/ 02.11.18 dl

$p(x, y)$

Правила работы с вер-ю

① Правило суммы: $\int p(a, b)db = p(a)$

② Правило произв.: $p(a, b) = p(a|b)p(b)$

$$p(x, y, z), \quad p(y|x) = \frac{p(x,y)}{p(x)} = \frac{\int p(x,y,z)dz}{\int p(x,\tilde{y},\tilde{z})d\tilde{y}d\tilde{z}}$$

ML: $p(x|\theta) \to \max_\theta$

Bayes: $p(\theta)$ - априорное распр-е

$$p(x|\theta), \quad p(\theta|x) = \frac{p(x|\theta)p(\theta)}{\int p(x|\theta)p(\theta)d\theta}$$

$$\theta_{bayes} = \arg\min_{\theta'} \mathbb{E}_{p(\theta|x)} L(\theta, \theta')$$

$L(\theta, \theta') = [\theta \neq \theta']$, $\theta_{bayes} = \theta_{MAP} = \arg\max \theta(x)$

$L(\theta, \theta') = \|\theta - \theta'\|^2$, $\theta_{bayes} = \mathbb{E}_{p(\theta|x)} \theta$

$p(x, z|\theta)$, $p(\theta)$

$$p(\theta|x) = \frac{p(x,\theta)}{p(x)} = \frac{\int p(x,z|\theta)dz \, p(\theta)}{p(x)}$$

Вер. модель регрессии

$\{x_i, y_i\}_{i=1}^N$, $x_i \in \mathbb{R}^D$, $y_i \in \mathbb{R}$, $\in \{-1, +1\}$

$y(x) = w^T x$ - регр.

$= \operatorname{sign}(w^T x)$ - класс.

$$p(y_i \mid x_i, w) = \mathcal{N}(y_i \mid w^T x_i, \beta^2)$$

$$= \sigma(y_i w^T x_i)$$

$$p(w \mid \alpha) = \mathcal{N}(w \mid 0, \alpha^2 I) \quad \text{prior}$$

$$p(y, w \mid x, \alpha, \beta) = \left[\prod_{i=1}^{N} p(y_i \mid x_i, w, \beta)\right] p(w \mid \alpha)$$

$$p(w \mid y, x, \alpha, \beta) \to \max_w \iff$$

$$p(y \mid x, w, \beta)\, p(w \mid \alpha) \to \max_w \iff$$

$$-\sum_{i=1}^{N} \log p(y_i \mid x_i, w, \beta) - \log p(w \mid \alpha) \to \min_w$$

регр. $\frac{1}{2\beta^2} \sum_{i=1}^{N} (y_i - w^T x_i)^2 + \frac{1}{2\alpha^2} \|w\|^2 \to \min_w$

$$\underbrace{p(y \mid x, \alpha, \beta)}_{} \to \max_{\alpha, \beta}$$

неполное правдоподобие, evidence

$$= \int p(y, w \mid x, \alpha, \beta)\, dw$$

Байесовское ансамблирование

$$p(y_{test} \mid x_{test}, y, x, \alpha, \beta) =$$

$$= \int p(y_{test} \mid x_{test}, w, \beta)\, p(w \mid y, x, \alpha, \beta)\, dw$$

EM - алгоритм

$$p(x, z \mid \theta)$$

$$p(x \mid \theta) \to \max$$

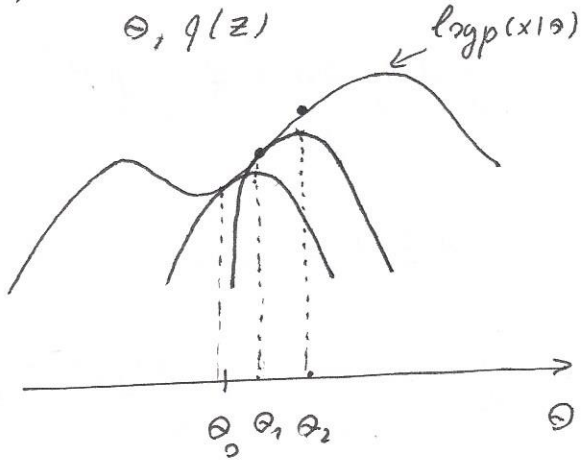
$$= \int p(x, z \mid \theta)\, dz$$

$$\log p(x \mid \theta) \overset{\forall q(z)}{=} \int \log p(x \mid \theta)\, q(z)\, dz = \int \log \frac{p(x, z \mid \theta)}{p(z \mid x, \theta)}\, q(z)\, dz =$$

