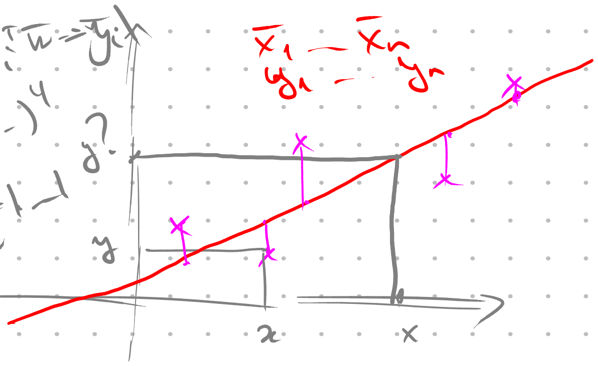
$\sum |\bar{x}_i^T \bar{w} - y_i|$

$\sum (\ldots)^4$

$\sum e^{|\ldots|}$



$\hat{y} = w_0 + w_1 x$ $\qquad \bar{x} \mapsto \begin{pmatrix} \bar{x} \\ 1 \end{pmatrix}$

$\hat{y} = w_0 + w_1 x_1 + \ldots + w_d x_d$

$\hat{y} = \bar{w}^T \bar{x}$

Метод наименьших квадратов

$$L = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y_i - \bar{w}^T \bar{x}_i)^2 \longrightarrow \min$$

$$\bar{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \qquad \underset{n \times d}{X = \begin{pmatrix} - \bar{x}_1 - \\ \vdots \\ - \bar{x}_n - \end{pmatrix}} \qquad \underset{d \times 1}{\begin{pmatrix} w_1 \\ \vdots \\ w_d \end{pmatrix}}$$

$$\begin{pmatrix} \vdots \\ y_i - \bar{w}^T \bar{x}_i \\ \vdots \end{pmatrix}^T \begin{pmatrix} \vdots \\ y_i - \bar{w}_i \bar{x}_i \\ \vdots \end{pmatrix}$$

$$\nabla_{\bar{w}} f = \begin{pmatrix} \partial f / \partial w_1 \\ \vdots \\ \partial f / \partial w_d \end{pmatrix}$$

$$L = (\bar{y} - X\bar{w})^T (\bar{y} - X\bar{w}) \xrightarrow[\bar{w}]{} \min$$

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

$$\bar{w}^T \bar{a} = w_1 a_1 + \ldots + w_d a_d \qquad \bar{w}^T \bar{w} = w_1^2 + w_2^2 + \ldots + w_d^2$$

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

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

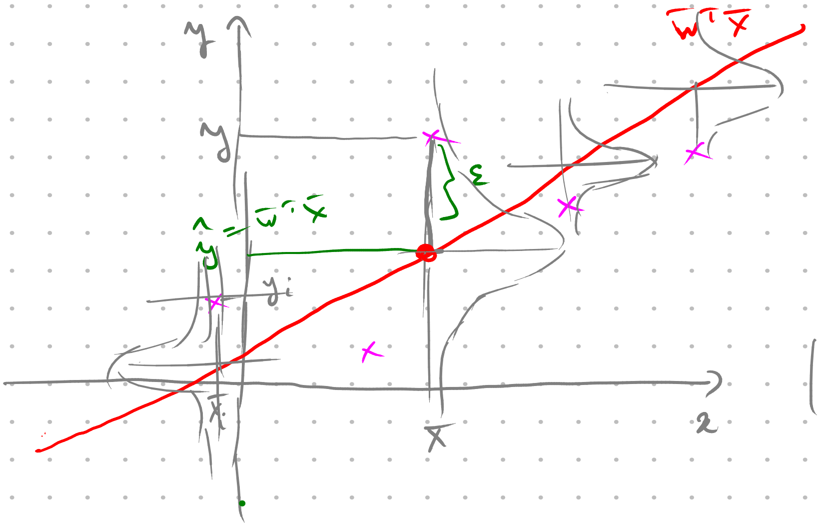
$$= 2 a_{kk} w_k + \sum_{j \neq k} a_{kj} w_j + \sum_{i \neq k} a_{ik} w_i =$$

$$= \sum_j a_{kj} w_j + \sum_i a_{ik} w_i$$

$$L = (\bar{y} - X\bar{w})^T (\bar{y} - X\bar{w}) = \bar{y}^T \bar{y} + \bar{w}^T X^T X \bar{w} - 2\bar{w}^T X^T \bar{y} - \cancel{\bar{y}^T X \bar{w}}$$

