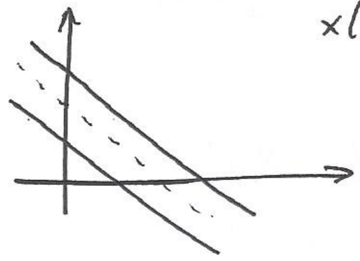
23.09.16 была лек   и 30.09.16 $\boxed{RVM}$ IV   V

$$Ax = b \mid A^T \qquad A^T A x = A^T b$$

$$x = A^{-1} b \qquad x = (A^T A)^{-1} A^T b \qquad A^T A \geq 0$$



$$x(\lambda) = x = (A^T A + \lambda I)^{-1} A^T b$$

регуляризация

$$\lambda \to 0, \ x(\lambda) \to x_{\text{псевд.}}$$

$$\frac{\partial f(A)}{\partial A} = \left(\frac{\partial f}{\partial a_{ij}}\right)_{i,j}, \quad \frac{\partial A(\alpha)}{\partial \alpha} = \left(\frac{\partial a_{ij}(\alpha)}{\partial \alpha}\right)_{i,j}$$

$$\det A = \sum_{i,j} a_{ij} M_{ij}, \quad \frac{\partial \det A}{\partial a_{ij}} = M_{ij}$$

$$\frac{\partial \log \det A}{\partial A} = A^{-1}$$

$$\frac{\partial f(A(\alpha))}{\partial \alpha} = \frac{\partial f}{\partial A} \cdot \frac{\partial A}{\partial \alpha} \ ? = \sum_{i,j} \frac{\partial f}{\partial a_{ij}} \frac{\partial a_{ij}}{\partial \alpha} = \operatorname{tr}\left(\frac{\partial f}{\partial A}\right)^T \frac{\partial A}{\partial \alpha}$$

$$\operatorname{tr}(A^T B) = \langle A, B \rangle$$

$$\frac{\partial f(A(\alpha))}{\partial \alpha} = \left\langle \frac{\partial f}{\partial A}, \frac{\partial A}{\partial \alpha} \right\rangle$$

Relevance Vector Machine (RVM)

$$p(t, \theta \mid x) = p(t \mid \theta, x) p(\theta) = N(t \mid \theta^T x, \beta^{-1}) N(\theta \mid 0, \alpha^{-1} I)$$

$$(X, T) = (x_i, t_i)_{i=1}^{n} \qquad x_i, \theta \in \mathbb{R}^m$$

$$p(\theta \mid X, T) = \frac{p(T \mid X, \theta) p(\theta)}{\int p(T \mid X, \theta) p(\theta) d\theta} = N(\theta \mid \mu, \Sigma)$$

$$p(T \mid X, \theta) p(\theta) = \prod_{i=1}^{n} \sqrt{\frac{\beta}{2\pi}} e^{-\frac{\beta}{2}(t_i - x_i^T \theta)^2} \cdot \prod_{j=1}^{m} \sqrt{\frac{\alpha}{2\pi}} e^{-\frac{\alpha}{2}\theta_j^2} =$$

$$= \left(\frac{\beta}{2\pi}\right)^{\frac{n}{2}} \left(\frac{\alpha}{2\pi}\right)^{\frac{m}{2}} e^{-\frac{\beta}{2}\|T - X\Theta\|^2} \cdot e^{-\frac{\alpha}{2}\|\Theta\|^2} = \frac{(\beta^{\frac{n}{2}} \alpha^{\frac{m}{2}})}{(2\pi)^{\frac{n}{2}+\frac{m}{2}}} e^{-\frac{\beta}{2}(T^T T - 2T^T X\Theta + \Theta^T X^T X \Theta) - \frac{\alpha}{2}\Theta^T\Theta}$$

$$\|T - X\Theta\|^2 = (T - X\Theta)^T (T - X\Theta) = T^T T - 2T^T X\Theta + \Theta^T X^T X \Theta$$

