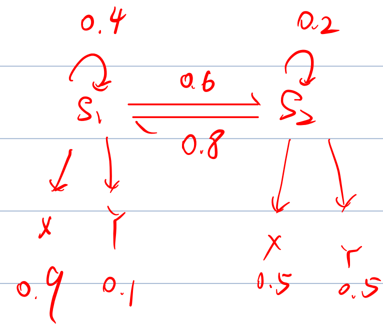
$\pi = [0.3 \quad 0.7]$
$\quad\quad\ s_1 \quad s_2$



$O = XY$

$\delta_1(s_1) = 0.3 \times 0.4 = 0.27$

$\delta_1(s_2) = 0.7 \times 0.5 = 0.35$

$\varphi_2(s_1) = \underset{i}{\operatorname{argmax}} [\delta_1(i) a_{ij}]$

$= \underset{i}{\operatorname{argmax}} [\delta_1(i) a_{i s_1}]$

$\underline{\varphi_2(s_1) = s_2}$

$\delta_2(s_1) = 0.35 \times 0.8 \times 0.1 = 0.028$

$\underline{\varphi_2(s_2) = s_1}$

$\delta_2(s_2) = 0.27 \times 0.6 \times 0.5 = 0.081$

$\underline{\varphi_2(s_2) = s_1}$

$s_1 \, s_2$

$\beta_t(i) = P(O_{t+1}, \cdots, O_T \mid q_t = S_i, \lambda)$

已算出 $\beta_{t+1}(i)$ $i = 1, 2, \cdots, N$ 如何求 $\beta_t(i)$?

① $\beta_{t+1}(j) \cdot \underline{b_j(O_{t+1})} = P(\underline{O_{t+1}}, \cdots, O_T \mid q_{t+1} = j)$

② $a_{ij} \cdot \beta_{t+1}(j) \cdot b_j(O_{t+1})$

$$= \frac{P(q_{t+1}=j, q_t=i, \lambda)}{P(q_t=i, \lambda)} \boxed{\frac{P(O_{t+1}, \cdots, O_T, q_{t+1}=j, \lambda)}{P(q_{t+1}=j, \lambda)}}$$

$$= \frac{P(q_{t+1}=j, q_t=i, \lambda)}{P(q_t=i, \lambda)} \boxed{\frac{P(O_{t+1}, \cdots, O_T, q_t=i, q_{t+1}=j, \lambda)}{P(q_{t+1}=j, q_t=i, \lambda)}}$$

$$= \frac{P(O_{t+1}, \cdots, O_T, q_{t+1}=j, q_t=i, \lambda)}{P(q_t=i, \lambda)}$$

③ $\beta_t(i) = \sum ②$

因为已知 q_{t+1} 下，O_{t+1}、$O_{t+2} \ldots O_T$ 发生的概率就等于已知 q_t. q_{t+1} 下 O_{t+1}、$O_{t+2} \ldots O_T$ 发生的概率，因为只要知道 q_{t+1} 就足够影响 $t+1$ 及以后的观测了（马尔可夫性）

前向：$\alpha_t(i) = P(O_1,\dots,O_t, q_t = S_i \mid \lambda)$

① $\alpha_t(j)\, a_{ji} = \dfrac{P(O_1,\dots,O_t, q_t = S_j, \lambda)}{P(\lambda)} \dfrac{P(q_{t+1} = S_i, q_t = S_j, \lambda)}{P(q_t = S_j, \lambda)}$

$= \alpha_t(j) \dfrac{P(q_{t+1} = S_i, q_t = S_j, \lambda, O_1, \dots, O_t)}{P(q_t = S_j, \lambda, O_1, \dots, O_t)}$

因为已知 ≥t 状态，≤t 观测与状态无关，概率可以抵消

② $b_i(O_{t+1})\, \alpha_t(j)\, a_{ji} = \dfrac{P(O_1,\dots,O_{t+1}, q_t = S_j, q_{t+1} = S_i, \lambda)}{P(\lambda)}$

③ $\alpha_{t+1}(i) = \sum_j$ ②

$$\alpha_t(i)\beta_t(i) = P(O_1 O_2 \dots O_t, q_t = S_i|\lambda)\, P(O_{t+1} \dots O_T \mid q_t = S_i, \lambda)$$

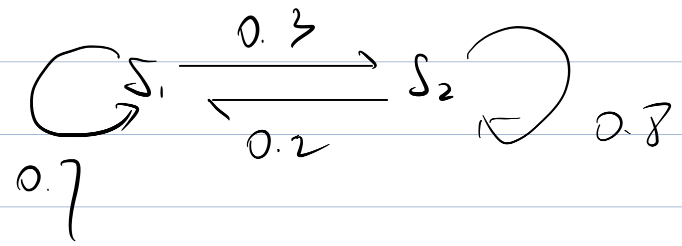
$$= \frac{P(O_1 O_2 \dots O_t, q_t = S_i, \lambda)}{P(\lambda)} \quad \frac{P(O_{t+1} \dots O_T, q_t = S_i, \lambda)}{P(q_t = S_i, \lambda)}$$

$$= \frac{\cancel{P(O_1 O_2 \dots O_t, q_t = S_i, \lambda)}}{P(\lambda)} \quad \frac{P(O_1 \dots O_t\, O_{t+1} \dots O_T\, q_t = S_i, \lambda)}{\cancel{P(q_t = S_i, O_1 \dots O_t, \lambda)}}$$

