π → x_1 → x_2 → x_3 → … → x_{t-1} → x_t → x_{t+1} → … → x_T

y_1 y_2 y_3 … y_{t-1} y_t y_{t+1} … y_T

f a n e m a

$x_1 \sim \pi(x_1)$

$x_t \sim p(x_t | x_{t-1})$

$p(x_t | x_{t-1}, x_{t-2}, \dots, x_1) = p(x_t | x_{t-1})$

$x_t \in \{1, \dots, n\}$

$p(x_t = i | x_{t-1} = j) = a_{ji}$

$A = (a_{ij})_{i,j=1}^{n}$

$p(x_1 = i_1, \dots, x_T = i_T) = \pi(i_1) \cdot a_{i_1 i_2} \dots a_{i_{T-1} i_T}$

$D = \{x_n\}_{n=1}^{N}$

$p(D|\lambda) \to \max$

$\lambda = (\pi, A)$

$q_{n1} \dots q_{nT}$

$p(\pi)$

$p(A)$

$$\pi_i^{ML} = p(x_1 = i) = \frac{\#\{q_{n1} = i\}}{N}$$

$$a_{ij}^{ML} = \frac{\#\{q_{nt} = i, q_{n,t+1} = j\}}{\#\{q_{nt} = i, q_{n,t+1} = x\}}$$

$$b_i^{ML}(k) = \frac{\#\{q_{n1} = i, d_{nt} = k\}}{\#\{q_{nt} = i\}}$$

λ

π: $p(x_1 = i) = \pi_i$

A: $p(x_t = j | x_{t-1} = i) = a_{ij}$

B: $p(y_t = k | x_t = i) = b_i(k)$

$y \in \{1, \dots, V\}$

$$p(Q, D|\lambda) = p(\underbrace{x = q_1 q_2 \dots q_T}_{Q}, \underbrace{y = d_1 d_2 \dots d_T}_{D} | \pi, A, B) =$$

$$= \pi_{q_1} b_{q_1}(d_1) a_{q_1 q_2} b_{q_2}(d_2) a_{q_2 q_3} \dots a_{q_{T-1} q_T} b_{q_T}(d_T)$$

$D = \{d_{n1} d_{n2} \dots d_{nT}\}_{n=1}^{N}$

$\lambda \to \max$

$$p(D|\lambda) = \sum_Q p(Q, D|\lambda) =$$

$$= \sum_{q_1 \dots q_T} \pi_{q_1} \dots b_{q_T}(d_T)$$

$\pi \to x_1 \to x_2 \to \dots \to x_{t-1} \to x_t \to \dots \to x_T$

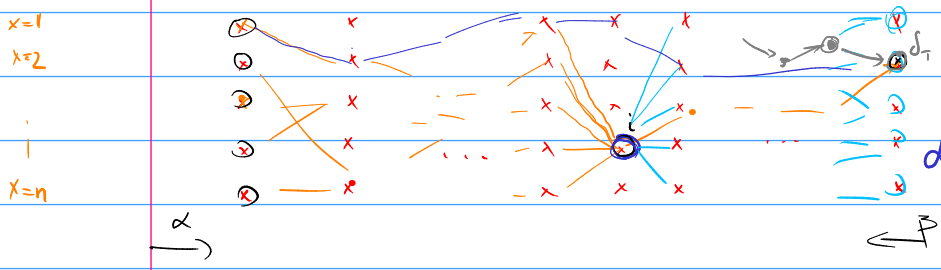
y_1 y_2 … y_{t-1} y_t … y_T

① $p(D|\lambda) = ?$

② $Q^* = \underset{Q}{\operatorname{argmax}}\, p(Q, D|\lambda)$ inference

③ $\lambda^* = \underset{\lambda}{\operatorname{argmax}}\, p(D|\lambda)$ learning

$p(D|\lambda) = \sum_{q_1 \dots q_T} \pi_{q_1} \dots b_{q_T}(d_T)$ $\quad O(n^T)$



