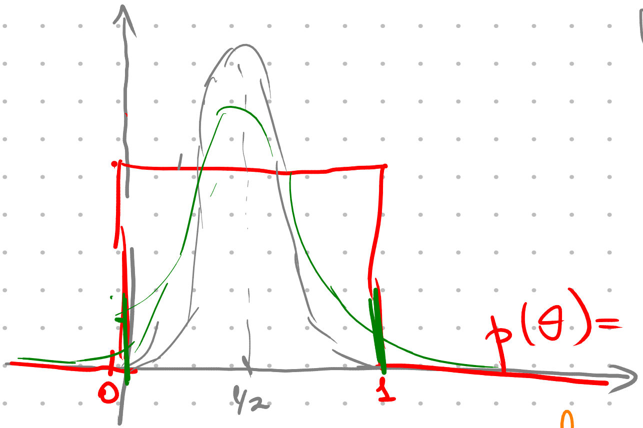
$D = htthhh \qquad \theta = p(\text{«орёл»})$

$p(D|\theta) = \theta^n (1-\theta)^m \qquad \theta_{ML} = \dfrac{n}{n+m}$



$$\boxed{p(\theta|D) \propto p(\theta)p(D|\theta)}$$

$\theta_{MAP} = \theta_{ML}$

$p(\theta) = \begin{cases} 1, & \theta \in [0,1] \\ 0, & \theta \notin [0,1] \end{cases}$

$p(\text{орёл}|D) = \int p(\text{орёл}, \theta|D)\,d\theta = \int \underbrace{p(\text{орёл}|\theta)}_{\theta}\, p(\theta|D)\,d\theta$

$p(\text{орёл}|D) = \dfrac{n+1}{n+m+2}$

$p(\theta) = \begin{cases} \propto \dfrac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(\theta - \frac{1}{2})^2} \\ 0, \quad \theta \notin [0,1] \end{cases} \qquad \times \quad p(D|\theta) = \boxed{\theta^n (1-\theta)^m}$

бета-распределение:

$p(\theta|\alpha,\beta) = \dfrac{1}{B(\alpha,\beta)} \theta^{\alpha-1}(1-\theta)^{\beta-1}, \quad \theta \in [0,1]$

на $[0,1]$

$p(\theta|D) \propto p(D|\theta)\,p(\theta|\alpha,\beta) \propto \theta^n (1-\theta)^m \cdot \theta^{\alpha-1}(1-\theta)^{\beta-1} =$

$= \theta^{n+\alpha-1}(1-\theta)^{m+\beta-1}$

$p(\theta|D) = \mathrm{Beta}(\theta \,|\, \alpha+n, \beta+m)$

$p(\theta)$	$\times$ $p(D|\theta)$	$\propto$	$p(\theta|D)$
$\mathrm{Beta}(\theta|\alpha,\beta)$	$\theta^n(1-\theta)^m$		$\mathrm{Beta}(\theta|\alpha+n,\beta+m)$

$D', \; n', m'$ – новые данные

$$p(\theta \mid \underbrace{D, D'}) \propto p(\theta)\, p(D, D' \mid \theta)$$

$$\propto \underline{p(\theta \mid D)} \cdot p(D' \mid \theta)$$

Conjugate priors

$p(\theta \mid \bar{\alpha})$ - сем. сопр. апр. р. для сем. пр. $p(D \mid \theta)$, если

$\forall D \quad p(\theta \mid \bar{\alpha}) \cdot p(D \mid \theta) \propto p(\theta \mid \bar{\alpha}')$

