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

болятся → y ; features → $\bar{x}$

выборка: $s=1$, $s=0$

вероятность: $\pi = p(y=1)$

$$\gamma_0 = p(s=1|y=0) = \frac{p(s=1,y=0)}{p(y=0)} = \frac{n_0/N}{(1-\pi)N/N} = \frac{n_0}{(1-\pi)N}$$

$$\gamma_1 = p(s=1|y=1) = n_1/\pi N$$

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

$$p(y=1|\bar{x},s=1) = \frac{p(s=1|y=1,\bar{x})\,p(y=1|\bar{x})}{p(s=1|y=1,\bar{x})p(y=1|\bar{x}) + p(s=1|y=0,\bar{x})p(y=0|\bar{x})} =$$

$$= \frac{\gamma_1\sigma(\eta(\bar{x}))}{\gamma_1\sigma(\eta(\bar{x})) + \gamma_0(1-\sigma(\eta(\bar{x})))} =$$

$$= \frac{\gamma_1/\gamma_0\, e^{\eta(\bar{x})}}{\frac{\gamma_1}{\gamma_0}\cdot e^{\eta(\bar{x})} + \frac{\gamma_0\cdot 1}{\gamma_0}} = \sigma\left(\underline{\log\frac{\gamma_1}{\gamma_0}} + \underline{\eta(\bar{x})}\right)$$

по n_0, n_1 обучаю

$$\sigma(\eta^*(\bar{x})) = p(y=1|\bar{x},s=1)$$

$$\eta(\bar{x}) = \eta^*(\bar{x}) - \log\frac{\gamma_1}{\gamma_0} = \eta^*(\bar{x}) - \log\frac{n_1}{n_0} + \log\frac{\pi}{1-\pi}$$

Presence-only:

n_p - positive $z=1$, $y=1$

n_u - unknown $z=0$, $y=?$

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

$$p(y=0|s=1) = \frac{(1-\pi)n_u}{n_p+n_u}$$

$$\gamma_1 = p(s=1|y=1) = \frac{p(y=1|s=1)\,p(s=1)}{p(y=1)} = \frac{n_p+\pi n_u}{\pi(n_p+n_u)}\cdot p(s=1)$$

$$\gamma_0 = p(s=1|y=0) = \frac{(1-\pi)n_u}{(1-\pi)(n_p+n_u)}\,p(s=1) = \frac{n_u}{n_p+n_u}\,p(s=1)$$

$$\sigma(\eta^*_{naive}(\bar{x})) = p(\underline{z}=1|\bar{x},s=1)$$

$$\eta_{naive}(\bar{x}) = \eta^*_{naive}(\bar{x}) - \log\frac{n_p+\pi n_u}{\pi n_u}$$

$z\ ,\ y$ $\qquad \underline{L(\bar\eta \mid \bar y, \bar z, X)} = \prod_i p(y_i \mid s_i=1, \bar x_i) =$

$$= \prod_i \sigma(\eta^*(\bar x_i))^{y_i}\left(1-\sigma(\eta^*(\bar x_i))\right)^{1-y_i}$$

$$\eta^* = \eta(\bar x_i) + \log\frac{n_p+\pi n_u}{\pi n_u}$$

$$L(\bar\eta \mid \bar z, X) = \prod_i p(z_i \mid s_i=1, \bar x_i) =$$

$$= \prod_i \underline{p(z_i=1 \mid s_i=1, \bar x_i)}^{z_i}\left(1-p(z_i=1 \mid s_i=1, \bar x_i)\right)^{1-z_i} = \circledast$$

$$p(z=1 \mid s=1, \bar x) = \overline{p(z=1 \mid y=1, s=1}\,)\,\overbrace{p(y=1 \mid s=1, \bar x)}^{\sigma(\eta^*(\bar x))} +$$

$$+ \cancel{p(z=1 \mid y=0, s=1, \bar x)}\; \underbrace{p(y=0 \mid s=1, \bar x)}_{1-\eta^*(\bar x)} =$$

$$= \sigma(\eta^*(\bar x)) \cdot \frac{p(z=1, y=1 \mid s=1)}{p(y=1 \mid s=1)} = \sigma(\eta^*(\bar x)) \cdot \frac{n_p / n_p+n_u}{(n_p+\pi n_u)/n} = \frac{n_p}{n_p+\pi n_u}\,\eta^*(\bar x)$$

