EM $p(X|\bar\theta) \xrightarrow{\bar\theta} \max$ $\qquad p(X|\bar\theta) = \prod_n \left(\sum_k \pi_k p_k(\bar x_n) \right)$

$p(X,Z|\bar\theta) \xrightarrow{\bar\theta} \max$ $\qquad p(X,Z|\bar\theta) = \prod_n \prod_k \left(\pi_k p_k(\bar x_n) \right)^{z_{nk}}$

$$Q(\bar\theta, \bar\theta^{(m)}) = \mathbb{E}_{p(Z|X,\bar\theta^{(m)})} \left[\log p(X,Z|\bar\theta) \right] =$$

$$= \int p(Z|X,\bar\theta^{(m)}) \log p(X,Z|\bar\theta)\, dZ \xrightarrow{\theta} \max$$

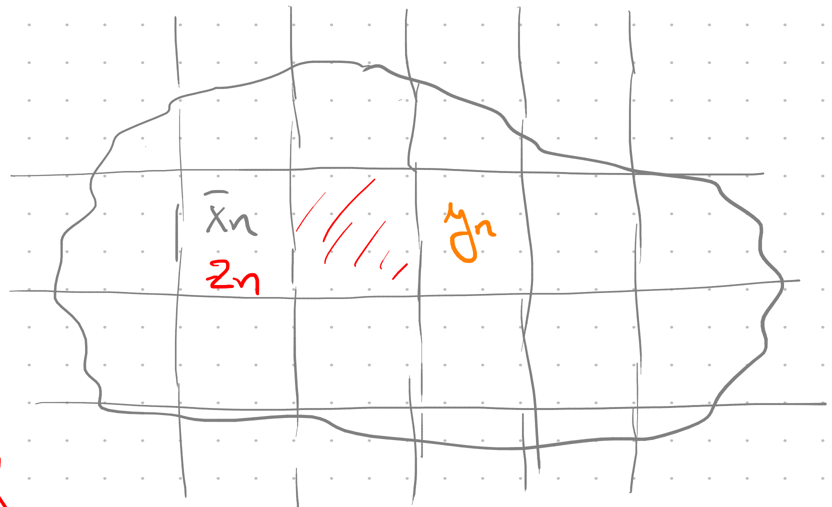
$$\bar\theta^{(m+1)} = \arg\max_{\bar\theta} Q(\bar\theta, \bar\theta^{(m)})$$

Presence-only data

$z=1$ – „суслик водится"

$z=0$ – „не водится"

$p(z|\bar x)$



$$p(z|\bar x) = \sigma(\bar w^T \bar x) = \frac{1}{1+e^{-\bar w^T \bar x}}$$

$y=1$ – „видели суслика"

$y=0$ – „не видели"

$$p(z|\bar x) = \sigma(\eta(\bar x)) = \frac{1}{1+e^{-\eta(\bar x)}}$$

Если $y=1$, то $z=1$

Но если $y=0$, то $z=?$

Prospective vs retrospective studies

	$d=0$	$d=1$	
$x=0$	n_{00}, π_{00}	n_{01}, π_{01}	π_{0*}
$x=1$	n_{10}, π_{10}	n_{11}, π_{11}	π_{1*}
	π_{*0}	n_{*1}, π_{*1}	

$n = n_{00} + \dots + n_{11}$

$$\pi_{ij} = \frac{n_{ij}}{n}$$

$$p(d=1) = \pi_{*1}$$

Prospective study / Cohort study

$$p(d=1|x=0) = \frac{\pi_{01}}{\pi_{0*}} = \cancel{\sigma(\eta(\bar{x}))}$$

$$p(d=1|x=1) = \frac{\pi_{11}}{\pi_{1*}} = \sigma(\eta(\cancel{\bar{x}'}))$$

Retrospective study

$$\pi_0 = p(s=1|d=0)$$

$$\pi_1 = p(s=1|d=1)$$

$$\frac{\pi_{01}}{\pi_{0*}} = p(d=1|x=0, s=1) = \frac{p(d=1|x=0)\overbrace{p(s=1|d=1, x=0)}^{\pi_1}}{p(d=1|x=0)\underbrace{p(s=1|d=1,x=0)}_{\pi_1} + p(d=0|x=0)\underbrace{p(s=1|d=0,x=0)}_{\pi_0}}$$

