《大数据机器学习》第 8 次作业

姓名：刘培源　　学号：2023214278

题目 1： 在习题 10.1 中，试用维特比算法求最优路径 $I^* = (i_1^*, i_2^*, i_3^*, i_4^*)$。

10.1　给定盒子和球组成的隐马尔可夫模型 $\lambda = (A, B, \pi)$，其中，

$$A = \begin{bmatrix} 0.5 & 0.2 & 0.3 \\ 0.3 & 0.5 & 0.2 \\ 0.2 & 0.3 & 0.5 \end{bmatrix}, \quad B = \begin{bmatrix} 0.5 & 0.5 \\ 0.4 & 0.6 \\ 0.7 & 0.3 \end{bmatrix}, \quad \pi = (0.2,\ 0.4,\ 0.4)^{\mathrm{T}}$$

设 $T = 4$，$O = (\text{红}, \text{白}, \text{红}, \text{白})$，试用后向算法计算 $P(O|\lambda)$。

答： 我的代码实现如下：

```python
import numpy as np

def viterbi(A, B, pi, O, N):
    """
    A: 状态转移概率矩阵, B: 观测概率矩阵
    pi: 初始状态概率向量, O: 观测序列(红和白)
    T: 观测序列长度, N: 状态数
    return: 最优路径及其概率
    """
    # 将观测序列转换为数值
    o = np.array([0 if o == '红' else 1 for o in O], dtype=np.int64)
    T = len(o)

    # 初始化delta、psi和第一时间步
    delta = np.zeros((T, N))
    psi = np.zeros((T, N), np.int64)
    delta[0] = pi * B[:, o[0]]

    # 递推计算delta和psi
    for t in range(1, T):
        for i in range(N):
            delta[t, i] = np.max(delta[t-1] * A[:, i]) * B[i, o[t]]
            psi[t, i] = np.argmax(delta[t-1] * A[:, i])

    # 回溯找到最优路径
    path = np.zeros(T, dtype=np.int64)
    path[T-1] = np.argmax(delta[T-1])
    for t in range(T-2, -1, -1):
        path[t] = psi[t + 1, path[t + 1]]

    return path + 1, np.max(delta[T-1])

# 测试
N = 3
O = ['红', '白', '红', '白']
A = np.array([[0.5, 0.2, 0.3], [0.3, 0.5, 0.2], [0.2, 0.3, 0.5]])
B = np.array([[0.5, 0.5], [0.4, 0.6], [0.7, 0.3]])
pi = np.array([0.2, 0.4, 0.4])

I, P = viterbi(A, B, pi, O, N)
print(f'最优路径：{I}, 最优路径的概率：{P:.4f}')
```

代码的输出如下：

```
最优路径：[3 2 2 2]，最优路径的概率：0.003024
```

所以，最优路径 $I^* = (i_1^*, i_2^*, i_3^*, i_4^*) = (3, 2, 2, 2)$

题目 2： 试用前向概率和后向概率推导

$$P(O \mid \lambda) = \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_t(i) a_{ij} b_j\left(o_{t+1}\right) \beta_{t+1}(j), \quad t = 1, 2, \cdots, T-1$$

答： 将 $P(O \mid \lambda)$ 展开推导如下：

$$\begin{aligned}
P(O \mid \lambda) &= P\left(o_1, o_2, \ldots, o_T \mid \lambda\right) \\
&= \sum_{i=1}^{N} P\left(o_1, \ldots, o_t, i_t = q_i \mid \lambda\right) P\left(o_{t+1}, \ldots, o_T \mid i_t = q_i, \lambda\right) \\
&= \sum_{i=1}^{N} \sum_{j=1}^{N} P\left(o_1, \ldots, o_t, i_t = q_i \mid \lambda\right) P\left(o_{t+1}, i_{t+1} = q_j \mid i_t = q_i, \lambda\right) P\left(o_{t+2}, \ldots, o_T \mid i_{t+1} = q_j, \lambda\right) \\
&= \sum_{i=1}^{N} \sum_{j=1}^{N} \left[P\left(o_1, \ldots, o_t, i_t = q_i \mid \lambda\right) P\left(o_{t+1} \mid i_{t+1} = q_j, \lambda\right) P\left(i_{t+1} = q_j \mid i_t = q_i, \lambda\right)\right]
\end{aligned} \tag{1}$$

图1a和1b展示了前向概率和后向概率定义

定义 10.2（前向概率） 给定隐马尔可夫模型 λ，定义到时刻 t 部分观测序列为 $o_1, o_2, \cdots, o_t$ 且状态为 q_i 的概率为前向概率，记作

$$\alpha_t(i) = P(o_1, o_2, \cdots, o_t, i_t = q_i|\lambda) \tag{10.14}$$

可以递推地求得前向概率 $\alpha_t(i)$ 及观测序列概率 $P(O|\lambda)$。

(a) 前向概率定义

定义 10.3（后向概率） 给定隐马尔可夫模型 λ，定义在时刻 t 状态为 q_i 的条件下，从 $t+1$ 到 T 的部分观测序列为 $o_{t+1}, o_{t+2}, \cdots, o_T$ 的概率为后向概率，记作

$$\beta_t(i) = P(o_{t+1}, o_{t+2}, \cdots, o_T|i_t = q_i, \lambda) \tag{10.18}$$

可以用递推的方法求得后向概率 $\beta_t(i)$ 及观测序列概率 $P(O|\lambda)$。

(b) 后向概率定义

图 1: 前向概率和后向概率定义

根据以上，我们可以推导

$$a_{ij} = P\left(i_{t+1} = q_j \mid i_t = q_i\right), \quad i = 1, 2, \cdots, N; \quad j = 1, 2, \cdots, N$$

$$b_j(k) = P\left(o_t = v_k \mid i_t = q_j\right), \quad k = 1, 2, \cdots, M; \quad j = 1, 2, \cdots, N$$

将这些代入到公式1中，我们可以得到：

$$\begin{aligned}
P(O \mid \lambda) &= \sum_{i=1}^{N} \sum_{j=1}^{N} \left[P\left(o_1, \ldots, o_t, i_t = q_i \mid \lambda\right) P\left(o_{t+1} \mid i_{t+1} = q_j, \lambda\right) P\left(i_{t+1} = q_j \mid i_t = q_i, \lambda\right)\right] \\
&= \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_t(i) a_{ij} b_j\left(o_{t+1}\right) \beta_{t+1}(j), \quad t = 1, 2, \ldots, T-1
\end{aligned}$$

综上所述，命题得证!