$$= \int q(z) \log \frac{p(x, z \mid \theta)\, q(z)}{p(z \mid x, \theta)\, q(z)}\, dz =$$

$$= \int q(z)\log\frac{p(x,z|\theta)}{q(z)}\,dz + \underbrace{\int q(z)\log\frac{q(z)}{p(z|x,\theta)}\,dz}_{KL(q(z)\,\|\,p(z|x,\theta))} \geq$$

$$\geq \mathbb{E}_{q(z)}\log p(x,z|\theta) - \mathbb{E}_{q(z)}\log q(z) \to \max_{\theta,\, q(z)}$$

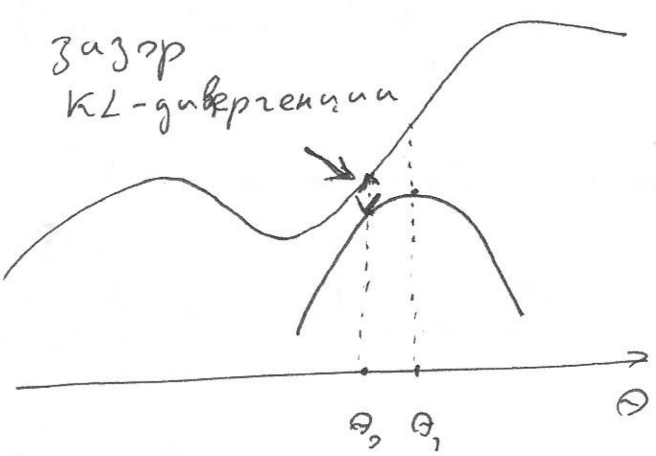


E-шаг: $\quad ELBO(q,\theta) \to \max_q$

$$q(z) = p(z|x,\theta)$$

M-шаг: $\quad ELBO(q,\theta) \to \max_\theta$

$$\mathbb{E}_{q(z)}\log p(x,z|\theta) \to \max_\theta$$



Relevance Vector Regression

$\{x_i, y_i\}_{i=1}^N, \; x_i \in \mathbb{R}^D, \; y_i \in \mathbb{R}$

$$y(x) = w^T x$$

$$p(y_i|x_i,w,\beta) = \mathcal{N}(y_i|x_i^T w, \beta^2)$$

$$p(w|\alpha) = \mathcal{N}(w|0, \mathrm{diag}(\alpha^2))$$

$$p(y|X,\alpha,\beta) \to \max_{\alpha,\beta}, \quad \alpha_j \to 0 \Rightarrow w_{MAP,j} \to 0$$

$$\log p(y|X,\alpha,\beta) \geq \mathbb{E}_{q(w)}\log p(y,w|X,\alpha,\beta) - \mathbb{E}_{q(w)}\log q(w) =$$

$$= \sum_{i=1}^N \mathbb{E}_{q(w)}\log p(y_i|x_i,w,\beta) + \underbrace{\mathbb{E}_{q(w)}\log p(w|\alpha) - \mathbb{E}_{q(w)}\log q(w)}_{-KL(q(w)\,\|\,p(w|\alpha))} =$$

$$= \{q(w) = \mathcal{N}(w|\mu,\Sigma)\} = \sum_{i=1}^N \mathbb{E}_{\mathcal{N}(w|\mu,\Sigma)}\log \mathcal{N}(y_i|x_i^T w, \beta^2) -$$

$$- KL(\mathcal{N}(w|\mu,\Sigma)\,\|\,\mathcal{N}(w|0,\mathrm{diag}(\alpha^2))) \to \max_{\alpha,\beta,\mu,\Sigma}$$