$$\|\Theta\|^2 = \Theta^T \Theta$$

стр →

$$= \frac{(\beta^{\frac{n}{2}}\alpha^{\frac{m}{2}})}{(2\pi)^{\frac{n}{2}+\frac{m}{2}}} e^{\left(-\frac{\beta}{2} T^T T + \beta T^T X \Theta - \frac{1}{2}\Theta^T \underbrace{(\beta X^T X + \alpha I)}_{\Sigma^{-1}} \Theta\right)}$$

$$\underline{\Sigma = (\beta X^T X + \alpha I)^{-1}} \qquad (x^2)' = 2x$$

$$F(\Theta) = \beta T^T X \Theta - \frac{\beta}{2}\Theta^T X^T X \Theta - \frac{1}{2}\Theta^T(\alpha I)\Theta = \beta T^T X \Theta - \frac{1}{2}\Theta^T(\beta X^T X + \alpha I)\Theta$$

$$F'(\Theta) = \beta (T^T X)^T - (\beta X^T X + \alpha I)\Theta = 0$$

$$\Theta = (..)\,\beta (T^T X)^T, \quad \Theta = (\beta X^T X + \alpha I)^{-1}\beta (T^T X)^T$$

$$\Theta_{MP} = (\beta X^T X + \alpha I)^{-1}\beta X^T T$$

$$\mu = (\beta X^T X + \alpha I)^{-1}\beta X^T T \qquad \{X\Theta \simeq T\}$$

$$\underline{\mu = \beta \Sigma X^T T = \Theta_{MP}} \qquad \Theta_{ML} = (X^T X)\, X^T T \quad (\alpha = 0)$$

$$\Theta_{MP} \equiv 0, \quad \alpha \to \infty$$

$\frac{n}{m} \gg 1 \Rightarrow$ переобучение!

$$p(t,\Theta | x) = N(t | \Theta^T x, \beta^{-1})\, N(\Theta | 0, A^{-1})$$

$$A = \begin{bmatrix} \alpha_1 & & 0 \\ & \ddots & \\ 0 & & \alpha_m \end{bmatrix}$$

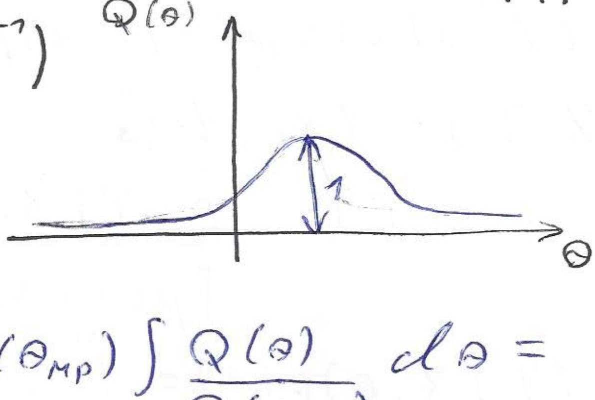
<u>Обоснованность</u>

$$p(T | X, A, \beta) = \int \overbrace{p(\cancel{X} T | X, \Theta, \beta)\, p(\Theta | A)}^{Q(\Theta)}\, d\Theta \xrightarrow[A,\beta]{} \max$$

$$\parallel$$

$$N(T | 0, (\beta^{-1} I + X A^{-1} X^T)^{-1})$$

$$\int_{-\infty}^{\infty} e^{-\frac{1}{2}x^T \Sigma^{-1} x}\, dx = (2\pi)^{\frac{m}{2}}\sqrt{\det \Sigma}$$



$$\int Q(w)\,dw, \quad \int Q(\Theta)\,d\Theta = Q(\Theta_{MP}) \int \frac{Q(\Theta)}{Q(\Theta_{MP})}\, d\Theta =$$

$$= Q(\Theta_{MP})(2\pi)^{\frac{m}{2}}\sqrt{\det(-\nabla\nabla \log Q(\Theta))^{-1}} =$$

$$= Q(\Theta_{MP})(2\pi)^{\frac{m}{2}}\sqrt{\det \Sigma}$$

$$Q(\Theta_{MP}) \cancel{\alpha}\, (2\pi)^{\frac{m}{2}} \frac{1}{\det\sqrt{\beta X^T X + A}} \to \max_{\beta, A}$$