$$\frac{\pi_{01}}{\pi_{0*}} = \frac{\pi_1 \cdot \overbrace{p(d=1|x=0)}^{p}}{\pi_1 \cdot p(d=1|x=0) + \pi_0(1 - p(d=1|x=0))}$$

$$\pi_{01}(\pi_1 \cdot p + \pi_0(1-p)) = \pi_1 \cdot p \cdot \pi_{0*}$$

$$\pi_0 \cdot \pi_{01} = p \cdot (\overbrace{\pi_1 \cdot \pi_{0*} - \pi_1 \cdot \pi_{01}}^{\pi_1 \cdot \pi_{00}} + \pi_0 \cdot \pi_{01})$$

$$\sigma(\eta(\bar{x})) = p = \frac{\pi_0 \pi_{01}}{\pi_0 \pi_{01} + \pi_1 \pi_{00}}$$

$$\sigma(\eta^*(\bar{x})) = \frac{\pi_{01}}{\pi_{01} + \pi_{00}}; \quad 1 - \sigma(\eta^*(\bar{x})) = \frac{\pi_{00}}{\pi_{01} + \pi_{00}}$$

$$\sigma(\eta(\bar{x})) = \frac{\pi_0 \cdot \sigma(\eta^*(\bar{x}))}{\pi_0 \cdot \sigma(\eta^*(\bar{x})) + \pi_1 (1 - \sigma(\eta^*(\bar{x})))} =$$

$$= 1/(1+e^{-\dots}) \qquad \frac{e^{-\dots}}{1+e^{-\dots}}$$

$$= \frac{\pi_0}{\pi_0 + \pi_1 \cdot e^{-\eta^*(\bar{x})}} = \frac{1}{1+e^{-\left(\eta^*(\bar{x}) - \log\frac{\pi_1}{\pi_0}\right)}}$$

нужно могу обучить

$$\boxed{\eta(\bar{x}) = \eta^*(\bar{x}) - \log\frac{\pi_1}{\pi_0}}$$

$z = 1$ – вообще ли

$y = 1$ – будем ли

$s = 1$ – выборка

$$\pi = p(z=1)$$

$$\sigma(\eta_{naive}(\bar{x})) = p(y=1 \mid \bar{x}, s=1)$$

Дальше:

n_p – positive

n_u – unknown

$N = n_p + n_u$

$z=1$: $n_p + \pi \cdot n_u$

$z=0$: $(1-\pi)\, n_u$

$$p(y=1 \mid \bar{x}, s=1) = \sum_z p(y=1, z \mid \bar{x}, s=1) =$$

$$= p(y=1, z=1 \mid \bar{x}, s=1) + \cancel{p(y=1, z=0 \mid \bar{x}, s=1)}$$

$$= p(y=1 \mid \cancel{\bar{x}}, s=1, z=1) \cdot p(z=1 \mid \bar{x}, s=1) = (*)$$

$$\| \quad p(y=1 \mid s=1, z=1) = \frac{p(y=1, z=1 \mid s=1)}{p(z=1 \mid s=1)} = \frac{n_p/N}{(n_p + \pi n_u)/N} =$$

$$= \frac{n_p}{n_p + \pi n_u}$$

$$\circledast = \frac{n_p}{n_p + \pi \cdot n_u} \cdot p(z=1 \mid \bar{x}, s=1) = \frac{n_p}{n_p + \pi n_u} \cdot \sigma\left(\eta(\bar{x}) + \log \frac{\pi_1}{\pi_0}\right)$$

$$\pi_0 = p(s=1 \mid z=0) = \frac{p(s=1)\, p(z=0 \mid s=1)}{p(z=0)} = \frac{(\cancel{1-\pi}) \cdot n_u}{(\cancel{1-\pi}) N} \quad p(s=1)$$

