白板系列(10)　EM算法　算法收敛性证明

$\boxed{\text{GitHub: CBJerry993}}$

有隐变量混合模型参数估计

生成模型▲

$z \xrightarrow{\text{生成}} x$

找θ, Learning问题▲

坐标上升法思想▲ (SMO)

MLE : $p(x|\theta)$　$\theta = \arg\max\limits_{\theta} p(x|\theta) \xrightarrow{\text{使用log}} \arg\max\limits_{\theta} \underbrace{\log p(x|\theta)}_{\text{log-likelihood}}$

EM = $\theta^{(t+1)} = \arg\max\limits_{\theta} \int_z \underbrace{\log p(x,z|\theta)}_{\text{对数完全概率}} \cdot \underbrace{p(z|x,\theta^{(t)})}_{\text{后验概率}} dz$ 迭代

期望 $E_{z|x,\theta^{(t)}}[\log p(x,z|\theta)]$

$\theta^{(t)} \longrightarrow \theta^{(t+1)}$　$\log p(x|\theta^{(t)}) \le \log p(x|\theta^{(t+1)}) \longrightarrow$ 要证明

已知: $\log p(x|\theta) = \log p(x,z|\theta) - \log p(z|x,\theta)$　左边对z积分 $= \int_z p(z|x,\theta^{(t)}) \cdot \log p(x|\theta) \cdot dz$

右边 $= \underbrace{\int_z p(z|x,\theta^{(t)}) \cdot \log p(x,z|\theta) dz}_{Q(\theta,\theta^{(t)}) \to EM} - \underbrace{\int_z p(z|x,\theta^{(t)}) \cdot \log p(z|x,\theta) dz}_{H(\theta,\theta^{(t)})}$

与θ无关,提前 $= \log p(x|\theta) \underbrace{\int_z p(z|x,\theta^{(t)}) \cdot dz}_{=1} = \log p(x|\theta)$

令:

由定义 ~~θ^(t+1)~~ $Q(\theta^{(t+1)},\theta^{(t)}) \ge Q(\theta^{(t)},\theta^{(t)})$

只要证明 $H(\theta^{(t+1)},\theta^{(t)}) \le H(\theta^{(t)},\theta^{(t)})$

相减 $= \int_z p(z|x,\theta^{(t)}) \cdot \log p(z|x,\theta^{(t+1)}) dz$

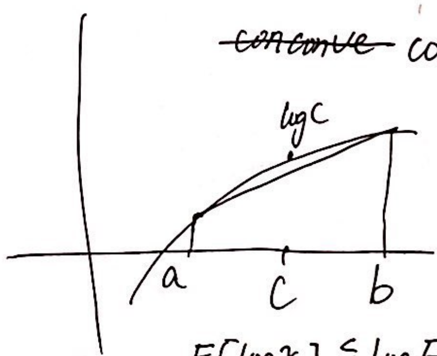
$- \int_z p(z|x,\theta^{(t)}) \cdot \log p(z|x,\theta^{(t)}) dz$

$= \int_z p(z|x,\theta^{(t)}) \cdot \log \frac{p(z|x,\theta^{(t+1)})}{p(z|x,\theta^{(t)})} dz$

略过 $\left\{ \begin{array}{l} = -KL(p(z|x,\theta^{(t)}) \,||\, p(z|x,\theta^{(t+1)})) \quad \text{KL距离} \ge 0 \\ \le 0 \end{array} \right.$

~~concave~~ convex

凸函数



$E[\log x] \le \log E(x)$

$\le \log \underbrace{\int_z p(z|x,\theta^{(t+1)}) dz}_{=1} \le \log 1 \le 0$　收敛性证毕!

↳ 约掉 $p(z|x,\theta^{(t)})$

白板系列(10)　　EM算法　　公式导出 ELBO + KL Divergence

. evidence lower bound 下界

$\boxed{\text{GitHub: CBJerry993}}$

EM公式 $\theta^{(t+1)} = \arg\max_{\theta} \int \log P(x,z|\theta) \cdot P(z|x,\theta^{(t)})\, dz$

E-step: $P(z|x,\theta^{(t)}) \xrightarrow{\text{写}} E_{z|x,\theta^{(t)}}[\log P(x,z|\theta)]$

M-step: $\theta^{(t+1)} = \arg\max_{\theta} E_{z|x,\theta^{(t)}}[\log P(x,z|\theta)]$

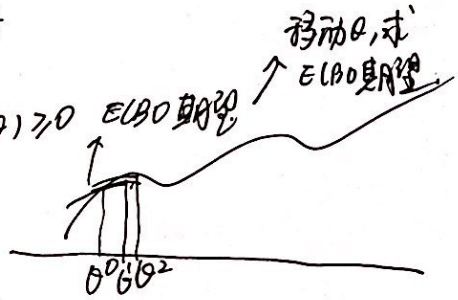
上节已证明收敛性 $\log P(x|\theta^{(t+1)}) \geq \log P(x|\theta^{(t)})$

$\log P(x|\theta) = \log P(x,z|\theta) - \log P(z|x,\theta)$　呈全一句验

$= \log \frac{P(x,z|\theta)}{q(z)} - \log \frac{P(z|x,\theta)}{q(z)}$　$q(z) \neq 0$

左边积分对 q $= \int_z q(z) \cdot \log P(x|\theta)\, dz = \log P(x|\theta) \underbrace{\int_z q(z) \cdot dz}_{=1} = \log P(x|\theta)$