$\Phi(A,\beta)$, $G(A,\beta,\Theta)$

$$\| \log p(T \mid X, A, \beta) = -\frac{1}{2}\|T - X\Theta_{MP}\|^2 +$$

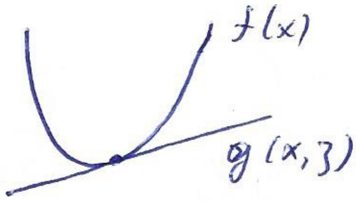
$$+ \frac{m}{2}\log\beta - \frac{n}{2}\log 2\pi + \frac{1}{2}\log\det A - \frac{m}{2}\log 2\pi -$$

$$- \frac{1}{2}\Theta_{MP}^T A \Theta_{MP} + \frac{m}{2}\log(2\pi) - \frac{1}{2}\log\det(\beta X^T X + A)$$

$$p(T \mid X, \Theta_{MP}, \beta) = \left(\frac{\beta}{2\pi}\right)^{\frac{n}{2}} e^{-\frac{\beta}{2}\|T - X\Theta_{MP}\|^2}$$

$$p(\Theta \mid A) = \frac{\sqrt{\det A}}{(2\pi)^{\frac{m}{2}}} \exp\left(-\frac{1}{2}\Theta_{MP}^T A \Theta_{MP}\right)$$

Вариационные нижние оценки



$$f(x) \to \max_x \iff \begin{cases} g(x, \xi_0) \to \max_x \\ g(x_0, \xi) \to \max_\xi \end{cases}$$

$\exists\, g(x, \xi)$!

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

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

1) $G(A,\beta,\Theta) \le \Phi(A,\beta)$

$$G(A,\beta,\Theta) = -\frac{1}{2}\|T - X\Theta\|^2 - \frac{1}{2}\Theta^T A \Theta + \text{—} \ldots \text{—}$$

$\Theta_{MP} = \arg\max Q(\Theta)$

2) $\forall\, (A,\beta)^\Theta \ \exists\, \Theta = \Theta_{MP} : G(A,\beta,\Theta) = \Phi(A,\beta)$

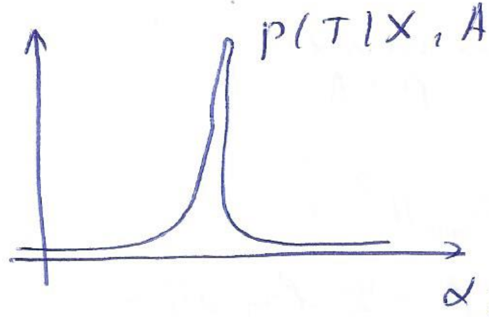
$G(A,\beta,\Theta)$ — вариационная нижняя оценка.

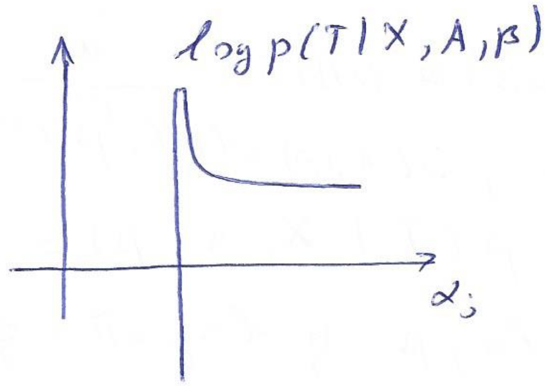
$$\frac{\partial G(A,\beta,\Theta)}{\partial \alpha_i} = \frac{1}{2\alpha_i} - \frac{1}{2}\Theta_{MP,i}^{*2} - \frac{1}{2}\Sigma_{ii} = 0$$

