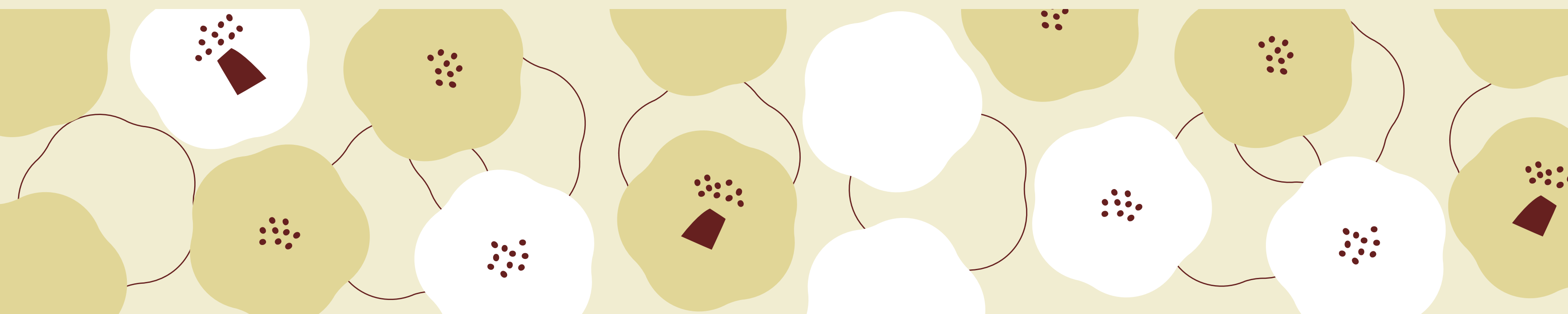
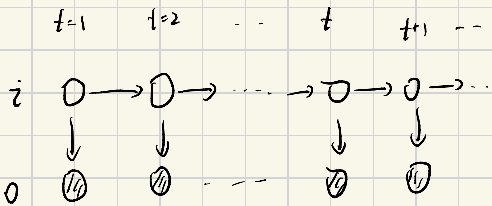


HMM 隐马尔可夫模型

频率派 → 统计机器学习

贝叶斯派 → 概率图模型



$I=(i_1,i_2,\dots,i_T)$ — 状态序列

$O=(o_1,o_2,\dots,o_T)$ — 观测序列

$Q=\{q_1,q_2,\dots,q_N\}$ — 状态集合

$V=\{v_1,v_2,\dots,v_M\}$ — 观测集合

隐马尔可夫模型

$\lambda=(\pi,A,B)$

$\pi=(\pi_i)=(P(i_1=q_i)), i=1,2,\dots,N$ - 初始状态向量

$A=[a_{ij}]_{N\times N}$ $\quad a_{ij}=P(i_{t+1}=q_j|i_t=q_i)$ — 状态转移矩阵

$B=[b_j(k)]_{N\times M}$ $\quad b_j(k)=P(o_t=v_k|i_t=q_j)$ — 观测概率矩阵

① 齐次马尔可夫假设

$$P(i_{t+1}|i_t,o_t,i_{t-1},o_{t-1},\dots)=P(i_{t+1}|i_t)$$

② 观测独立假设

$$P(o_t|i_T,i_{T-1},\dots,i_1,o_{T-1},\dots,o_1)=P(o_t|i_t)$$

三个问题

① Evaluation Given λ，求 $P(O|\lambda)$ → 前向后向算法

② Learning $\lambda_{MLE}=\underset{\lambda}{\arg\max}\, P(O|\lambda)$ → EM

③ decoding $\hat{I}=\underset{I}{\arg\max}\, P(I|O,\lambda)$

HMM Evaluation：求 $p(O;\lambda)$

1. 直接计算法

$$p(O;\lambda)=\sum_I p(O|I;\lambda)\cdot p(I;\lambda)$$

$$p(I;\lambda)=p(i_1,i_2,i_3,\dots,i_{t-1},i_t)$$

$$=\underbrace{p(i_t|i_1,i_2,i_3,\dots,i_{t-1})}_{\text{马尔可夫假设}}\cdot p(i_1,i_2,i_3,\dots,i_{t-1})$$

$$=a_{i_{t-1}i_t}\cdot\underbrace{p(i_1,i_2,i_3,\dots,i_{t-1})}_{\text{递归化简}}$$

$$=a_{i_{t-1}i_t}\cdot a_{i_{t-2}i_{t-1}}\cdots a_{i_1i_2}\cdot\pi_{i_1}$$

$$p(O|I;\lambda)=p(O_1,O_2,\dots,O_t|i_1,i_2,\dots,i_t;\lambda)$$

$$=\underbrace{p(O_t|O_1,O_2,\dots,O_{t-1},i_1,i_2,\dots,i_t;\lambda)}_{\text{观测独立假设}\quad p(O_t|i_t;\lambda)}$$

$$\cdot\,p(O_1,O_2,\dots,O_{t-1}|i_1,i_2,\dots,i_t;\lambda)$$

$$=b_{i_t}(O_t)\cdot b_{i_{t-1}}(O_{t-1})\cdots b_{i_1}(O_1)$$

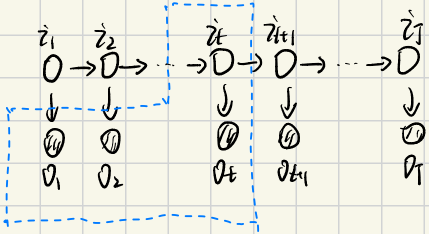
$$\therefore p(O;\lambda)=\sum_I p(I;\lambda)\cdot p(O|I;\lambda)$$

$$=\sum_{i_1,i_2,\dots,i_t}\pi_{i_1}a_{i_1}a_{i_2}\cdots a_{i_{t-2}i_{t-1}}a_{i_{t-1}i_t}\cdot b_{i_1}(O_1)b_{i_2}(O_2)\cdots b_{i_t}(O_t)$$

上面的公式要对所有 i_t 的状态求和，每个 i_t 有 N 个状态.

