机器学习第六周作业

樊泽羲 2200010816

Q1：推导HMM中的后向概率递推公式

A1：

$$
\begin{aligned}
\beta_t(i) &= P(o_{t+1},\ldots,o_T|i_t=q_i,\lambda) \\
&= \sum_{j=1}^{N} P(o_{t+2},\ldots,o_T|i_{t+1}=q_j,\lambda)P(o_{t+1}|i_{t+1}=q_j,\lambda)P(i_{t+1}=q_j|i_t=q_i,\lambda) \\
&= \sum_{j=1}^{N} \beta_{t+1}b_j(o_{t+1})a_{ij}, 1 \le t \le T-1
\end{aligned}
$$

特别地：

$$
\beta_{T-1}(i) = \sum_{j=1}^{N} \beta_T(j)b_j(o_T)a_{ij}
$$

$$
\beta_{T-1}(i) = \sum_{j=1}^{N} P(o_T|i_t=q_j,\lambda)a_{ij} = \sum_{j=1}^{N} b_j(o_T)a_{ij}
$$

因此：

$$
\beta_T(j) := 1, 1 \le j \le N
$$

并且我们有：

$$
\begin{aligned}
P(O|\lambda) &= P(o_1,\ldots,o_T|\lambda) \\
&= \sum_{j=1}^{N} P(i_1=q_i|\lambda)P(o_2,\ldots,o_T|i_1=q_i,\lambda)P(o_1|i_1=q_i,\lambda) \\
&= \sum_{j=1}^{N} \pi_i\beta_1(i)b_i(o_1) \qquad \blacksquare
\end{aligned}
$$

Q2：推导Viterbi算法中路径最大概率$\delta_t(i)$的递推公式

A2:

设$(i_1^\star,\ldots,i_T^\star) = argmax_I P(O|I,\lambda)$为最优路径

断言：从$i_t^\star$出发，到达T时刻的最优路径为$(i_{t+1}^\star,\ldots,i_T^\star)$

若否，假设存在更优的路径$(\widehat{i_{t+1}},\ldots,\widehat{i_T})$

则：

$$
P(O|(i_1^\star,\ldots,i_t^\star,\widehat{i_{t+1}},\ldots,\widehat{i_T}),\lambda) > P(O|(i_1^\star,\ldots,i_t^\star,i_{t+1}^\star,\ldots,i_T^\star),\lambda)
$$

但这与$(i_1^\star,\ldots,i_T^\star) = argmax_I P(O|I,\lambda)$的定义矛盾

从而：只需利用动态规划贪心计算$\delta_t(i)$再向前回溯

$$\delta_t(i) := t\text{时刻通过某路径到达}i\text{节点并观测到}o_t\text{的最大概率}$$
$$\Rightarrow \delta_t(i) = max_{i_1,...,i_{t-1}} P(i_t = i, i_{t-1}, \ldots, i_1, o_t, \ldots, o_1|\lambda)$$

因此：

$$\begin{aligned}\delta_{t+1}(i) &= max_{i_1,...,i_t} P(i_{t+1}, i_t, \ldots, i_1, o_{t+1}, \ldots, o_1|\lambda) \\ &= max_j max_{i_1,...,i_{t-1}} P(i_t = j, i_{t-1}, \ldots, i_1, o_t, \ldots, o_1|\lambda) a_{ji} b_i(o_{t+1}) \\ &= max_j \delta_t(j) a_{ji} b_i(o_{t+1}) \qquad \blacksquare\end{aligned}$$

Q3：书后习题10.3

A3：

(1).初始化

$$\begin{aligned}\delta_1(i) &= \pi_i b_i(o_1) \\ &= \pi_i b_i(\text{红}) \\ \delta_1(1) &= 0.2 \cdot 0.5 = 0.1 \\ \delta_1(2) &= 0.4 \cdot 0.2 = 0.08 \\ \delta_1(3) &= 0.4 \cdot 0.7 = 0.28\end{aligned}$$

(2).t=2

$$\begin{aligned}\delta_2(i) &= max_{1\le j\le 3} \delta_1(j) a_{ji} b_i(o_2) \\ &= max_{1\le j\le 3} \delta_1(j) a_{ji} b_i(\text{白}) \\ \psi_2(i) &= argmax_{1\le j\le 3} \delta_1(j) a_{ji} \\ \Rightarrow \delta_2(1) &= 0.025, \delta_2(2) = 0.0504, \delta_2(3) = 0.042 \\ \psi_2(1) &= 1, \psi_2(2) = 3, \psi_2(3) = 3\end{aligned}$$

(3).t=3

$$\begin{aligned}\delta_3(i) &= max_{1\le j\le 3} \delta_2(j) a_{ji} b_i(o_3) \\ &= max_{1\le j\le 3} \delta_2(j) a_{ji} b_i(\text{红}) \\ \psi_3(i) &= argmax_{1\le j\le 3} \delta_2(j) a_{ji} \\ \Rightarrow \delta_3(1) &= 0.00756, \delta_3(2) = 0.01008, \delta_3(3) = 0.0147 \\ \psi_3(1) &= 2, \psi_3(2) = 2, \psi_3(3) = 3\end{aligned}$$

(4).t=4

$$\begin{aligned}\delta_4(i) &= max_{1\le j\le 3} \delta_3(j) a_{ji} b_i(o_4) \\ &= max_{1\le j\le 3} \delta_3(j) a_{ji} b_i(\text{白}) \\ \psi_4(i) &= argmax_{1\le j\le 3} \delta_3(j) a_{ji} \\ \Rightarrow \delta_4(1) &= 0.00189, \delta_4(2) = 0.003024, \delta_4(3) = 0.0006804 \\ \psi_4(1) &= 1, \psi_4(2) = 2, \psi_4(3) = 1\end{aligned}$$

(5).求解最优路径

$$\begin{aligned}P^\star &= max_{1\le i\le 3} \delta_4(i) = 0.003024 \\ i_4^\star &= argmax_{1\le i\le 3} \delta_4(i) = 2 \\ \Rightarrow t = 3 : i_3^\star &= \psi_4(i_4^\star) = 2 \\ \Rightarrow t = 2 : i_2^\star &= \psi_3(i_3^\star) = 2 \\ \Rightarrow t = 1 : i_1^\star &= \psi_2(i_2^\star) = 3\end{aligned}$$

从而：$(i_1^\star, i_2^\star, i_3^\star, i_4^\star) = (3, 2, 2, 2)$

Q4：书后习题10.5

A4:

Viterbi算法和隐马尔可夫模型中的前向概率的计算方法之间有几个关键的区别:

1.路径依赖上来看：

Viterbi算法寻找单个最有可能产生观测序列的隐藏状态路径。它的工作原理是在每一步找到具有最大联合概率的路径
HMMs中的前向概率计算的是在给定点之前生成观测序列的所有可能状态路径的总概率。它对所有可能的路径进行求和，而不是只找到最大值

2.从算法上看：

Viterbi算法的目的是寻找最优，因此使用了动态规划算法

HMMs中的前向概率目的是求和，因此采用的是递推求和的算法