$$\pi_1 = p(s=1 \mid z=1) = \frac{p(s=1, z=1)}{p(z=1)} = \frac{p(s=1)\, p(z=1 \mid s=1)}{\pi} =$$

$$= p(s=1) \cdot \frac{n_p + \pi \cdot n_u}{N \cdot \pi}$$

$$\frac{\pi_1}{\pi_0} = \frac{n_p + \pi \cdot n_u}{\pi \cdot n_u}$$

$$\sigma(\eta_{naive}(\bar{x})) = \frac{n_p}{n_p + \pi \cdot n_u} \cdot \sigma\left(\eta(\bar{x}) + \log \frac{n_p + \pi \cdot n_u}{\pi \cdot n_u}\right)$$

$$p(y \mid X, \eta) = \prod_{i=1}^{N} p(y_i \mid \bar{x}, \eta, s=1) =$$

$$= \prod_{i=1}^{N} \sigma(\eta_{naive}(\bar{x}))^{[y_i=1]} \cdot \left(1 - \sigma(\eta_{naive}(\bar{x}))\right)^{[y_i=0]}$$

$$= \prod_{i=1}^{N} \left(\frac{n_p}{n_p + \pi \cdot n_u} \sigma\left(\eta(\bar{x}) + \log \frac{n_p + \pi \cdot n_u}{\pi \cdot n_u}\right)\right)^{[y_i=1]} \cdot \left(1 - \ldots\ldots\right)^{[y_i=0]} \longrightarrow \max_{\eta}$$

$$p(y, z \mid X, \bar{\eta}) = \prod_{i=1}^{N} \sigma(\eta(\bar{x}_i))^{[z_i=1]} \cdot (1 - \sigma(\eta(\bar{x}_i)))^{[z_i=0]} \cdot$$

$$\boxed{\cdot \prod_i p(y_i \mid z_i)} = \text{const}$$

$$\text{EM:}\quad \bar\eta_{naive} = \bar\eta^{(0)} \to \bar\eta^{(1)} \to \ldots \to \bar\eta^{(m)} \to \bar\eta^{(m+1)} \to \ldots$$

$$Q(\bar\eta, \bar\eta^{(m)}) = \mathbb{E}_{z|x,y,\bar\eta^{(m)}}\left[\log p(y, z \mid x, \bar\eta)\right] \xrightarrow{\bar\eta} \max$$

$$= \mathbb{E}_z\left[\sum_i \left([z_i=1]\cdot \log\sigma + [z_i=0]\log(1-\sigma)\right)\right] =$$

$$= \sum_i \left(\underbrace{\mathbb{E}_{z|\bar\eta^{(m)}}[z_i=1]}_{\sigma(\bar\eta^{(m)}(\bar x_i))} \log\sigma(\bar\eta) + \underbrace{\mathbb{E}_{z|\bar\eta^{(m)}}[z_i=0]}_{1-\sigma(\bar\eta^{(m)}(\bar x_i))} \log(1-\sigma(\bar\eta))\right)$$

если $y_i=0$: $\sigma(\bar\eta^{(m)}(\bar x_i))$; $1-\sigma(\bar\eta^{(m)}(\bar x_i)) = 0$

