30.09.16 бммо V

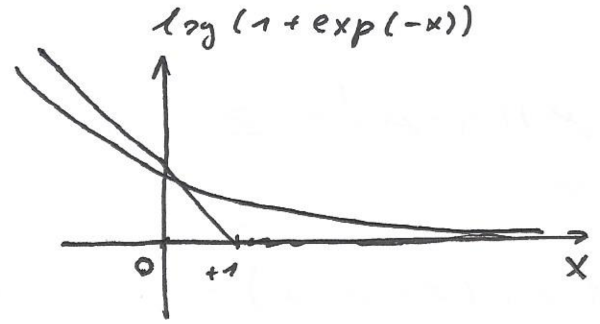
$x \in \mathbb{R}, \ t \in \{-1, 1\}$

$$p(T, w \mid X) = p(T \mid X, w)\, p(w) = \prod_{i=1}^{n} \frac{1}{1 + \exp(-t_i w^T x_i)} \cdot \mathcal{N}(w \mid 0, \alpha^{-1} I)$$

$$\log p(w \mid X, T) = \log \frac{p(T \mid X, w)\, p(w)}{\int p(T \mid X, w)\, p(w)\, dw} \propto \log \prod_{i=1}^{n} \frac{1}{1 + \exp(-t_i w^T x_i)} +$$

$$+ \log p(w) + \text{const} \longrightarrow \max_w$$ (байес для бедных)

$$\sum_{i=1}^{n} \log(1 + \exp(-t_i w^T x_i)) + \frac{\alpha}{2} \|w\|^2 \longrightarrow \min_w$$



базисные ф-ии

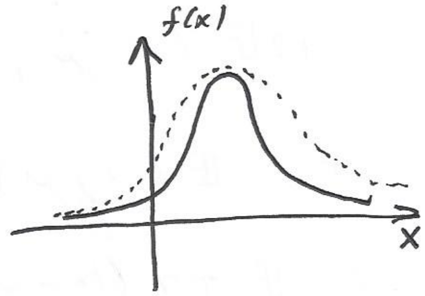
$x_i \longrightarrow (\varphi_1(x_i) \ldots \varphi_k(x_i))$

$\varphi_j(x_i) = \exp(-\gamma \|x_i - x_j\|^2)$

$\varphi_j(x) = \exp(-\gamma \|x - x_j\|^2)$

$$p(T, w \mid X) = p(T \mid X, w) \cdot p(w \mid A) = \prod_{i=1}^{n} \frac{1}{1 + \exp(-t_i w^T x_i)} \cdot$$

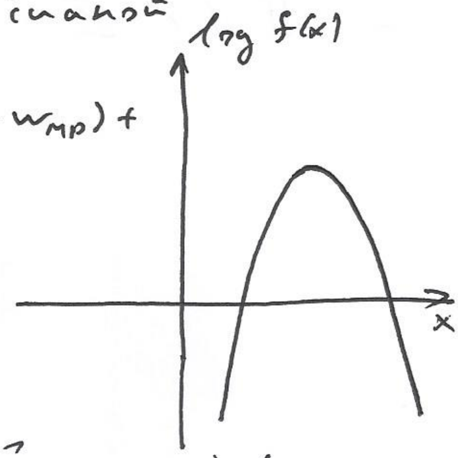
$$\cdot \mathcal{N}(w \mid 0, A^{-1}), \quad A = \begin{pmatrix} \alpha_1 & & \\ & \ddots & \\ & & \alpha_d \end{pmatrix}$$



приближение гауссианой

$$p(T \mid X, A) = \int p(T \mid X, w)\, p(w \mid A)\, dw =$$

$$= \int \underbrace{\prod_{i=1}^{n} \frac{1}{1 + \exp(-t_i w^T x_i)} \cdot \mathcal{N}(w \mid 0, A^{-1})}_{Q(w)} dw$$



$$\log Q(w) = \log Q(w_{MP}) + \underbrace{\nabla \log Q(w_{MP})}_{0} (w - w_{MP}) +$$

$$+ \frac{1}{2} (w - w_{MP})^T \nabla\nabla \log Q(w_P)(w - w_{MP})$$

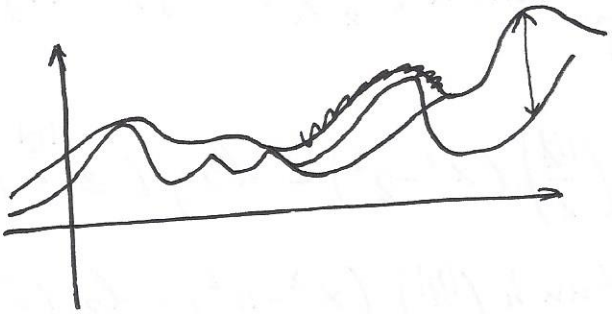
$$-\nabla\nabla \log Q(w_P) = \hat{\Sigma}^{-1}$$

$$\int Q(w)\, dw = \int Q(w_{MP}) \exp\left(-\frac{1}{2}(w - w_{MP})^T \Sigma^{-1} (w - w_{MP})\right) dw =$$

