

## EM 算法的推导 (ELBO 角度)

### 1. EM 算法

E-Step:

$$Q(\theta, \theta^{(i)}) = E_{Z|Y, \theta^{(i)}}[\log P(Y, Z|\theta)] = \sum_Z P(Z|Y, \theta^{(i)}) \log P(Y, Z|\theta)$$

【利用当前时刻所估算参数，计算 Z 的分布，然后求期望】

M-Step

$$\theta^{(i+1)} = \arg \max_{\theta} Q(\theta, \theta^{(i)})$$

【最大化期望】

### 2. EM 算法有效性

核心：迭代计算是单调递增（似然函数  $\log P(Y|\theta)$  越来越大），且有上界。

以下证明单调递增  $\log P(Y|\theta^{(i+1)}) \geq \log P(Y|\theta^{(i)})$

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

两边同时乘以引入的一个分布  $P(Z|Y, \theta^{(i)})$ ，再对 Z 取累加

左边

$$\sum_Z P(Z|Y, \theta^{(i)}) \log P(Y|\theta) = \log P(Y|\theta) \sum_Z P(Z|Y, \theta^{(i)}) = \log P(Y|\theta)$$

右边

$$\sum_Z P(Z|Y, \theta^{(i)}) \log P(Y, Z|\theta) - \sum_Z P(Z|Y, \theta^{(i)}) \log P(Z|Y, \theta)$$

令

$$Q(\theta, \theta^{(i)}) \equiv \sum_Z P(Z|Y, \theta^{(i)}) \log P(Y, Z|\theta)$$

$$H(\theta, \theta^{(i)}) \equiv - \sum_Z P(Z|Y, \theta^{(i)}) \log P(Z|Y, \theta)$$

则

$$\log P(Y|\theta) = Q(\theta, \theta^{(i)}) + H(\theta, \theta^{(i)})$$

所以

$$\log P(Y|\theta^{(i+1)}) = Q(\theta^{(i+1)}, \theta^{(i)}) + H(\theta^{(i+1)}, \theta^{(i)})$$

根据 M-step ( $\theta^{(i+1)}$  为让...取最大值的参数)

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

$\theta^{(i+1)}$  让  $Q(\theta, \theta^{(i)})$  取最大, 对右边的  $\theta$  任意取值 (此处取为  $\theta^{(i)}$ ), 不等式依然成立

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

另外

$$\begin{aligned} H(\theta^{(i+1)}, \theta^{(i)}) - H(\theta^{(i)}, \theta^{(i)}) &= - \sum_Z P(Z|Y, \theta^{(i)}) \log \frac{P(Z|Y, \theta^{(i+1)})}{P(Z|Y, \theta^{(i)})} \\ &= KL(P(Z|Y, \theta^{(i)}) || P(Z|Y, \theta^{(i+1)})) \geq 0 \\ H(\theta^{(i+1)}, \theta^{(i)}) &\geq H(\theta^{(i)}, \theta^{(i)}) \end{aligned}$$

所以

$$\begin{aligned} \log P(Y|\theta^{(i+1)}) &= Q(\theta^{(i+1)}, \theta^{(i)}) + H(\theta^{(i+1)}, \theta^{(i)}) \geq Q(\theta^{(i)}, \theta^{(i)}) + H(\theta^{(i)}, \theta^{(i)}) \\ &= \log P(Y|\theta^{(i)}) \end{aligned}$$

### 3. 函数 Q 的由来

根据极大似然估计, 需要最大化观测数据的似然函数  $\log P(Y|\theta)$  【可以证明, 梯度下降法不好求解, 因此才需要采用 EM 算法】

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

引入分布  $Q(Z)$ , 上式减法两侧同时减  $Q(Z)$

$$\log P(Y|\theta) = \log \frac{P(Y, Z|\theta)}{Q(Z)} - \log \frac{P(Z|Y, \theta)}{Q(Z)}$$

再求关于  $Q(Z)$  的期望

Left:

$$\sum_Z Q(Z) \log P(Y|\theta) = \log P(Y|\theta) \sum_Z Q(Z) = \log P(Y|\theta)$$

Right:

$$\sum_Z Q(Z) \log \frac{P(Y, Z|\theta)}{Q(Z)} - \sum_Z Q(Z) \log \frac{P(Z|Y, \theta)}{Q(Z)}$$

定义  $ELBO$  (Evidence Lower Bound)

$$ELBO = \sum_Z Q(Z) \log \frac{P(Y, Z|\theta)}{Q(Z)}$$

则

$$\log P(Y|\theta) = ELBO + KL(Q(Z) || P(Y, Z|\theta)) \geq ELBO$$

$ELBO$ 是 $\log P(Y|\theta)$ 一个下界。

$\log P(Y|\theta)$ 的优化不能直接计算取得，需要采用分布迭代算法。在迭代计算时，当前参数 $\theta^{(i)}$ ，需要得到下一个参数。有一种策略是优化下界 $ELBO$ ， $ELBO$ 和 $Q(Z)$ 相关， $Q(Z)$ 的最佳取值为 $\theta = \theta^{(i)}$ 时， $\log P(Y|\theta^{(i)}) = ELBO$ ，这样最大化 $ELBO$ 就能尽可能地让 $\log P(Y|\theta)$ 取到最大值

【

等号成立的条件：

$$Q(Z) = P(Y, Z|\theta^{(i)}) \Rightarrow KL(Q||P) = 0$$

即 $Q$ 取当前参数 $\theta^{(i)}$ 下的后验条件概率 $P(Z|Y, \theta^{(i)})$

这里需要注意， $Q(Z)$ 和 $\theta$ 无关， $P(Y, Z|\theta)$ 和 $\theta$ 有关， $P(Y, Z|\theta^{(i)})$ 和 $\theta$ 无关

】

也就是说，对于当前参数 $\theta^{(i)}$ ，下一个参数的取得可以最大化下界 $ELBO$ 。

$$\hat{\theta} = \arg \max_{\theta} ELBO = \arg \max_{\theta} \sum_Z Q(Z) \log \frac{P(Y, Z|\theta)}{Q(Z)}$$

$$\hat{\theta} = \arg \max_{\theta} \sum_Z P(Z|Y, \theta^{(i)}) \log \frac{P(Y, Z|\theta)}{P(Z|Y, \theta^{(i)})}$$

$$\hat{\theta} = \arg \max_{\theta} \sum_Z P(Z|Y, \theta^{(i)}) [\log P(Y, Z|\theta) - \log P(Z|Y, \theta^{(i)})]$$

$\log P(Z|Y, \theta^{(i)})$ 与 $\theta$ 无关，因此可以改写成

$$\hat{\theta} = \arg \max_{\theta} \sum_Z P(Z|Y, \theta^{(i)}) \log P(Y, Z|\theta)$$

这就是 M-Step，其中

$$Q(\theta, \theta^{(i)}) = \sum_Z P(Z|Y, \theta^{(i)}) \log P(Y, Z|\theta)$$

得到新的 $\theta$ 后  $ELBO$ 也随之得到更新，迭代得以继续。