GMM-HMM

Hanwen Liu 522030910109

2024 年 6 月 4 日

有 GMM-HMM 参数:
转移概率 $A : a_{ij} = P(q_t = j|q_{t-1} = i), 1 \leq j \leq N)$
状态输出分布 $B : b_j(o_t) = p(o_t|q_t = j) = \sum_{m=1}^{M} c_{jm}\mathcal{N}(o_t|\mu_{jm}, \sum_{jm})$
对数似然函数:$\mathcal{L}(\theta) = \sum_{r=1}^{R} logp(O^{(r)}|\theta) = \sum_{r=1}^{R} log(\sum_q p(O^{(r)}, q|\theta)$
需要最大化似然函数，根据 PPT 上的 Jenson's Inequality, 有

$$\mathcal{L}(\theta) \geq \sum_{r=1}^{R} H(P(q|O^{(r)}, \hat{\theta})) + \mathcal{Q}(\theta, \hat{\theta})$$

定义辅助函数 $Q(\theta, \hat{\theta}) = \sum_{r=1}^{R} \sum_q P(q|O^{(r)}, \hat{\theta})logp(O^{(r)}, q|\theta)$ 因此得到了对数似然函数的一个下界
在辅助函数中,$\sum_q P(q|O^{(r)}, \hat{\theta}) = \sum_{j=1}^{N} P(q_t = j|O^{(r)}, \hat{\theta}) = \sum_{i=1}^{N} \sum_{j=1}^{N} P(q_{t-1} = i, q_t = j|O^{(r)}, \hat{\theta})$
将软分配占用率记为：$\gamma_{(i,j)}(t) = P(q_{t-1} = i, q_t = j|O^{(r)}, \hat{\theta}), \gamma_j(t) = P(q_t = j|O^{(r)}, \hat{\theta})$

因此，辅助函数

$$
\begin{aligned}
\mathcal{Q}(\theta,\hat{\theta}) &= \sum_{r=1}^{R}\sum_{q}P(q|O^{(r)},\hat{\theta})logp(O^{(r)},q|\theta) \\
&= \sum_{r=1}^{R}\sum_{q}P(q|O^{(r)},\hat{\theta})[\sum_{t=1}^{T}logP(q_t|q_{t-1},\theta)+\sum_{t=1}^{T}logp(o_t|q_t,\theta)] \\
&= \sum_{r=1}^{R}[\sum_{t=1}^{T}\sum_{q}P(q|O^{(r)},\hat{\theta})logP(q_t|q_{t-1},\theta)+\sum_{t=1}^{T}\sum_{q}P(q|O^{(r)},\hat{\theta})logp(o_t|q_t,\theta)] \\
&= \sum_{r=1}^{R}[\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\gamma_{(i,j)}(t)loga_{ij}+\sum_{t=1}^{T}\sum_{j=1}^{N}\gamma_j(t)logp(o_t|q_t=j,\theta)] \\
&= \sum_{r=1}^{R}\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\gamma_{(i,j)}(t)loga_{ij}+\sum_{r=1}^{R}\sum_{t=1}^{T}\sum_{j=1}^{N}\gamma_j(t)logp(o_t|q_t=j,\theta) \\
&= \sum_{r=1}^{R}\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\gamma_{(i,j)}(t)loga_{ij}+\sum_{r=1}^{R}\sum_{t=1}^{T}\sum_{j=1}^{N}\gamma_j(t)logb_j(o_t)
\end{aligned}
$$

可以得到 PPT 上的两部分:$\mathcal{Q}_A(\theta,\hat{\theta}),\mathcal{Q}_B(\theta,\hat{\theta})$
分别求解，对于第一部分:$\mathcal{Q}_A(\theta,\hat{\theta})=\sum_{r=1}^{R}\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\gamma_{(i,j)}(t)loga_{ij}$
要求最大化下界，所以可以转化为优化问题:

$$
\begin{gathered}
max\ \mathcal{Q}_A(\theta,\hat{\theta}) \\
s.t.\ \sum_{j=1}^{N}a_{ij}=1
\end{gathered}
$$

用拉格朗日求解:

$$
\begin{aligned}
&L=\sum_{r=1}^{R}\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\gamma_{(i,j)}(t)loga_{ij}+\lambda(1-\sum_{j=1}^{N}a_{ij}) \\
&\frac{\partial L}{\partial a_{ij}}=0 \\
&\sum_{r=1}^{R}\sum_{t=1}^{T}\frac{\gamma_{i,j}(t)}{a_{ij}}=\gamma
\end{aligned}
$$

可以解得

$$
\begin{aligned}
\lambda &= \sum_{r=1}^{R}\sum_{t=1}^{T}\sum_{i=1}^{N}\gamma_{(i,j)}(t) \\
a_{ij} &= \frac{\sum_{r=1}^{R}\sum_{t=1}^{T}\gamma_{(i,j)}(t)}{\sum_{r=1}^{R}\sum_{t=1}^{T}\sum_{i=1}^{N}\gamma_{(i,j)}(t)}
\end{aligned}
$$

对于第二部分 $\mathcal{Q}_B(\theta,\hat{\theta})$，将 GMM 展开，得到:

$$\mathcal{Q}_B(\theta,\hat{\theta})=\sum_{r=1}^{R}\sum_{t=1}^{T}\sum_{j=1}^{N}\gamma_j(t)log\sum_{m=1}^{M}\frac{c_{jm}}{(2\pi)^{D/2}|\Sigma_{jm}|^{1/2}}exp[-1/2(o_t-\mu_{jm})^T\Sigma_{jm}^{-1}(o_t-\mu_{jm})]$$