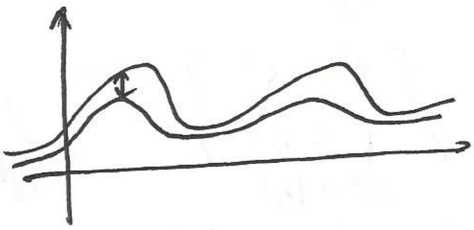
$$= (2\pi)^{\frac{d}{2}} \sqrt{\det \Sigma}\; Q(w_{MP})$$

$$\log p(T|X,A) = \log Q(w_{MP}) - \frac{1}{2}\log\det(-\nabla\nabla\log Q(w_{MP})) \to \max_{w_{MP}}$$

$$\alpha_j^{new} = \frac{1-\alpha_j^{old}\Sigma_{jj}}{\theta_j^2}\ ; \quad w_{MP} = \arg\max_w p(T|X,w)\,p(w|A)$$



приближение

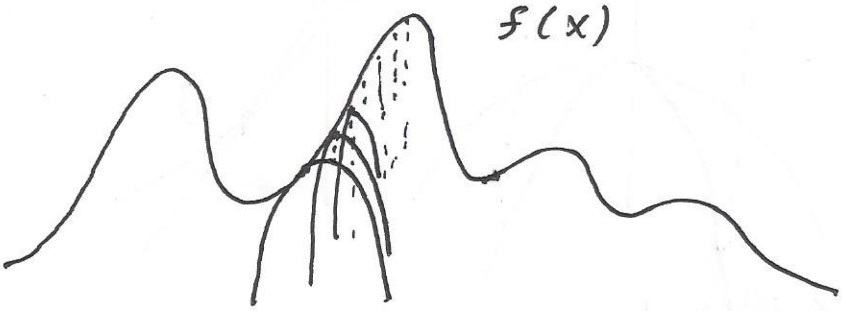


максимизация нижней оценки

обоснованность:

$$\log Q(w_{MP}) - \frac{1}{2}\log\det(-\nabla\nabla\log Q(w_{MP})) \to \max_A$$

максимизация



$\{g(x,\xi)\}$

1) $g(x,\xi) \le f(x)\ \forall x,\xi$

2) $\forall x_0\ \exists \xi_0:\ g(x_0,\xi_0) = f(x_0)$

итерационная оптимизация вариационных нижних оценок

$$\int Q(w,A)\,dw \ge \int h(w,\xi,A)\,dw = I(\xi,A) \to \max$$

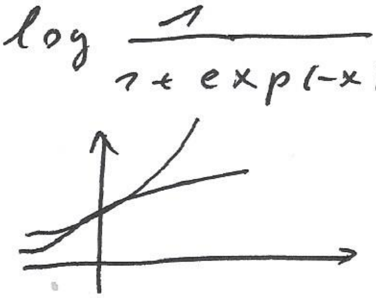
произведение гауссиан $\leftrightarrow$ произведение сигмоидных функций.

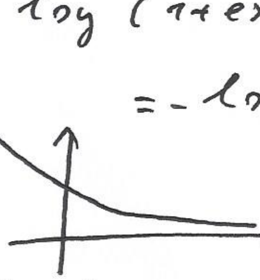
континуум вариационных нижних оценок

$$\log\frac{1}{1+\exp(-x)} = -\log(1+\exp(-x)) = \text{вогнутая}$$

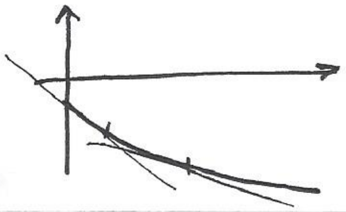
$$= -\log\left(e^{-\frac{x}{2}}\left(e^{\frac{x}{2}}+e^{-\frac{x}{2}}\right)\right) =$$

$$= \frac{x}{2} - \underbrace{\log\left(e^{\frac{|x|}{2}}+e^{-\frac{|x|}{2}}\right)}_{\text{чётная ф-ция}}$$