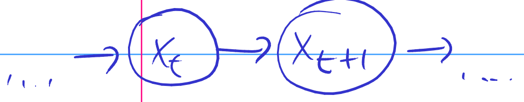
$$\circledast = \prod_i \left(\underbrace{\frac{n_p}{\cancel{n_p+\pi n_u}} \cdot \frac{\cancel{\frac{n_p+\pi n_u}{\pi n_u}} \cdot e^{\eta(\bar x_i)}}{1+\frac{n_p+\pi n_u}{\pi n_u}\cdot e^{\eta(\bar x_i)}}}\right)^{z_i}\left(1-\underline{\qquad}\right)^{1-z_i} \xrightarrow[\bar\eta]{} \max$$

E-war: $\sigma(\eta^{*(t+1)}(\bar x_i)) = E_{(t)}[y_i] = \sigma(\underline{\eta^{(t)}(\bar x_i)})$

$\eta^T \bar x_i \qquad (\pi = p(y=1))$

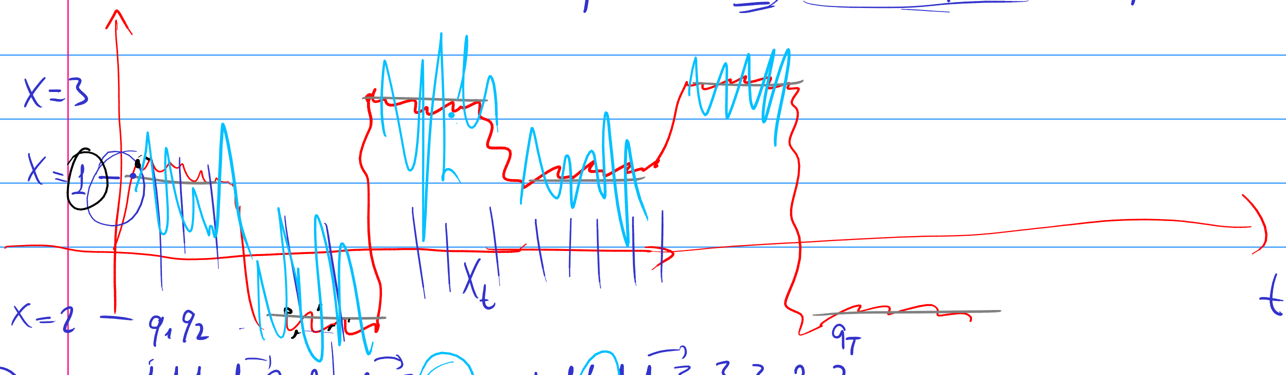
M-war: $\eta^{(t+1)} = \eta^{*(t+1)} - \log\frac{n_p}{n_p+\pi n_u}$

Markov chains



$X_1, X_2, \dots, X_{t-1}, X_t, X_{t+1}, \dots$

$$p(X_{t+1} \mid X_t, X_{t-1}, X_{t-2}, \dots, X_1) = p(X_{t+1} \mid X_t)$$



D = 1 1 1 1 2 2 | 2 3 3 3 1 1 1 1 3 3 3 2 2

$\boxed{a_{ij}} = p(X_t = j \mid X_{t-1} = i)$ $\quad A = (a_{ij})_{i,j=1}^{n}$ $\quad$ n - число сост

$\pi_i = p(X_1 = i)$

$$p(D \mid A, \pi) = \prod_{n=1}^{N} p(X_1 = q_{n1})\, p(X_2 = q_{n2} \mid X_1 = q_{n1})\, p(X_3 = q_3 \mid X_2 = q_2) \ldots p(X_T = q_{nT} \mid X_{T-1} = q_{nT-1})$$

$$= \prod_{n=1}^{N} \pi_{q_{n1}} a_{q_{n1} q_{n2}} a_{q_{n2} q_{n3}} \ldots a_{q_{nT-1}, q_{nT}} \xrightarrow[\pi, A]{} \max$$

