HMM模型--参数学习

在上一个文档中，对HMM模型应用的三种问题，分为概率计算、参数学习、预测。在本文中对参数学习问题进行介绍。

3. 参数学习

HMM的参数是$\lambda=(A,B,\pi)$，分别表示状态转移概率，观测概率，以及初始的状态概率。根据拿到的数据是有观测和状态序列，还是仅有观测数据，分为有监督学习和无监督学习了。

3.1 有监督学习

利用有监督学习去估计HMM的参数非常简单，利用Bernoulli大数定理的结论：频率的极限就是概率，进行HMM参数估计。

假设给定数据集D，包含有S个样本，每个样本为长度相等的状态序列I和观测序列O。

$$D=\{(O_1,I_1),(O_2,I_2),...,(O_S,I_S)\}$$

1. 估计转移概率a_{ij}：状态从t时刻的i转移到$t+1$时刻的j

   统计样本集里，前一时刻位于状态i，后一时刻位于状态j的频数，记作A_{ij}。则

$$\hat{a}_{ij}=\frac{A_j}{\sum_{j=1}^{N}A_{ij}},\quad i=1,2,\cdots,N;j=1,2,\cdots,N$$

2. 观测概率$b_j(k)$：在状态j的情况下得到了观测k

   统计样本集里，状态为j，并且观测为k的频数，记作B_{jk}，则

$$\hat{b}_j(k)=\frac{B_{jk}}{\sum_{k=1}^{M}B_{jk}},\quad j=1,2,\cdots,N;k=1,2,\cdots,M$$

3. 初始状态π_i

   统计样本集里，初始状态为i的样本个数，记作S_i

$$\hat{\pi}_i=\frac{S_i}{S}$$

3.2 Baum-Welch算法

如果没有标注数据，则为非监督学习，采用BW算法。假设给定训练数据集D，包含有S个样本

$$D=\{O_1,O_2,...,O_S\}$$

每个样本为长度为T的观测序列，没有状态序列。将观测序列数据看成观测数据O，状态序列数据看成隐数据I，则观测数据O的似然，可以表示成

$$P(O|\lambda)=\sum_I P(O|I,\lambda)P(I|\lambda)=\sum_I P(O,I|\lambda)$$

对含有隐状态的极大似然估计问题，采用EM算法求解。

在EM算法的E步，定义Q函数。假设当前的模型估计的参数为$\bar{\lambda}$，在Q函数中的变量是λ。因此Q函数可以表示为

$$\begin{aligned}
Q(\lambda,\bar{\lambda}) &= \sum_I (\log P(O,I|\lambda))\, P(I|O,\bar{\lambda}) \\
&= \sum_I (\log P(O,I|\lambda))\, \frac{P(I,O|\bar{\lambda})}{P(O|\bar{\lambda})} \\
&\propto \sum_I (\log P(O,I|\lambda))\, P(I,O|\bar{\lambda})
\end{aligned}$$

这是根据EM算法推导出的Q函数。但是在BW算法中，直接将Q函数定义成

$$Q(\lambda,\bar{\lambda})=\sum_I (\log P(O,I|\lambda))\, P(I,O|\bar{\lambda})$$

其中的$P(O,I|\lambda)$把参数代入，展开可表达为

$$P(O,I|\lambda)=\pi_{i_1} b_{i_1}(o_1) a_{i_1 i_2} b_{i_2}(o_2)\ldots a_{i_{T-1} i_T} b_{i_T}(o_T)$$

代入Q函数中，可得

$$Q(\lambda,\bar{\lambda})=\sum_I (\log \pi_{i_1}) P(I,O|\bar{\lambda})+\sum_I \left(\sum_{t=1}^{T-1}\log a_{i_t i_{t+1}}\right) P(I,O|\bar{\lambda})+\sum_I \left(\sum_{t=1}^{T}\log b_{i_t}(o_t)\right) P(I,O|\bar{\lambda})$$

接下来为EM算法的M步，对Q函数进行极大似然估计。把Q函数对模型参数A,B,π求偏导数。

观察Q函数，为三项相加的形式。之所以整理成上面的样子，是因为每一项都只和一个参数有关，因此求导过程只需要考虑其中的一项即可。

首先是π，Q函数的第一项，可以写成

$$\sum_I (\log \pi_{i_1}) P(I,O|\bar{\lambda})=\sum_{i=1}^{N}\log \pi_i P(O,i_1=i|\bar{\lambda})$$

由于π是概率，因此满足$\sum_{i=1}^{N}\pi_i=1$。因此要把约束，利用拉格朗日乘子法，放进目标函数中。因此拉格朗日函数为

$$\sum_{i=1}^{N}\log \pi_i P(O,i_1=i|\bar{\lambda})+\gamma\left(\sum_{i=1}^{N}\pi_i-1\right)$$

求对π_i的偏导数，令其等于0。

$$\frac{\partial}{\partial \pi_i}\left[\sum_{i=1}^{N}\log \pi_i P(O,i_1=i|\bar{\lambda})+\gamma\left(\sum_{i=1}^{N}\pi_i-1\right)\right]=0$$