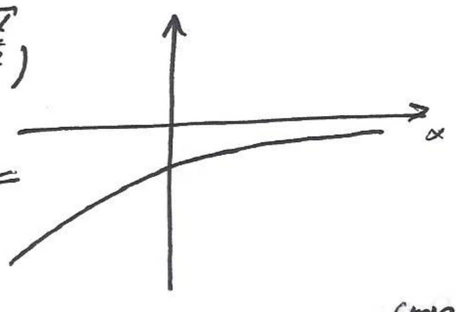


$$-\log\left(e^{\frac{|x|}{2}}+e^{-\frac{|x|}{2}}\right) = \left\{\begin{matrix} t = x^2 \\ |x| = \sqrt{t}\end{matrix}\right\} = -\log\left(e^{-\frac{\sqrt{t}}{2}}+e^{\frac{\sqrt{t}}{2}}\right)$$



$\Leftarrow$

$\Leftarrow$



$$f(x)\ ;\ g(x,\xi) = f(\xi) + f'(\xi)(x-\xi)$$

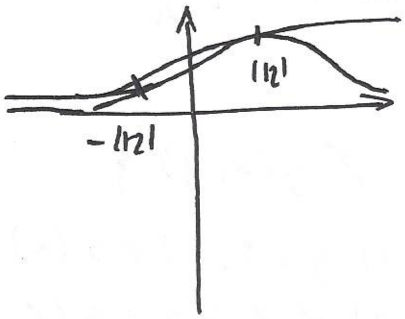
$$\log'\left(e^{-\frac{\sqrt{t}}{2}} + e^{\frac{\sqrt{t}}{2}}\right) = \frac{e^{\frac{\sqrt{t}}{2}} - e^{\frac{\sqrt{t}}{2}}}{e^{-\frac{\sqrt{t}}{2}} + e^{\frac{\sqrt{t}}{2}}} \cdot \frac{1}{4\sqrt{t}}$$

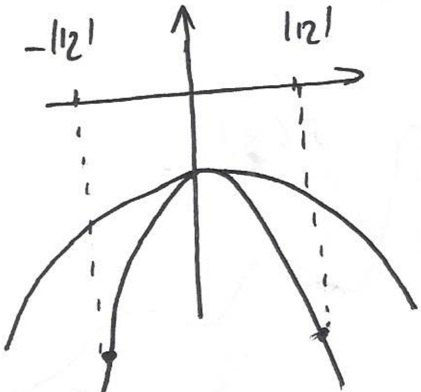
$$-\log\left(e^{-\frac{\sqrt{t}}{2}} + e^{\frac{\sqrt{t}}{2}}\right) \geqslant -\frac{1}{4\sqrt{\xi}} \tanh\left(\frac{\sqrt{\xi}}{2}\right)(t-\xi) - \log\left(e^{-\frac{\sqrt{\xi}}{2}} + e^{\frac{\sqrt{\xi}}{2}}\right) =$$

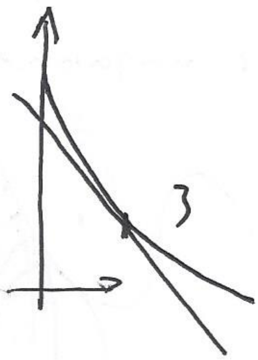
$$= \left\{\begin{array}{l} t = x^2 \\ \xi = \eta^2 \\ \sqrt{\xi} = |\eta| \end{array}\right\} = -\frac{1}{4|\eta|} \tanh\left(\frac{|\eta|}{2}\right)(x^2-\eta^2) - \log\left(e^{-\frac{|\eta|}{2}} + e^{\frac{|\eta|}{2}}\right)$$

$$-\log(1+e^{-x}) \geqslant \frac{x}{2} - \frac{1}{4|\eta|}\tanh\left(\frac{|\eta|}{2}\right)(x^2-\eta^2) - \log\left(e^{-\frac{|\eta|}{2}} + e^{\frac{|\eta|}{2}}\right)$$

$$\frac{1}{1+e^{-x}} \geqslant \frac{1}{e^{\frac{|\eta|}{2}} + e^{-\frac{|\eta|}{2}}} \cdot \exp\left(\frac{x}{2} - \frac{1}{4|\eta|}\tanh\left(\frac{|\eta|}{2}\right)(x^2-\eta^2)\right)$$