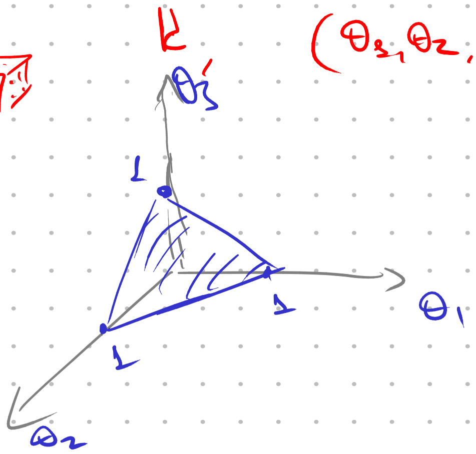
Equivalent sample size

$$\theta^{\alpha-1}(1-\theta)^{\beta-1} \qquad \frac{\alpha-1}{\alpha+\beta-2}$$

$$p(\theta) = \mathrm{Beta}(\theta \mid 11, 11) \iff p(\theta) = \mathrm{Beta}(1,1)$$

$$\underline{p(D \mid \theta)} = \theta^{10}(1-\theta)^{10}$$

$$\mathrm{Beta}\left(\theta \mid \tfrac{1}{2}, \tfrac{1}{2}\right) \propto \theta^{-\frac{1}{2}}(1-\theta)^{-\frac{1}{2}}$$



$$(\theta_1, \theta_2, \ldots, \theta_K) = \bar{\theta} \qquad 0 \le \theta_k \le 1$$

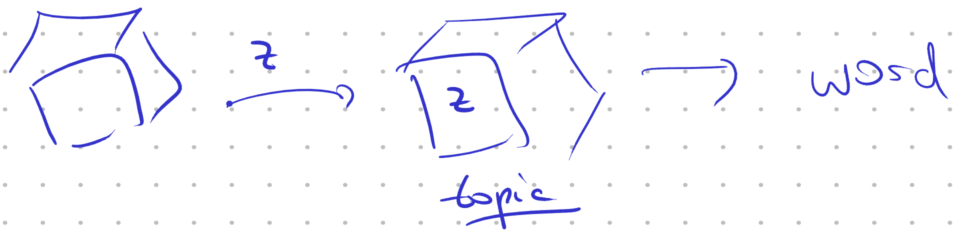
$$\theta_K = 1 - \sum_{1}^{K-1} \theta_k \qquad \sum \theta_k = 1$$

$$p(D \mid \bar{\theta}) = \theta_1^{n_1}\, \theta_2^{n_2} \ldots \theta_K^{n_K}$$

$$p(\bar{\theta} \mid \bar{\alpha}) = \frac{1}{\mathrm{Dir}(\bar{\alpha})}\, \theta_1^{\alpha_1-1}\, \theta_2^{\alpha_2-1} \ldots \theta_K^{\alpha_K-1}$$

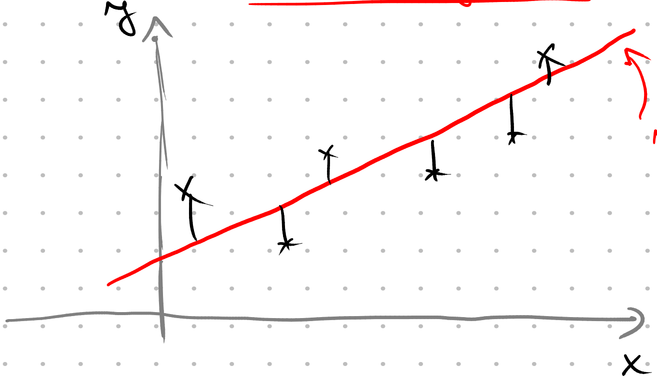
$$p(\bar{\theta} \mid D) \propto p(\bar{\theta})\, p(D \mid \bar{\theta}) \propto \theta_1^{\alpha_1+n_1-1} \ldots \theta_K^{\alpha_K+n_K-1}$$

$$\underline{p\left(\bar{\theta} \mid \left(\tfrac{1}{10}, \ldots, \tfrac{1}{10}\right)\right)} = \frac{1}{\mathrm{Dir}(\bar{\alpha})}\, \theta_1^{-\frac{9}{10}}\, \theta_2^{-\frac{9}{10}} \ldots \theta_K^{-\frac{9}{10}}$$



