07.10.16 бммо лекция

VI

$x_1 \dots x_n \sim p(x|\theta)$, $\ln p(x|\theta) \to \max_\theta$

$t_1 \dots t_n$, $p(x, t|\theta)$ экспоненциальный класс

латентные переменные

$\ln p(X, \cancel{T}|\theta)$, $\ln p(X|\theta) \to \max_\theta$

EM - алгоритм

$$\ln p(X|\theta) = \int q(T) \log p(X|\theta) dT = \int q(T) \ln \frac{p(X,T|\theta)}{p(T|X,\theta)} dT =$$

$$= \int q(T) \ln \frac{p(X,T|\theta)}{p(T|X,\theta)} \frac{q(T)}{q(T)} dT = \int q(T) \ln \frac{p(X,T|\theta)}{q(T)} dT +$$

$$+ \int q(T) \ln \frac{q(T)}{p(T|X,\theta)} dT \leftarrow KL(q(T) \| p(T|X,\theta)) \qquad \uparrow \mathcal{L}(q,\theta)$$

$\geq \mathcal{L}(q,\theta)$

$$KL(p\|q) = \int p(x) \ln \frac{p(x)}{q(x)} dx$$

$KL \geq 0$, $= 0 \Leftrightarrow p(x) = q(x)$ п. в.

оптимизация вариационной нижней оценки $\mathcal{L}(q,\theta)$ по q и по θ

E-step: $\mathcal{L}(q,\theta) \to \max_q$, M-step: $\mathcal{L}(q,\theta) \to \max_\theta$

оптимизация в функциональном пр-ве

$$\log p(X|\theta) = \mathcal{L}(q,\theta) + KL(q\|p)$$

$$\mathcal{L}(q|\theta) \to \max_q \Leftrightarrow KL(q(T) \| p(T|X,\theta)) \to \min_q \Leftrightarrow q(T) = p(T|X,\theta)$$

$$\mathcal{L}(q|\theta) = \int q(T) \ln \frac{p(X,T|\theta)}{q(T)} dT \to \max_\theta$$

$$\Leftrightarrow \underbrace{\mathbb{E}_q \underbrace{\ln p(X,T|\theta)}_{\text{вогнутая ф-ия}}}_{\text{выпуклая комбинация вогнутых ф-ий}} \to \max_\theta$$

$\theta x_1 + (1-\theta) x_2$ выпуклая комбинация

$$\sum_{j=1}^{n} \theta_j x_j, \quad \theta_j \geq 0, \quad \sum_{j=1}^{n} \theta_j = 1$$

$$\int \theta(j)\, x(j)\, dj = \mathbb{E}_{\theta(j)} x(j), \quad \theta(j) \geq 0, \quad \int \theta(j)\, dj = 1$$

$$q(T) = p(T \mid X, \theta) = \frac{p(X, T \mid \theta)}{\int p(X, T \mid \theta)\, dT} = \frac{p(X \mid T, \theta)\, p(T \mid \theta)}{\int p(X \mid T, \theta)\, p(T \mid \theta)\, dT}$$

$$KL\left(q(T) \,\|\, p(T \mid X, \theta)\right) \to \min_{\boxed{q \in Q}} \iff q(T) \approx p(T \mid X, \theta)$$

вариационный вывод

$Q = \Delta$ пр-во δ-ф-ий

$$KL\left(\delta(x - x_0) \,\|\, p(x)\right) = \int \delta(x - x_0) \ln \frac{\delta(x - x_0)}{p(x)}\, dx =$$

$$= -\int \delta(x - x_0) \ln p(x)\, dx + \underbrace{\int \delta(x - x_0) \ln \delta(x - x_0)\, dx}_{const} =$$

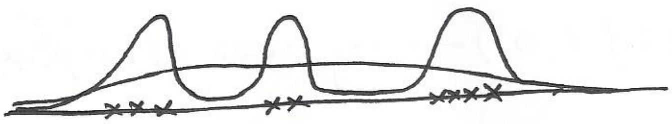
$$= const - \ln p(x_0) \to \min_{x_0}, \quad x_0 = \arg\max_x p(x)$$

$$\mathrm{mod}\, p(x) = p(x_0)$$

$$T_{MP} = \arg\max_T p(T \mid X, \theta) = \arg\max_T p(X, T \mid \theta)$$

$$\mathbb{E}_q \ln p(X, T \mid \theta) = \mathbb{E}_{\delta(T - T_{MP})} \ln p(X, T \mid \theta) = \ln p(X, T_{MP} \mid \theta)$$

