Lecture-10 (EM and GMM)

Friday, 8 September 2023 4:55 PM

$$p(x_1;\theta)\, p(x_2;\theta) \cdots \quad p(x_n;\theta)$$

$$\arg\max_{\theta} \prod_{i=1}^{N} p(x_i;\theta)$$

$$\underbrace{\sum_{i=1}^{N} \log p(x_i;\theta)}$$

Negative log likelihood.

$$\theta = \begin{cases} \mu = \{\mu_1, \mu_2 \cdots \mu_K\} \\ \Sigma = \{\Sigma_1, \Sigma_2 \cdots \Sigma_K\} \\ \pi = \{\pi_1, \pi_2 \cdots \pi_K\} \end{cases}$$

$z_1, z_2 \ldots z_n$

$$p_{\theta}(x_i) = \sum_{k=1}^{K} \pi_k \mathcal{N}(x; \mu_k, \Sigma_k)$$

where $\sum_k \pi_k = 1$, $\pi_k \geq 0$

Goal: Find θ

$$\theta^* = \arg\max_{\theta} \prod_{i=1}^{N} p(x_i;\theta)$$

Spoiler: E-step: for each i, compute r_{ij}

$$P(z_i \mid x_i;\theta) = Q_i(z_i = j)$$

$$= \frac{p(x_i \mid z_i = j;\theta)\, p(z_i = j;\theta)}{\sum_{l=1}^{K} p(x_i \mid z_i = l;\theta)\, p(z_i = l;\theta)} =$$

$\mathcal{N}(x_i; \mu_j, \Sigma_j)$

π_j

M-step: For each j,

$$\mu_j = \frac{\sum_{i=1}^{N} Q_i(z_i = j)\, x_i}{\sum_{i=1}^{N} Q_i(z_i = j)}$$

$$\Sigma_j = \frac{\sum_{i=1}^{N} Q_i(z_i = j)(x_i - \mu_j)(x_i - \mu_j)^T}{\sum_{i=1}^{N} Q_i(z_i = j)}$$

$$f(\theta) = \sum_{i=1}^{N} \log(p(x_i;\theta))$$

$$\pi_j = \frac{\sum_{i=1}^{N} Q_i(z_i = j)}{N}$$

$$f(\theta) = \sum_{i=1}^{N} \log p(x_i;\theta) \quad \ldots \sum p(A,B)$$

$$l(\theta) = \sum_{i=1}^{N} \log p(x_i ; \theta)$$

$p(A) = \sum_{B} p(A, B)$

$p(A) = \sum_{B} p(A, B)$

$$= \sum_{i=1}^{N} \log\left(\sum_{z_i} \underbrace{p(x_i, z_i)} \right)$$

$$= \sum_{i=1}^{N} \log\left(\sum_{z_i} \underbrace{p(x_i \mid z_{i=k} ; \theta)}_{\mathcal{N}(\mu_k, \Sigma_k)} \underbrace{p(z_{i=k} ; \theta)}_{\pi_k} \right)$$

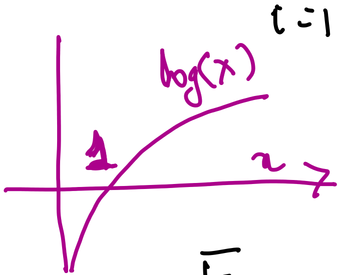
Directly optimizing for θ (μ_s, Σ_s) is hard

$$\frac{\partial f(\theta)}{\partial \mu_j} = 0 \qquad \sum_{i=1}^{N} \underline{\hspace{6cm}}$$

$$= \sum_{i=1}^{N} \log\left(\sum_{z_i} Q_i(z_i) \left[\frac{p(x_i, z_i)}{Q_i(z_i)} \right] \right)$$

Assume $Q_i(z_i)$ is a prob. mass fn $\quad \sum_{z_i} Q(z_i) = 1$

$$= \sum_{i=1}^{N} \log\left(\mathop{E}_{z_i \sim Q_i(z_i)} \left[\frac{p(x_i, z_i)}{Q_i(z_i)} \right] \right)$$



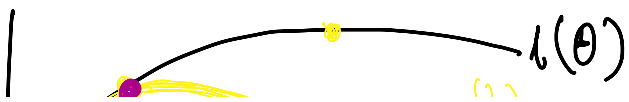
$\sum_x x\, p(x) = E[X]$

$\rightarrow \sum_x g(x)\, p(x) = E[g(X)]$

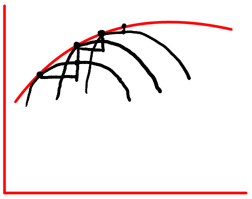
For a concave function f and a random variable X, $f(E[X]) \geqslant E[f(X)]$

Jensen's inequality

$$\geqslant \sum_{i} \mathop{E}_{z_i \sim Q_i} \left[\log\left(\frac{p(x_i, z_i)}{Q_i(z_i)} \right) \right]$$



To guarantee convergence, want the



convergence,
bound to be tight

$$f(E[X]) = E[f(X)]$$

Tight $\Rightarrow$

If f is strictly concave,

$$f(E[X]) = E[f(X)] \iff X \text{ is a constant variable}$$

$$\Rightarrow \frac{p(x_i, z_i)}{Q_i(z_i)} = \text{constant}$$

Take $Q_i(z_i) = \dfrac{p(x_i, z_i)}{\sum_{z_i} p(x_i, z_i)}$ $\quad \dfrac{p(z_i|x_i)}{\cancel{p(x_i)}}$

$$\boxed{= p(z_i \mid x_i; \theta)} \leftarrow$$