$$\mathbb{E}_{N(w|\mu,\Sigma)} \log N(y_i | x_i^T w, \beta^2) =$$

$$= \mathbb{E}_{N(w|\mu,\Sigma)} \left[-\frac{1}{2}\log 2\pi - \frac{1}{2}\log\beta^2 - \frac{1}{2\beta^2}(y_i - w^T x_i)^2 \right] =$$

$$= -\frac{1}{2}\log 2\pi - \frac{1}{2}\log\beta^2 - \frac{1}{2\beta^2}\Big(y_i^2 - 2y_i \underbrace{\mathbb{E} w^T x_i}_{\mu^T x_i} + \underbrace{\mathbb{E} x_i^T w w^T x_i}_{x_i^T(\mathbb{E} w w^T) x_i}\Big) \circeq$$

$$// \Sigma = \mathbb{E}(w-\mu)(w-\mu)^T \Rightarrow \mathbb{E} w w^T = \Sigma + \mu\mu^T$$

$$\circeq -\frac{1}{2}\log 2\pi - \frac{1}{2}\log\beta^2 - \frac{1}{2\beta^2}(y_i^2 - 2y_i\mu^T x_i + x_i^T \mu\mu^T x_i + x_i^T \Sigma x_i) =$$

$$= \log N(y_i | \mu^T x_i, \beta^2) + x_i^T \Sigma x_i$$

$$KL(N(w|\mu,\Sigma) \| N(w|0, \mathrm{diag}(\alpha^2))) =$$

$$= \mathbb{E}_{N(w|\mu,\Sigma)} \log \frac{N(w|\mu,\Sigma)}{N(w|0,\mathrm{diag}(\alpha^2))} =$$

$$= \mathbb{E}_{N(w|\mu,\Sigma)} \left[-\frac{D}{2}\cancel{\log 2\pi} - \frac{1}{2}\log\det\Sigma - \frac{1}{2}(w-\mu)^T \Sigma^{-1}(w-\mu) + \right.$$

$$\left. + \cancel{\log 2\pi}\,\frac{D}{2} + \frac{1}{2}\sum_i \log\alpha_i^2 + \frac{1}{2}\sum_i \frac{w_i^2}{\alpha_i^2} \right] \circeq$$

$$// \mathbb{E}_{N(w|\mu,\Sigma)} \check{tr}\,(w-\mu)^T\Sigma^{-1}(w-\mu) = \mathbb{E}_{N(w|\mu,\Sigma)} tr\,\Sigma^{-1}(w-\mu)(w-\mu)^T =$$

$$= tr\,\Sigma^{-1}\mathbb{E}(w-\mu)(w-\mu)^T = tr\, I = D$$

$$// \mathbb{E}_{N(w|\mu,\Sigma)} w_i^2 = \Sigma_{ii} + \mu_i^2$$

$$\circeq -\frac{1}{2}\log\det\Sigma - \frac{D}{2} + \frac{1}{2}\sum_i \log\alpha_i^2 + \frac{1}{2}\sum_i \frac{\Sigma_{ii} + \mu_i^2}{\alpha_i^2} \to \min_{\alpha_i}$$

$$\frac{\partial}{\partial\alpha_i^2} = \frac{1}{2\alpha_i^2} - \frac{\Sigma_{ii} + \mu_i^2}{2\alpha_i^4}, \quad \alpha_i^2 = \Sigma_{ii} + \mu_i^2$$

$$- \frac{1}{2} \log \det \Sigma - \frac{D}{2} + \frac{1}{2} \sum_i \log(\Sigma_{ii} + \mu_i^2) +$$

$$+ \frac{1}{2} \sum_i \cancel{\frac{\Sigma_{ii} + \mu_i^2}{\Sigma_{ii} + \mu_i^2}} =$$