$$p(\theta|D) \propto p(\theta)\, p(D|\theta)$$

Linear regression



$y = w_0 + w_1 x$

$D = \{(x_n, y_n)\}_{n=1}^{N}$

$D = \{(\bar{x}_n, y_n)\}_{n=1}^{N}$

$y \sim w_0 + w_1 x_1 + \dots + w_d x_d$

$\boxed{y \sim \bar{w}^T \bar{x}}$ $= \begin{pmatrix} \bar{x} \\ 1 \end{pmatrix}$

$$L = \sum_{n=1}^{N} (y_n - \bar{w}^T \bar{x}_n)^2 \longrightarrow \min$$

$\bar{x}_i^T \bar{w}$

$$\begin{pmatrix} y_1 - \bar{w}^T \bar{x}_1 \\ \vdots \\ y_N - \bar{w}^T \bar{x}_N \end{pmatrix}^T \begin{pmatrix} y_1 - \bar{w}^T \bar{x}_1 \\ \vdots \\ y_N - \bar{w}^T \bar{x}_N \end{pmatrix} =$$

$N \times 1$ $\quad N \times d \cdot d \times 1$

$$= (\bar{y} - X\bar{w})^T (\bar{y} - X\bar{w}) = \bar{y}^T \bar{y} - (X\bar{w})^T \bar{y} - \bar{y}^T X \bar{w} + (X\bar{w})^T X \bar{w} =$$

$\bar{w}^T (X^T \bar{y})$

$$\begin{pmatrix} - x_1 - \\ - x_2 - \\ \vdots \\ - x_N - \end{pmatrix} \begin{pmatrix} \bar{w} \end{pmatrix}$$

$$= \bar{w}^T X^T X \bar{w} - 2 \bar{y}^T X \bar{w} + \cancel{\bar{y}^T \bar{y}} \xrightarrow[\bar{w}]{} \min$$

$$\nabla_{\bar{w}} L = 0$$

$$\nabla_{\bar{w}} L = \begin{pmatrix} \partial L / \partial w_1 \\ \dots \\ \partial L / \partial w_d \end{pmatrix}$$

$w_1 a_1 + \dots + w_d a_d$

$$\nabla_{\bar{w}} (\bar{w}^T \bar{a}) = \begin{pmatrix} a_1 \\ \vdots \\ a_d \end{pmatrix} = \bar{a}$$

$$\nabla_{\bar{w}} (\bar{w}^T \bar{w}) = 2\bar{w}$$

$$\nabla_{\bar{w}} (\bar{w}^T A \bar{w}) = (A + A^T)\bar{w}$$

$$\partial \left(\sum_{i=1}^{d} \sum_{j=1}^{d} a_{ij} w_i w_j \right) / \partial w_k = \sum_{j \neq k} a_{kj} w_j + \sum_{i \neq k} a_{ik} w_i + 2 a_{kk} w_k =$$

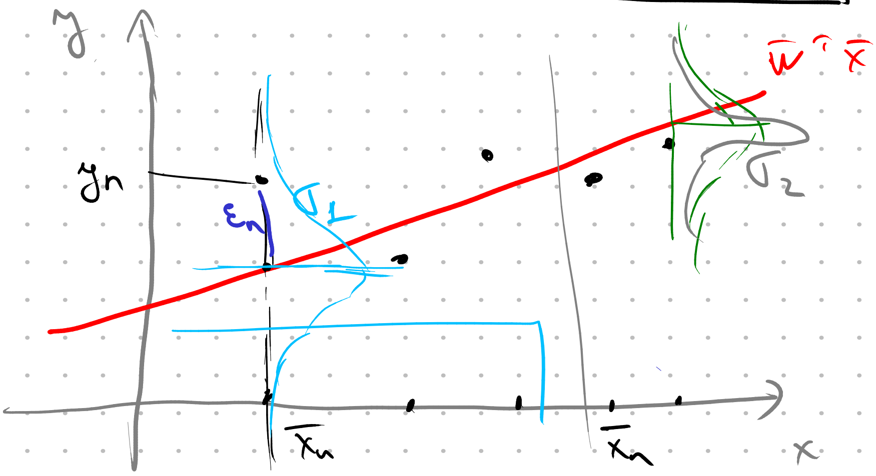
$$= \bar{a}_{k*}^T \bar{w} + \bar{a}_{*k}^T \bar{w}$$

