Lecture 15 Week 13.2 27/03

EM Algorithm

Mixture Example

$x \sim f(x|\theta)$ $\quad \theta = (\pi_1, \mu_1, \mu_2)$

$$f(x|\theta) = \sum_{k=1}^{2} \pi_k \, \phi(x; \mu_k, 1)$$

for $\phi(x; \mu_k, 1) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(x-\mu_k)^2\right\}$.

What is the full likelihood?

$x_1, \dots, x_n \overset{iid}{\sim} f(x|\theta)$

$$L(\theta) = \prod_{i=1}^{n} \sum_{k=1}^{2} \pi_k \, \phi(x_i; \mu_k, 1)$$

$$l(\theta) = \sum_{i=1}^{n} \log\left(\sum_{k=1}^{2} \pi_k \, \phi(x_i; \mu_k, 1)\right)$$

$$\propto \sum_{i=1}^{n} \log\left(\sum_{k} \pi_k \exp\left\{-\frac{1}{2}(x_i-\mu_k)^2\right\}\right)$$

Find μ_1, μ_2, π_1 that maximize this expression:

$$\underline{\frac{\partial l}{\partial \mu_l} = \sum_{i=1}^{n} \frac{1}{\sum_k \pi_k \, \phi(x_i; \mu_k, 1)} \cdot \pi_l \exp\left\{-\frac{1}{2}(x_i-\mu_l)^2\right\}(x_i-\mu_l) = 0}$$

Idea: $\quad X_1, \dots, X_n \qquad z_i = k \iff x_i \sim \phi(x_i; \mu_k, 1)$.

$\quad z_1, \dots, z_n$

If we include $z_1, \dots, z_n$, what is the resulting complete likelihood? For 1 observation:

$$f(x_i, z_i \mid \theta) = f(x_i \mid z_i, \theta)\, f(z_i \mid \theta)$$

$$= f(x_i \mid z_i = 1, \theta)\, f(z_i = 1 \mid \theta)\, \mathbb{1}\{z_i = 1\}$$
$$+ f(x_i \mid z_i = 2, \theta)\, f(z_i = 2 \mid \theta)\, \mathbb{1}\{z_i = 2\}$$

$$= \phi(x_i; \mu_1, 1)\, \pi_1\, \mathbb{1}\{z_i = 1\} + \phi(x_i; \mu_2, 1)\, \pi_2\, \mathbb{1}\{z_i = 2\}$$

$$= \sum_{k=1}^{2} \phi(x_i; \mu_k, 1)\, \pi_k\, \mathbb{1}\{z_i = k\}.$$

For all observations:

$$L(\theta) = \prod_{i=1}^{n} \sum_{k=1}^{2} \phi(x_i; \mu_k, 1)\, \pi_k\, \mathbb{1}\{z_i = k\}$$

$$= \prod_{i: z_i = 1} \phi(x_i; \mu_1, 1)\, \pi_1 \times \prod_{i: z_i = 2} \phi(x_i; \mu_2, 1)\, \pi_2$$

$$\ell(\theta) = \sum_{i: z_i = 1} \left[\log(\phi(x_i; \mu_1, 1)) + \log \pi_1\right]$$
$$+ \sum_{i: z_i = 2} \left[\log(\phi(x_i; \mu_2, 1)) + \log \pi_2\right].$$

What is the conditional density of z_i given x_i?

$$f(z_i = \ell \mid x_i) = \frac{f(x_i \mid z_i = \ell)\, f(z_i = \ell)}{f(x_i)} \qquad \text{(Bayes' Rule)}$$

$$= \frac{\phi(x_i\,;\, \mu_\ell, 1)\, \pi_\ell}{\sum_{k=1}^{2} \phi(x_i\,;\, \mu_k, 1)\, \pi_k} \equiv \delta_{z_i}(\ell)$$

"defined as"

$$\frac{\partial \ell}{\partial \mu_\ell} = \sum_{i=1}^{n} \underbrace{\frac{1}{\sum_k \pi_k\, \phi(x_i\,;\, \mu_k, 1)} \cdot \pi_\ell \exp\left\{-\frac{1}{2}(x_i - \mu_\ell)^2\right\}}_{= \delta_{z_i}(\ell)} (x_i - \mu_\ell) = 0$$

$$= \sum_{i=1}^{n} \delta_{z_i}(\ell)\,(x_i - \mu_\ell) = 0.$$

- If we knew $\delta_{z_i}(\ell)$, we could calculate $\hat{\mu}_\ell$:

$$\hat{\mu}_\ell = \frac{\sum_i \delta_{z_i}(\ell)\, x_i}{\sum_i \delta_{z_i}(\ell)}.$$

- If we knew μ_ℓ and π_l, we could calculate $\delta_{z_i}(\ell)$.

