最大期望算法

吴天阳　张卓立

XJTU

强基数学

2022 年 11 月 14 日

收敛性

问题:

1. EM 算法是否收敛?
2. 如果收敛, 能否收敛到全局最大值?

设有 m 个样本观察数据, $x=(x^{(1)},\cdots,x^{(m)})$, 模型参数 θ, 在观测数据中有未观察到的隐含数据 $z=(z^{(1)},\cdots,z^{(m)})$, 令 $Q_i\left(z^{(i)}\right)=\mathbb{P}\left(z^{(i)}\mid x^{(i)};\theta\right)$, θ_j 表示根据 EM 算法迭代得到的参数 θ 的各个估计值.

要证明 EM 算法的收敛性, 只需证明对数似然的值在迭代过程中单调递增, 即

$$\sum_{i=1}^{m}\log\mathbb{P}\left(x^{(i)};\theta^{j+1}\right)\geqslant\sum_{i=1}^{m}\log\mathbb{P}\left(x^{(i)};\theta^{j}\right).$$

由于

$$L\left(\theta,\theta^{j}\right)=\sum_{i=1}^{m}\sum_{z^{(i)}}\mathbb{P}\left(z^{(i)}\mid x^{(i)};\theta^{j}\right)\log\mathbb{P}\left(x^{(i)},z^{(i)};\theta\right),\tag{1}$$

令

$$H(\theta,\theta^{j})=\sum_{i=1}^{m}\sum_{z^{(i)}}\mathbb{P}\left(z^{(i)}\mid x^{(i)};\theta^{j}\right)\log\mathbb{P}\left(z^{(i)}\mid x^{(i)};\theta\right)\tag{2}$$

式 $(1)-(2)$ 得到:

$$\sum_{i=1}^{m}\log\mathbb{P}\left(x^{(i)};\theta\right)=L(\theta,\theta^j)-H(\theta,\theta^j),$$

将 θ^j,θ^{j+1} 分别代入上式, 相减可得

$$\begin{aligned}&\sum_{i=1}^{m}\log\mathbb{P}\left(x^{(i)};\theta^{j+1}\right)-\sum_{i=1}^{m}\log\mathbb{P}\left(x^{(i)};\theta^{j}\right)\\=&\left[L\left(\theta^{j+1},\theta^j\right)-L\left(\theta^j,\theta^j\right)\right]-\left[H\left(\theta^{j+1},\theta^j\right)-H\left(\theta^j,\theta^j\right)\right].\end{aligned}\tag{3}$$

下面证明式 (3) 右端非负.

收敛性

问题 1：EM 算法是否收敛?

由于 θ^{j+1} 使得 $L(\theta,\theta^j)$ 极大, 所以

$$L\left(\theta^{j+1},\theta^j\right)-L\left(\theta^j,\theta^j\right)\geqslant 0,$$

又因为

$$\begin{aligned}
H\left(\theta^{j+1},\theta^j\right)-H\left(\theta^j,\theta^j\right)&=\sum_{i=1}^m\sum_{z^{(i)}}\mathbb{P}\left(z^{(i)}\mid x^{(i)};\theta^j\right)\log\frac{\mathbb{P}\left(z^{(i)}\mid x^{(i)};\theta^{j+1}\right)}{\mathbb{P}\left(z^{(i)}\mid x^{(i)};\theta^j\right)}\\
&\leqslant\sum_{i=1}^m\log\left[\sum_{z^{(i)}}\mathbb{P}\left(z^{(i)}\mid x^{(i)};\theta^{j+1}\right)\right.\\
&\qquad\left.\cdot\frac{\mathbb{P}\left(z^{(i)}\mid x^{(i)};\theta^{j+1}\right)}{\mathbb{P}\left(z^{(i)}\mid x^{(i)};\theta^j\right)}\right]\\
&=\sum_{i=1}^m\log\sum_{z^{(i)}}\mathbb{P}\left(z^{(i)}\mid x^{(i)};\theta^{j+1}\right)=0,
\end{aligned}\tag{4}$$

其中式 (4) 用到 Jensen 不等式. 综上, 式 (3) 得证, 即 EM 算法具有收敛性.

收敛性

问题 2: 如果收敛, 能否收敛到全局最大值?

根据问题 1 的推导, EM 算法可以保证收敛到一个稳定点, 但不能保证是全局最优的, 所以它是局部最优的算法.
若 $L(\theta, \theta^j)$ 是凸的, 则可以收敛到全局最大值.