если $y_i=1$: $z_i = 1$ $= 1$

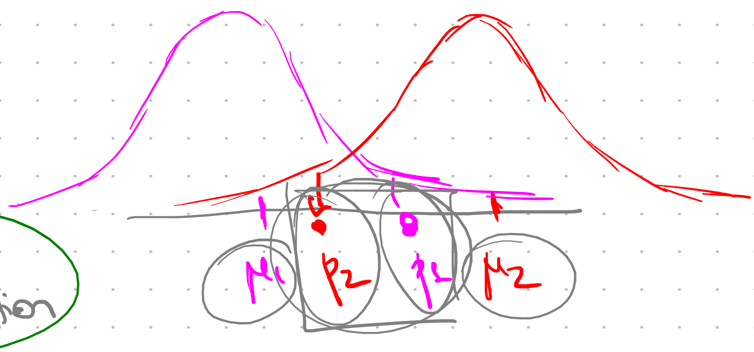
pseudolabels

$$\bar x \sim \pi\, p_1(\bar x) + (1-\pi)\, p_2(\bar x)$$

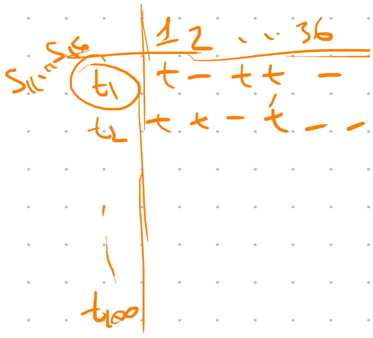
Рейтинг-система для ЧГК

~~Bradley-Terry~~ / Elo

Expectation Propagation



1) Команд много
2) Оценки меняются
3) Много делёжек мест 36 300 команд



Вопрос q

команда t — $(s_1 \dots s_t)$

y_{tq} = "ком. t отв. на вопр. q"

s_i — skill

c_q — complexity

$$p(x_{iq}=1 \mid s_i, c_q) = \sigma(\mu + s_i + c_q)$$

1) Baseline: $x_{iq} = y_{tq}$ для $t \ni i$ $\Rightarrow$ Log. Regr.

2) $y_{tq} = 1 \iff \exists i \in t: \ x_{iq} = 1$

$p(\underline{y} \mid \bar{s}, \bar{c}) \to \max$

$p(\cancel{y}, x \mid \bar{s}, \bar{c}) \to \max$

x_{iq}: если $y_{tq} = 0$, то $x_{iq} = 0$

если $y_{tq} = 1$, то $\exists i \in t: \ x_{iq} = 1$

EM: $\bar{s}^{(0)}, \bar{c}^{(0)}, \mu^{(0)} \in$ baseline

$i \in t$

E-шаг:
$$E[x_{iq}]_{x \mid \bar{s}^{(m)}, \bar{c}^{(m)}} = \begin{cases} 0, & \text{если } y_{tq} = 0 \\ p(x_{iq} = 1 \mid \exists j \in t \ x_{jq} = 1), & \text{если } y_{tq} = 1 \end{cases}$$

$$p(x_{iq}=1 \mid \exists j \in t \ x_{jq} = 1, \bar{\theta}^{(m)}) = \frac{p(x_{iq}=1 \mid \bar{\theta}^{(m)})}{p(\exists j \in t \ x_{jq}=1 \mid \bar{\theta}^{(m)})} =$$

$$= \frac{\sigma(\mu^{(m)} + s_i^{(m)} + c_q^{(m)})}{1 - \prod_{j \in t} \left(1 - \sigma(\mu^{(m)} + s_j^{(m)} + c_q^{(m)})\right)}$$

M-шаг: обучить $\sigma(\mu + s_i + c_q) \approx E[z_{iq}]$

Markov chain

$X_1, X_2, \dots, X_t, X_{t+1}, \dots$ $\qquad X_t \in \{1, \dots, L\}$

$$p(X_{t+1} | X_1 \dots X_t) = p(X_{t+1} | X_t)$$

$$p(x_{t+1} = j \mid x_t = i) = a_{ij}$$

$$p(x_1 = i) = \pi_i$$

$$\left(\bar{\pi}, A = \left(- a_{ij} -\right)\right) = \bar{\theta}$$

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

$$p(D | \bar{\theta}) = \prod_{n=1}^{N} p(d_{n1} d_{n2} \dots d_{nT} \mid \bar{\pi}, A) =$$

$$= \prod_{n=1}^{N} \pi_{d_{n1}} \cdot a_{d_{n1} d_{n2}} a_{d_{n2} d_{n3}} \dots a_{d_{n,T-1} d_{n,T}} \xrightarrow{\bar{\pi}, A} \max$$

$$\pi_i^{ML} = \frac{\#\{x_1 = i\} + 1}{N + L} \qquad a_{ij}^{ML} = \frac{\#\{x_t = i, x_{t+1} = j\} + 1}{\#\{x_t = i\} + L}$$

w_{t-1} $\quad w_t$ $\quad w_{t+1}$?

