优化后似然更大

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

$P(x) = \frac{P(x,z)}{P(z|x)}$

- $\theta^{(g+1)} = \arg\max_{\theta} \int_z \log P(X,z|\theta)\, P(z|X,\theta^{(g)})\, dz.$

保证 $\log P(X|\theta^{(g+1)}) \geq \log P(X|\theta^{(g)})$.

推导: $\log P(X|\theta) = \log P(X,z|\theta) - \log P(z|X,\theta)$.

$E[\log P(X|\theta)] = E[\log P(X,z|\theta)] - E[\log P(z|X,\theta)]$.

$P(z|X,\theta^{(g)})$ ←分布→ $P(z|X,\theta^{(g)})$

$\int_z \log P(x|\theta)\, P(z|X,\theta^{(g)})\, dz$

$$\log P(x|\theta) = \int \log P(x,z|\theta)\, P(z|x,\theta^{(g)})\, dz - \int \log P(z|x,\theta)\, P(z|x,\theta^{(g)})\, dz$$

$$= Q(\theta,\theta^{(g)}) - H(\theta,\theta^{(g)}).$$

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

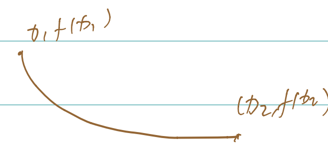
$\Rightarrow H(\theta^{(g)},\theta^{(g)}) \geq H(\theta^{(g+1)},\theta^{(g)})$.

证明

$$\boxed{\begin{aligned}
& H(\theta^{(g)},\theta^{(g)}) - H(\theta,\theta^{(g)}) \geq 0 \\
&= \int \log P(z|x,\theta^{(g)})\, P(z|x,\theta^{(g)})\, dz - \int \log P(z|x,\theta)\, P(z|x,\theta^{(g)})\, dz \\
&= \int \log\left(\frac{P(z|x,\theta^{(g)})}{P(z|x,\theta)}\right) \cdot P(z|x,\theta^{(g)})\, dz \\
&= \int -\log \frac{P(z|x,\theta)}{P(z|x,\theta^{(g)})} \cdot P(z|x,\theta^{(g)})\, dz \\
&\geq \int -\log \frac{P(z|x,\theta)}{\cancel{P(z|x,\theta^{(g)})}}\, \cancel{P(z|x,\theta^{(g)})}\, dz \\
&\geq \int -\log P(z|x,\theta)\, dz \\
&\geq -\log 1 = 0
\end{aligned}}$$

凸函数

杰森不等式




$(1-p)f(x_1) + pf(x_2) \geq f[(1-p)x_1 + px_2]$

可以得到

$$\log P(x|\theta^{(g+1)}) \geq \log P(x|\theta^{(g)})$$

# EM 算法的 Tips

- EM算法不能保证对每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能保证是凸的；