жёсткий EM-алгоритм.



$$p(x \mid \theta) \to p(x, t \mid \theta) = p(x \mid t, \theta)\, p(t \mid \theta) =$$

$$t \in \{1 \ldots K\}$$

$$= \mathcal{N}(x \mid \mu_t, \sigma_t^2) \cdot \prod_{k=1}^{K} \pi_k^{[t=k]}$$

$$\pi_k > 0, \quad \sum \pi_k = 1$$

$$\Theta = (\{\mu_k\}, \{\sigma_k^2\}, \pi_k)$$

-step: $q(T) = p(T \mid X, \theta) = \dfrac{p(X, T \mid \theta)}{\sum_T p(X, T \mid \theta)} =$

$$= \frac{\prod_{i=1}^{n} p(x_i, t_i \mid \theta)}{\sum_T \prod_{i=1}^{n} p(x_i, t_i \mid \theta)} = \frac{\prod_{i=1}^{n} p(x_i, t_i \mid \theta)}{\prod_{i=1}^{n} \sum_{t_i=1}^{K} p(x_i, t_i \mid \theta)} = \prod_{i=1}^{n} \frac{p(x_i, t_i \mid \theta)}{\sum_{t_i=1}^{K} p(x_i, t_i \mid \theta)} =$$

$$= \prod_{i=1}^{n} q(t_i)$$

$$q(t_i = l) = \frac{p(x_i, l \mid \theta)}{\sum_{t_i=1}^{K} p(x_i, t_i \mid \theta)} = \frac{N(x_i \mid \mu_l, \sigma^2_l) \cdot \pi_l}{\sum_{k=1}^{K} N(x_i \mid \mu_k, \sigma^2_k) \pi_k} = \gamma_{il} =$$

$$= p(t_i = l \mid x_i, \theta)$$

M-шаг

$$\mathbb{E}_q \ln p(X, T \mid \theta) = \mathbb{E}_q \sum_{i=1}^{n} \ln p(x_i, t_i \mid \theta) = \sum_{i=1}^{n} \mathbb{E}_{q(t_i)} \ln p(x_i, t_i \mid \theta) =$$

$$= \sum_{i=1}^{n} \mathbb{E}_{q(t_i)} \left(\ln N(x_i \mid \mu_{t_i}, \sigma^2_{t_i}) + \sum_{k=1}^{K} [t_i = k] \ln \pi_k \right) =$$

$$= \sum_{i=1}^{n} \sum_{k=1}^{K} \left(\gamma_{ik} \ln N(x_i \mid \mu_{t_i}, \sigma^2_{t_i}) + \gamma_{ik} \ln \pi_k \right) \to \max_{\mu, \sigma^2, \pi}$$

PCA

$$p(x, t \mid \theta) = p(x \mid t, \theta) p(t) = N(x \mid \mu + wt, \sigma^2 I) N(t \mid 0, I)$$

$x \in \mathbb{R}^D$, $t \in \mathbb{R}^d$, $X = (x_1 \dots x_n)$, $p(X \mid \theta) \to \max_\theta$

$$p(X \mid \theta) = \prod_{i=1}^{n} p(x_i \mid \theta) = \prod_{i=1}^{n} \int p(x_i, t_i \mid \theta)\, dt_i = \prod_{i=1}^{n} \int p(x_i \mid t_i, \theta) \cdot$$

$$\cdot p(t_i)\, dt_i$$

E-step: $p(T \mid X, \theta) = \dfrac{\prod_{i=1}^{n} p(x_i \mid t_i, \theta) p(t_i)}{\int \prod_{i=1}^{n} p(x_i \mid t_i, \theta) p(t_i)\, dt_i} = \prod_{i=1}^{n} \dfrac{p(x_i \mid t_i, \theta) p(t_i)}{\int p(x_i \mid t_i, \theta) p(t_i)\, dt_i}$