1

ЗАДАЧИ
02.11.18 dl

$$= - \frac{1}{2} \log \det \Sigma + \frac{1}{2} \sum_i \log(\Sigma_{ii} + \mu_i^2)$$

$$X = \{x_1 \dots x_N\} \quad \text{i.i.d}$$

$$N_k = \sum_{n=1}^{N} \mathbb{I}[x_n = k], \quad$$ семинар

$$p(x|\theta) = \prod_{k=1}^{K} \theta_k^{N_k}$$

$$\text{s.t.} \quad \sum_k \theta_k = 1, \quad \sum_k N_k = N$$

$$p(x|\theta) = \prod_k \theta_k^{N_k}$$

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

$$p(\theta) = \mathcal{D}ir(\theta|\alpha) = \frac{\mathbb{I}[\theta \in S_k]}{B(\alpha_1 \dots \alpha_k)} \prod_k \theta_k^{\alpha_k - 1}$$

$$\alpha_k > 0, \quad B(\alpha_1 \dots \alpha_k) = \int_{S_k} \prod_k \theta_k^{\alpha_k - 1} d\theta$$

5

$$p(x) = \int p(x|\theta)\, p(\theta)\, d\theta =$$

$$= \int \prod_k \theta_k^{N_k} \frac{\mathbb{I}[\theta \in S_K]}{B(\alpha_1 \dots \alpha_K)} \prod_k \theta_k^{\alpha_k - 1} d\theta =$$

$$= \int_{\theta \in S_K} \prod_k \theta_k^{N_k + \alpha_k - 1} \frac{1}{B(\alpha_1 \dots \alpha_K)} d\theta =$$

$$= \frac{B(N_1 + \alpha_1 \dots N_K + \alpha_K)}{B(\alpha_1 \dots \alpha_K)}$$

$$p(\theta|x) = \frac{p(x|\theta)\, p(\theta)}{p(x)} = \frac{1}{B(N_1 + \alpha_1 \dots N_K + \alpha_K)}$$

$$\cdot \prod_k \theta_k^{N_k + \alpha_k - 1} \cdot \mathbb{I}[\theta \in S_K] =$$

$$= \mathcal{D}ir(\theta \mid \cancel{N_k + \alpha_k} (N_1 + \alpha_1 \dots N_K + \alpha_K))$$

$$\cancel{\mathbb{E}_{p(\theta)}} \; \mathbb{E}_{p(\theta|x)} \theta_j = \frac{N_j + \alpha_j}{\sum_k (N_k + \alpha_k)}$$

Смеси распределений

$p(x)$ — смесь, если

$$p(x) = \sum_k \pi_k \, p(x|\theta_k)$$

$$\pi_k \geq 0, \quad \sum_k \pi_k = 1$$

$$z \sim Cat(\pi_1 \dots \pi_K)$$

$$x \sim p(x|\theta)$$

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

E-step $\quad q^* = p(z|x,\theta)$

M-step $\quad \arg\max_\theta \mathbb{E}_{p(z|x,\theta)} \log p(x,z|\theta)$

$$p(z|x,\theta) = \frac{p(z,x|\theta)}{p(x|\theta)} \leftarrow \text{hard}$$

$X = \{x_1 \dots x_N\}$ i.i.d.

$$p(X|\theta) = \prod_n \sum_k \pi_k \, p(x_n|\theta_k)$$

$$\arg\max_\theta \log p(X|\theta) =$$

$$= \arg\max_\theta \sum_n \log\Big(\sum_k \pi_k \, p(x_n|\theta_k)\Big)$$

$\Theta = \{ \{\pi_k\}, \{\theta_k\} \}$

// $\theta_k = \arg\max (X^z | \theta_k)$

знаяя принадлежности объектов кластерам, можно проводить покомпонентную оптимизацию

$z_n \in \{0, 1\}^k, \quad \sum_k z_{nk} = 1$

$$p(X|Z,\Theta) = \prod_n \prod_k [p(x_n|\theta_k)]^{z_{nk}}$$

$$p(Z|\Theta) = \prod_n \prod_k \pi_k^{z_{nk}} \quad \text{Categorical}$$

$$p(Z|X,\Theta) = \frac{p(X|Z,\Theta)\,p(Z)}{p(X|\Theta)} =$$

$$= \frac{\prod_n \prod_k [p(x_n|\theta_k)\pi_k]^{z_{nk}}}{\prod_n \sum_k \pi_k p(x_n|\theta_k)} = \{\text{св-во}\} =$$