利用 Jensen's Inequality, 得到:
$\mathcal{Q}_B(\theta,\hat{\theta})\geq K+\sum_{r=1}^{R}\sum_{t=1}^{T}\sum_{j=1}^{N}\sum_{m=1}^{N}\gamma_{jm}(t)\{logc_{jm}-1/2[log|\Sigma_{jm}|+(o_t-\mu_{jm})^T\Sigma_{jm}^{-1}(o_t-\mu_{jm})]\}$
定义右半部分为 $\mathcal{Q}'_B(\theta,\hat{\theta})$, 要求最大化
写出拉格朗日方程:$L=\mathcal{Q}'_B(\theta,\hat{\theta})+\lambda(1-\sum_{m=1}^{M}c_{jm})$
求解拉格朗日方程,可以解得 λ,c_{jm} 的最优解分别为 $\sum_{r=1}^{R}\sum_{t=1}^{T}\sum_{m=1}^{M}\gamma_{jm}(t)$ 和 $\frac{\sum_{r=1}^{R}\sum_{t=1}^{T}\gamma_{jm}(t)}{\sum_{r=1}^{R}\sum_{t=1}^{T}\sum_{m=1}^{M}\gamma_{jm}(t)}$
因此,根据拉格朗日,可以解得 μ_{jm},Σ_{jm} 的解分别为 $\frac{\sum_{r=1}^{R}\sum_{t=1}^{T}\gamma_{jm}(t)o_t}{\sum_{r=1}^{R}\sum_{t=1}^{T}\gamma_{jm}(t)}$, $\frac{\sum_{r=1}^{R}\sum_{t=1}^{T}\gamma_{jm}(t)(o_t-\mu_{jm})(o_t-\mu_{jm})^T}{\sum_{r=1}^{R}\sum_{t=1}^{T}\gamma_{jm}(t)}$
根据前后向概率在 PPT 中的公式，概率可以进行递归计算，由此可以计算软分配占用率:

$$\begin{aligned}
\gamma_j(t)&=P(q_t=j|O_1^T,\hat{\theta})\\
&=\frac{p(O_1^T,q_t=j|\hat{\theta})}{p(O_1^T|\hat{\theta})}\\
&=\frac{p(O_1^T,O_{t+1}^T,q_t=j|\hat{\theta})}{p(O_1^T|\hat{\theta})}\\
&=\frac{p(O_1^T,q_t=j|\hat{\theta})p(O_{t+1}^T,q_t=j|\hat{\theta})}{p(O_1^T|\hat{\theta})}\\
&=\frac{\alpha_j(t)\beta_j(t)}{\alpha_N(T+1)}\\
\gamma_{(i,j)}(t)&=\frac{\alpha_i(t-1)\hat{\alpha}_{ij}b_j(o_t)\beta_j(t)}{\alpha_N(T+1)}(PPT)\\
\gamma_{jm}(t)&=\gamma_j(t)\gamma_m(t)(PPT)
\end{aligned}$$

综上，整个算法可以表达为:

$$In\ k's\ iteration\ while\ params\ do\ not\ converge:$$

$$updata\ params:$$

$$\alpha_j^k(t) = \sum_{i=1}^{N} b_j^{k-1}(o_t)\alpha_{ij}^{k-1}\alpha_i^k(t-1)$$

$$\beta_j^k(t) = \sum_{i=1}^{N} b_j^{k-1}(o_{t+1})\alpha_{ji}^{k-1}\beta_i^k(t+1)$$

$$\gamma_j^k(t) = \frac{\alpha_j^k(t)\beta_j^k(t)}{\alpha_N(T+1)}$$

$$\gamma_{(i,j)}^k(t) = \frac{\alpha_i^k(t-1)\alpha_{ij}^{k-1}b_j^{k-1}(o_t)\beta_j^k(t)}{\alpha_N^k(T+1)}$$

$$\gamma_{jm}(t) = \gamma_j^k(t)\gamma_m^k(t)$$

$$a_{ij}^k = \frac{\sum_{r=1}^{R}\sum_{t=1}^{T}\gamma_{(i,j)}^k(t)}{\sum_{r=1}^{R}\sum_{t=1}^{T}\sum_{i=1}^{N}\gamma_{(i,j)}^k(t)}$$

$$c_{jm}^k = \frac{\sum_{r=1}^{R}\sum_{t=1}^{T}\gamma_{jm}^k(t)}{\sum_{r=1}^{R}\sum_{t=1}^{T}\sum_{m=1}^{M}\gamma_{jm}^k(t)}$$

$$\mu_{jm}^k = \frac{\sum_{r=1}^{R}\sum_{t=1}^{T}\gamma_{jm}^k(t)o_t}{\sum_{r=1}^{R}\sum_{t=1}^{T}\gamma_{jm}^k(t)}$$

$$\Sigma_{jm}^k = \frac{\sum_{r=1}^{R}\sum_{t=1}^{T}\gamma_{jm}^k(t)(o_t-\mu_{jm}^k)(o_t-\mu_{jm}^k)^T}{\sum_{r=1}^{R}\sum_{t=1}^{T}\gamma_{jm}^k(t)}$$