EM算法

推导：$\{x_i\}_{i=1\sim N}$ 样本，$\{z_i\}_{i=1\sim N}$ 为 x_i 的隐含变量 (Latent Variable)，θ 为待求隐含变量

目的：$\underset{\theta}{\text{Maximize}} \sum_{i=1}^{N} \log p(x_i|\theta)$

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

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

设 $Q_i(z_i)$ 为 z_i 的概率分布，$\sum_{z_i} Q_i(z_i) = 1$

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

根据 Jenson's Inequality

(Jenson's Inequality：$f(x)$ 是凹函数，则

$$f[EX] \geq E(f(x))$$

)

有：

$$E(\theta) \geq \sum_{i=1}^{N} \sum_{z_i} Q_i(z_i) \log \frac{p(x_i, z_i|\theta)}{Q_i(z_i)}$$

当且仅当 $Q_i(z_i)$ 与 $p(x_i, z_i|\theta)$ 成比例时，等号成立。

所以，当 $Q_i(z_i)$ 取如下值时

$$Q_i(z_i) = \frac{p(x_i, z_i|\theta)}{\sum_{z_i} p(x_i, z_i|\theta)}$$

$E(\theta)$ 取最大，且最大值为 $\sum_{i=1}^{N} \sum_{z_i} Q_i(z_i) \log \frac{p(x_i, z_i|\theta)}{Q_i(z_i)}$

基于以上推导，列出EM算法步骤

① 给定 $\{x_1, x_2, \dots x_N\}$，随机选取 θ_0。

② E-step：在第 k 步中获得 z_i 的分布

$$Q_i(z_i) = \frac{p(x_i, z_i | \theta_k)}{\sum_{z_i} p(x_i, z_i | \theta_k)}$$

③ 固定 $Q_i(z_i)$，求 θ_{k+1}

$$\theta_{k+1} = \arg\max_{\theta} \sum_{i=1}^{N} \sum_{z_i} Q_i(z_i) \log \frac{p(x_i, z_i | \theta_k)}{Q_i(z_i)}$$

④ 重复②③直至收敛

算法收敛性证明：

设 $M(\theta) = \sum_{i=1}^{N} \sum_{z_i} Q_i(z_i) \log \frac{p(x_i, z_i | \theta)}{Q_i(z_i)}$

则 第②步做完后，$E(\theta_k) = M(\theta_k)$

第③步做完后 $M(\theta_{k+1}) \geq M(\theta_k)$

接着做第②步 $E(\theta_{k+1}) = M(\theta_{k+1}) \geq M(\theta_k) = E(\theta_k)$

因此 $E(\theta)$ 在循环中不断增大，但 $E(\theta)$ 有上界 0，

所以必然收敛。

EM算法举例

① K均值算法

已知 $\{x_1, x_2 \cdots x_N\}$ 待求 $\theta = \{\mu_1, \mu_2 \cdots \mu_K$

设 $\{z_i\}_{1\sim N}$ $\forall i$ $z_i = \{1, 2, 3, \cdots K\}$

设 $$p(x, z|\theta) = \begin{cases} \frac{1}{(\sqrt{2\pi})^d} \exp\left\{-\frac{\|\cancel{x_i} - \mu_z\|^2}{2}\right\} & \text{当 } z = \arg\min_j \|x - \mu_j\| \text{ 时} \\ 0 & \text{其他} \end{cases}$$

第一步：随机选取 $\{\mu_1, \mu_2 \cdots \mu_K\}$

第二步(E-step)：

$$Q_i(z_i) = \frac{p(x_i, z_i|\theta_k)}{\sum_{z_i} p(x_i, z_i|\theta_k)}$$

$$= \begin{cases} 1, & \text{当 } z_i = \arg\min_j \|x_i - \mu_j\| \text{ 时} \\ 0, & \text{其他} \end{cases}$$

第三步(M-step)：

$$\theta_{k+1} = \arg\max_\theta \sum_{i=1}^N \sum_{z_i} Q_i(z_i) \log \frac{p(x_i, z_i|\theta_k)}{Q_i(z_i)}$$

因此

$$\mu_j^{(new)} = \arg\max_{\mu_j} \sum_{\substack{i=1 \\ (z_i=j)}}^N \log \frac{p(x_i, j|\theta_k)}{}$$

设 $E(\mu_j)=\sum_{\substack{i=1\sim N\\ z_i=j}} \log p(x_i, j|\theta_k)$

$$=\sum_{\substack{i=1\sim N\\ z_i=j}} -\frac{\|x_i-\mu_j\|^2}{2} - 常数$$

$$\frac{\partial E(\mu_j)}{\partial \mu_j}=\sum_{\substack{i=1\sim N\\ z_i=j}} -(x_i-\mu_j)=0$$

因此 $\mu_j=\dfrac{\sum_{\substack{i=1\sim N\\ z_i=j}} x_i}{\sum_{\substack{i=1\sim N\\ z_i=j}} 1}$

② 高斯混合模型

已知 $\{x_1, x_2, \cdots x_N\}$，$\theta=\{\pi_1, \pi_2, \cdots \pi_K, \mu_1, \mu_2 \cdots \mu_K, \Sigma_1, \Sigma_2 \cdots \Sigma_K\}$