复杂度 $O(TN^T)$ 指数阶，不可行

2.前向算法



记 $\alpha_t(i) = P(O_1, O_2, \dots, O_t, i_t = q_i; \lambda)$

$\alpha_T(i) = P(O, i_T = q_i; \lambda)$

$P(O;\lambda) = \sum_{i=1}^{N} P(O, i_T = q_i; \lambda)$ 对 i_T 的N个状态求和

$= \sum_{i=1}^{N} \alpha_T(i)$

下面就探寻 $\alpha_t(i)$ 是否可以从 $\alpha_{t+1}(j)$ 递推得到

$$\alpha_{t+1}(j) = P(O_1, O_2, \dots, O_t, O_{t+1}, i_{t+1} = q_j; \lambda)$$

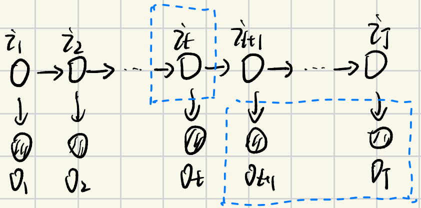
$$= \sum_{i=1}^{N} P(O_1, O_2, \dots, O_t, O_{t+1}, i_{t+1} = q_j, i_t = q_i; \lambda)$$

$$= \sum_{i=1}^{N} \underbrace{P(O_{t+1} \mid O_1, \dots, O_t, i_{t+1} = q_j, i_t = q_i; \lambda)}_{\text{独立观测假设}} \cdot P(O_1, \dots, O_t, i_{t+1} = q_j, i_t = q_i; \lambda)$$

$$= \sum_{i=1}^{N} \underbrace{P(O_{t+1} \mid i_{t+1} = q_j)}_{b_j(O_{t+1})} \cdot \underbrace{P(i_{t+1} = q_j \mid O_1, \dots, O_t, i_t = q_i; \lambda)}_{a_{ij}} \cdot \underbrace{P(O_1, \dots, O_t, i_t = q_i; \lambda)}_{\alpha_t(i)}$$

$$= \sum_{i=1}^{N} \alpha_t(i)\, a_{ij} \cdot b_j(O_{t+1})$$

3. 后向算法



记 $\beta_t(i) = P(O_{t+1}, O_{t+2}, \dots, O_T \mid i_t = q_i; \lambda)$

$\Rightarrow \beta_1(i) = P(O_2, \dots, O_T \mid i_1 = q_i; \lambda)$

$$P(O;\lambda) = P(O_1, O_2, \dots, O_T; \lambda)$$

$$= \sum_{i=1}^{N} P(O_1, O_2, \dots, O_T, i_1 = q_i; \lambda)$$

$$= \sum_{i=1}^{N} P(O_1, O_2, \dots, O_T \mid i_1 = q_i; \lambda)\, \underbrace{P(i_1 = q_i)}_{\pi_i}$$

$$= \sum_{i=1}^{N} P(O_1 \mid O_2, \dots, O_T, i_1 = q_i; \lambda) \cdot P(O_2, \dots, O_T \mid i_1 = q_i; \lambda) \cdot \pi_i$$

$$= \sum_{i=1}^{N} P(O_1 \mid i_1 = q_i; \lambda) \cdot \beta_1(i) \cdot \pi_i$$

$$= \sum_{i=1}^{N} B_i(O_1) \cdot \beta_1(i) \cdot \pi_i$$

$$\beta_t(i) = P(O_{t+1}, \dots, O_T \mid i_t = q_i)$$

$$= \sum_{j=1}^{N} P(O_{t+1}, \dots, O_T, i_{t+1} = q_j \mid i_t = q_i)$$

$$= \sum_{j=1}^{N} P(O_{t+1}, \dots, O_T \mid i_{t+1} = q_j, i_t = q_i) \cdot P(i_{t+1} = q_j \mid i_t = q_i)$$

概率图性质

$$= \sum_{j=1}^{N} P(O_{t+1}, \dots, O_T \mid i_{t+1} = q_j) \cdot a_{ij}$$

$$= \sum_{j=1}^{N} P(O_{t+1} \mid O_{t+2}, \dots, O_T, i_{t+1} = q_j) \cdot P(O_{t+2}, \dots, O_T \mid i_{t+1} = q_j) \cdot a_{ij}$$

$$= \sum_{j=1}^{N} P(O_{t+1} \mid i_{t+1} = q_j) \cdot \beta_{t+1}(j) \cdot a_{ij}$$

$$= \sum_{j=1}^{N} b_j(O_{t+1}) \cdot \beta_{t+1}(j) \cdot a_{ij}$$

利用前向后向概率可以将观测序列写成:

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