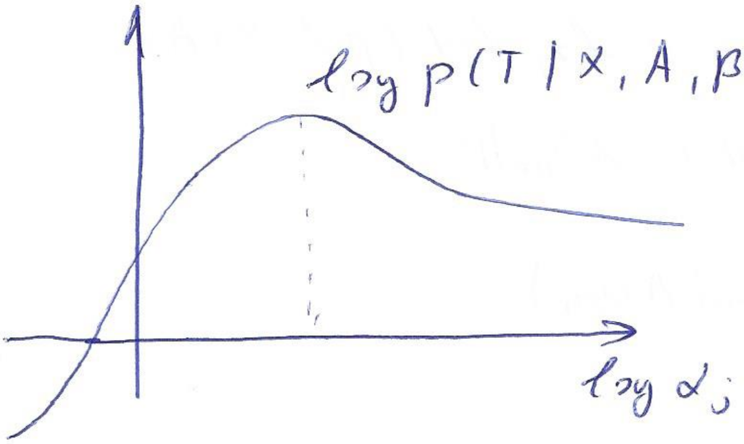
$$\frac{\log\det\Sigma^{-1}}{\partial\alpha_i} = tr\left(\frac{\partial}{\partial\Sigma^{-1}}\log\det\Sigma^{-1}\right)^T \frac{\partial\Sigma^{-1}}{\partial\alpha_i} =$$

$$= tr\left(\Sigma^T; \begin{bmatrix} 0 & & 0 \\ & 1 & \\ 0 & & 0 \end{bmatrix}\right) = tr\, \Sigma_{ii}$$

$$\alpha_i^{t+1} = \frac{1}{(\Theta_i^t)^2 + \Sigma_{ii}}$$







$$\alpha_j^{t+1} = \frac{1}{(\theta_j^t)^2 + \Sigma_{jj}}$$

$$\frac{\partial G}{\partial \log \alpha_j} = \frac{\partial G}{\partial \alpha_j} \frac{\partial \alpha_j}{\partial \log \alpha_j} = \frac{\partial G}{\partial \alpha_j} \frac{1}{\frac{\partial \log \alpha_j}{\partial \alpha_j}} = \alpha_j \frac{\partial G}{\partial \alpha_j}$$

$$\cancel{\ldots}\ 1 - \theta_j^2 \alpha_j^{new} - \alpha_j^{all} \Sigma_{jj} = 0 \quad , f(x^{all}) = x^{new}$$

$$\alpha_j^{new} = \frac{1 - \alpha_j^{all} \Sigma_{jj}}{\theta_j^2} \ ; \ \beta^{new} = \frac{1 - \sum_{j=1}^{n} (1 - \alpha_j^{all} \Sigma_{jj})}{\|T - X\theta\|^2}$$

Automatic Relevance Determination (ARD)

23. 09. 10 днмо сем

$f(x_1 \ldots x_n) : \mathbb{R}^n \to \mathbb{R} \quad , \frac{\partial f}{\partial x} = \left(\frac{\partial f}{\partial x_i}\right) \in \mathbb{R}^n$

$f : \mathbb{R}^n \to \mathbb{R}^m \quad \left(\frac{\partial f_i}{\partial x_j}\right) \in \mathbb{R}^{m \times n}$

$A : \mathbb{R} \to \mathbb{R}^{m \times n} \quad \frac{\partial A(x)}{\partial x} \in \mathbb{R}^{m \times n} \quad \frac{\partial}{\partial x} x^T a = a$

$f(A) \ \mathbb{R}^{m \times n} \to \mathbb{R} \quad , \frac{\partial f}{\partial A} \in \mathbb{R}^{m \times n}$

$$\frac{\partial}{\partial x} \|Ax - b\|^2 = \frac{\partial}{\partial x} \langle Ax - b, Ax - b \rangle = \frac{\partial}{\partial x} (Ax-b)^T(Ax-b) =$$

$$= \frac{\partial}{\partial x} \left(x^T A^T A x - 2 b^T A x + b^T b \right) =$$

$$= \frac{\partial}{\partial x} \left(-2x^T A^T b + x^T A^T A x \right) = -2A^T b + 2A^T A x =$$