M-step: $\mathbb{E}_{p(T \mid X, \theta)} \ln p(X, T \mid \theta) \to \max_\theta$

$$\ominus \prod_{i=1}^{n} \frac{N(x_i | \mu + w t_i, \sigma^2 I)\, N(t_i | 0, I)}{\int N(x_i | \mu + w t_i, \sigma^2 I)\, N(t_i | 0, I)\, dt_i} =$$

$$= \prod_{i=1}^{n} \frac{N(x_i | \mu + w t_i, \sigma^2 I)\, N(t_i | 0, I)}{N(x_i | \mu, \sigma^2 I + w w^T)} =$$

$$= \prod_{i=1}^{n} N\left(t_i \,|\, (\sigma^2 I + w^T w)^{-1} w^T (x_i - \mu), (I + \sigma^{-2} w^T w)^{-1}\right)$$

M-step

$$\mathbb{E}_{T|x,\Theta}\, p(x, T | \Theta) = \sum_{i=1}^{n} \mathbb{E}_{t_i | x_i, \Theta} \left(\ln p(x_i | t_i, \Theta) + \ln p(t_i)\right) =$$

$$= \sum_{i=1}^{n} \mathbb{E}\Big(-\frac{D}{2} \ln 2\pi - D \ln \sigma - \frac{1}{2\sigma^2} (x_i - \mu - w t_i)^T (x_i - \mu - w t_i) -$$

$$- \frac{d}{2} \ln 2\pi - \frac{1}{2} t_i^T t_i\Big) = \sum_{i=1}^{n} \Big(- D \ln \sigma - \frac{1}{2\sigma^2} (x - \mu)^T (x - \mu) +$$

$$+ \frac{1}{\sigma^2} (x_i - \mu)^T w\, \mathbb{E} t_i - \frac{1}{2\sigma^2} \underbrace{\mathbb{E}\, t_i^T w^T w t_i}_{\mathrm{tr}\, w^T w\, \mathbb{E} t_i t_i^T} \Big) + const \;\#$$

$$\# \; \frac{\partial}{\partial w} \mathbb{E} \ln p(x, T | \Theta) = \sum_{i=1}^{n} \left(\frac{1}{\cancel{\sigma^2}} (x_i - \mu)(\mathbb{E} t_i)^T - \frac{1}{\cancel{\sigma^2}} w\, \mathbb{E} t_i t_i^T\right) = 0$$

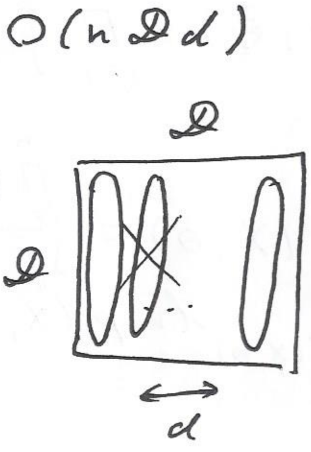
$$\sum_{i=1}^{n} w\, \mathbb{E} t_i t_i^T = \sum_{i=1}^{n} (x_i - \mu)(\mathbb{E} t_i)^T$$

$$w = \left(\sum_{i=1}^{n} (x_i - \mu)(\mathbb{E} t_i)^T\right) \left(\sum_{i=1}^{n} \mathbb{E} t_i t_i^T\right)^{-1}$$

PCA $O(nD^2 + D^3) \sim O(nD^2) \Rightarrow O(nDd)$

$t \in \mathbb{R}^D$, $w \in \mathbb{R}^{D \times D}$

$$p(w | \alpha) = \prod_{i=1}^{D} N(w_i | 0, \alpha_i^{-1} I)$$



Эффект АМД

07.10.16 бммо сем

$$p(x) = \sum_{k=1}^{K} \pi_k p_k(x|\Theta), \quad \sum_{i=1}^{K} \pi_i = 1, \quad \pi_i \geq 0 \ \forall i$$

$$p(x, t) = \prod_{k=1}^{K} [\pi_k p_k(x|\Theta)]^{t_k}, \quad t \in \{0,1\}^k, \ \sum_{i=1}^{K} t_i = 1$$

$$x_1 \dots x_n \sim p(x)$$

ммп : $\pi_{мл}, \Theta_{мл}$?

