EM 算法及其推广

–算法介绍和收敛性证明

罗俊勋

School of Mathmatical
Zhejiang University

2024 年 11 月 06 日

EM 算法介绍

适用场景

适用于模型已知，参数未定的情况：如果没有隐变量就直接使用最大似然估计，若不然考虑使用 EM 算法.

- 统计男女生身高，已知其服从正态分布，但不同性别的均值和方差未知．(数据混在一起)
  - 根据名字判断性别，但有些名字是中性的.
  - 根据身高判断性别，但有些人的身高不符合性别特征.
- 三枚质量不均的硬币 A, B, C，正面出现的结果为 π, p, q．每一次实验抛掷两次，第一次抛 A, 如果正面则抛 B，否则抛 C，只记录最后一次的结果．正面为 1，反面为 0.

EM 算法流程和 Q 函数定义

算法流程

算法接受变量数据 Y, 隐变量数据 Z, 联合分布 $P(Y, Z|\theta)$, 条件分布 $P(Z|Y, \theta)$. 输出参数 θ.

- 初始化参数 $\theta^{(0)}$.
- E 步: $Q(\theta, \theta^{(i)}) = \mathbb{E}_Z[\log P(Y, Z|\theta)|Y, \theta^{(i)}]$
- M 步: $\theta^{(i+1)} = \arg\max_{\theta} Q(\theta, \theta^{(i)})$
- 重复 E, M 步骤，直到满足收敛条件

Q 函数

完全数据的对数似然函数 $\log P(Y, Z|\theta)$ 关于给定观测数据 Y，和当前参数 $\theta^{(i)}$, 下对未观测数据 Z 的期望

$$Q(\theta, \theta^{(i)}) = \mathbb{E}_Z[\log P(Y, Z|\theta)|Y, \theta^{(i)}]$$

算法导出

要通过迭代求出 $L(\theta)=P(Y|\theta)$ 的极值，我们希望 $L(\theta^{(i+1)})\geq L(\theta^{(i)})$

$$
\begin{aligned}
L(\theta)-L\left(\theta^{(i)}\right) &= \log\left(\sum_Z P\left(Z\mid Y,\theta^{(i)}\right)\frac{P(Y\mid Z,\theta)P(Z\mid\theta)}{P\left(Z\mid Y,\theta^{(i)}\right)}\right)-\log P\left(Y\mid\theta^{(i)}\right) \\
&\geq \sum_Z P\left(Z\mid Y,\theta^{(i)}\right)\log\frac{P(Y\mid Z,\theta)P(Z\mid\theta)}{P\left(Z\mid Y,\theta^{(i)}\right)}-\log P\left(Y\mid\theta^{(i)}\right) \\
&= \sum_Z P\left(Z\mid Y,\theta^{(i)}\right)\log\frac{P(Y\mid Z,\theta)P(Z\mid\theta)}{P\left(Z\mid Y,\theta^{(i)}\right)P\left(Y\mid\theta^{(i)}\right)}
\end{aligned}
$$

令 $B(\theta,\theta^{(i)})=L(\theta^{(i)})+\sum_Z P\left(Z\mid Y,\theta^{(i)}\right)\log\frac{P(Y|Z,\theta)P(Z|\theta)}{P\left(Z|Y,\theta^{(i)}\right)P\left(Y|\theta^{(i)}\right)}$ 则 $L(\theta^{(i)})\geq B(\theta,\theta^{(i)})$，当且仅当 $\theta=\theta^{(i)}$ 时取等

