Can do Gradient ascent (Less stable here, and not well suited to more complicated models, Complete likelihood for individual $i$

$$\left[\pi N(x_i \mid \mu_1, \sigma_1^2)\right]^{z_i} \left[(1-\pi) N(x_i \mid \mu_2, \sigma_2^2)\right]^{1-z_i}$$

Complete likelihood

$$\prod_{i=1}^{n} \left[\pi N(x_i \mid \mu_1, \sigma_1^2)\right]^{z_i} \left[(1-\pi) N(x_i \mid \mu_2, \sigma_2^2)\right]^{1-z_i}$$

Complete log-likelihood

$$\sum_{i=1}^{n} \left[ z_i \log N(x_i \mid \mu_1, \sigma_1^2) + (1-z_i) \log N(x_i \mid \mu_2, \sigma_2^2) \right] + \pi \sum z_i + (1-\pi) \sum (1-z_i)$$

EM Algorithm

E-step: $E[z_i \mid x_i] = w_i$

M-step: Maximize

$$\sum_{i=1}^{n} \left( w_i \log N(x_i \mid \mu_1, \sigma_1^2) + (1-w_i) \log N(x_i \mid \mu_2, \sigma_2^2) \right) + \log(\pi) \sum w_i + \log(1-\pi) \sum (1-w_i)$$

iterate

Bayes Rule

$$E(z_i \mid x_i) \overset{\swarrow}{=} \frac{\pi N(x_i \mid \mu_1, \sigma_1^2)}{\pi N(x_i \mid \mu_1, \sigma_1^2) + (1-\pi) N(x_i \mid \mu_2, \sigma_2^2)} = w_i$$

<u>M-step:</u>

can do optimizations independently

$$\max_{\mu_1, \sigma_1^2} \sum w_i \log N(x_i \mid \mu_1, \sigma_1^2)$$

$$\max_{\mu_2, \sigma_2^2} \sum w_i \log N(x_i \mid \mu_2, \sigma_2^2)$$

$$\max_{\pi} \log(\pi) \sum w_i + \log(1-\pi) \sum (1 - w_i)$$

$\mu_1$:

$$\frac{\partial}{\partial \mu} \Big( \sum w_i (x_i - \mu_1)^2$$

$$-2 \sum w_i (x_i - \mu_1) \overset{\text{set}}{=} 0$$

$$\Rightarrow \sum w_i x_i = \mu_1 \sum w_i$$

$$\Rightarrow \hat{\mu}_1 = \frac{\sum w_i x_i}{\sum w_i}$$

$\pi$:

$$\frac{\partial}{\partial \pi} \Big( \log(\pi) \sum w_i + \log(1-\pi) \sum (1 - w_i)$$

$$\frac{\sum w_i}{\pi} - \frac{\sum (1 - w_i)}{1 - \pi} \overset{\text{set}}{=} 0$$

$$\Rightarrow \pi \sum (1 - w_i) = (1 - \pi) \sum w_i$$

$$\Rightarrow \pi n = \sum w_i$$

$$\Rightarrow \pi = \frac{1}{n} \sum w_i$$ opt...

• Complete likelihood

$$\mathcal{L}(\underset{\text{observed}}{X}, \underset{\text{latent}}{Z} \mid \underset{\text{parameters}}{\theta})$$

likelihood

$$\sum_{Z} \mathcal{L}(X, Z \mid \theta)$$ ← want to maximize this

EM: $\hat{\theta}^{new} = \underset{\theta}{\arg\max}\ E\left[\log\left[\mathcal{L}(X, Z \mid \theta)\right]\right]$

maximize expected complete log likelihood

• Ex.) $Z_i$ = genotype for individual $i$
$X_i$ = data for individual $i$

Given $Pr(X_i \mid Z_i = k)$ "Genotype likelihood"

Let $Y_i$ be a 1-of-$(K+1)$ indicator vector for genotype

$\pi_k = Pr(\text{genotype } k)$

Complete likelihood

$$\sum_{i=1}^{n} \left[Pr(X_i \mid Z_i = 0)\,\pi_0\right]^{Y_{i0}} \left[Pr(X_i \mid Z_i = 1)\,\pi_1\right]^{Y_{i1}} \cdots \left[Pr(X_i \mid Z_i = K)\,\pi_K\right]^{Y_{iK}}$$

do EM on this to estimate $\pi_0, \pi_1, \ldots, \pi_K$

$$E(Y_{ik} \mid X_i) = P_r(Z_i = k \mid x_i)$$

$$= \frac{P_r(x_i \mid Z_i = k)\,\pi_k}{\sum_{h=0}^{k} P_r(x_i \mid Z_i = h)\,\pi_h} =: w_{ik}$$

- Maximize $\sum_{k=0}^{k} \log(\pi_k) \underbrace{\sum_{i=1}^{n} w_{ik}}_{=: w_k}$

$$\frac{d}{d\pi_j}\left( \sum_{k=0}^{k} \log(\pi_k)\, w_k + \lambda\left(\sum_{k}^{k} \pi_k - 1\right)\right)$$

$$\frac{w_j}{\pi_j} + \lambda \overset{set}{=} 0$$

$$\Rightarrow \pi_j = \frac{w_j}{-\lambda}$$

$$\sum \pi_k = 1$$

$$\Rightarrow -\lambda = \sum w_k$$

$$\hat{\pi}_j = \frac{w_j}{\sum_{k=0}^{k} w_k} \quad //$$