因为 $O_{t+1} \dots O_T$ 与 $O_1 \dots O_t$ 独立 且 q_t 与 $O_1 \dots O_t$ 独立

$$= \frac{P(O_1 \dots O_T, q_t = S_i, \lambda)}{P(\lambda)} = P(O_1 \dots O_T, q_t = S_i \mid \lambda)$$

$x = AcT$



$\alpha_1(1) = \pi_1 b_1(O_1) = \frac{1}{2} \times 0.4 = \frac{1}{5}$

$\alpha_1(2) = \pi_2 b_2(O_1) = \frac{1}{2} \times \frac{1}{10} = \frac{1}{20}$

$\alpha_2(1) = \alpha_1(1) a_{11} b_1(O_2) + \alpha_1(2) a_{21} b_1(O_2)$

$= \frac{1}{5} \times \frac{7}{10} \times \frac{1}{10} + \frac{1}{20} \times \frac{1}{5} \times \frac{1}{10} = \frac{15}{1000} = \frac{3}{200}$

$\alpha_2(2) = \alpha_1(1) a_{12} b_2(O_2) + \alpha_1(2) a_{22} b_2(O_2)$

$= \frac{1}{5} \times \frac{3}{10} \times \frac{2}{5} + \frac{1}{20} \times \frac{4}{5} \times \frac{2}{5} = \frac{12+8}{500} = \frac{1}{25}$

$\alpha_3(1) = \alpha_2(1) a_{11} b_1(O_3) + \alpha_2(2) a_{21} b_1(O_3)$

$= \frac{3}{200} \times \frac{7}{10} \times \frac{1}{10} + \frac{1}{25} \times \frac{1}{5} \times \frac{1}{10} = \frac{37}{20000}$

$\alpha_3(2) = \alpha_2(1) a_{12} b_2(O_3) + \alpha_2(2) a_{22} b_2(O_3)$

$= \frac{3}{200} \times \frac{3}{10} \times \frac{2}{5} + \frac{1}{25} \times \frac{4}{5} \times \frac{2}{5} = \frac{148}{10000}$

$\delta_1(1) = \pi_1 b_1(O_1) = \frac{1}{2} \times \frac{2}{5} = \frac{1}{5}$ $\quad \varphi_1(1) = 0$

$\delta_1(2) = \pi_2 b_2(O_1) = \frac{1}{2} \times \frac{1}{10} = \frac{1}{20}$ $\quad \varphi_1(2) = 0$

$\delta_2(1) = \max \begin{cases} \delta_1(1) \times a_{11} = \frac{1}{5} \times \frac{7}{10} = \frac{7}{50} \checkmark \\ \delta_1(2) \times a_{21} = \frac{1}{20} \times \frac{1}{5} = \frac{1}{100} \end{cases}$ $\quad b_2(O_2)$

$= \frac{7}{50} \times \frac{1}{10} = \frac{7}{500}$

$\underline{\varphi_2(1) = 1}$ $\quad \delta_2(2) = \max \begin{cases} \delta_1(1) \times a_{12} = \frac{1}{5} \times \frac{3}{10} = \frac{3}{50} \checkmark \\ \delta_1(2)\, a_{12} = \frac{1}{20} \times \frac{4}{5} = \frac{4}{100} \end{cases}$ $\cdot\, b_2(O_2)$

$\varphi_2(1) = 1$ $\quad = \frac{3}{50} \times \frac{2}{5} = \frac{3}{125}$ $\quad \varphi_2(2) = 1$

$$\delta_3(1) = \max \begin{cases} \delta_2(1)\, a_{11} = \frac{}{500} \times \frac{}{10} = \frac{}{5000} \\ \delta_2(2)\, a_{21} = \frac{3}{125} \times \cancel{8}\ \frac{2}{10} = \frac{3}{625} \end{cases} b(O_3) \cdot \frac{1}{10}$$

$$= \frac{49}{50000}$$

$$\psi_3(1) = 1$$

$$\delta_3(2) = \max \begin{cases} \delta_2(1)\, a_{12} = \frac{7}{500} \times \frac{3}{10} = \frac{21}{5000} \\ \delta_2(2)\, a_{22} = \frac{3}{125} \times \frac{4}{5} = \frac{12}{625} \checkmark \end{cases}$$

$$b_3(O_2) \quad \frac{2}{5} = \frac{24}{125 \times 5}$$

$$\psi_3(2) =$$

δ_2 δ_2 δ_1

$\psi_2(2) = 1$ δ_1 $\psi_1($

2.1

(1) $\pi = [0.5 \quad 0.5]$ S_1: 畅销 S_2: 滞销

$$A = \begin{bmatrix} 0.6 & 0.4 \\ 0.5 & 0.5 \end{bmatrix} \qquad B = \begin{bmatrix} 0.4 & 0.4 & 0.2 \\ 0.2 & 0.3 & 0.5 \end{bmatrix}$$

