# EM 期望最大

Expectation maximum

EM：$\longrightarrow$ 用于解决具有隐变量时，混合模型的参数估计。

简单的模型可直接求出解析解：MLE：$P(X|\theta)$

$$\theta_{MLE} = \arg\max_{\theta} \underbrace{\log P(X|\theta)}_{\text{log-likelihood}}$$

但混合模型的解析解很难求出来。
$\longrightarrow$ 迭代算法

EM：

$$\theta^{t+1} = \arg\max_{\theta} \int_z \underbrace{\log P(x, z|\theta)}_{\text{log - complete data}} \cdot \underbrace{P(z|x, \theta^{(t)})}_{\text{z的后验}} \, dz$$

收敛性证明：(合法性)

$\theta^{(t)} \longrightarrow \theta^{(t+1)}$

$\log P(x|\theta^{(t)}) \leq \log P(x|\theta^{(t+1)})$ $\longrightarrow$ need to be proved.

$$\log P(x|\theta) = \log(x, z|\theta) - \log P(z|x, \theta)$$

同时求期望：左边：$\int_z P(z|x, \theta^{(t)}) \cdot \log P(x|\theta) \, dz$

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

$= \log P(x|\theta)$

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

$Q(\theta^{(t+1)}, \theta^{(t)}) \geq Q(\theta, \theta^{(t)})$

$Q(\theta^{(t+1)}, \theta^{(t)}) \geq Q(\theta^{(t)}, \theta^{(t)})$

证：$H(\theta^{(t+1)}, \theta^{(t)}) \le H(\theta^{(t)}, \theta^{(t)})$

$$H(\theta^{(t+1)}, \theta^{(t)}) - H(\theta^{(t)}, \theta^{(t)})$$

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

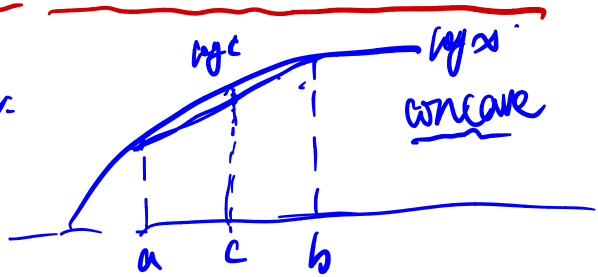
$$E_{z\sim P(z|x,\theta^{(t)})}\left[\log \frac{P(z|x,\theta^{(t+1)})}{P(z|x,\theta^{(t)})}\right]$$

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

$\rightarrow$ KL散度 $> 0$

$$= -KL(P(z|x,\theta^{(t)}) \,||\, P(z|x,\theta^{(t+1)})) \le 0$$

log function:




$E[\log x] \le \log E[x]$ ~ Jensen 不等式

个 KL 散度 ……

EM算法的导出!  $X$: observed data. $Z$: unobserved data (latent variable)

$(X,Z)$: complete data.  $\theta$: parameter

Recap:

EM公式: $\theta^{(t+1)} = \arg\max_{\theta} \int_z \underbrace{\log P(X,Z|\theta)}_{变量} \underbrace{P(Z|X,\theta^{(t)})}_{常数} dz$

E-step: $P(Z|X,\theta^{(t)}) \rightarrow E_{Z|X,\theta^{(t)}}[\log P(X,Z|\theta)]$

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

推导: $\log P(X|\theta) = \log P(X,Z|\theta) - \log P(Z|X,\theta)$  引入 $q(Z)$

$$= \log \frac{P(X,Z|\theta)}{q(Z)} - \log \frac{P(Z|X,\theta)}{q(Z)}, \quad q(Z) \neq 0$$

$P(X,Z|\theta) = P(X|\theta) P(Z|X,\theta)$

$P(X|\theta) = \frac{P(X,Z|\theta)}{P(Z|X,\theta)}$

$\log P(X|\theta) = \log P(X,Z|\theta) - \log P(Z|X,\theta)$

对 $q(Z)$ 求期望:

左边 $= \int_z q(Z) \log P(X|\theta) dz = \log P(X|\theta) \overset{1}{\boxed{\int_z dz}} = \log P(X|\theta)$

右边: $\underbrace{\int_z q(Z) \log \frac{P(X,Z|\theta)}{q(Z)} dz}_{\text{ELBO}} - \underbrace{\int_z q(Z) \log \frac{P(Z|X,\theta)}{q(Z)} dz}_{KL[q(Z)\,||\,P(Z|X,\theta)]}$.

ELBO: Evidence lower bound

of $\log P(X|\theta)$

KL-Divergence

$\log P(X|\theta) = ELBO + KL(q\|P)$ → z的后验

→ $KL \geq 0 \Leftrightarrow$ "=" $q = P$

∴ $\log P(X|\theta) \geq ELBO$

当 $q=P$ 时 $\log P(X|\theta) = ELBO$ → 下界

$$\theta^{(t+1)} = \arg\max_{\theta} ELBO$$

不太理解

$$= \arg\max_{\theta} \int_z q(z) \log \frac{P(X,z|\theta)}{q(z)} dz$$

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

$$= \arg\max_{\theta} \int_z P(z|X,\theta^{(t)}) [\log P(X,z|\theta) - \underline{\log P(z|X,\theta^{(t)})}] dz$$

$$= \arg\max_{\theta} \int P(z|X,\theta^{(t)}) \cdot \log P(X,z|\theta)$$

EM算法

与θ无关.

ELBO + Jensen equality.

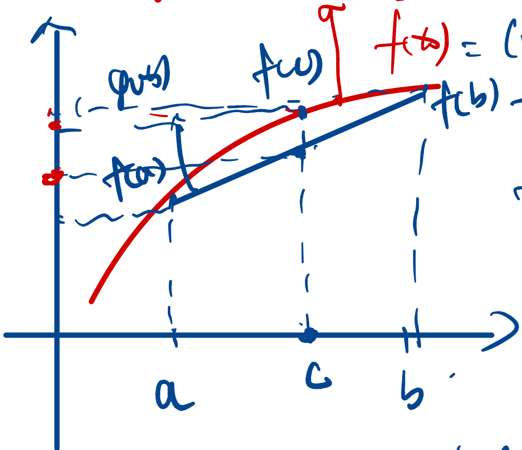
EM 公式导出:

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

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

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

Jensen inequality: concave function




$f(x)$ = concave function

$\Downarrow$

for $x_i$ in $(a,b)$

$f(x_i) \geq g(x_i)$

Jensen equality:

$t \in [0,1]$

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

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

$$\geq t f(a) + (1-t) f(b)$$

当 $t = \frac{1}{2}$ 时

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

$$f(z) \geq E[f]$$

$$\geq \boxed{E_{q(z)}\left[\log \frac{P(x,z|\theta)}{q(z)}\right]} \rightsquigarrow \text{ELBO}$$

$$\text{"="} \iff \frac{P(x,z|\theta)}{q(z)} = C \rightsquigarrow \text{Constant}$$

we wanna chose a $q(z)$ letting the bound tight.

so

$$q(z) = \frac{1}{c} P(x,z|\theta)$$

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

$$1 = \frac{1}{c} P(x|\theta)$$

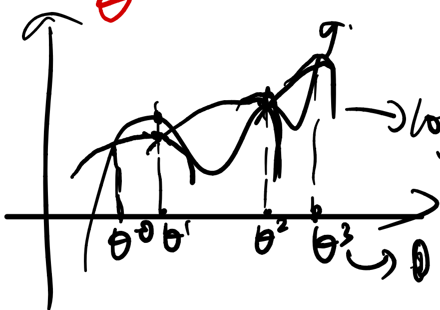
$$\Downarrow$$

$$c = P(x|\theta)$$

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

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

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




$\rightarrow \theta^3$ maximize $\log P(x|\theta)$

# EM - Extention

→ 解决参数估计问题, MLE

① 从狭义 EM → 广义 EM

② 狭义 EM 是 广义 EM 的一个特例

③ real EM

$\hat{\theta} = \arg\max P(X|\theta)$

$= \arg\max \log P(X|\theta)$

X由z生成, 将P(X)分解处理

复杂. 需要假设.