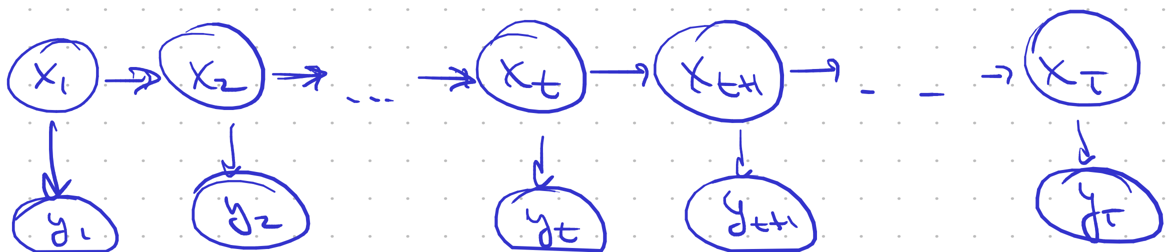
Moi dogu conax zaunax

$p(w_{t+1} | w_t)$ – unigram

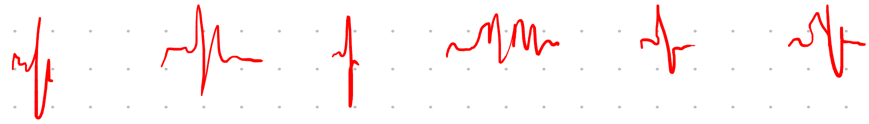
$p(w_{t+1} | w_t, w_{t-1})$ – bigram

$$p\left((\cancel{w_t}, w_{t+1}) \mid (w_{t-1}, w_t)\right)$$

Hidden Markov models



$f \to a \to n' \to e: \to m \to ə$



B-DATE → DATE → BG → B-LOC → BG → BG → BG → BG

14 февраля в Японии — Белый день .

$$\begin{tikzcd} x_1 \arrow[r] \arrow[d] & \dots \arrow[r] & x_t \arrow[r] \arrow[d] & x_{t+1} \arrow[r] \arrow[d] & \dots \arrow[r] & x_T \arrow[d] \\ y_1 & & y_t & y_{t+1} & & y_T \end{tikzcd}$$

$x_t \in \{1..n\}$

$y_t \in \{1..m\}$

$$p(X,Y) = p(x_1)\,p(y_1|x_1)\,p(x_2|x_1)\,p(y_2|x_2)\,p(x_3|x_2) \dots p(x_T|x_{T-1})\,p(y_T|x_T)$$

$\pi_i = p(x_1 = i)$ $\quad \bar{\pi}$

$a_{ij} = p(x_{t+1} = j \mid x_t = i)$ $\quad A$

$b_i(k) = p(y_t = k \mid x_t = i)$ $\quad B$

$\bar{\theta} = (\pi, A, B)$

$(\boxed{\cancel{Q = q_1 \dots q_T}}, D = d_1 \dots d_T)$

$$p(Q, D \mid \bar{\theta}) = \pi_{q_1} b_{q_1}(d_1)\, a_{q_1 q_2} b_{q_2}(d_2) \dots a_{q_{T-1} q_T} b_{q_T}(d_T)$$

$$\mathcal{D} = \{(d_1^{(s)} \dots d_T^{(s)})\}_{s=1}^{S}$$

$$p(\mathcal{D} \mid \bar{\theta}) = \prod_{s=1}^{S} p(D^{(s)} \mid \bar{\theta}) = \prod_{s=1}^{S} \left(\sum_{Q} p(D^{(s)}, Q \mid \bar{\theta}) \right)$$