设 $\{z_i\}_{i=1\sim N}$，$\forall i$ $z_i=\{1, 2, 3, \cdots, K\}$

设 $p(x, z|\theta)=(2\pi)^{-\frac{d}{2}}|\Sigma_z|^{-\frac{1}{2}}\pi_z \exp\{-\frac{1}{2}(x-\mu_z)^T\Sigma_z^{-1}(x$

第一步：随机选取 $\theta_0=\{\pi_1, \pi_2 \cdots \pi_K, \mu_1, \mu_2 \cdots \mu_K, \Sigma_1, \Sigma_2 \cdots \Sigma_K$

第二步(E-step)：

$$Q_i(z_i)=\frac{p(x_i, z_i|\theta_k)}{\sum_{z_i} p(x_i, z_i|\theta_k)}$$

设 $\nu_{ij} = Q_i(j) = \dfrac{p(x_i, j|\theta_k)}{\sum_{j=1}^{K} p(x_i, j|\theta_k)}$

$$= \frac{\pi_j N(x_i, j|\theta_k)}{\sum_{j=1}^{K} \pi_j N(x_i, j|\theta_k)}$$

其中 $N(x_i, j|\theta) = (2\pi)^{-\frac{d}{2}} |\Sigma_j|^{-\frac{1}{2}} \exp\{-\frac{1}{2}(x-\mu_j)^T \Sigma_j^{-1}(x-\mu_j$

第三步(M-step)

$$\theta_{k+1} = \arg\max_{\theta} \sum_{i=1}^{N} \sum_{z_i} Q_i(z_i) \log \frac{p(x_i, z_i|\theta_k)}{Q_i(z_i)}$$

设 $E(\theta) = \sum_{i=1}^{N} \sum_{z_i} Q_i(z_i) \log \dfrac{p(x_i, z_i|\theta)}{Q_i(z_i)}$

$$= \sum_{i=1}^{N} \sum_{j=1}^{K} \nu_{ij} \log \frac{\pi_j N(x_i, j|\theta)}{\nu_{ij}}$$

$$= \sum_{i=1}^{N} \sum_{j=1}^{K} \nu_{ij} [\log \pi_j - \frac{1}{2} \log|\Sigma_j| - \frac{1}{2}(x_i-\mu_j)^T \Sigma_j^{-1}(x_i-\mu$$

$$\cancel{\frac{\partial E(\theta)}{\partial \mu_j} = \sum_{i=1}^{N} \sum_{j=1}^{K} \nu_{ij} \Sigma_j^{-1}(x_i-\mu_j) = 0}$$

$$\cancel{\text{因此 } \mu_j^{(new)} = \sum_{i=1}^{N} \sum_{j=1}^{K} \nu_{ij} x_i}$$

$$\frac{\partial E(\theta)}{\partial \mu_j} = -\sum_{i=1}^{N} \nu_{ij} \Sigma_j^{-1}(x_i - \mu_j) = 0$$

得出 $\mu_j = \dfrac{\sum_{i=1}^{N} \nu_{ij} x_i}{\sum_{i=1}^{N} \nu_{ij}}$

$$\frac{\partial E(\theta)}{\partial \Sigma_j^{-1}} = \sum_{i=1}^{N} \gamma_{ij}\left[+\frac{1}{2}\Sigma_j - \frac{1}{2}(x_i-\mu_j)(x_i-\mu_j)^T\right] = 0$$

得出 $$\Sigma_j = \frac{\sum_{i=1}^{N}\gamma_{ij}(x_i-\mu_j)(x_i-\mu_j)^T}{\sum_{i=1}^{N}\gamma_{ij}}$$

设 $M(\theta) = E(\theta) + \lambda\left(\sum_{j=1}^{K}\pi_j - 1\right)$

（因为 $\sum_{j=1}^{N}\pi_j = 1$，上式用拉格朗日乘子法求解）

$$\frac{\partial M(\theta)}{\partial \pi_j} = \sum_{i=1}^{N}\gamma_{ij}\cdot\frac{1}{\pi_j} + \lambda = 0$$

即 $$\sum_{i=1}^{N}\gamma_{ij} + \lambda\pi_j = 0$$

上式对所有 j 求和，并利用 $\sum_{j=1}^{K}\gamma_{ij} = 1$

有：$$\sum_{j=1}^{N}\left(\sum_{i=1}^{N}\gamma_{ij} + \lambda\pi_j\right) = 0$$

即 $$N + \lambda = 0 \Longrightarrow \lambda = -N$$

因此：$$\sum_{i=1}^{N}\gamma_{ij} - N\pi_j = 0 \Rightarrow$$

$$\pi_j = \frac{1}{N}\sum_{i=1}^{N}\gamma_{ij}$$

所以
$$\begin{cases} \pi_j^{(new)} = \frac{1}{N}\sum_{i=1}^{N}\gamma_{ij} \\ \mu_j^{(new)} = \frac{\sum_{i=1}^{N}\gamma_{ij}x_i}{\sum_{i=1}^{N}\gamma_{ij}} \\ \Sigma_j^{(new)} = \frac{\sum_{i=1}^{N}\gamma_{ij}(x_i-\mu_j^{(new)})(x_i-\mu_j^{(new)})^T}{\sum_{i=1}^{N}\gamma_{ij}} \end{cases}$$

第四步：重复②③直到收