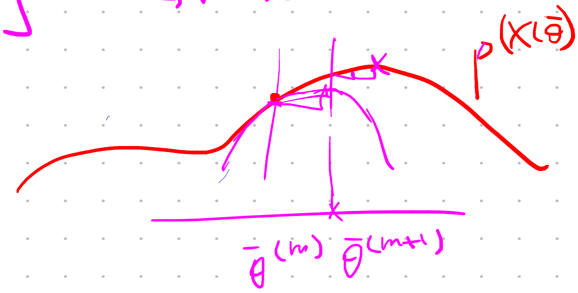
① Расширения и варианты ЕМ-алгоритма

$X = \{\bar{x}_n\}_{n=1}^N \qquad p(X|\bar\theta) \xrightarrow{\bar\theta} \max$ – трудно

$Y = \{(\bar{x}_n, \bar{z}_n)\}_{n=1}^N, \qquad p(Y|\bar\theta) = p(X,Z|\bar\theta) \xrightarrow{\bar\theta} \max$ – легко

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

$$\bar\theta^{(m+1)} = \underset{\bar\theta}{\operatorname{argmax}}\, Q(\bar\theta, \bar\theta^{(m)})$$



(a) Чтобы ЕМ работал, дост., чтобы

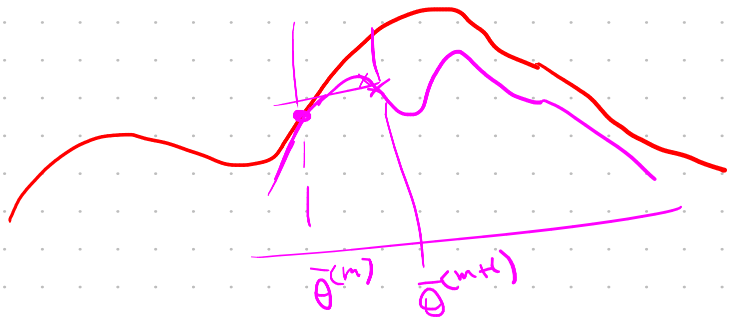
$$p(\bar{x}|\bar{y},\bar\theta) = p(\bar{x}|\bar{y})$$

$$\bar{x} = f(\bar{y})$$

$$\bar\theta \to \bar{y} \to \bar{x}$$

(b) Generalized EM

$$\forall\ \bar\theta^{(m+1)}:\ Q(\bar\theta^{(m+1)}, \bar\theta^{(m)}) > Q(\bar\theta^{(m)}, \bar\theta^{(m)})$$



(c) Randomized versions of EM

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

Stochastic EM

$R=1$ Monte Carlo EM

$$\approx \frac{1}{R}\sum_{r=1}^{R} \log p(X, Z^{(r)}|\bar\theta), \text{ где } Z^{(r)} \sim p(Z|X,\bar\theta^{(m)})$$

$$\mathbb{E}\left[\log p(X,Z|\bar\theta)\right] = \sum_{n=1}^{N} \mathbb{E}\left[\log p(\bar{x}_n,\bar{z}_n|\bar\theta)\right]$$

$\bar{z}_n \sim p(\bar{z}_n|\bar{x}_n, \bar\theta^{(m)})$

(d) Priors

$\bar\theta_{MAP}:\ \log p(\bar\theta|X) \to \max$

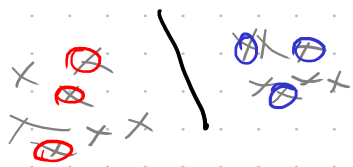
$$= \text{Const} + \log p(\bar\theta) + \log p(X|\bar\theta) \xrightarrow{\bar\theta} \max$$

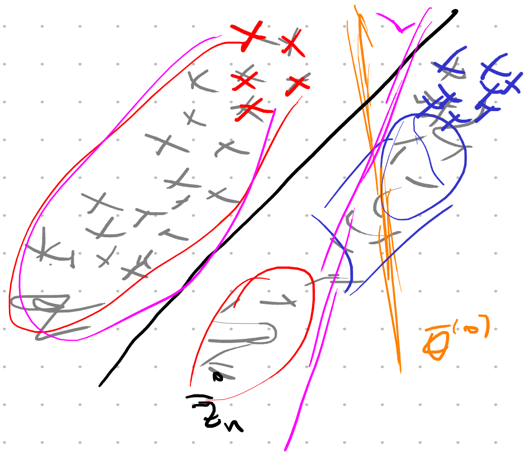
$$L(\bar\theta, \bar\theta^{(m)}) \le \log p(X|\bar\theta)$$

$$L'(\bar\theta, \bar\theta^{(m)}) = L(\bar\theta, \bar\theta^{(m)}) + \log p(\bar\theta) \le \log p(X|\bar\theta) + \log p(\bar\theta)$$

$$Q'(\bar\theta, \bar\theta^{(m)}) = Q(\bar\theta, \bar\theta^{(m)}) + \log p(\bar\theta)$$