$$= 2A^T(Ax - b)$$

$$\frac{\partial}{\partial A} \det A = \frac{\partial}{\partial A} \sum_{i,j} a_{ij} M_{ij}$$

$$\frac{\partial}{\partial x}(x^T A x) = \frac{\partial}{\partial x} \sum_{i,j} x_i A_{ij} x_j$$

$$\frac{\partial}{\partial x_k} \sum_{i,j} x_i A_{ij} x_j = \frac{\partial}{\partial x_k} \sum_{i \neq k} x_i A_{ij} x_j + \frac{\partial}{\partial x_k} \sum_{j \neq k} x_i A_{ij} x_j +$$

$$+ \frac{\partial}{\partial x_k} \sum_{i,j=k} x_k A_{kk} x_k = \sum_{i \neq k} x_i A_{ik} + \sum_{j \neq k} A_{kj} x_j + 2x_k A_{kk} =$$

$$= \cancel{2\sum_{i \neq k} x_i A_{ik} + 2x_k A_{kk} = 2 \sum x_i A_{ik}}$$

$$= \sum_i x_i A_{ik} + \sum_i x_i A_{ki} = (Ax + A^T x)_k$$

$$\frac{\partial}{\partial x} x^T A x = \cancel{2Ax}\ (A + A^T)x$$

$$\frac{\partial}{\partial A_{ij}} \det A = \frac{\partial}{\partial a_{ij}} \sum_k a_{ik} M_{ik} = M_{ij}$$

$$A^{-1}_{ij} = \frac{1}{|A|} M_{ji} \qquad \frac{\partial}{\partial A} \det A = A^{-T} \det A$$

$$M_{ji} = A^{-1}_{ij} |A|$$

$$\frac{\partial}{\partial A} \operatorname{tr} AB = B^T, \quad \sum_k (AB)_{kk} = \sum_k \sum_j A_{kj} B_{jk}$$

$$\frac{\partial}{\partial A_{ij}} \sum_{p,q} A_{pq} B_{qp} = B_{ji}$$

$$\frac{\partial}{\partial A} x^T A y = \frac{\partial}{\partial A} A y x^T = (y x^T)^T = x y^T$$

$$\frac{d}{dA} \log \det A = \frac{1}{\det A} (\det A)'_A = A^{-T} \frac{\det A}{\det A} = A^{-T}$$

$$\mathcal{A}(x) : \mathbb{R} \to \mathbb{R}^{m \times n}$$

$$\frac{\partial}{\partial x} \log \det \mathcal{A}(x) = \cancel{\partial} \frac{1}{\det \mathcal{A}(x)} (\det \mathcal{A}(x))'_{\mathcal{A}(x)} \cdot (\mathcal{A}(x))'_x =$$

$$= \frac{1}{\det \mathcal{A}(x)} \mathcal{A}^{-T}(x) \det \mathcal{A}(x) \ \mathcal{A}'(x) = \mathcal{A}^{-T}(x) \frac{d\mathcal{A}(x)}{dx}$$

$$\cancel{\frac{\partial}{\partial y}} \frac{d}{dx} \log \det \mathcal{A}(x) = \cancel{\frac{d}{\det \mathcal{A}(x)}} \frac{1}{\det \mathcal{A}(x)} \sum_{ij} \frac{d}{d\mathcal{A}_{ij}} \cancel{\det \mathcal{A}_{ij}(x)} \frac{d}{dx} \mathcal{A}_{ij}(x) =$$

$$\cancel{= n \mathcal{A}^{-T}(x) \frac{d\mathcal{A}(x)}{dx}}$$

$$m \times n \qquad n \times m$$

$$= \frac{1}{\det \mathcal{A}(x)}$$

$$\mathcal{A} : \mathbb{R} \to \mathbb{R}^{n \times n} \qquad \mathcal{A}^{-1} \mathcal{A} = I$$

$$\frac{\partial}{\partial x} \mathcal{A}^{-1}(x) \qquad \frac{\partial}{\partial x} \mathcal{A}^{-1}(x) \mathcal{A}(x) = 0$$

