机器学习基础作业 6

2025 年 4 月 26 日

问题 1. 试用前向概率和后向概率推导

$$P(\mathbf{O}|\lambda) = \sum_{l=1}^{N}\sum_{k=1}^{N}\alpha_t(l)a_{lk}b_k(O_{t+1})\beta_{t+1}(k)$$

证明.

$$P(\mathbf{O}|\lambda) = \sum_{l=1}^{N}\sum_{k=1}^{N}P(\mathbf{O}, X_t = s_l, X_{t+1} = s_k|\lambda)$$

注意到:

$$\begin{aligned}\alpha_t(l)a_{lk} &= P(O_1, O_2, \cdots, O_t, X_t = s_l|\lambda)P(X_{t+1} = s_k|X_t = s_l, \lambda)\\ &= P(O_1, O_2, \cdots, O_t, X_t = s_l, X_{t+1} = s_k|\lambda)\end{aligned}$$

$$\begin{aligned}b_k(O_{t+1})\beta_{t+1}(k) &= P(O_{t+1}|X_{t+1} = s_k, \lambda)P(O_{t+2}, O_{t+3}, \cdots, O_T|X_{t+1} = s_k, \lambda)\\ &= P(O_{t+1}, O_{t+2}, \cdots, O_T|X_{t+1} = s_k, \lambda)\\ &= P(O_{t+1}, O_{t+2}, \cdots, O_T|X_t = s_l, X_{t+1} = s_k, \lambda)\end{aligned}$$

最后这一行是因为马尔可夫性质, X_{t+1} 已知时, O_{t+1} 与 X_t 无关. 因而二者相乘即有:

$$\begin{aligned}P(\mathbf{O}|\lambda) &= \sum_{l=1}^{N}\sum_{k=1}^{N}P(O_1, O_2, \cdots, O_t, X_t = s_l, X_{t+1} = s_k, O_{t+1}, O_{t+2}, \cdots, O_T|\lambda)\\ &= \sum_{l=1}^{N}\sum_{k=1}^{N}P(O_1, O_2, \cdots, O_t, X_t = s_l, X_{t+1} = s_k|\lambda)P(O_{t+1}, O_{t+2}, \cdots, O_T|X_t = s_l, X_{t+1} = s_k, \lambda)\\ &= \sum_{l=1}^{N}\sum_{k=1}^{N}\alpha_t(l)a_{lk}b_k(O_{t+1})\beta_{t+1}(k)\end{aligned}$$

□

问题 2. 证明维特比算法中 δ 的递推公式.

证明.

$$\begin{aligned}\delta_t(i) &= \max_{X_1, X_2, \cdots, X_{t-1}} P(X_1, X_2, \cdots, X_{t-1}, O_1, O_2, \cdots, O_t, X_t = s_i|\lambda)\\ &= \max_{X_1, X_2, \cdots, X_{t-2}} \max_{1\le j\le N} P(X_1, X_2, \cdots, X_{t-1} = s_j, O_1, O_2, \cdots, O_t, X_t = s_i|\lambda)\\ &= \max_{1\le j\le N} \max_{X_1, X_2, \cdots, X_{t-2}} P(X_1, X_2, \cdots, X_{t-2}, O_1, O_2, \cdots, O_{t-1}, X_{t-1} = s_j, X_t = s_i|\lambda)P(O_t|X_t = s_i)\\ &= \max_{1\le j\le N} \max_{X_1, X_2, \cdots, X_{t-2}} P(X_1, X_2, \cdots, X_{t-2}, O_1, O_2, \cdots, O_{t-1}, X_{t-1} = s_j|\lambda)P(X_t = s_i|X_{t-1} = s_j)b_i(O_t)\\ &= \max_{1\le j\le N} \delta_{t-1}(j)b_i(O_t)a_{ji} = \max_{1\le j\le N}\left(\delta_{t-1}(j)a_{ji}\right)b_i(O_t)\end{aligned}$$

其中第三行用到了马尔可夫性质, 当 X_t 已知时观测结果 O_j 与 $X_1, \cdots, X_{t-1}$ 无关.

□

问题 3. 在下述隐马尔可夫模型 $\lambda = (A, B, \pi)$ 中，可能的观测值集合为 $\{\nu_1, \nu_2, \nu_3\}$，可能的状态集为 $\{q_1, q_2, q_3\}$，$A = \begin{bmatrix} 0.6 & 0.2 & 0.2 \\ 0.3 & 0.5 & 0.2 \\ 0.3 & 0.3 & 0.4 \end{bmatrix}$, $B = \begin{bmatrix} 0.2 & 0.5 & 0.3 \\ 0.6 & 0.2 & 0.2 \\ 0.3 & 0.3 & 0.4 \end{bmatrix}$, $\pi = (0.6, 0.2, 0.2)^T$，观察序列 $O = (\nu_3, \nu_2, \nu_1)$，用维特比算法求最优状态序列.

证明. 初值:

$$\delta_1(1) = \pi_1 b_1(O_1) = 0.6 \cdot 0.3 = 0.18$$
$$\delta_1(2) = \pi_2 b_2(O_1) = 0.2 \cdot 0.2 = 0.04$$
$$\delta_1(3) = \pi_3 b_3(O_1) = 0.2 \cdot 0.4 = 0.08$$

第二项:

$$\delta_2(1) = \max_{1\le j\le 3} (\delta_1(j) a_{j1})\, b_1(O_2) = \max(0.18 \cdot 0.6, 0.04 \cdot 0.3, 0.08 \cdot 0.3) \cdot 0.5 = 0.054$$
$$\delta_2(2) = \max_{1\le j\le 3} (\delta_1(j) a_{j2})\, b_2(O_2) = \max(0.18 \cdot 0.2, 0.04 \cdot 0.5, 0.08 \cdot 0.3) \cdot 0.2 = 0.0072$$
$$\delta_2(3) = \max_{1\le j\le 3} (\delta_1(j) a_{j3})\, b_3(O_2) = \max(0.18 \cdot 0.2, 0.04 \cdot 0.2, 0.08 \cdot 0.4) \cdot 0.3 = 0.0108$$

且最优历史为 $\Psi_2(1) = 1, \Psi_2(2) = 1, \Psi_2(3) = 1$. 第三项:

$$\delta_3(1) = \max_{1\le j\le 3} (\delta_2(j) a_{j1})\, b_1(O_3) = \max(0.054 \cdot 0.6, 0.0072 \cdot 0.3, 0.0108 \cdot 0.3) \cdot 0.2 = 0.00648$$
$$\delta_3(2) = \max_{1\le j\le 3} (\delta_2(j) a_{j2})\, b_2(O_3) = \max(0.054 \cdot 0.2, 0.0072 \cdot 0.5, 0.0108 \cdot 0.3) \cdot 0.6 = 0.00648$$
$$\delta_3(3) = \max_{1\le j\le 3} (\delta_2(j) a_{j3})\, b_3(O_3) = \max(0.054 \cdot 0.2, 0.0072 \cdot 0.2, 0.0108 \cdot 0.4) \cdot 0.3 = 0.00324$$

且最优历史为 $\Psi_3(1) = 1, \Psi_3(2) = 1, \Psi_3(3) = 1$. 注意 $\delta_3(1)$ 和 $\delta_3(2)$ 同为最大值, 因此最优路径有两条: (q_1, q_1, q_1) 和 (q_1, q_1, q_2).

□