(2) Semi-supervised learning





Pseudolabels

(3) Bradley-Terry models

$1,\dots,n$ $\quad i \succ j \quad j \succ i$ $\qquad i \longrightarrow \gamma_i \in \mathbb{R}$

если $\gamma_i > \gamma_j$, то $p(i \succ j) > 1/2$

Предположение:

$$\boxed{p(i \succ j) = \frac{\gamma_i}{\gamma_i + \gamma_j} \qquad p(j \succ i) = \frac{\gamma_j}{\gamma_i + \gamma_j}}$$

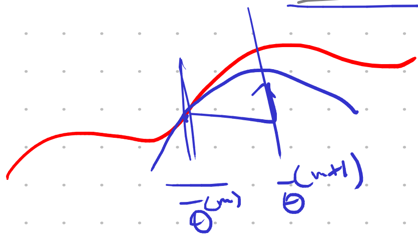
$$D = \{(i,j)\}^{i \succ j} = \{w_{ij}\}_{i,j=1}^{n}, \quad w_{ij} = \#\{i \succ j\}$$

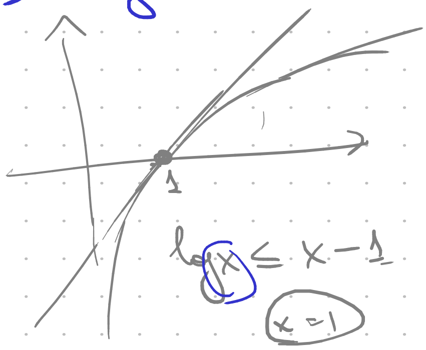
$$p(D|\bar\gamma) = \prod_{(i,j)\in D} \frac{\gamma_i}{\gamma_i+\gamma_j} = \prod_{i,j=1}^{n} \left(\frac{\gamma_i}{\gamma_i+\gamma_j}\right)^{w_{ij}} \xrightarrow[\gamma]{} \max$$

$$\log p(D|\bar\gamma) = \sum_{i,j=1}^{n} w_{ij}\left(\log\gamma_i - \underline{\log(\gamma_i+\gamma_j)}\right) \xrightarrow[\bar\gamma]{} \max$$

MM-algorithm

minorization-maximization





$\log(x) \le x - 1$

$x=1$

$$\log\frac{\gamma_i+\gamma_j}{\gamma_i^{(m)}+\gamma_j^{(m)}} \le \frac{\gamma_i+\gamma_j}{\gamma_i^{(m)}+\gamma_j^{(m)}} - 1 + \log(\text{———})$$

$$L(\bar\gamma, \bar\gamma^{(m)})$$

$$\log p(D|\bar\gamma) \ge \sum_{i,j=1}^{n} w_{ij}\left(\log\gamma_i - \frac{\gamma_i+\gamma_j}{\gamma_i^{(m)}+\gamma_j^{(m)}} + 1 - \log(\gamma_i^{(m)}+\gamma_j^{(m)})\right) \xrightarrow[\bar\gamma]{} \max$$

$$\frac{\partial L}{\partial \gamma_k} = \sum_{i,j} w_{ij}\left(\frac{1}{\gamma_k}\cdot[i=k] - \frac{[i=k \text{ OR } j=k]}{\gamma_i^{(m)}+\gamma_j^{(m)}}\right) = 0$$

$$\sum_j \frac{w_{kj}}{\gamma_k} - \sum_j \frac{w_{kj}}{\gamma_k^{(m)}+\gamma_j^{(m)}} - \sum_i \frac{w_{ik}}{\gamma_i^{(m)}+\gamma_k^{(m)}} = 0$$

#побед k $\qquad$ #партий i, k

$$\frac{\sum_j W_{kj}}{\gamma_k} = \sum_{i=1}^{n} \frac{W_{ki}+W_{ik}}{\gamma_i^{(m)}+\gamma_k^{(m)}}$$

$$\boxed{\gamma_k^{(m+1)} = \frac{Wins(k)}{\sum_{i=1}^{n} \frac{Games(i,k)}{\gamma_i^{(m)}+\gamma_k^{(m)}}}}$$

— добавим ничья $\quad \theta > 1$

$$p(i>j) = \frac{\gamma_i}{\gamma_i+\theta\gamma_j} \qquad p(j>i) = \frac{\gamma_j}{\gamma_j+\theta\gamma_i}$$

$$p(i \approx j) = 1 - \frac{\gamma_i}{\gamma_i+\theta\gamma_j} - \frac{\gamma_j}{\gamma_j+\theta\gamma_i}$$

i, j, k, l

γ_i — γ_l

— белые / чёрные $\quad a>1 \quad p(i>j,\ i \text{ белые}) = \frac{a\cdot\gamma_i}{a\cdot\gamma_i+\gamma_j}$

