

优化后似然更大

$$\log p(\mathbf{t} | \theta^{(g+1)}) \geq \log p(\mathbf{t} | \theta^{(g)})$$

$$\theta^{(g+1)} = \arg \max_{\theta} \int_{\mathbf{z}} \log p(\mathbf{x}, \mathbf{z} | \theta) p(\mathbf{z} | \mathbf{x}, \theta^{(g)}) d\mathbf{z}$$

保记 $\log p(\mathbf{x} | \theta^{(g+1)}) \geq \log p(\mathbf{x} | \theta^{(g)})$

推导: $\log p(\mathbf{x} | \theta) = \log p(\mathbf{x}, \mathbf{z} | \theta) - \log p(\mathbf{z} | \mathbf{x}, \theta)$

$$E[\log p(\mathbf{x} | \theta)] = E[\log p(\mathbf{x}, \mathbf{z} | \theta)] - E[\log p(\mathbf{z} | \mathbf{x}, \theta)]$$

$p(\mathbf{z} | \mathbf{x}, \theta^{(g)}) \leftarrow \text{分布} \rightarrow p(\mathbf{z} | \mathbf{x}, \theta^{(g)})$

$$\int_{\mathbf{z}} \log p(\mathbf{t} | \theta) p(\mathbf{z} | \mathbf{x}, \theta^{(g)}) d\mathbf{z}$$

$$\begin{aligned} \log p(\mathbf{t} | \theta) &= \int \log p(\mathbf{t}, \mathbf{z} | \theta) p(\mathbf{z} | \mathbf{t}, \theta^{(g)}) d\mathbf{z} \\ &\quad - \int \log p(\mathbf{z} | \mathbf{t}, \theta) p(\mathbf{z} | \mathbf{t}, \theta^{(g)}) d\mathbf{z} \\ &= Q(\theta, \theta^{(g)}) - H(\theta, \theta^{(g)}) \end{aligned}$$

IF: $H(\theta^{(g)}, \theta^{(g)}) \geq H(\theta, \theta^{(g)}) \quad \forall \theta$

$\Rightarrow H(\theta^{(g)}, \theta^{(g)}) \geq H(\theta^{(g+1)}, \theta^{(g)})$

证明

$$H(\theta^{(g)}, \theta^{(g)}) - H(\theta, \theta^{(g)}) \geq 0$$

$$= \int \log p(\mathbf{z} | \mathbf{t}, \theta^{(g)}) p(\mathbf{z} | \mathbf{t}, \theta^{(g)}) d\mathbf{z} - \int \log p(\mathbf{z} | \mathbf{t}, \theta) p(\mathbf{z} | \mathbf{t}, \theta^{(g)}) d\mathbf{z}$$

$$= \int \log \left(\frac{p(\mathbf{z} | \mathbf{t}, \theta^{(g)})}{p(\mathbf{z} | \mathbf{t}, \theta)} \right) \cdot p(\mathbf{z} | \mathbf{t}, \theta^{(g)}) d\mathbf{z}$$

$$= \int -\log \frac{p(\mathbf{z} | \mathbf{t}, \theta)}{p(\mathbf{z} | \mathbf{t}, \theta^{(g)})} \cdot p(\mathbf{z} | \mathbf{t}, \theta^{(g)}) d\mathbf{z}$$

$$\geq \int -\log \frac{p(\mathbf{z} | \mathbf{t}, \theta)}{p(\mathbf{z} | \mathbf{t}, \theta)} p(\mathbf{z} | \mathbf{t}, \theta^{(g)}) d\mathbf{z}$$

$$\geq \int -\log p(\mathbf{z} | \mathbf{t}, \theta) d\mathbf{z}$$

$$\geq -\log 1 = 0$$

可以得到

$$\log p(\mathbf{t} | \theta^{(g+1)}) \geq \log p(\mathbf{t} | \theta^{(g)})$$

凸函数

杰森不等式

$\mathbf{t}_1 + (1-p)\mathbf{t}_2$

$(p\mathbf{t}_1 + (1-p)\mathbf{t}_2)$

$$(1-p)f(\mathbf{t}_1) + pf(\mathbf{t}_2)$$

$$\geq f[(1-p)\mathbf{t}_1 + p\mathbf{t}_2]$$

EM 算法的 Tips

- EM 算法不能保证对每一个应用场景的问题都能得到全局最优解，如果目标函数是凸函数，那么可以保证。但若不是，比如聚类问题中，若对文本分类中的余弦相似度计算函数就不可能保证是凸的。