因为对第i个π求导数，因此只需考虑$\sum$中的第i项，得到

$$\frac{P(O, i_1 = i|\bar{\lambda})}{\pi_i} + \gamma = 0$$

整理得到

$$\pi_i \gamma = -P(O, i_1 = i|\bar{\lambda})$$

两边对i求和，即

$$\begin{aligned} \sum_{i=1}^{N} \pi_i \gamma &= \sum_{i=1}^{N} -P(O, i_1 = i|\bar{\lambda}) \\ &= \gamma \sum_{i=1}^{N} \pi_i = -P(O|\bar{\lambda}) = \gamma \end{aligned}$$

将γ带回$\pi_i \gamma = -P(O, i_1 = i|\bar{\lambda})$得到

$$\pi_i = \frac{P(O, i_1 = i|\bar{\lambda})}{P(O|\bar{\lambda})}$$

从而可以得到pi_i的更新值的计算公式。

接着是第二项$a_{i_t i_{t+1}}$。Q函数的第二项可以写成

$$\sum_I \left(\sum_{t=1}^{T-1} \log a_{i_t i_{t+1}} \right) P(I, O|\bar{\lambda}) = \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{t=1}^{T-1} \log a_{ij} P(O, i_t = i, i_{t+1} = j|\bar{\lambda})$$

注意到转移概率a_{ij}具有约束条件$\sum_{j=1}^{N} a_{ij} = 1$，因此同样利用拉格朗日乘子法，写出拉格朗日函数

$$\sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{t=1}^{T-1} \log a_{ij} P(O, i_t = i, i_{t+1} = j|\bar{\lambda}) + \zeta \left(\sum_{j=1}^{N} a_{ij} - 1 \right)$$

对a_{ij}求偏导，令其等于0

$$\frac{\partial}{\partial a_{ij}} \left[\sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{t=1}^{T-1} \log a_{ij} P(O, i_t = i, i_{t+1} = j|\bar{\lambda}) + \zeta \left(\sum_{j=1}^{N} a_{ij} - 1 \right) \right] = 0$$

只需考虑t时刻等于i，$t+1$时刻等于j的情况，因此得到

$$\frac{\sum_{t=1}^{T-1} P(O, i_t = i, i_{t+1} = j|\bar{\lambda})}{a_{ij}} + \zeta = 0$$

得到

$$a_{ij} \zeta = -\sum_{t=1}^{T-1} P(O, i_t = i, i_{t+1} = j|\bar{\lambda})$$

两边对j求和，得到

$$\begin{aligned}\sum_{j=1}^{N} a_{ij}\zeta &= -\sum_{t=1}^{T-1}\sum_{j=1}^{N} P(O, i_t = i, i_{t+1} = j|\bar{\lambda}) \\ &= \zeta = -\sum_{t=1}^{T-1} P(O, i_t = i|\bar{\lambda})\end{aligned}$$

将ζ带回$a_{ij}\zeta = -\sum_{t=1}^{T-1} P(O, i_t = i, i_{t+1} = j|\bar{\lambda})$，整理得到

$$a_{ij} = \frac{\sum_{t=1}^{T-1} P(O, i_t = i, i_{t+1} = j|\bar{\lambda})}{\sum_{t=1}^{T-1} P(O, i_t = i|\bar{\lambda})}$$

最后对$b_j(k)$求偏导，构造拉格朗日函数，整理得到

$$b_j(k) = \frac{\sum_{t=1}^{T} P(O, i_t = j|\bar{\lambda}) I(o_t = v_k)}{\sum_{t=1}^{T} P(O, i_t = j|\bar{\lambda})}$$

注意，只有在$o_t = v_k$的时候，$b_j(o_t)$对$b_j(k)$的偏导数才不为零，$I(o_t = v_k)$表示如果括号里面的条件不满足，则该函数为0。

在第二节的结尾，我们定义了$\xi_t(i, j) = P(i_t = i, i_{t+1} = j|O, \lambda)$以及$\gamma_t(i) = \frac{P(i_t = i, O|\lambda)}{\sum_{i=1}^{N} P(i_t = i, O|\lambda)}$，代入$a_{ij}, b_j(k), \pi_i$，整理得到

$$\begin{aligned}a_{ij} &= \frac{\sum_{t=1}^{T-1} \xi_t(i, j)}{\sum_{t=1}^{T-1} \gamma_t(i)} \\ b_j(k) &= \frac{\sum_{t=1, o_t = v_k}^{T} \gamma_t(j)}{\sum_{t=1}^{T} \gamma_t(j)} \\ \pi_i &= \gamma_1(i)\end{aligned}$$

因此，整体的BW算法，为

1. 先初始化$pi_i^{(0)}, A^{(0)}, B^{(0)}$，作为HMM模型初始化的参数。
2. 利用上面介绍的方法，反复交替计算$a_{ij}, b_j(k), \pi_i$和计算γ和ξ。
3. 得到最终参数$\lambda^{(n+1)} = (A^{(n+1)}, B^{(n+1)}, \pi^{(n+1)})$。