15.02.19 нейробайес I

Стохастический варвывод

Известно распр-е $p(x|\theta)$ с точностью до пар-ов.

$$\mathbb{E}_{p(x|\theta)} \frac{\partial \log p(x|\theta)}{\partial \theta} = \int p(x|\theta) \frac{\partial \log p(x|\theta)}{\partial \theta} dx =$$

$$= \int p(x|\theta) \frac{1}{p(x|\theta)} \frac{\partial}{\partial \theta} p(x|\theta) dx = \frac{\partial}{\partial \theta} \underbrace{\int p(x|\theta) dx}_{1} = 0$$

$$\frac{\partial^2 \log p(x|\theta)}{\partial \theta^2} = \frac{\partial}{\partial \theta} \left(\frac{1}{p(x|\theta)} \frac{\partial p(x|\theta)}{\partial \theta} \right) =$$

$$= -\left(\frac{1}{p(x|\theta)}\right)^2 \left(\frac{\partial p(x|\theta)}{\partial \theta}\right)^2 + \frac{1}{p(x|\theta)} \frac{\partial^2 p(x|\theta)}{\partial \theta^2} =$$

$$= -\left(\frac{\partial \log p(x|\theta)}{\partial \theta}\right)^2 + \frac{1}{p(x|\theta)} \cancel{\frac{\partial^2 p(x|\theta)}{\partial \theta^2}} \quad \Big| \cdot \int dx$$

0 матожидание

$$\mathbb{E} \frac{\partial^2 \log p(x|\theta)}{\partial \theta^2} = -\mathbb{E}\left(\frac{\partial \log p(x|\theta)}{\partial \theta}\right)^2 = \underbrace{-F(\theta)}_{\text{информация Фишера}}$$

Экспоненциальный класс

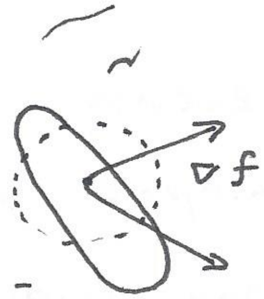
$$p(x|\theta) = \frac{f(x)}{g(\theta)} \exp(\theta^T u(x))$$

$$\frac{\partial^2 \log p(x|\theta)}{\partial \theta^2} = \frac{\partial^2}{\partial \theta^2} (\log f(x) + \theta^T u(x) - \log g(\theta)) =$$

$$= -\frac{\partial^2 \log g(\theta)}{\partial \theta^2} = -F(\theta)$$

$$\operatorname{grad} f(x) \propto \arg\max_{\Delta x} f(x+\Delta x)$$

$$\text{s.t. } \|\Delta x\|_E < \varepsilon$$

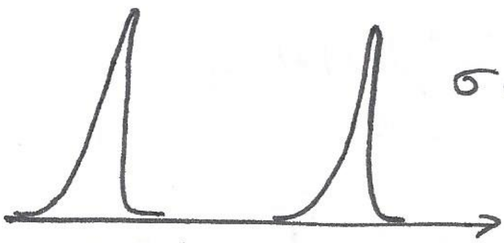


Понятие градиента неинвариантно к репараметризации переменных

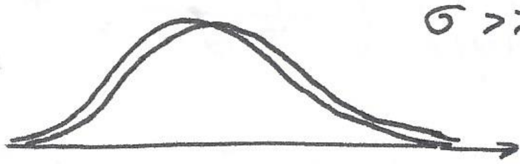
$$\operatorname{grad} \log p(x|\theta) \propto \arg\max_{\Delta\theta} \log p(x|\theta+\Delta\theta)$$

$$\text{s.t. } \|\Delta\theta\|_E < \varepsilon$$

$$N(x|0,\sigma^2),\quad N(x|\Delta\mu,\sigma^2),\quad \|\Delta\mu\| < \varepsilon = 0.01$$



$\sigma \ll \mu$



$\sigma \gg \Delta\mu$

решение:

$$\text{s.t. } KL\left(p(x|\theta)\,\|\,p(x|\theta+\Delta\theta)\right) < \varepsilon$$