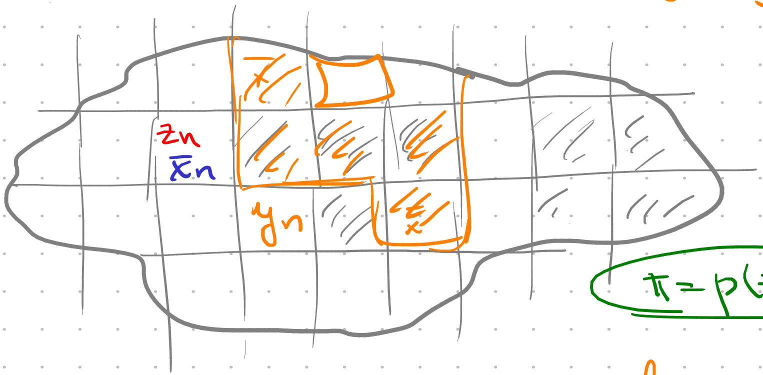
(4) Presence-only data

$z=1$ — суслик водится

$z=0$ — суслик не водится

$\bar{x}$ — признаки $\qquad x \rightsquigarrow z$

$p(z|\bar{x})$



$\pi = p(z=1)$

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

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

$y=1$ — «видели суслика»

$y=0$ — «не видели»

$y=1 \Rightarrow z=1$ ✓

$y=0 \Rightarrow z=?$

$p(z=1|y=0) = \pi$

(5) Prospective vs. retrospective studies

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

Prospective study / cohort study

$$p(d=1|x=0) = \boxed{\frac{\pi_{01}}{\pi_{00}+\pi_{01}}} = \sigma(\eta(\bar{x}))$$

$$p(d=1|x=1) = \frac{\pi_{11}}{\pi_{10}+\pi_{11}} = \sigma(\eta'(\bar{x}))$$

Retrospective study

$s=1$ — «взяли в выборку»

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

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

$$p(s=1) = \frac{\#\text{выб}}{\#\text{поп}}$$

$$\boxed{\frac{\pi_{01}}{\pi_{00}+\pi_{01}}} = p(d=1|x=0,s=1) = \frac{p(d=1|x=0)\,\overbrace{p(s=1|d=1,\cancel{x=0})}^{\pi_1}}{p(d=1|x=0)\,p(s=1|d=1,\cancel{x=0}) + p(d=0|x=0)\,\underbrace{p(s=1|d=0,\cancel{x=0})}_{\pi_0}}$$

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

$$\frac{\pi_{01}}{\pi_{00}+\pi_{01}} = \frac{\pi_1 \cdot p}{\pi_0 + (\pi_1 - \pi_0)\cdot p} \qquad p = \frac{\pi_0 \pi_{01}}{\pi_1(\pi_{00}+\cancel{\pi_{01}}) - \pi_{01}(\cancel{\pi_1} - \pi_0)}$$

$$p = \frac{\pi_0 \pi_{01}}{\pi_1 \cdot \pi_{00} + \pi_0 \cdot \pi_{01}} \qquad \boxed{p = p(d=1|x=0)}$$

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

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

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

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

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

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

$$\underbrace{\overbrace{\eta(\bar x)}^{\bar w^T \bar x}}_{\text{нужно}} = \underbrace{\eta^*(\overbrace{\bar x}^{\bar w^T \bar x})}_{\text{исходная модель легко обучить}} - \boxed{\log\frac{\pi_1}{\pi_0}}$$

$$\cancel{\eta = f(\eta^*)}$$

$$\sigma(\eta^*(\bar x)) \approx p(d=1|x=0, s=1)$$

$$\sigma(\eta(\bar x)) \approx p(d=1|x=0)$$

$$p(d=1|x=0) = \sigma(w_0)$$

$$p(d=1|x=1) = \sigma(w_0 + \boxed{w_1})$$

4 cont'd

y – „выдали" s – „выборка"

z – „бодится"

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

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

Данные: n_p – positive, n_u – unknown, $N = n_p + n_u$

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

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

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

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

$$\approx \sigma(\eta^*(\bar{x}))$$

$$\sigma(\eta_{naive}(\bar{x})) = \frac{n_p}{n_p + \pi n_u} \cdot \sigma\left(\eta(\bar{x}) + \log \frac{\pi_1}{\pi_0}\right)$$

$$= \eta(\bar{x}) + \log \frac{\pi_1}{\pi_0}$$

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

$$\pi_1 = p(s=1 \mid z=1) = \frac{p(s=1)\, p(z=1 \mid s=1)}{p(z=1)} = \frac{n_p + \pi n_u}{N \cdot \pi}\, p(s=1)$$