$$= \prod_n \prod_k \left[\frac{p(x_n|\theta_k)\pi_k}{\sum_k \pi_k p(x_n|\theta_k)} \right]^{z_{nk}} = \prod_n \prod_n \gamma_{nk}^{z_{nk}}$$

$$p(Z|X,\Theta) = \prod_{n,k} p(z_{nk}|x_n,\theta_k)$$

св-во факторизации

$$p(x_n|\theta_k) = \mathcal{N}(x_n|\mu_k, \Sigma_k) =$$

$$= \frac{1}{(2\pi)^{\frac{D}{2}} \det^{\frac{1}{2}}\Sigma_k} \exp\left(-\frac{1}{2}(x_n-\mu_k)^T \Sigma_k^{-1}(x_n-\mu_k)\right)$$

$$\mathbb{E}_{p(z|x,\theta)} \log p(x,z|\theta) = \mathbb{E}_{p(z|x,\theta)} \sum_n \sum_k \{ z_{nk} \cdot$$

$$\cdot \log \pi_k \cdot \mathcal{N}(x_n|\mu_k, \Sigma_k)\} =$$

$$= \sum_n \sum_k \{ \underbrace{\mathbb{E}\, z_{nk}}_{\gamma_{nk}} \cdot \log(\pi_k \cdot \mathcal{N}(x_n|\mu_k, \Sigma_k))\} =$$

$$= \sum_{n,k} \gamma_{nk} \log \pi_k + \sum_{n,k} \gamma_{nk} \cdot \left(-\frac{D}{2}\log 2\pi -\right.$$

$$\left. - \frac{1}{2}\det \Sigma_k - \frac{1}{2}(x_n-\mu_k)^T \Sigma_k^{-1}(x_n-\mu_k)\right)$$

$$\longrightarrow \max_{\pi,\mu,\Sigma} \quad \text{s.t.} \quad \sum_k \pi_k = 1$$

$$\mathcal{L}(\pi, \mu) = \sum_{n,k} \gamma_{nk} \log \pi_k - \mu\left(\sum_k \pi_k - 1\right)$$

$$\frac{\partial}{\partial \pi_j}\mathcal{L} = \frac{\sum_n \gamma_{nj}}{\pi_j} - \mu = 0, \quad \pi_j = \frac{\sum_n \gamma_{nj}}{\mu} \Big| \sum_j$$

$$\mu = N, \quad \pi_j = \frac{\sum_n \gamma_{nj}}{N}$$

$$\frac{\partial}{\partial \mu_k}\left(\sum_n \gamma_{nk} \langle \Sigma_k^{-1}(x_n-\mu_k), x_n-\mu_k \rangle\right) =$$

$$= \frac{\partial}{\partial \mu_k}\sum_n \left(\gamma_{nk} \langle \Sigma_k^{-1} x_n, x_n \rangle - 2 \langle \Sigma_k^{-1}\mu_k, x_n \rangle + \langle \Sigma_k^{-1}\mu_k, \mu_k \rangle\right) =$$

$$= \sum_n \gamma_{nk}\left(-2\,\Sigma_k^{-1} x_n + 2\,\Sigma_k^{-1}\mu_k\right) = 0$$

$$\sum_n \gamma_{nk} x_n = \sum_n \gamma_{nk}\mu_k$$

$$\mu_k = \frac{\sum_n \gamma_{nk} x_n}{\sum_n \gamma_{nk}}$$

$$\frac{\partial}{\partial \Sigma_k^{-1}}\Big(\sum_n \gamma_{nk}\big[\log\det \Sigma_k^{-1} - x_n^T \Sigma_k^{-1} x_n +$$

$$+ 2\,\mu_k^T \Sigma_k^{-1} x_n - \mu_k^T \Sigma_k^{-1}\mu_k\big]\Big) =$$

$$= \sum_n \gamma_{nk}\left(\Sigma_k - (x_n-\mu_k)(x_n-\mu_k)^T\right) = 0$$

$$\Sigma_k = \frac{\sum_n \gamma_{nk}(x_n-\mu_k)(x_n-\mu_k)^T}{\sum_n \gamma_{nk}}$$