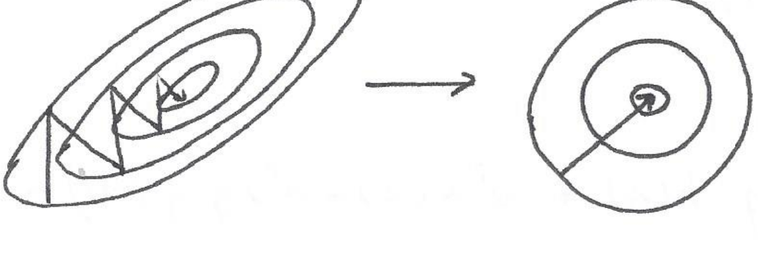
KL-дивергенция уже инвариантна отн. реп.

$$\operatorname{nat\,grad} \log p(x|\theta) \propto \arg\max_{\Delta\theta} \log p(x|\theta+\Delta\theta)$$

$p(\theta)$

$$\| \quad F(\theta)^{-1} \operatorname{grad} \log p(x|\theta)$$

натуральный градиент



$$p(x,z,\theta) = \overbrace{p(x,z|\theta)}\,\overbrace{p(\theta)} = \underbrace{p(x|z,\theta)\,p(z|\theta)\,p(\theta)}_{\text{сопряжение}}$$

$$p(X,Z,\theta) = \prod_n p(x_n, z_n|\theta)\,p(\theta)$$

$X = (x_1, \dots x_n)$

условное сопряжение, есть
1) сопр. на θ при изв. z
2) сопр. на z при изв. θ

$$p(z,\theta|x) \approx q(z)q(\theta) = \arg\min_q KL(q(\theta)q(z) \| p(\theta,z|x))$$

$$\left.\begin{aligned} p(x,z|\theta) &= f(\theta)\,\rho(x,z)\exp(\theta^T h(x,z)) \\ p(\theta) &= f(\theta)^{\nu_0} \exp(\theta^T \eta_0)\frac{1}{g(\nu_0,\eta_0)} \end{aligned}\right\} \text{сопряжение экcп. сем.}$$

$$\log q(z) = \mathbb{E}_{q(\theta)} \log p(x,z,\theta) + const =$$

$$= \mathbb{E}_{q(\theta)}\Big(\sum_{i=1}^n [\underbrace{\cancel{\log f(\theta)}}_{const} + \log \rho(x_i,z_i) + \theta^T h(x_i,z_i)] +$$

$$+ \nu_0 \underbrace{\cancel{\log f(\theta)}}_{const} + \underbrace{\cancel{\theta^T\eta_0}}_{const}\Big) + const =$$

$$= \sum_{i=1}^n [\log \rho(x_i,z_i) + \mathbb{E}\,\theta^T h(x_i,z_i)] + const =$$

$$= \sum_{i=1}^n \log q_i(z_i) \Rightarrow q(Z) = \prod_{i=1}^n q_i(z_i)$$

$$\log q(\theta) = \mathbb{E}_{q(z)} \log p(x,z,\theta) + const =$$

$$= \mathbb{E}_{q(z)}\Big(\sum_{i=1}^n [\log f(\theta) + \underbrace{\cancel{\log \rho(x_i,z_i)}}_{const} + \theta^T h(x_i,z_i)] +$$

$$+ \theta^T h(x_i,z_i) + \nu_0 \log f(\theta) + \theta^T \eta_0\Big) + const =$$

$$= n\log f(\theta) + \theta^T\Big(\sum_{i=1}^n \mathbb{E}_{z_i} h(x_i,z_i)\Big) + \nu_0 \log f(\theta) +$$

$$+ \theta^T\eta_0 + const = (n+\nu_0)\log f(\theta) + \theta^T\Big(\eta_0 + \sum_{i=1}^n \mathbb{E}_{z_i} h(x_i,z_i)\Big) + c$$

$$q(\theta) = f^{\gamma}(\theta)\exp(\theta^T \eta)\,\frac{1}{g(\eta,\gamma)}$$

$$\gamma = \gamma_0 + n\,,\quad \eta = \eta_0 + \sum_{i=1}^{n} \mathbb{E}_{z_i} h(x_i, z_i)$$