D = $q_1 q_2 \ldots q_{T_1}$

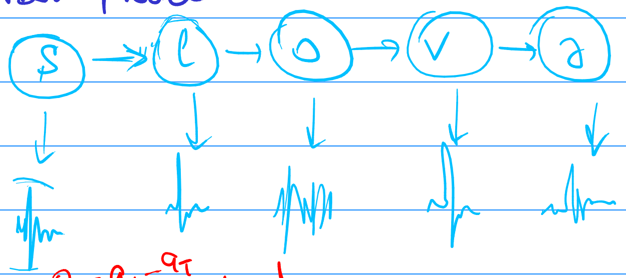
$\ldots$

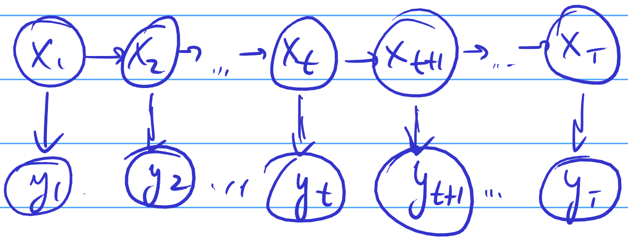
$q_{N1} q_{N2} \ldots q_{NT_N}$

$$\ln p(D \mid A, \pi) = \sum_n \left(\log \pi_{q_{n1}} + \sum_t \log a_{q_{nt}, q_{n,t+1}} \right)$$

$$\pi_i^* = \frac{\#\{q_{n1} = i\}}{N} ; \quad a_{ij}^* = \frac{\#\{q_{nt} = i, q_{n,t+1} = j\}}{\#\{q_{nt} = i\}}$$

Hidden Markov Model





$Q = q_1 \ldots q_T$

$D = d_1 \ldots d_T$

$D = \{d_1 \ldots d_T\}$

$$p(X, y) = p(x_1)\, p(y_1 \mid x_1)\, p(X_2 \mid x_1)\, p(y_2 \mid X_2)\, p(X_3 \mid X_2) \ldots p(X_T \mid X_{T-1})\, p(y_T \mid X_T)$$

$\lambda = (\overset{n\times 1}{\pi}, \overset{n\times n}{A}, \overset{n\times m}{B})$

$\pi_i = p(X_1 = i)$

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

$b_i(k) = p(y_t = v_k \mid X_t = i)$

emission probabilities

$v_1, \ldots, v_m$

$$p(Q, D \mid \lambda) = \pi_{q_1} b_{q_1}(d_1)\, a_{q_1 q_2}\, b_{q_2}(d_2)\, a_{q_2 q_3} - a_{q_{T-1}, q_T} b_{q_T}(d_T)$$

$$b_i^*(v_k) = \frac{\#\{q_{nt} = i, d_{nt} = v_k\}}{\#\{q_{nt} = i\}}$$

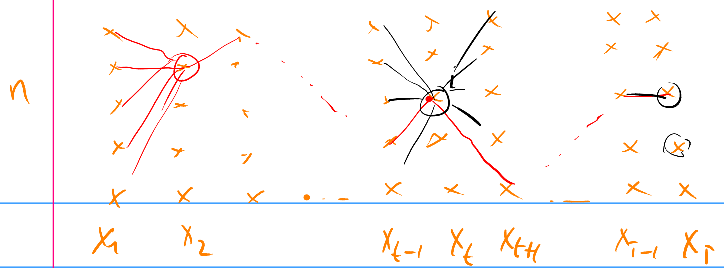
$$p(D \mid \lambda) = \sum_Q p(Q, D \mid \lambda) = \sum_{q_1 \ldots q_T} \left(\pi_{q_1} \ldots b_{q_T}(d_T) \right)$$

① $p(D \mid \lambda) = ?$ $\quad \alpha_t(i)$, $\beta_t(i)$ ✓

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

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

③ $p(D \mid \lambda) \xrightarrow[\pi, A, B]{} \max$



Q – путь слева направо

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

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

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

$$\alpha_1(i) = p(X_1 = i, d_1|\lambda) = \pi_i \cdot b_i(d_1)$$

$$\alpha_{t+1}(i) = \left(\sum_{j=1}^{n} \alpha_t(j) \cdot a_{ji}\right) \underline{b_i(d_t)}$$

$$p(D|\lambda) = \sum_i p(X_T = i, D|\lambda) = \sum_{i=1}^{n} \alpha_T(i)$$

$$\alpha_{t+1}(i) = p(X_{t+1} = i, d_1 \dots d_{t+1} | \lambda) =$$

$$= \sum_{j=1}^{n} p(X_{t+1} = i, X_t = j, d_1 \dots d_{t+1}|\lambda) =$$

$$= \sum_{j=1} p(X_t = j, d_1 \dots d_t|\lambda) p(X_{t+1} = i | X_t = j, \lambda)$$