$\alpha_t(i) = p(d_1, d_2, \dots, d_t, q_t = i | \lambda)$

$\alpha_t(i) = \sum_{j=1}^{n} \alpha_{t-1}(j) a_{ji} b_i(d_t)$ $\quad O(T \cdot n^2)$

$\alpha_T(i) = p(D, q_T = i | \lambda)$

$p(D|\lambda) = \sum_i \alpha_T(i)$

$t=1$ $t=2$ $t-1$ t $t+1$

d_1 d_2 d_{t-1} d_t d_{t+1} d_T

$\alpha_1(i) = p(d_1, q_1 = i | \lambda) = \pi_i \cdot b_i(d_1)$

$\alpha_t(i) = p(d_{1..t}, q_t = i | \lambda) = p(d_{1..t-1}, d_t, q_t = i | \lambda) =$

$$= \sum_{j=1}^{n} p(d_{1..t-1}, d_t, q_t = i, q_{t-1} = j | \lambda) =$$

$$= \sum_{j=1}^{n} \underbrace{p(d_{1..t-1}, q_{t-1} = j | \lambda)}_{\alpha_{t-1}(j)} \cdot \underbrace{p(q_t = i | q_{t-1} = j, \lambda)}_{a_{ji}} \cdot \underbrace{p(d_t | q_t = i, \lambda)}_{b_i(d_t)}$$

$\beta_t(i) = p(d_{t+1}, \dots, d_T | q_t = i, \lambda)$

$\beta_t(i) = p(d_{t+1..T} | q_t = i, \lambda) =$

$$= \sum_{j=1}^{n} p(d_{t+1..T}, q_{t+1} = j | q_t = i, \lambda) =$$

$$= \sum_{j=1}^{n} p(d_{t+2..T} | q_{t+1} = j, \cancel{q_t = i}, \lambda) \cdot \cancel{p(d_{t+1} | q_{t+1} = j, q_t = i, \lambda)} \cdot p(q_{t+1} = j | q_t = i, \lambda) \cdot$$

$\beta_{t+1}(j)$ $\quad b_j(d_{t+1})$ $\quad a_{ij}$

$$\beta_t(i) = \sum_{j=1}^{t+1} a_{ij} \cdot b_j(d_{t+1}) \cdot \beta_{t+1}(j)$$

$\beta_1(i) = p(d_2 \dots d_T | q_1 = i, \lambda)$

$$p(D|\lambda) = \sum_{i=1}^{n} \pi_i b_i(d_1) \cdot \beta_1(i)$$

② $\rightarrow Q^* = \arg\max_Q p(Q \mid D, \lambda)$

$\rightarrow q_t^* = \arg\max_{q_t} p(q_t \mid D, \lambda)$

$$\gamma_t(i) = p(q_t = i \mid D, \lambda) = \frac{p(q_t = i, d_{1..t}, d_{t+1..T} \mid \lambda)}{p(D \mid \lambda)} = \frac{\overbrace{p(d_{1..t}, q_t = i \mid \lambda)}^{\alpha_t(i)} \cdot \overbrace{p(d_{t+1..T} \mid q_t = i, \lambda)}^{\beta_t(i)}}{p(D \mid \lambda)} = \frac{\alpha_t(i)\beta_t(i)}{p(D \mid \lambda)}$$

$\gamma_t(i) \propto \alpha_t(i)\beta_t(i)$

$$\delta_t(i) = \max_{q_1..q_{t-1}} p(q_1..q_{t-1}, q_t = i, d_1..d_t \mid \lambda) =$$

$$= \max_{q_1..q_{t-1}} p(q_1..q_{t-2}, q_{t-1}, d_1..d_{t-1} \mid \lambda) \cdot \underbrace{p(q_t = i \mid q_{t-1}, d_1..d_{t-1}, \lambda)}_{a_{q_{t-1} i}} \cdot \underbrace{p(d_t \mid q_t, \lambda)}_{b_i(d_t)} =$$

$$= \max_{q_{t-1}} \Big[\underbrace{\Big(\max_{q_1..q_{t-2}} p(q_1..q_{t-2} q_{t-1} .. \mid \lambda) \Big)}_{\delta_{t-1}(q_{t-1}),\ q_{t-1} = j} \cdot a_{q_{t-1} i}\, b_i(d_t) \Big]$$