(2) $O = [H \quad M \quad L]$

$\delta_1(1) = \pi_1 b_1(O_1) = \frac{1}{2} \times \frac{2}{5} = \frac{1}{5}$ $\quad \varphi_1(1) = 0$

$\delta_1(2) = \pi_2 b_2(O_1) = \frac{1}{2} \times \frac{1}{5} = \frac{1}{10}$ $\quad \varphi_1(2) = 0$

$$\delta_2(1) = \max_{1\le i\le 2} [\delta_1(i) a_{i1}] b_1(O_2)$$

$$= \max \begin{cases} \frac{1}{5} \times \frac{3}{5} \checkmark \\ \frac{1}{10} \times \frac{1}{2} \end{cases} \times \frac{2}{5} = \frac{6}{125} \qquad \varphi_2(1) = 1$$

$$\delta_2(2) = \max_{1\le i\le 2} [\delta_1(i) a_{i2}] b_2(O_2)$$

$$= \max \begin{cases} \frac{1}{5} \times \frac{2}{5} \checkmark \\ \frac{1}{10} \times \frac{1}{2} \end{cases} \times \frac{3}{10} = \frac{3}{125} \qquad \varphi_2(2) = 1$$

$$\delta_3(1) = \max_{1\le i\le 2} [\delta_2(i) a_{i1}] b_1(O_3)$$

$$= \max \begin{cases} \frac{6}{125} \times \frac{3}{5} \checkmark \\ \frac{3}{125} \times \frac{1}{2} \end{cases} \times \frac{1}{5} = \frac{18}{3125} \qquad \varphi_3(1) = 1$$

$$\delta_3(2) = \max_{1\le i\le 2} [\delta_2(i) a_{i2}] b_2(O_3)$$

$$= \max \begin{cases} \frac{6}{125} \times \frac{2}{5} \\ \frac{3}{125} \times \frac{1}{2} \checkmark \end{cases} \times \frac{1}{2} = \frac{3}{500} \qquad \varphi_3(2) = 2$$

$$\frac{3}{500} > \frac{18}{3125} \longrightarrow i_3^* = 2 \longrightarrow i_2^* = \varphi_3(2) = 2 \longrightarrow i_1^* = \varphi_2(2) = 1$$

∴ 推测前三个月的市场行情为 $\boxed{\text{畅销、滞销、滞销}}$

2.2 $$h_t = Q\Lambda^t Q^T h_0$$

(1) Λ是对角阵，故Λ^t中元素$\Lambda_{ii} = (\lambda_i)^t$

若$\lambda_i < 1$ 则Λ_{ii}会以指数形式趋于0 $\longrightarrow$ 梯度逐渐消失

(2) $h_0 = q_i$

$$h_t = (q_1 \cdots q_n)\begin{bmatrix} \lambda_1^t & & & \\ & \lambda_2^t & & \\ & & \ddots & \\ & & & \lambda_n^t \end{bmatrix}\begin{bmatrix} q_1^T \\ \vdots \\ q_n^T \end{bmatrix} q_i$$

$$= \begin{bmatrix} \lambda_1^t q_1 & \lambda_2^t q_2 & \cdots & \lambda_n^t q_n \end{bmatrix}\begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$$ $\longrightarrow$ 只有第i行为1

$$= \lambda_i^t q_i$$

$$\therefore \frac{\partial L}{\partial h_0} = \frac{\partial L}{\partial h_t}\frac{\partial h_t}{\partial h_0} = \lambda_i^t \frac{\partial L}{\partial h_t}$$

(3) $h_0 = \sum_{i=1}^n k_i q_i$ 由(2)中的分析可知

$$h_t = \sum_{i=1}^n k_i \lambda_i^t q_i$$

$$\frac{\partial L}{\partial h_0} = \sum_{i=1}^n \frac{\partial L}{\partial q_i} = \sum_{i=1}^n \frac{\partial L}{\partial h_t}\frac{\partial h_t}{\partial q_i} = \sum_{i=1}^n k_i \lambda_i^t \frac{\partial L}{\partial h_t}$$

$\longrightarrow 0$

若所有λ_i绝对值均小于0，则λ_i^t随t增大而指数级减小

因此L对较早时刻的h的梯度趋近于0，当输入序列在时间上跨度较长时，这一特点将导致新产生的L的梯度无法回传到较早时刻的隐藏状态h，导致无法进行长距离相关信息建模.