$$p(X, T | \pi, \Theta) = \cancel{\prod_{n=1}^{n} \prod_{k=1}^{K}} \prod_{n} \prod_{k} [\pi_k p_k(x_n|\Theta)]^{t_{nk}}$$

E-step: $\cancel{p}\ q(T) = p(T | X, \pi, \Theta) = \prod_n q_n(t_n)$

$$q_n(t_n) \propto \prod_k [\pi_k p_k(x_n|\Theta)]^{t_{nk}}$$

$$q_n(t_{nk}=1) = \frac{\pi_k p_k(x_n|\Theta)}{\sum_k \pi_k p_k(x_n|\Theta)} = \gamma_{nk}$$

M-step: $\mathbb{E}_q \ln p(x, T | \pi, \Theta) = \mathbb{E}_q \sum_n \sum_k t_{nk}(\ln p_k(x_n|\Theta) + \ln \pi_k)$

$$= \sum_{n,k} (\mathbb{E}_q t_{nk})(\ln \pi_k + \ln p_k(x_n|\Theta)) =$$

$$= \sum_{n,k} \gamma_{nk} (\ln \pi_k + \ln p_k(x_n|\Theta)) \longrightarrow \max_{\pi, \Theta}$$

$$\pi_k = \frac{\sum_n \gamma_{nk}}{N}$$

\# $x_1 \dots x_n \sim \gamma p_1(x) + (1-\gamma) p_2(x)$

	1	2	3
p_1	α	$(1-\alpha)$	0
p_2	0	$(1-\beta)$	β

$\alpha_0 = \beta_0 = \gamma_0 = \frac{1}{2}$ $\quad N_1 = 30, \ N_2 = 20, \ N_3 = 60$

γ, α, β ?

$t \in [x \in p_2]$, $\Theta = (\alpha, \beta, \gamma)$

$$p(x, t | \Theta) = \gamma^{1-t} p_1^{1-t}(x)(1-\gamma)^t p_2^t(x) = \gamma p_1(x) \left(\frac{p_2(x)(1-\gamma)}{p_1(x)\gamma}\right)^t$$

$$p(X, T | \Theta) = \prod_n p(x_n, t_n | \Theta) = \prod_n \gamma p_1(x_n) \left(\frac{p_2(x_n)(1-\gamma)}{p_1(x_n)\gamma}\right)^{t_n}$$

$$\gamma_{n1} = \frac{\gamma p_1(x_n|\theta)}{\gamma p_1(x_n|\theta)+(1-\gamma)p_2(x_n|\theta)} = q_n(t_n=0)$$

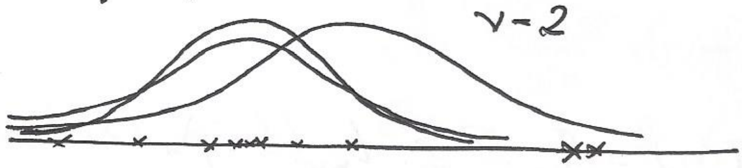
$$q_n(t_n=0)\Big|_{x_n=1} = \frac{\gamma d}{\gamma \alpha} = 1$$

$$\mathbb{E}_q \ln p(X,T|\theta) = \mathbb{E}_q \sum_n \ln\gamma p_1(x_n) + t_n(\ln p_2(x_n) \cancel{1-\gamma}) +$$
$$+\ln(1-\gamma) - \ln p_1(x_n) - \ln\gamma) = \sum_n(\mathbb{E}_q t_n(\ln p_2(x_n)+$$
$$+\ln(1-\gamma)-\ln p_1(x_n)-\ln\gamma)+\ln\gamma p_1(x_n)) \to \max_{\alpha,\beta,\gamma}$$

Распределение Стьюдента 

$x\in\mathbb{R}$, $\mathcal{T}(x|\mu,\sigma^2,\nu) \propto \left(1+\frac{1}{\nu}\left(\frac{x-\mu}{\sigma}\right)^2\right)^{-\frac{\nu+1}{2}}$

$\mathbb{E}x=\mu$, $\mathbb{D}x = \sigma^2\frac{\nu}{\nu-2}$



устойчивость Стьюдента.