можем обучить

надо

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

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

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

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

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

$$\sigma(\eta^*(\bar{x}_i))^{[z_i=1]} \left(1 - \sigma(\eta^*(\bar{x}_i))\right)^{[z_i=0]}$$

const

$\bar\eta^{(0)} = \bar\eta_{naive} \longrightarrow \bar\eta^{(1)} \rightsquigarrow \ldots -$

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

$$= \text{Const} + \sum_{i=1}^{N}\left[\mathbb{E}[z_i=1]\log\sigma(\eta(\bar x_i)) + \mathbb{E}[z_i=0]\log(1-\sigma(\eta(\bar x_i)))\right] \to \max_{\eta}$$

$\bar\eta^{(m)}$: если $y_i = 1$, то $\mathbb{E}[z_i=1] = 1$

иначе $\mathbb{E}[z_i=1] = \sigma(\eta^{(m)}(\bar x_i))$

M-step: – оценить $\mathbb{E}[z_i=1]$

– обучить $\bar\eta^{*}$ как лог-регр. $p(z=1)$

– $\bar\eta^{(m+1)} = \bar\eta^{*} - \log\dfrac{n_p + \pi n_u}{\pi n_u}$

⑥ Рейтинг спортивного ЧГК

1) Команд много ~~BT~~

2) Игроки все время переходят

3) Много делена мест ~~TrueSkill~~

(Nikolenko, Sirotkin, 2011) ICML

TrueSkill + +

– #1 — 42 / 48

– #2 — 41

– #3–5 — 40

– #6–14 — 38

	1	2	3	4	–	–	–	–	–	48	Итого
#1	+	+		+	+		+			+	
#2	+		+	+							
⋮	⋮										
#N											

вопрос q — y_{tq} = „ком. t отв. на вопрос q"

команда t

$\parallel$

$(s_1, s_2, \ldots, s_k)$

a_t – skill команды t

c_q – сложность q / „лёгкость"

$y_{tq} \sim \sigma(\mu + a_t + c_q)$

i – skill игрока i

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

z_{iq} = „игрок i отв. на вопрос q"

— Если $y_{tq}=0$, то $z_{iq}=0$ $\forall i\in t$

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

EM:
- init обычной лог-регр., где $z_{iq}=y_{tq}$ $\forall i\in t$

  $\bar\theta^{(0)}=(\bar s^{(0)}, \bar c^{(0)}, \mu^{(0)})$

- repeat:
  - E-step:

$$\mathbb{E}[z_{iq}]=\begin{cases}0, & \text{если } y_{tq}=0,\ i\in t\\ p(z_{iq}=1 \mid \exists j\in t\ z_{jq}=1, \bar\theta^{(m)}), & \text{если } y_{tq}=1\end{cases}$$

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

$$= \frac{\sigma(\mu^{(m)}+s_i^{(m)}+c_q^{(m)})}{1-\prod_{j\in t} p(z_{jq}=0 \mid \bar\theta^{(m)})} = \frac{\sigma(\mu^{(m)}+s_i^{(m)}+c_q^{(m)})}{1-\prod_{j\in t}(1-\sigma(\mu^{(m)}+s_j^{(m)}+c_q^{(m)}))}$$

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

7 Markov chains

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

$p(X_{t+1} \mid X_1 \dots X_t) = p(X_{t+1} \mid X_t)$ — Markov property

$X_t \in \{1, 2, \dots, L\}$

$\pi_i = p(X_1=i)$ $\qquad$ $\bar\theta=(\bar\pi, A)$

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

$D=\{d_1, d_2, \dots, d_N\}_{n=1}^{N}$, $d_n=(d_{n1}, d_{n2}, \dots, d_{nT})$, d_{nt} — знач. X_t в n-м ед. данных

$$p(D \mid \bar\theta) = \prod_{n=1}^{N} p(d_{n1} d_{n2} \dots d_{nT} \mid \bar\pi, A) = \prod_n \pi_{d_{n1}} \underbrace{a_{d_{n1}d_{n2}}} \dots a_{d_{n,T-1} d_{nT}} \to \max_{\bar\pi, A}$$

$$\pi_i^{ML/MAP} = \frac{\#\{X_1=i\}+1}{N+L}, \quad a_{ij}^{ML/MAP} = \frac{\#\{X_t=i, X_{t+1}=j\}+1}{\#\{X_t=i\}+L}$$

Мой дядя самых честных правил, когда не в шутку —

unigram $p(w_{t+1}|w_t)$ backoff

bigram $p(w_{t+2}|w_{t+1}, w_t)$ $p((w_{t+2}, w_{t+1})|(w_{t+1}, w_t))$

n-gram $p(w_{t+n}|w_{t+n-1}, \dots, w_t)$