$$\nabla_{\bar w} L = 2X^T X \bar w - 2X^T \bar y = 0$$

$$\boxed{X^T X \bar w = X^T \bar y}$$

Moore-Penrose pseudoinverse

$$\boxed{\bar w^* = (X^T X)^{-1} X^T \bar y}$$



$\cancel{p(\bar x)}$ discriminative

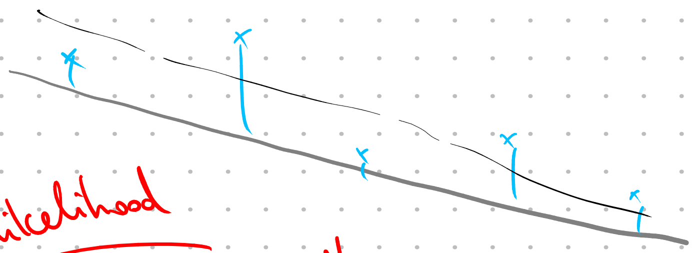
$p(y|\bar w, \bar x)$

$p(D|\bar w, X)$

" $\{y_n\}$

$y = (\bar w^T \bar x + \varepsilon)$

$\varepsilon \sim \mathcal{N}(0, \sigma^2)$



σ_n

likelihood

$$p(D|\bar w, X) = \prod_{n=1}^{N} p(y_n|\bar w, \bar x_n) = \prod_{n=1}^{N} \mathcal{N}(y_n | \bar w^T \bar x_n, \sigma^2) =$$

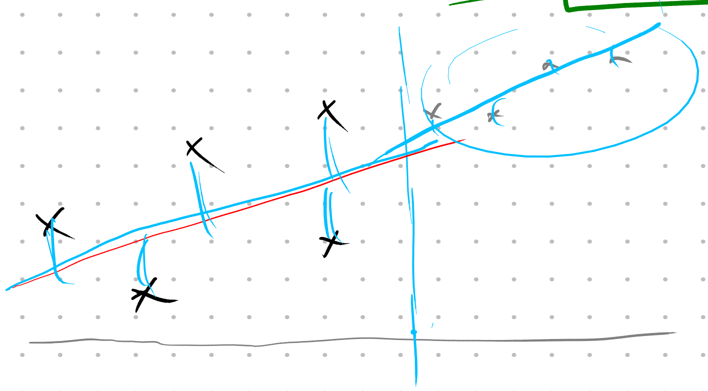
$$= \prod_{n=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(y_n - \bar w^T \bar x_n)^2} \xrightarrow[\bar w]{} \max$$