$x\in\mathbb{R}^D$, $\mathcal{T}\propto\left(1+\frac{1}{\nu}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)^{-\frac{\nu+D}{2}}$

$$\mathcal{T}(x|\mu,\Sigma,\nu) = \int_0^\infty \mathcal{N}(x|\mu,\tfrac{1}{z}\Sigma)\, G(z|\tfrac{\nu}{2},\tfrac{\nu}{2})\,dz,\ z\in\mathbb{R}^+$$

$x_1\dots x_N \sim \mathcal{T}(x|\mu,\Sigma,\nu)$ $\left\{G(x|a,b)=\frac{b^a}{\Gamma(a)}x^{a-1}\exp(-bx)\right\}$

$$p(x,z|\mu,\Sigma,\nu)=\mathcal{N}(x|\mu,\tfrac1z\Sigma)\,G(z|\tfrac\nu2,\tfrac\nu2)$$

E: $q(z)=p(z|X,\theta)=\prod_n q_n(z_n)$

$$q_n(z_n)=\mathcal{N}(x_n|\mu,\tfrac{1}{z_n}\Sigma)\,G(z_n|\tfrac\nu2,\tfrac\nu2) \cancel{\propto} =$$

$$= \frac{1}{(2\pi)^{\frac{D}{2}}\det^{\frac12}\frac{1}{z_n}\Sigma}\exp\left(-\frac12(x_n-\mu)^T\left(\frac{1}{z_n}\Sigma\right)^{-1}(x_n-\mu)\right)\cdot$$

$$\cdot\frac{(\frac\nu2)^{\frac\nu2}}{\Gamma(\frac\nu2)}\, z_n^{\frac\nu2-1}\exp\left(-\frac\nu2 z_n\right) =$$

$$\frac{z_n^{\frac{D}{2}+\frac{\nu}{2}-1}}{(2\pi)^{\frac{N}{2}}\,\frac{1}{z_n}\det^{\frac{1}{2}}\Sigma}\,\frac{\left(\frac{\nu}{2}\right)^{\frac{\nu}{2}}}{\Gamma\left(\frac{\nu}{2}\right)}\cdot\exp\left(-\frac{z_n}{2}\left(\nu+(x_n-\mu)^T\Sigma^{-1}(x_n-\mu)\right)\right)=$$

$$=G\left(z_n \,\middle|\, \frac{D+\nu}{2},\ \frac{1}{2}\left(\nu+(x_n-\mu)^T\Sigma^{-1}(x_n-\mu)\right)\right)$$

M: $\mathbb{E}_q \ln p(x,z|\Theta)=\mathbb{E}_q\sum_n\Big(\frac{D}{2}\ln z_n-\frac{1}{2}\ln\det\Sigma-$

$-\frac{z_n}{2}(x_n-\mu)^T\Sigma^{-1}(x_n-\mu)+\left(\frac{\nu}{2}-1\right)\ln z_n-\frac{\nu z_n}{2}+$

$+\frac{\nu}{2}\ln\frac{\nu}{2}-\ln\Gamma\left(\frac{\nu}{2}\right)\Big)$

$$\frac{\partial}{\partial\mu}\sum_n\mathbb{E}_q\left(-\frac{z_n}{2}(x_n-\mu)^T\Sigma^{-1}(x_n-\mu)\right)=-\sum_n\mathbb{E}_q z_n\Sigma^{-1}(x_n-\mu)=0$$

$$-\sum_n\mathbb{E}_q z_n\Sigma^{-1}x_n+\sum_n\mathbb{E}_q z_n\Sigma^{-1}\mu=0,\quad \mu=\frac{\sum_n x_n\mathbb{E}_q z_n}{\sum_n\mathbb{E}_q z_n}$$

13.10.16

$$p(X_k|d_k,\Theta)=\prod_{ij}\begin{cases}N(X_k(i,j)\,|\,F(i-d_k^h,\ j-d_k^w),\ s^2) & \text{если } (i,j)\in faceArea(d_k)\\ N(X_k(i,j)\,|\,B(i,j),\ s^2), & \text{иначе}\end{cases}$$