$$p(y_{t+1} = d_{t+1} | X_{t+1} = i, \lambda) =$$

$$\beta_t(i) = p(d_{t+1}, d_{t+2}, \dots, d_T / X_t = i, \lambda)$$

$$\beta_T(i) = 1$$

$$\beta_t(i) = \sum_{j=1}^{n} \beta_{t+1}(j)\, a_{ij}\, b_{\gamma}(d_{t+1})$$

$$p(D|\lambda) = \sum_i p(X_1 = i, D|\lambda) = \sum_i \pi_i \cdot \beta_1(i)$$

$$\beta_t(i) = \sum_{j=1}^{n} p(X_{t+1} = j, d_{t+1} \dots d_T | X_t = i, \lambda) =$$

$$= \sum_{j=1}^{n} p(d_{t+2} \dots d_T | X_{t+1} = j, \lambda) \cdot$$

$$\cdot p(d_{t+1} | X_{t+1} = j, \lambda)\, p(X_{t+1} = j | X_t = i, \lambda)$$

② $\underline{\gamma_t(i)} = p(X_t = i | D, \lambda)$

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

$$\gamma_t(i) = \frac{p(X_t = i, D|\lambda)}{p(D|\lambda)} = \frac{p(X_t = i, d_1 \dots d_t, d_{t+1} \dots d_T|\lambda)}{p(D|\lambda)}$$

$$= \frac{p(X_t = i, d_1 \dots d_t|\lambda) \cdot p(d_{t+1} \dots d_T | X_t = i, \lambda)}{p(D|\lambda)}$$

$$\underline{Q^*} = \underset{q_1 \dots q_T}{\operatorname{argmax}}\ \overline{p(Q|D,\lambda)}$$

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

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

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

$$= \max_{q_1 \dots q_{t-2}} \max_j \underline{p(q_t = i, q_{t-1} = j, q_{t-2} \dots q_1, d_1 \dots d_t | \lambda)}$$

$$\shortparallel$$

$$p(q_{t-1} = j, q_1 \dots q_{t-2}, d_1 \dots d_{t-1} | \lambda) \cdot$$

$$\cdot p(q_t = i | q_{t-1} = j, \lambda)\, p(d_t | q_t = i, \lambda)$$

Baum-Welch algorithm

③ $p(D|\lambda) \xrightarrow[\lambda]{} \max$

$p(\underline{Q}, D|\lambda) \xrightarrow[\lambda]{} \max$

E-war: $\lambda, D \rightsquigarrow \boxed{\alpha_t(i), \beta_t(i), \gamma_t(i), \xi_t(i,j)}$

$$\underline{\xi_t(i,j)} = p(q_t = i, q_{t+1} = j | D, \lambda) \propto \alpha_t(i) a_{ij} b_j(d_{t+1}) \cdot \beta_{t+1}(j)$$

M-war: $\pi_i^* = \dfrac{\mathbb{E}[\#\{q_{n1} = i\}]}{N} = \dfrac{\sum_n \gamma_{n1}(i)}{N}$

$$a_{ij}^* = \frac{\mathbb{E}[\#\{q_{nt} = i, q_{nt+1} = j\}]}{\mathbb{E}[\#\{q_{nt} = i\}]} = \frac{\sum_n \sum_t \xi_t(i,j)}{\sum_n \sum_t \gamma_t(i)}$$

$$b_j(v_k) = \frac{\mathbb{E}[\#\{q_{nt} = i, d_{nt} = v_k\}]}{\mathbb{E}[\# q_{nt} = i]} = \frac{\sum_{n, t:\, d_{nt} = v_k} \gamma_t(i)}{\sum_n \sum_t \gamma_t(i)}$$

Непрерывные наблюдаемые:

в ⟶ $p(y|\bar{x},\bar{\theta})$

Обучить θ^* по данным с весами $\underline{\gamma_t(i)}$

$b_i(d_t) \to \underline{p(d_t|x=i,\theta)}$

$\boxed{\mu_i}\Sigma_i \Rightarrow \bar{\mu}_i^* = \frac{\sum_{n,t}\gamma_{nt}(i)\,d_{nt}}{\sum_{nt}\gamma_{nt}(i)}$;

Время пребывания

A $\underline{p(\text{в сост. } i \text{ буду } k \text{ шагов}|\lambda)} = \boxed{a_{ii}^{k-1}(1-a_{ii})}$

$p(\ldots k \text{ шагов}) = Gam(k|\ldots)$ ↑

$c_i(k) = p(\text{в сост. } i \text{ остаёмся } k \text{ шагов})$

$\alpha_t^+(i)$, $\alpha_t^-(i)$