EM Algorithm:

- Start from $\theta^{(0)}$
- for $i = 0$ to convergence
- <u>E-step</u>: evaluate $\delta_{z_i}(\ell)$ from $\theta^{(i)}$
- <u>M-step</u>: estimate $\theta^{(i+1)}$ based on the E-step

What is the $Q(\theta)$ function?

Continuing with the mixture example:

$\theta = (\mu_1, \mu_2, \pi_1)$

$$Q(\theta) = E_{\underset{\text{"with respect to"}}{\vec{Z}}}\left[\ell(\vec{X}, \vec{Z}) \mid \vec{X}, \theta^{(j)}\right]$$

$$= E_{\vec{Z}}\left[\sum_{i=1}^{n}\sum_{k=1}^{2}\left(\log \pi_k^{(j)} + \log \phi(x_i; \mu_k^{(j)}, 1)\right)\cdot \mathbb{1}\{z_i = k\} \,\middle|\, \theta^{(j)}, \vec{x}\right]$$

$$= \sum_{i=1}^{n}\sum_{k=1}^{2}\left(\log \pi_k^{(j)} + \log \phi(x_i; \mu_k, 1)\right) E\left[\mathbb{1}\{z_i = k\} \mid \theta^{(j)}, \vec{x}\right]$$

$$E\left[\mathbb{1}\{z_i = k\} \mid \theta^{(j)}, \vec{x}\right] = P(z_i = k \mid \vec{x}, \theta^{(j)})$$

$$= \delta_{z_i}(\ell).$$

Example (Genes):

T t

genes

$$\underbrace{TT \quad Tt \quad tT}_{\text{dark hair}} \quad \underbrace{tt}_{\text{light hair}}$$

Observe: (n_D, n_L) $\qquad n = n_D + n_L$

Want to estimate: p_T, $p_t = 1 - p_T$

$\theta = p_T$

Complete dataset: $Y = (\overbrace{n_{TT}, n_{tT}}^{n_{TT}+n_{tT}=n_D}, \overset{n_L}{\overset{\shortparallel}{n_{tt}}})$

$$Y|\theta \sim \text{multinomial}(n, (p_T^2, 2p_t p_T, p_t^2))$$

$$\ell(\theta) = n_{TT} \log p_T^2 + n_{tT} \log(2p_t p_T) + n_{tt} \log p_t^2 \quad \leftarrow$$

$$+ \underbrace{\log \binom{n}{n_{TT}\ n_{tT}\ n_{tt}}}_{\text{constant ("multinomial coefficient")}}$$

E-step:

$$E\left[\ell(\theta;\vec{z}) \mid \vec{x}, \theta^{(j)}\right] = Q(\theta)$$

M step:

maximize $Q(\theta)$ wrt θ

Note: If n_D is observed, then (n_{TT}, n_{tT}) is binomial (multinomial) with parameters n_D, $\left(\frac{p_T^2}{2p_t p_T + p_T^2}, \frac{2p_t p_T}{2p_t p_T + p_T^2}\right)$

$$Q(\theta) = E\left[\ell(\theta) \mid \vec{x}, \theta^{(j)}\right]$$

$$E\left[n_{TT} \mid \vec{x}, \theta^{(j)}\right] = \underbrace{n_D \cdot \frac{(p_T^{(j)})^2}{(p_T^{(j)})^2 + 2p_T^{(j)} p_t^{(j)}}}_{n_{TT}^{(j)}}$$

$$E\left[n_{tT} \mid \vec{x}, \theta^{(j)}\right] = \underbrace{n_D \cdot \frac{2p_t^{(j)} p_T^{(j)}}{(p_T^{(j)})^2 + 2p_T^{(j)} p_t^{(j)}}}_{n_{tT}^{(j)}}$$

$$E[n_{tt} \mid \vec{x}, \theta^{(j)}] = n_{tt} = n_L$$

E-Step

$$Q(\theta) = n_{TT}^{(j)} \log p_T^2 + n_{tT}^{(j)} \log 2 p_T p_t + \overset{n_L}{\overset{\parallel}{n_{tt}}} \log p_t^2$$

M-Step

Maximize $Q(\theta)$

$$\frac{\partial Q}{\partial p_T} = 0$$

$$\boxed{p_T = \frac{2 n_{TT}^{(j)} + n_{tT}^{(j)}}{2n}}$$