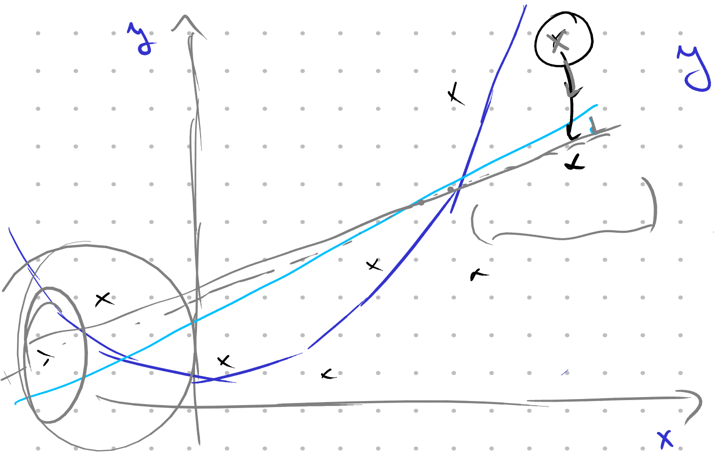
σ_n $\ll$ const σ_n

$$\ln p(D|\bar w, X) = \sum_{n=1}^{N} \left(-\frac{1}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} (y_n - \bar w^T \bar x_n)^2 \right) =$$

$$= \text{const} - \frac{1}{2\sigma^2} \boxed{\sum_{n=1}^{N} (y_n - \bar w^T \bar x_n)^2} \xrightarrow[\bar w]{} \max$$

$$\xrightarrow[\bar w]{} \min$$

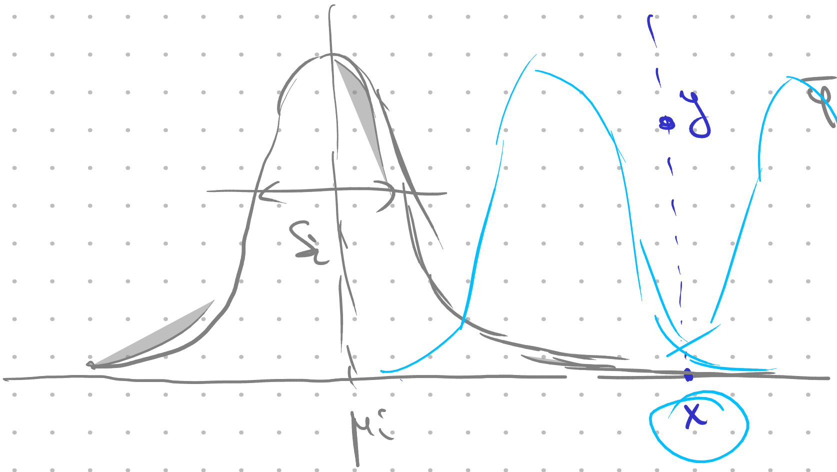




$$y \sim w_0 + w_1 x + w_2 x^2 = \begin{pmatrix} w_0 \\ w_1 \\ w_2 \end{pmatrix}^T \begin{pmatrix} 1 \\ x \\ x^2 \end{pmatrix}$$

$$y \sim \bar{w}^T \bar{\varphi}(\bar{x})$$

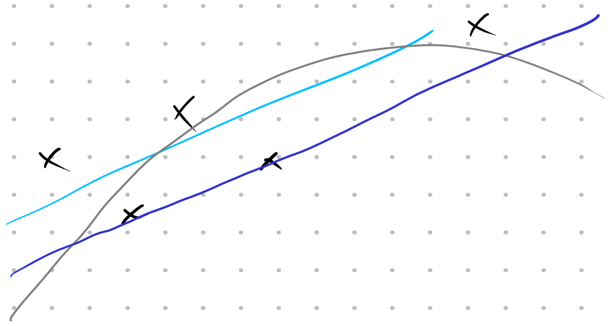
$$\bar{x} \longmapsto \bar{\varphi}(\bar{x})$$



$$\bar{\varphi} = \begin{pmatrix} \vdots \\ \varphi_i(\bar{x}) \\ \vdots \end{pmatrix}$$

$$\varphi_i(x) = e^{-\frac{1}{2s_i^2}(x-\mu_i)^2}$$

$$\boxed{y \sim \sum_i w_i \varphi_i(x)}$$



$$p(d_1, d_2|\theta) = p(d_1|\theta)\, p(d_2|\theta)$$

$$p(d_1|d_2)$$

$$p(x|\theta)$$

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

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

Model selection

$$p(D|\theta) = \prod \boxed{p(d|\theta)}$$

модель