$$\prod_{i=1}^{n} \sigma(+t_i w^T x_i) \geqslant \prod_{i=1}^{n} \ell(t_i w^T x_i, \eta_i)$$

$$\int \sigma(t_i w^T x_i)\, N(w \mid 0, A^{-1})\, dw \geqslant \int \prod_{i=1}^{K} \ell(t_i w^T x_i, \eta_i)\, N(w \mid 0, \vec{A})\, dw =$$

$$= \Phi(A, \vec{\eta}) \longrightarrow \max_{A,\eta}$$

30.09.16 Бишоп сем

$p(x) = N(x \mid \mu, \Sigma)$

$p(y \mid x) = N(y \mid \cancel{\mu} A x, \Gamma)$

$p(y)$? $\mathbb{E} y = A\mu$, $\mathbb{D} y = A^{a} \Sigma A^{T} + \mathcal{E}$

$y = Ax + \mathcal{E}$, $\mathcal{E} \sim N(\mathcal{E} \mid 0, \Gamma)$

$\cancel{p} \ \{x_n, t_n\}_{n=1}^{N}$, $x \in \mathbb{R}^d$, $t \in \mathbb{R}$

$p(t, w \mid X, A, \beta) = \prod_i N(t_i \mid x_i^T w, \beta^{-1}) \cdot N(w \mid 0, A^{-1})$

$A = \alpha I$, $A = diag(\alpha_1 \dots \alpha_d)$

$\log p(t_{t2} \mid X_{t2}, A, \beta) \longrightarrow \max\limits_{A, \beta}$

$= \int p(t_{t2} \mid X_{t2}, w, \beta) \, p(w \mid A, \beta, X_{t2}, t_{t2}) \, dw$

evidence $T_{t2} \mid X_{t2}, A, \beta \sim N(0, \beta^{-1} I + X_{t2} A^{-1} X_{t2}^{T})$

$-\frac{1}{2} \log\det(\beta^{-1} I + X A^{-1} X^T) - \frac{1}{2} T^T (\beta^{-1} I + X A^{-1} X^T)^{-1} T$

поправка малого радиуса:

$(A + UCV)^{-1} = A^{-1} - A^{-1} U (C^{-1} + V A^{-1} U)^{-1} V A^{-1}$

$\det(X + AB) = \det X \cdot \det(I + B X^{-1} A)$

$\log\det(\underbrace{\beta^{-1} I}_{X'} + \underbrace{X}_{A'} \underbrace{A^{-1} X^T}_{B'}) = \log\det \beta^{-1} I + \log\det(I + \beta X^T X A^{-1}) =$

$= \log \beta^{-n} + \log\det(I + \beta X^T X A^{-1}) = -n \log\beta + \log\det(I + \beta X^T X A^{-1}) =$

$= -n \log\beta + \log\det(A + \beta X^T X^{m}) A^{-1} = -n \log\beta - \log\det\Sigma -$

$- \log\det A$ $\quad \{ \mu = \beta \Sigma X^T T, \ \Sigma = (\beta X^T X + A)^{-1}$

$-\frac{\beta}{2} \| X\mu - T \|^2 - \frac{1}{2} \mu^T A \mu = -\frac{\beta}{2} \| X \beta \Sigma X^T T - T \|^2 - \frac{1}{2} \beta^2 T^T X \Sigma A \cdot$

$\cdot \Sigma X T = -\frac{\beta}{2} (\beta X \Sigma X^T T - T)^T (\beta X \Sigma X^T T - T) - \frac{1}{2} \beta^2 T^T X \Sigma A \Sigma X T \dots$

$\frac{1}{2} \log\det\Sigma + \frac{n}{2} \log\beta + \frac{1}{2} \log\det A - \frac{\beta}{2} \| X\mu - T \|^2 - \frac{1}{2} \mu^T A \mu$

$$p(t \mid X, X_{tr}, t_{tr}, A, \beta) = \int p(t \mid X, w, \beta) \cdot p(w \mid X_{tr}, t_{tr}, A, \beta)\, dw =$$

$$= \int N(t \mid Xw, \beta^{-1} I) \cdot N(w \mid \mu, \Sigma)\, dw =$$

$$= N(t \mid X\mu, \beta^{-1} I + X \Sigma X^T)$$

$\{ N(t \mid X\mu, \beta^{-1} I)$, если вместо интегрирования точечная оценка.