в случае если z известно $\mathbb{E}_{z_i}$ уходит, иначе его можно посчитать только имея условную сопряжённость

Пусть далее $\begin{matrix} n \gg 1 \\ d \gg 1 \end{matrix}$ $\; n \approx d$

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

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

$$\mathcal{L}(\eta) = \int q(\theta|\eta)q(z)[\log p(x,z|\theta) + \log p(\theta) -$$

$$-\log q(\theta|\eta) - \underbrace{\cancel{\log q(z)}}_{\text{const}}]\,d\theta\,dz =$$

$$= \int q(\theta|\eta)q(z)[\sum_{i=1}^{n}\{\cancel{\log f(\theta)} + \overbrace{\cancel{\log f(x_i, z_i)}}^{\text{const}} +$$

$$+\underline{\theta^T h(x_i, z_i)}\} + \cancel{\gamma_0 \log} f(\theta) + \underline{\theta^T\eta_0} - \overbrace{\cancel{\log g(\gamma_0, \eta_0)}}^{\text{const}} -$$

$$-\cancel{(\gamma_0 + n)\log} f(\theta) - \underline{\theta^T\eta} + \underline{\log g(\gamma_0 + n, \eta)}]\,dz\,d\theta =$$

$$= \int q(\theta|\eta)\cancel{q(z)}[\theta^T(\eta_0 - \eta + \sum_{i=1}^{n}\mathbb{E}_{z_i} h(x_i, z_i))]\,d\theta +$$

$$+ \log g(\gamma_0 + n, \eta) + \text{const} = \text{const} + \log g(\gamma_0 + n, \eta) +$$

$$+ (\eta_0 - \eta + \sum_{i=1}^{n}\mathbb{E}_{z_i} h(x_i, z_i))\underbrace{\mathbb{E}\theta}_{\frac{\partial \log g(\gamma_0 + n,\,\eta)}{\partial \eta}}$$

$$\cancel{\frac{\partial \log g(\nu_0+n,\eta)}{\partial \eta}} + \underbrace{\frac{\partial^2 \log g(\nu_0+n,\eta)}{\partial \eta^2}}_{F(\eta)} \cdot$$

$$\left(\eta_0 - \eta + \sum_{i=1}^{n} \mathbb{E}_{z_i} h(x_i, z_i)\right) - \cancel{\frac{\partial \log g(\nu_0+n,\eta)}{\partial \eta}} = \frac{\partial L}{\partial \eta}$$

$$\text{nat grad}_\eta \, \mathcal{L} = \cancel{F^{-1}(\eta)}\cancel{F(\eta)}\left(\eta_0 - \eta + \sum_{i=1}^{n} \mathbb{E}_{z_i} h(x_i, z_i)\right) =$$

$$= \eta_0 - \eta + \sum_{i=1}^{n} \mathbb{E}_{z_i} h(x_i, z_i) \approx \eta_0 - \eta + n\, \mathbb{E}_{z_j} h(x_j, z_j)$$

$$j \sim \mathcal{U}\{1 \dots n\}$$

1) $j \sim \mathcal{U}\{1 \dots n\}$

2) $q_j(z_j)$

3) $\eta_{t+1} = \eta_t + \varkappa_t\left(\eta_0 - \eta_t + n\, \mathbb{E}_{z_j} h(x_j, z_j)\right) =$

$$= (1 - \varkappa_t)\eta_t + \varkappa_t\left(\eta_0 + n\, \mathbb{E}_{z_j} h(x_j, z_j)\right)$$

семинар

$$\mathbb{E}_p (\nabla \log p)(\nabla \log p)^T = -\mathbb{E}_p \nabla^2 \log p = F$$