右边~~~~~ $= \underbrace{\int_z q(z) \log \frac{P(x,z|\theta)}{q(z)} dz}_{\text{ELBO}} - \underbrace{\int_z q(z) \log \frac{P(z|x,\theta)}{q(z)} dz}_{KL(q(z) \| P(z|x,\theta)) \geq 0}$



$\log P(x|\theta) = \text{ELBO} + KL \geq \text{ELBO}$

那么变成最大化ELBO问题.　取等号 $q(z) = P(z|x,\theta^{(t)})$

$\hat{\theta} = \arg\max_{\theta} \text{ELBO} = \arg\max_{\theta} \int q(z) \log \frac{P(x,z|\theta)}{q(z)} dz$

$= \arg\max_{\theta} \int P(z|x,\theta^{(t)}) \log \frac{P(x,z|\theta)}{P(z|x,\theta^{(t)})} dz$

↳ 与θ无关

$= \arg\max \int P(z|x,\theta^{(t)}) \cdot \log P(x,z|\theta) \cdot dz$　推导出了EM公式!

固定θ,求期望
对期望,求θ
不断EM

白板系列(10) EM算法 公式导出 ELBO + Jensen inequality

GitHub:OBJkmy993

$$\log p(x|\theta) = \log \int_z p(x,z|\theta)\,dz$$

$$= \log \int_z \frac{p(x,z|\theta)}{q(z)} \cdot q(z) \cdot dz$$

$$= \log E_{q(z)}\left[\frac{p(x,z|\theta)}{q(z)}\right]$$

Jeson不等式 $\geqslant E_{q(z)}\left[\log \frac{p(x,z|\theta)}{q(z)}\right] \longrightarrow$ 下界 ELBO

当 $\frac{p(x,z|\theta)}{q(z)}$ 是常数，取等号 $= C$

$$q(z) = \frac{1}{C} p(x,z|\theta) \quad \text{———} \ ①$$

$$1 = \int_z q(z)dz = \int_z \frac{1}{C} p(x,z|\theta)dz$$

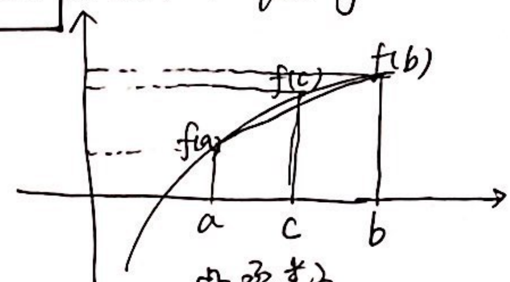
$$1 = \frac{1}{C}\int_z p(x,z|\theta)\cdot dz$$ 对联合概率积分变成边缘概率

$$1 = \frac{1}{C} p(x|\theta)$$

$\Rightarrow C = p(x|\theta)$ 代入①

$$q(z) = \frac{p(x,z|\theta)}{p(x|\theta)} = p(z|x,\theta)$$ 后验 ✓ 证毕!

Jenson inequlity



凸函数

$t \in [0,1]$

$c = ta + (1-t)b$

$$f(ta+(1-t)b) \geqslant f(a)\cdot t + f(b)\cdot(1-t)$$

当 $t=\frac{1}{2}$ 时

$$f\left(\frac{a+b}{2}\right) \geqslant \frac{1}{2}(f(a)+f(b))$$

$$\boxed{f(E) \geqslant E[f]}$$

$$\hat{\theta} = \arg\max_\theta p(x|\theta) = \arg\max_\theta \log p(x|\theta)$$

广义EM ▲

$$\log p(x|\theta) = \underbrace{ELBO}_{\mathcal{L}(q,\theta)} + KL(q\|p)$$

$$\begin{cases} ELBO = E_{q(z)}\left[\log \frac{p(x,z|\theta)}{q(z)}\right] \\ KL(q\|p) = \int q(z)\cdot \log \frac{q(z)}{p(z|x,\theta)}\cdot dz \end{cases}$$

狭义EM ▲

$$\begin{cases} \text{固定}\theta,\ \hat{q} = \arg\min_q KL(q\|p) = \arg\max_q \mathcal{L}(q,\theta) \longrightarrow E\text{-step}\quad \hat{q} = p(z|x,\theta^{(t)}) \\ \text{固定}\hat{q},\ \theta = \arg\max \mathcal{L}(\hat{q},\theta) \longrightarrow m\text{-step}\quad \theta^{(t+1)} = \arg\max_\theta E_{z\sim p(z|x,\theta^{(t)})}[\log p(x|z,\theta)] \end{cases}$$

$\because ELBO = E_{q(z)}[\log p(x,z|\theta)] + H[q(z)]$

广义EM ▲

$$\begin{cases} E \quad q^{(t+1)} = \arg\max_q \mathcal{L}(q,\theta^{(t)}) \\ M \quad \theta^{(t+1)} = \arg\max \mathcal{L}(q^{(t+1)},\theta) \end{cases}$$

$$\boxed{\text{变种 VBEM/VEM} \quad \text{MCEM}}$$

$$\mathcal{L}(q,\theta) = E_q[\log p(x,z) - \log q]$$

$$= E_q[\log p(x,z)] \underbrace{- E_q[\log q]}_{H[q]}$$

$H[q]$ 关于q的熵定义 $\longrightarrow \int q(z)\log\frac{1}{q(z)}\cdot dz$