$P(X) = \int_z P(X,z)\,dz$

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

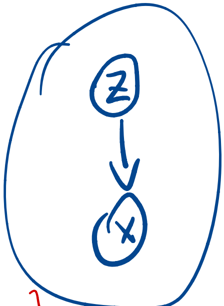
EM 用于 概率生成模型。

$X = \{x_i\}_{i=1}^N$ observed.

$Z = \{z_i\}_{i=1}^N$

$X, Z$

$\theta$




GMM

$z = 1 \quad 2 \quad \cdots \quad K$

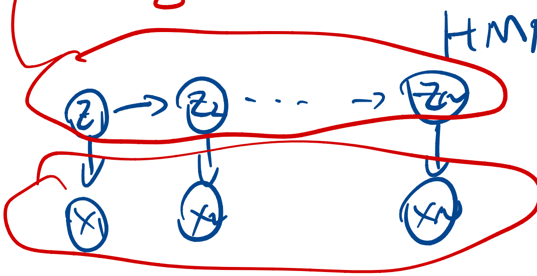
$\quad\; P_1 \quad P_2 \quad\quad P_K$

$P(X|z) \sim$ Gau

$P(X,z)$

z

HMM




$X : X_{1:N}$

目标函数：

$$\log P(X|\theta) = \underbrace{\mathrm{ELBO}}_{\mathcal{L}(q,\theta)} + \mathrm{KL\text{-}Divergence}(q\|p)$$

概分解 $q$ 　　　　$q$ 的后验

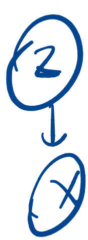
$$\begin{cases} \mathrm{ELBO} = E_{q(z)}\left[\log \dfrac{P(x,z|\theta)}{q(z)}\right]. \\ \mathrm{KL}(q,p) = \displaystyle\int q(z)\cdot \log \frac{q(z)}{P(z|x,\theta)}\, dz \end{cases}$$

$$\geq \mathcal{L}(q,\theta).$$

EM： E-step

$$\hat{q} = P(z|x,\theta^{(t)}).$$

我们直接让 $q=p$，但事实上 $P(z|x,\theta)$ 可能是 intractable 的.




GMM、HMM 的 $P(z|x,\theta)$ 是结构简单的，是 tractable 的，所以 EM works.

怎么办？

$$\begin{cases} \text{变分推断}: \boxed{\mathrm{VBEM}/\mathrm{VEM}}. \\ \mathrm{MCMC}: \boxed{\mathrm{MCEM}} \to \text{MCMC 去逼近后验分布} \end{cases}$$

因此 q 不一定能取到 P.

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

E-step

$\theta$ 固定，$\hat{q} = \arg\min_q KL(q\|P) = \arg\max \mathcal{L}(q,\theta)$

$\hat{q}$ 固定，$\theta = \arg\max_\theta \mathcal{L}(\hat{q},\theta)$

M-step.

广义 EM

E-step: $q^{(t+1)} = \arg\max_q \mathcal{L}(q,\theta^{(t)})$

M-step: $\theta^{(t+1)} = \arg\max_\theta \mathcal{L}(q^{(t+1)},\theta)$

$\mathcal{L}(q,\theta) = E_q[\log P(X,Z) - \log q]$

$= E_q[\log P(X,Z)] \underbrace{- E_q[\log q]}_{H[q]}$

$= E_q[\log P(X,Z)] + H[q]$ $H[q]$ 熵.