$$\frac{\partial}{\partial x} (\mathcal{A}^{-1}(x)) \mathcal{A}(x) + \mathcal{A}^{-1}(x) \frac{\partial}{\partial x} \mathcal{A}(x) = 0$$

$$\frac{\partial}{\partial x} \mathcal{A}^{-1}(x) = -\mathcal{A}^{-1}(x) \frac{\partial}{\partial x} \mathcal{A}(x) \mathcal{A}^{-1}(x)$$

$$\frac{\partial}{\partial A_{ij}} \operatorname{tr}(A^{-1} B) \neq \cancel{\frac{\partial}{\partial}} = \frac{\partial}{\partial A_{ij}} \sum_k (\mathcal{A}^{-1} \mathcal{B})_{kk} = \cancel{\frac{\partial}{\partial A_{ij}} \sum_k} \ (=)$$

$$= \sum_k \left(\frac{\partial}{\partial a_{ij}} \mathcal{A}^{-1} \mathcal{B} \right)_{kk} = \cancel{\sum_k (\mathcal{A}^{-1}(x) \frac{d}{dx} \mathcal{A}(x) \mathcal{A}^{-1}(x) \mathcal{B}(x))_{kk}}$$

$$\# \ \cancel{d} \frac{\partial}{\partial a_{ij}} \mathcal{A}^{-1} = -\mathcal{A}^{-1} \begin{vmatrix} 0 & & 0 \\ & 1 & \\ 0 & & 0 \end{vmatrix} \mathcal{A}^{-1}$$

$$(=) \ \cancel{- \sum_k \operatorname{tr} \mathcal{A}}$$

$$\frac{\partial}{\partial A_{ij}} \operatorname{tr}(A^{-1}B) = \frac{\partial}{\partial A_{ij}} \sum_k (A^{-1}B)_{kk} = \sum_k \left(\frac{\partial}{\partial A_{ij}} A^{-1}B\right)_{kk} =$$

$$= \sum_k \left(-A^{-1} \frac{\partial A}{\partial A_{ij}} A^{-1} B\right)_{kk} = \operatorname{tr}\left(-A^{-1} \frac{\partial A}{\partial A_{ij}} A^{-1} B\right) =$$

$$= \operatorname{tr}\Big(- \frac{\partial A}{\partial A_{ij}} \underbrace{A^{-1} B A^{-1}}_{C}\Big) = \sum_k -\left[\begin{pmatrix} 0 & & \\ & 1 & \\ 0 & & 0 \end{pmatrix} C \right]_{kk} =$$

$$= \sum_k -\left[\sum_p e_{kp} C_{pk} \right] = -C_{ji}$$

$$\frac{\partial}{\partial A} \operatorname{tr} A^{-1}B) = -(A^{-1}BA^{-1})^T$$

\# $X \sim N(x \mid \mu, \Sigma)$, $\quad \mathbb{E}(x-\mu)(x-\mu)^T = \Sigma$

$\qquad \mathbb{E}x = \mu$

$$\mathbb{E}x^T x = \mathbb{E}(x-\mu+\mu)^T(x-\mu+\mu) = \mathbb{E}((x-\mu)+\mu)^T((x-\mu)+\mu) =$$

$$= \mathbb{E}((x-\mu)+\mu, (x-\mu)+\mu) = \mathbb{E}(x-\mu, x-\mu) +$$

$$+ 2\mathbb{E}(x-\mu, \mu) + \mathbb{E}(\mu,\mu) = \mathbb{E}(x-\mu, x-\mu) +$$

$$+ 2\mathbb{E}(x,\mu) - \mathbb{E}(\mu,\mu) = \mathbb{E}(x-\mu)^T(x-\mu) + \mu^T\mu =$$

$$= \mathbb{E} \operatorname{tr}((x-\mu)^T(x-\mu)) + \mu^T\mu = \mathbb{E} \operatorname{tr}(x-\mu)(x-\mu)^T + \mu^T\mu =$$

$$= \mathbb{E} \operatorname{tr} \Sigma + \|\mu\|^2 = \operatorname{tr} \Sigma + \|\mu\|^2$$