Euclid: $\rho(x, x+dx) = \|dx\|_2^2 = dx^T I \, dx$

R^n, $dx^T A(x)\, dx$, $\tilde{\nabla} = A^{-1}\nabla$

$$D_{KL}^{sym}(\lambda \,\|\, \lambda + d\lambda) = D_{KL}(\lambda \,\|\, \lambda + d\lambda) + D_{KL}(\lambda + d\lambda \,\|\, \lambda) =$$

$$= d\lambda^T A(\lambda)\, d\lambda + O(\|\lambda\|^3)$$

$$\log p(x \mid \lambda + d\lambda) = \log p(x \mid \lambda) + \nabla \log p(x \mid \lambda)^T d\lambda + O(\|d\lambda\|^2)$$

$$p(x \mid \lambda + d\lambda) = p(x \mid \lambda) + \nabla p(x \mid \lambda)^T d\lambda + O(\|d\lambda\|^2) =$$

$$= p(x \mid \lambda) + p(x \mid \lambda) \nabla \log p(x \mid \lambda)^T d\lambda + O(\|d\lambda\|^2)$$

$$D_{KL}^{sym}(\lambda \| \lambda + d\lambda) = \int p(x \mid \lambda) \log \frac{p(x \mid \lambda)}{p(x \mid \lambda + d\lambda)} dx +$$

$$+ \int \log \frac{p(x \mid \lambda + d\lambda)}{p(x \mid \lambda)} p(x \mid \lambda + d\lambda) dx =$$

$$= \int (p(x \mid \lambda + d\lambda) - p(x \mid \lambda)) \log \frac{p(x \mid \lambda + d\lambda)}{p(x \mid \lambda)} dx =$$

$$= \int p(x \mid \lambda) \nabla \log p(x \mid \lambda)^T d\lambda \, \nabla \log p(x \mid \lambda)^T d\lambda \, dx =$$

$$= \int p(x \mid \lambda) \, d\lambda^T \nabla \log p(x \mid \lambda) \nabla \log p(x \mid \lambda)^T d\lambda \, dx =$$

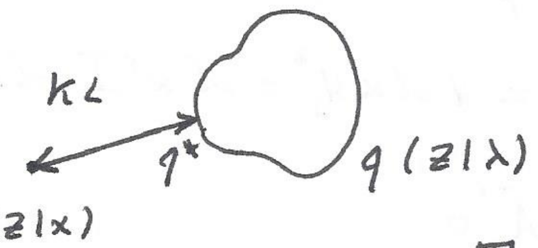
$$= d\lambda^T \left(\int p(x \mid \lambda) \nabla \log p(x \mid \lambda) \nabla \log p(x \mid \lambda)^T dx \right) d\lambda =$$

$$= d\lambda^T \underbrace{F(\lambda)}_{\text{инф. матрица Фишера}} d\lambda$$

информационное Риманово простр-во

LDA

$$p(\Theta, \Phi, Z \mid W)$$



$$p(\Theta, \Phi, Z, W) = \left[\prod_{t=1}^{T} p(\varphi_t \mid \beta) \right] \cdot$$

$$\cdot \left[\prod_{d=1}^{D} p(\theta_d \mid \alpha) \prod_{n=1}^{N_d} p(z_{dn} \mid \theta_d) \, p(w_{dn} \mid z_{dn}, \Phi) \right] =$$

$$= \left[\prod_{t=1}^{T} \mathrm{Dir}(\varphi_t \mid \beta) \right] \left[\prod_{d=1}^{D} \mathrm{Dir}(\theta_d \mid \alpha) \prod_{n=1}^{N_d} \prod_{t=1}^{T} [\theta_{dt} \varphi_{t w_{dn}}]^{[z_{dn}=t]} \right]$$

$$\prod_{n=1}^{N_d} \prod_{t=1}^{T} \left[\Theta_{dt} \, \varphi_{t w_{dn}} \right]^{[z_{dn} = t]} = \prod_{n=1}^{N_d} \Theta_{d z_{dn}} \, \varphi_{z_{dn} w_{dn}}$$

$$q(Z, \Theta, \Phi) = q(Z) \cdot q(\Theta, \Phi)$$

$$\log q(Z) = \mathbb{E}_{\Theta, \Phi \sim q} \log p(Z, \Theta, \Phi, w) + \text{const}$$

$$\log q(\Theta, \Phi) = \mathbb{E}_{Z \sim q} \log p(Z, \Theta, \Phi, w) + \text{const}$$

$$q(Z) = \mathrm{Cat}(\varkappa)$$

$$q(\Theta) = \mathrm{Dir}(\psi)$$

$$q(\Phi) = \mathrm{Dir}(\eta)$$