① $p(D|\bar{\theta}) = \sum_{Q} p(D,Q|\bar{\theta}) \to ?$

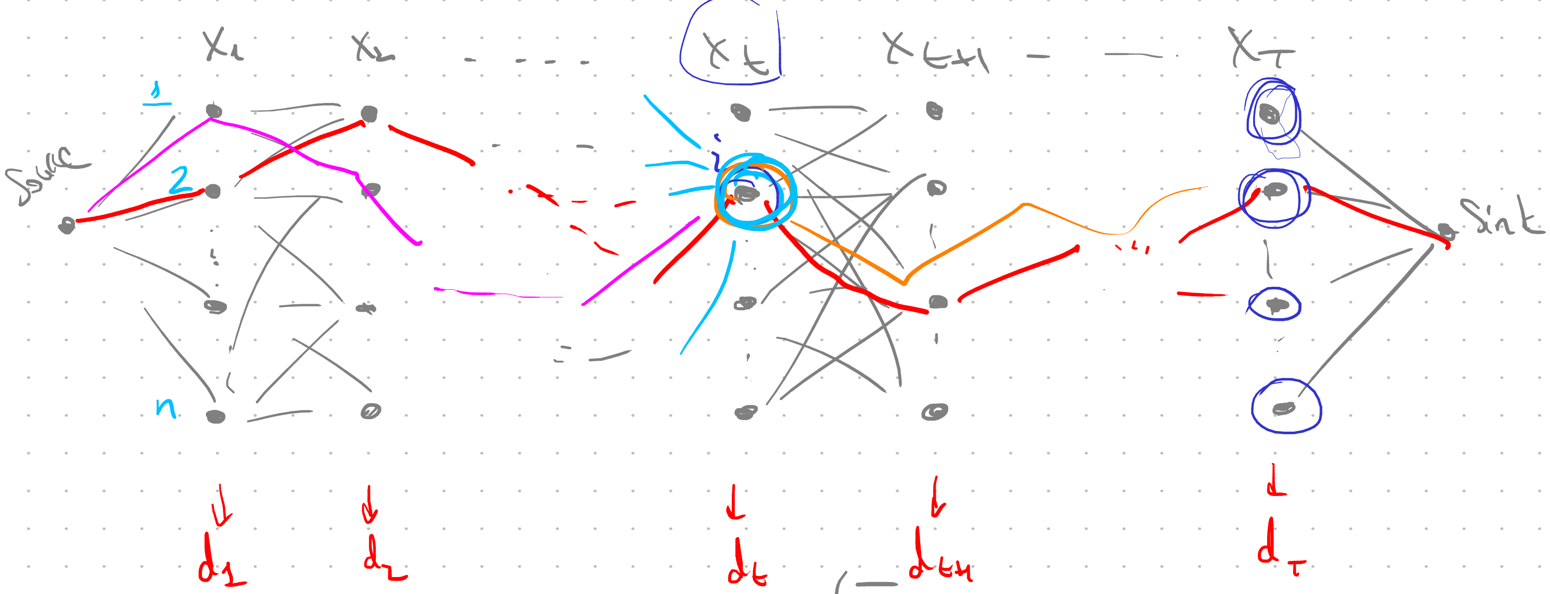
② $\bar{\theta} = (\bar{\pi}, A, B)$: $p(Q|D,\bar{\theta}) \xrightarrow{Q} \max$

③ $p(D|\bar{\theta}) \xrightarrow{\bar{\theta}} \max$

inference

learning

① $\underline{p(D|\bar{\theta})} = \sum_{Q} p(D,Q|\bar{\theta})$ $D = d_1 \dots d_T$



$(\bar{\pi}, A, B)$

$$\boxed{\alpha_t(i) = p(d_1 d_2 \dots d_t, q_t = i | \bar{\theta})}$$

$\underline{\alpha_T(i)}$: $p(D|\bar{\theta}) = p(d_1 d_2 \dots d_T|\bar{\theta}) = \sum_{i=1}^{T} \alpha_T(i)$