算法导出

当我们尝试去优化下界，也就是命 $\theta^{(i+1)} = \arg\max_{\theta} B(\theta, \theta^{(i)})$，时，有：

$$
\begin{aligned}
\theta^{(i+1)} &= \arg\max_{\theta} \left(L\left(\theta^{(i)}\right) + \sum_{Z} P\left(Z \mid Y, \theta^{(i)}\right) \log \frac{P(Y \mid Z, \theta) P(Z \mid \theta)}{P\left(Z \mid Y, \theta^{(i)}\right) P\left(Y \mid \theta^{(i)}\right)} \right) \\
&= \arg\max_{\theta} \left(\sum_{Z} P\left(Z \mid Y, \theta^{(i)}\right) \log(P(Y \mid Z, \theta) P(Z \mid \theta)) \right) \\
&= \arg\max_{\theta} \left(\sum_{Z} P\left(Z \mid Y, \theta^{(i)}\right) \log P(Y, Z \mid \theta) \right) \\
&= \arg\max_{\theta} Q\left(\theta, \theta^{(i)}\right)
\end{aligned}
$$

这就是我们在 M 步中做的事

收敛性证明

因为

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

令

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

并且

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

那么对数似然函数

$$\begin{aligned} L(\theta) &= \log P(Y|\theta) = \log P(Y, Z|\theta) - \log P(Z|Y, \theta) \\ &= \sum_Z P(Z|Y, \theta^{(i)}) \log P(Z, Y|\theta) - P(Z|Y, \theta^{(i)}) \sum_Z \log P(Z|Y, \theta) \\ &= Q(\theta, \theta^{(i)}) - H(\theta, \theta^{(i)}) \end{aligned}$$

收敛性证明

$L(\theta^{(i+1)}) - L(\theta^{(i)}) = \left[Q(\theta^{(i+1)}) - Q(\theta^{(i)})\right] - \left[H(\theta^{(i+1)}) - H(\theta^{(i)})\right]$
前半部分已经为非负，现在考虑后半部分

$$\begin{aligned} H\left(\theta^{(i+1)}, \theta^{(i)}\right) - H\left(\theta^{(i)}, \theta^{(i)}\right) &= \sum_Z \left(\log \frac{P\left(Z \mid Y, \theta^{(i+1)}\right)}{P\left(Z \mid Y, \theta^{(i)}\right)}\right) P\left(Z \mid Y, \theta^{(i)}\right) \\ &\leqslant \log\left(\sum_Z \frac{P\left(Z \mid Y, \theta^{(i+1)}\right)}{P\left(Z \mid Y, \theta^{(i)}\right)} P\left(Z \mid Y, \theta^{(i)}\right)\right) \\ &= \log\left(\sum_Z P\left(Z \mid Y, \theta^{(i+1)}\right)\right) = 0 \end{aligned}$$

这就得到了 $L(\theta^{(i+1)}) \geq L(\theta^{(i)})$
显然有 $L(\theta) \leq 1$ 根据单调有界定理，$\{L(\theta^{(i)})\}_{i=1}^{\infty}$ 收敛.

混合高斯

现有男女生共 100 人，已知男女生升高分别服从正态分布．求分布的各个参数

- 初始化 $\theta_0 = (\mu_b, \mu_g, \sigma_b, \sigma_b)$
- E-step：计算 $P(Z|Y,\theta)$
  - 每个人是男生的概率：$\vec{P_b} = f_{\mu_b,\sigma_b}(Y)$
  - 每个人是女生的概率：$\vec{P_g} = f_{\mu_g,\sigma_g}(Y)$
  - 按概率大小更新 Z：$Z = \vec{P_b} > \vec{P_g}$
- M-step：这里求最大很简单，直接让估计值等于样本的均值和方差
  - $\mu_b = \frac{\sum\limits_{Z==1} Y}{\sum\limits_{Z==1} 1}$
  - $\sigma_b = \sqrt{\frac{\sum\limits_{Z==1} (Y-\mu_b)^2}{\sum\limits_{Z==1} 1}}$
  - $\mu_g = \frac{\sum\limits_{Z==0} Y}{\sum\limits_{Z==0} 1}$
  - $\sigma_g = \sqrt{\frac{\sum\limits_{Z==0} (Y-\mu_g)^2}{\sum\limits_{Z==0} 1}}$
- 重复 E,M 步骤，直到满足收敛条件