$$\delta_t(i) = \max_j \left[\delta_{t-1}(j) \cdot a_{ji}\, b_i(d_t) \right]$$

$p(D \mid \lambda) \xrightarrow{\lambda} \max$

$\arg\max p(Q \mid D, \lambda)$

③ Baum-Welch algorithm

$p(D \mid \lambda) \xrightarrow{\lambda} \max$

E-war $\mathbb{E}[Q \mid \lambda^{(m)}] \leftarrow \alpha_t(i), \beta_t(i), \gamma_t(i), \ \varepsilon_t(i,j) = p(q_{t-1} = i, q_t = j \mid D, \lambda)$

$(\pi^{(m)}, A^{(m)}, B^{(m)}) = \lambda^{(m)}$

M-war – $\lambda^{(m+1)} := \arg\max_\lambda \mathbb{E}_{Y(m)} \log p(D, Q \mid \lambda)$

M-war: $\pi_i := \mathbb{E}\, \#\{q_1 = i\} / N = \frac{\sum_n \gamma_1^{(n)}(i)}{N}$

$$a_{ij} = \frac{\mathbb{E}\, \#\{q_t = i, q_{t+1} = j\}}{\mathbb{E}\, \#\{q_t = i\}} = \frac{\sum_n \sum_{t=1}^{T-1} \varepsilon_{t+1}(i,j)}{\sum_n \sum_{t=1}^{T-1} \gamma_t^{(n)}(i)}$$

$\lambda :=$

$$b_i(k) = \frac{\mathbb{E}\, \#\{q_t = i, d_t = k\}}{\mathbb{E}\, \#\{q_t = i\}} = \frac{\sum_{n, t: d_t^{(n)} = k} \gamma_t^{(n)}(i)}{\sum_n \sum_t \gamma_t^{(n)}(i)}$$

$$\gamma_t^{(n)}(i) = p(q_t = i \mid d^{(n)}, \lambda)$$

$p(i)$? $p(i)$ $p(i)$

$d_1 d_2 \ d \ 2 \ [k] \ 2 \ [k] \ 2\ 7\ 1\ [k]$

$\rightarrow X_{t-1} \rightarrow X_t \rightarrow X_{t+1} \rightarrow$

$B \downarrow p(y \mid x; \theta_y)$

y_t

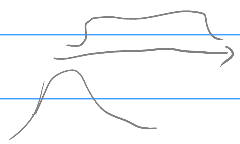
$\theta_y = \arg\max \prod p(d_n \mid \ldots)$

$A = (a_{ij})$

$x_t = 1 \xrightarrow{a_{11}} x_{t+1} = 1 \xrightarrow{a_{11}} x_{t+2} = 1$

$\rightarrow X_t \rightarrow X_{t+1} \rightarrow \ldots$

$p(\tau \text{ ward } \& \ x = i) = a_{ii}^{\tau-1}(1 - a_{ii})$



$\bar{d}$ MFCC

$\circ \rightarrow \circ \rightarrow \ldots \rightarrow \circ \rightarrow$

$|d_1|d_2| \ldots |d_t|$

$\in \mathbb{R}^{40}$ $\in \mathbb{R}^{40}$

e: m a | m a s

$\bullet \rightarrow \circ \rightarrow \circ \rightarrow \circ \rightarrow \bullet$ $\quad$ $\bullet \rightarrow \circ \rightarrow \circ \rightarrow \bullet$

$d_1 \ d_2 \ d_3$

$p(M|D) \propto p(D|M)\, p(M)$

$M_1, M_2, \ldots$ model combination

stacking

blending

$y \sim \sum \alpha_k M_k(x)$

boosting

$y \sim F_m(x) = \alpha_1 f_1(x) + \alpha_2 f_2(x) + \ldots + \alpha_m f_m(x)$

$f_m(x), \alpha_m \rightsquigarrow F_{m-1}(x)$

$F_m(x) = \sum_{k=1}^{m} \alpha_k f_k(x)$

AdaBoost

XGBoost - gradient boosting