$\alpha_1(i)$: $\alpha_1(i) = p(d_1, q_1 = i|\bar{\theta}) = \pi_i \cdot b_i(d_1)$

$$\alpha_{t+1}(i) = p(d_1 d_2 \dots d_{t+1}, q_{t+1} = i|\bar{\theta}) =$$

$$= \sum_{j=1}^{n} p(d_1 \dots d_t d_{t+1}, q_{t+1} = i, q_t = j|\bar{\theta}) =$$

$$= \sum_{j=1}^{n} p(d_1 \dots d_t, q_t = j|\bar{\theta}) \cdot p(d_{t+1}, q_{t+1} = i | \cancel{d_1 \dots d_t}, q_t = j, \bar{\theta})$$

$$\boxed{\alpha_{t+1}(i) = \sum_{j=1}^{n} \alpha_t(j) \cdot a_{ji} \cdot b_i(d_{t+1})}$$

$$\boxed{\beta_t(i) = p(d_{t+1},\dots,d_T \mid q_t = i, \bar\theta)} \qquad \beta_T(i) = 1$$

$$\beta_t(i) = \sum_{j=1}^{n} p(d_{t+1}\dots d_T, q_{t+1} = j \mid q_t = i, \bar\theta) =$$

$$= \sum_j p(d_{t+2},\dots d_T \mid q_{t+1} = j) \cdot p(d_{t+1} \mid q_{t+1} = j, \bar\theta)\, p(q_{t+1} = j \mid q_t = i, \bar\theta)$$

$$\beta_1(i) = p(d_2 \dots d_T \mid q_1 = i, \bar\theta) \qquad \boxed{\beta_t(i) = \sum_{j=1}^{n} a_{ij} \cdot b_j(d_{t+1}) \cdot \beta_{t+1}(j)}$$

$$p(D \mid \bar\theta) = \sum_{i=1}^{n} p(q_1 = i \mid \bar\theta) \cdot p(d_1 \mid q_1 = i) \cdot p(d_2 \dots d_T \mid q_1 = i) =$$

$$= \sum_{i=1}^{n} \pi_i\, b_i(d_1) \cdot \beta_1(i)$$

2 <u>Inference</u> $\qquad p(Q \mid D, \bar\theta) \xrightarrow{Q} \max$

$$p(q_t \mid D, \bar\theta) \xrightarrow[q_t]{} \max$$

$$\boxed{\gamma_t(i) = p(q_t = i \mid d_1 \dots d_T, \bar\theta)}$$

$$p(q_t = i \mid D, \bar\theta)$$

$$\gamma_t(i) = \frac{p(q_t = i, d_1 \dots d_T \mid \bar\theta)}{p(D \mid \bar\theta)} = \frac{p(q_t = i, d_1, \dots d_t \mid \bar\theta) \cdot p(d_{t+1} \dots d_T \mid q_t = i, \bar\theta)}{p(D \mid \bar\theta)}$$

$$\boxed{\gamma_t(i) = \frac{\alpha_t(i) \cdot \beta_t(i)}{p(D \mid \bar\theta)} \propto \alpha_t(i)\, \beta_t(i)}$$

<u>Viterbi algorithm</u>

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

$$\delta_1(i) = \alpha_1(i) = \pi_i\, b_i(d_1)$$

$$\delta_{t+1}(i) = \max_{j=1}^{n} \max_{q_1 - q_{t-1}} p(q_1 \ldots q_{t-1}, q_t = j, q_{t+1} = i, d_1 \ldots d_{t+1} | \bar{\theta})$$

$$= \max_{j} \left[\max_{q_1 - q_{t-1}} p(q_1 - q_{t-1}, q_t = j, d_1 \ldots d_t | \bar{\theta}) \right] \underline{p(q_{t+1} = i | q_t = j, \bar{\theta})} \, \underline{p(d_{t+1} | q_{t+1} = i, \bar{\theta})}$$

$$\delta_{t+1}(i) = \max_{j=1}^{n} \left[a_{ji} \cdot \underline{b_i(d_{t+1})} \cdot \delta_t(j) \right]$$