$\Theta=\{B,F,s^2\}$, $faceArea=\{(i,j)\,|\,d_k^h\le i\le d_k^h+h-1,\ d_k^w\le j\le d_k^w+w-1\}$

$X_k(i,j)$ – пиксель k-го изображения

$B\in\mathbb{R}^{H\times W}$ – маска чистого фона без лица преступника

$B(i,j)$ – пиксель маски

$F\in\mathbb{R}^{h\times w}$ – маска лица преступника, $F(i,j)$ – пиксель маски

$d_k=(d_k^h, d_k^w)$ – координаты верхнего левого угла маски лица на k-ом изображении, $d=(d_1\dots d_K)$ для выборки.

Распределение на неизвестные координаты лица:

$A\in\mathbb{R}^{H-h+1,\ W-w+1}$, $p(d_k|A)=A(d_k^h,d_k^w)$

$$\sum_{ij}A(i,j)=1$$

$$p(X, d \mid \Theta, A) = \prod_k p(X_k \mid d_k, \Theta)\, p(d_k \mid A)$$

$$p(X \mid \Theta, A) \to \max_{\Theta, A}$$

$$\mathcal{L}(q, \Theta, A) = \mathbb{E}_{q(d)} \ln p(X, d \mid \Theta, A) - \mathbb{E}_{q(d)} \ln q(d) \to \max_{q, \Theta, A}$$

$$E: \quad q(d) = p(d \mid X, \Theta, A) = \prod_k p(d_k \mid X_k, \Theta, A)$$

$$M: \quad \mathbb{E}_{q(d)} \ln p(X, d \mid \Theta, A) =$$

$$= \mathbb{E}_{q(d_k)} \sum_k \ln p(X_k, d_k \mid \Theta, A) =$$

$$= \sum_k \mathbb{E}_{q(d_k)} \ln p(X_k, d_k \mid \Theta, A) \to \max_{\Theta, A}$$

1) $$p(d_k \mid X_k, \Theta, A) = \frac{p(X_k \mid d_k, \Theta, A) \cdot p(d_k \mid \Theta, A)}{p(X_k \mid \Theta, A)} =$$

$$A(d_k^h, d_k^w) \quad \frac{p(X_k \mid d_k, \Theta, A) \cdot p(d_k \mid \Theta, A)}{\sum_k p(X_k, d_k \mid \Theta, A)}$$

$$q(d_k) = p(d_k \mid X_k, \Theta, A) = \frac{p(d_k, X_k \mid \Theta, A)}{p(X_k \mid \Theta, A)} = \frac{p(X_k \mid d_k, \Theta)\, p(d_k \mid A)}{p(X_k \mid \Theta, A)} =$$

$$= \frac{p(X_k \mid d_k, \Theta)\, p(d_k \mid A)}{\sum_{i,j} p(X_k \mid d_k, \Theta)\, p(d_k \mid A)} \qquad \begin{array}{l} 0 \le i \le H-h \\ 0 \le j \le W-w \end{array}$$

~~$\{i, j \in$ face Area$(d_k)\}$~~

2) $$\mathbb{E}_{q(d_k)} \ln p(X_k, d_k \mid \Theta, A) \to \max_{\Theta, A}$$

$$\mathbb{E}_{q(d_k)} (\ln p(X_k \mid d_k, \Theta) + \ln p(d_k \mid A)) =$$

$$= \mathbb{E}_{q(d_k)} \ln p(X_k \mid d_k, \Theta) + \mathbb{E}_{q(d_k)} \ln p(d_k \mid A)$$

$$\mathbb{E}_{q(d_k)} \ln p(X_k \mid d_k, \Theta) \to \max_{\Theta} \qquad \mathbb{E}_{q(d_k)} \ln A(d_k^h, d_k^w) \to \max_A$$

$$\mathbb{E}_{q(d_k)} \ln p(d_k \mid A) \to \max_A \qquad \sum_{i,j} q(d_k) \ln A(d_k^h, d_k^w) \to \max_A$$