EM (期望最大) 算法.

MLE : $P(X|\theta)$

$\theta_{MLE} = \arg\max\limits_{\theta} \log P(X|\theta)$ → log-likelihood

MLE可以直接求得解析解. 但若有隐变量, 解会很难求出.

EM算法是一种迭代优化策略, 每一次迭代都分为2步:

① 期望步 E-step (E步, Expectation)

② 极大步 M-step (M步, Maximization)

X: observed data

Z: unobserved data (latent variable)

(X, Z): complete data

θ: parameter

EM 公式:

$$\theta^{(t+1)} = \arg\max_{\theta} \int \log P(X,Z|\theta)\, P(Z|X,\theta^{(t)})\, dz$$

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

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

一. 公式推导 (ELBO + KL散度)

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

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

q(Z)是关于Z的分布.

将等式两边看作函数, 等式两边同时对q(Z)求期望.

$$\text{左边} = \int_Z q(Z) \log P(X|\theta)\, dz = \log P(X|\theta) \underbrace{\int_Z q(Z)\, dz}_{1} = \log P(X|\theta)$$

$$\text{右边} = \underbrace{\int_Z q(Z) \log \frac{P(X,Z|\theta)}{q(Z)}\, dz}_{\text{ELBO}} - \underbrace{\int_Z q(Z) \log \frac{P(Z|X,\theta)}{q(Z)}\, dz}_{KL(q(Z)\,\|\,P(Z|X,\theta))}$$

则 $\log P(X|\theta) = ELBO + KL(q||p)$

由于KL散度大于等于0，因此，$\log P(X|\theta) \geq ELBO$，即ELBO是似然函数 $\log P(X|\theta)$ 的下界。

当且仅当 $q(z) = p(z|x,\theta)$ 时，$KL(q||p)=0$，此时 $\log P(X|\theta) = ELBO$

在每次迭代中取 $q(z) = P(z|x,\theta^t)$，就可以保证 $\log p(x|\theta^t)$ 与ELMO相等，也就是：

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

当 $\theta = \theta^t$ 时，$\log P(X|\theta) = ELBO$

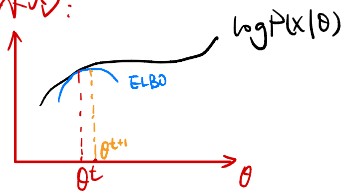
ELBO和 $\log P(X|\theta)$ 都是 θ 的函数，在第t次迭代时模型的参数为 θ^t.

此时：

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

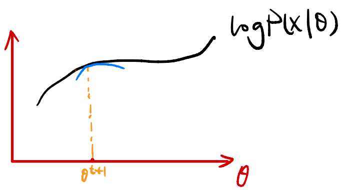
$\log P(X|\theta) = ELBO + KL \geq ELBO$. 只有当 $\theta = \theta^t$ 时，二者相等

故图像可表示为：



当 $\theta = \theta^{t+1}$ 时，参数为 θ^t 的ELBO取得最大值，如上图。

将参数更新为 θ^{t+1}，ELBO图像会随之更新，如图：



此时的ELBO函数只有在 $\theta = \theta^{t+1}$ 时，才会等于 $\log P(X|\theta)$。

因为在 θ^t 时，$\theta^{t+1} = \arg\max_\theta ELBO$ 且ELBO是 $\log P(X|\theta)$ 的下界，

所以 $\log P(X|\theta_{t+1})$ 一定 $\geq \log P(X|\theta_t)$，依据此规律一直迭代，完成参数估计。

$$\theta^{t+1} = \arg\max_\theta ELBO = \arg\max_\theta \int_z P(z|X,\theta^t) \log \frac{P(X,z|\theta)}{P(z|X,\theta^t)} dz$$

$$= \arg\max_\theta \int_z P(z|X,\theta^t) \log P(X,z|\theta) dz - \underbrace{\arg\max_\theta \int_z P(z|X,\theta^t) \log P(z|X,\theta^t) dz}_{\text{与}\theta\text{无关}}$$

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

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

公式推导(ELBO+Jensen不等式)

$$\log P(X|\theta) = \log \int_z P(X,z|\theta) dz$$

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

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

$$\geq \underbrace{E_{q(z)}\left[\log \frac{P(X,z|\theta)}{q(z)}\right]}_{ELBO}$$

当 $\frac{P(X,z|\theta)}{q(z)} = C$ 时等号成立

$$q(z) = \frac{1}{C} P(X,z|\theta)$$

$$1 = \int q(z) dz = \int_z \frac{1}{C} P(X,z|\theta) dz$$

$$1 = \frac{1}{C} P(X|\theta) \Longrightarrow C = P(X|\theta)$$

$$\therefore q(z) = \frac{1}{P(X|\theta)} \cdot P(X,z|\theta) = P(z|X,\theta)$$

此时与ELBO+KL散度相同.

三. EM算法的收敛性证明.

证明：当 $\theta^{(t)} \to \theta^{(t+1)}$ 时，有 $\log P(X|\theta^{(t)}) \leq \log P(X|\theta^{(t+1)})$

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

等式两边同时对 $P(z|X,\theta^{(t)})$ 求积分：

$$\text{左边} = \int_z P(z|X,\theta^{(t)}) \underbrace{\log P(X|\theta)}_{\text{与}z\text{无关}} dz = \log P(X|\theta) \underbrace{\int_z P(z|X,\theta^{(t)}) dz}_{1} = \log P(X|\theta)$$

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

令:

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

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

再证 $H(\theta^{(t+1)},\theta^{(t)}) \leq H(\theta^{(t)},\theta^{(t)})$ 即可.

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

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

$$= \underbrace{\int_z P(z|X,\theta^{(t)}) \log \frac{P(z|X,\theta^{(t+1)})}{P(z|X,\theta^{(t)})} dz}_{-KL散度} \leq 0$$

故 $H(\theta^{(t+1)},\theta^{(t)}) - H(\theta^{(t)},\theta^{(t)}) \leq 0$.

$\therefore \log(X|\theta^{(t+1)}) \geq \log(X|\theta^{(t)})$

EM是算法,不是模型,类似于GD.

广义EM

目标函数: $\log P(X|\theta) = ELBO + KL(q||p)$

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

ELBO与θ和q有关,可以写为关于θ和q的函数,记为 $L(q,\theta)$

在狭义EM中,我们是令 $q(z) = P(z|X,\theta)$ 的,但实际这个后验概率是可能无法求解的.

则当θ固定时, $\log P(X|\theta)$ 固定,我们希望 $KL(q||p)$ 越小越好,这样ELBO就会更大.

$$\begin{cases} \hat{q} = \arg\min_q KL(q||p) = \arg\max_q L(q|\theta) \\ 固定q, \ \theta = \arg\max_\theta L(\hat{q},\theta) \end{cases}$$

广义EM:

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

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

$$\text{其中 } ELBO = L(q, \theta) = E_{q(z)}[\log P(X, z|\theta) - \log q(z)]$$

$$= E_{q(z)}[\log P(X, z|\theta)] \underbrace{- E_{q(z)}[\log q(z)]}_{H[q]\text{ : q的熵}}$$

EM算法类似于坐标上升法(SMO)，如果在EM框架中，无法求解z的后验概率，那么需要采用一些变种的EM来估算这个后验：

① 基于平均场的变分推断，VBEM/VEM

② 基于蒙特卡洛的EM，MCEM