(3) $p(\mathcal{D} | \bar{\theta}) \xrightarrow[\pi, A, B]{} \max$

$$p(\mathcal{D} | \bar{\theta}) = \prod_{s=1}^{S} p(D^{(s)} | \bar{\theta}) = \prod_{s=1}^{S} \sum_{Q} p(D^{(s)}, Q | \bar{\theta}) \xrightarrow{\bar{\theta}} \max$$

$$Q(\bar{\theta}, \bar{\theta}^{(m)}) = E_{Q^{(s)} \sim p(Q^{(s)} | D^{(s)}, \bar{\theta})} \left[\log \prod_{s=1}^{S} \underline{p(D^{(s)}, Q^{(s)} | \bar{\theta})} \right] =$$

$$= E_{Q^{(s)}} \left[\sum_{s=1}^{S} \left(\underline{\log \pi_{q_1^{(s)}}} + \underline{\log b_{q_1^{(s)}}(d_1^{(s)})} + \log a_{q_1^{(s)} q_2^{(s)}} + \ldots \right.\right.$$

$$\left.\left. \ldots + \log a_{q_{T-1}^{(s)} q_T^{(s)}} + \log b_{q_T^{(s)}}(d_T^{(s)}) \right) \right] =$$

$$= \sum_{s=1}^{S} \underline{\sum_{q_1^{(s)} \ldots q_T^{(s)}}} \overline{p(q_1^{(s)} \ldots q_T^{(s)} | \bar{\theta}^{(m)}, D^{(s)})} \left(\boxed{\log \pi_{q_1^{(s)}}} + \ldots + \log b_{q_T^{(s)}}(d_T^{(s)}) \right)$$

$$\log \pi_i : \sum_{s} \sum_{q_2^{(s)} \ldots q_T^{(s)} : q_1^{(s)} = i} p(q_1^{(s)} \ldots q_T^{(s)} | \bar{\theta}^{(m)}, \underline{D^{(s)}}) = \sum_{s} \underbrace{p(q_1^{(s)} = i | D^{(s)}, \theta^{(m)})}_{= \gamma_1^{(s)}(i)}$$

$$\log \pi_i : \quad \sum_{s} \gamma_1^{(s)}(i)$$

$$\log a_{ij}: \sum_s \sum_{Q^{(s)}} p(Q^{(s)} | D^{(s)}, \bar{\theta}^{(m)}) \cdot \sum_{t=1}^{T-1} [q_t^{(s)} = i, q_{t+1}^{(s)} = j]$$

$$\log a_{ij} = \mathbb{E}_{\theta^{(m)}}[\#\text{repeated of } i \,\&\, j] = \sum_s \sum_{t=1}^{T-1} \xi_t^{(s)}(i,j)$$

$$\xi_t(i,j) = p(q_t = i, q_{t+1} = j | \theta, D) =$$

$$= \frac{p(q_t = i, q_{t+1} = j, d_1 \ldots d_t, d_{t+1}, d_{t+2} \ldots d_T | \theta)}{p(D|\theta)} \propto$$

$$\propto p(d_1 \ldots d_t, q_t = i | \bar{\theta}) \cdot p(q_{t+1} = j | q_t = i, \bar{\theta}) \, p(d_{t+1} | q_{t+1} = j, \theta) \cdot p(d_{t+2}, \ldots, d_T | q_{t+1} = j, \theta)$$

$$\boxed{\xi_t(i,j) = \alpha_t(i) \cdot a_{ij} \cdot b_j(d_{t+1}) \cdot \beta_{t+1}(j)}$$

$$\log b_i(k): \mathbb{E}_{\bar{\theta}^{(m)}}[\# \; q_t^{(s)} = i \;\&\; d_t^{(s)} = k] =$$

$$= \sum_s \sum_{t:\, d_t^{(s)} = k} \gamma_t^{(s)}(i)$$

$$\pi_i^{(m+1)} = \frac{\sum_s \gamma_1^{(s)}(i) + 1}{S + n} \qquad a_{ij}^{(m+1)} = \frac{\sum_s \sum_{t=1}^{T-1} \xi_t^{(s)}(i,j) + 1}{\sum_s \sum_{t=1}^{T-1} \gamma_t^{(s)}(i) + n}$$

$$b_i(k)^{(m+1)} = \frac{\sum_s \sum_{t:\, d_t^{(s)} = k} \gamma_t^{(s)}(i) + 1}{\sum_s \sum_t \gamma_t^{(s)}(i) + m}$$

Baum-Welch algorithm