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

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



$D = \{(\bar{x}_i, y_i)\}_{i=1}^n$

$p(\bar{w} | D)$

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

$$p(D|\bar{w}) = p(\bar{y} | \bar{w}, X) = \prod_{i=1}^{n} p(y_i | \bar{w}, \bar{x}_i)$$

$\cancel{p(y|\bar{w}, \bar{x}) = \bar{w}^T\bar{x}} + p(\varepsilon | \bar{w}, \bar{x})$

$\cancel{p(y|\bar{w}, \bar{x}) = \bar{w}^T\bar{x} + \mathcal{N}(\varepsilon | 0, \sigma^2)}$

$p(\bar{w})$

$p(\bar{w}|D) \propto p(\bar{w})\, p(D|\bar{w})$

$y = \bar{w}^T\bar{x} + \varepsilon, \quad \varepsilon \sim \mathcal{N}(\varepsilon | 0, \sigma^2)$

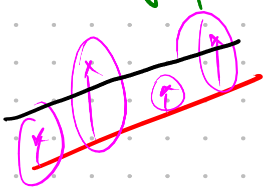
$$p(y|\bar{x}, D) = \int p(y|\bar{w}, \bar{x})\, p(\bar{w}|D)\, d\bar{w}$$

$$p(y|\bar{w}, \bar{x}) = \mathcal{N}(y | \bar{w}^T\bar{x}, \sigma^2)$$

$$p(\bar{y}|\bar{w}, X) \overset{3}{=} \prod_i p(y_i | \bar{w}, \bar{x}_i) \overset{1}{=} \prod_i \mathcal{N}(y_i | \bar{w}^T\bar{x}_i, \sigma^2) =$$

$$= \prod_{i=1}^{n} \left(\frac{1}{\sqrt{2\pi}\cdot\sigma}\right) \cdot e^{-\frac{1}{2\sigma^2}(y_i - \bar{w}^T\bar{x}_i)^2} \xrightarrow[\bar{w}]{} \max \quad \bar{w}_{ML}$$

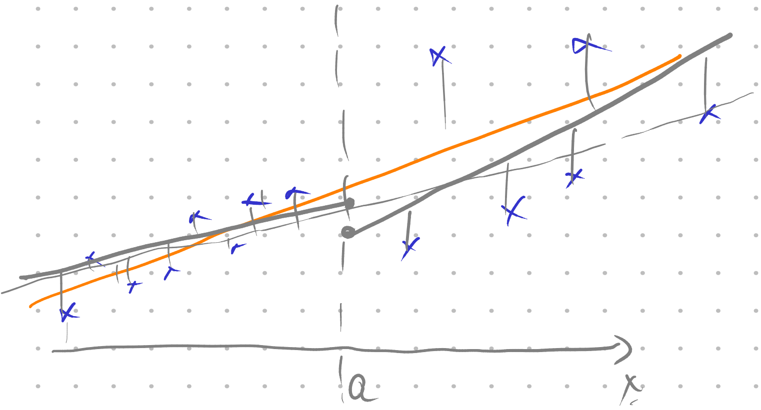
$$\log p(\bar{y}|\bar{w}, X) = \underbrace{-n \cdot \log\left(\sqrt{2\pi\sigma^2}\right)} - \sum_{i=1}^{n} \frac{1}{2\sigma^2}(y_i - \bar{w}^T\bar{x}_i)^2 \xrightarrow[\bar{w}]{} \max$$



$$\overset{4}{=} \text{const} - \frac{1}{2\sigma^2} \boxed{\sum_{i=1}^{n} (y_i - \bar{w}^T\bar{x}_i)^2} \xrightarrow[\bar{w}]{} \max$$

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

1. Гауссовский шум
2. Линейность $\bar{w}^T\bar{x}$
3. Независимость $y_i | \bar{w}$ ✓
4. Постоянная σ^2 — гомоскедастичность homoscedacity

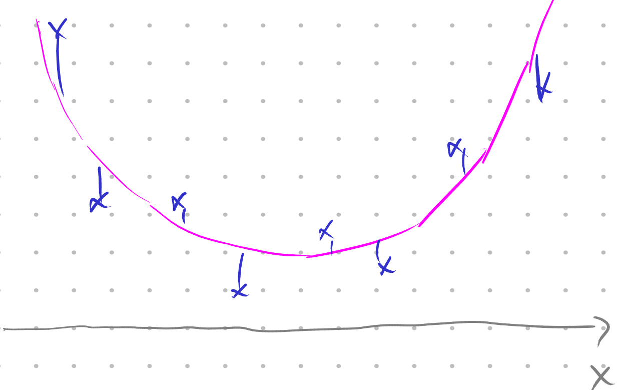


$$\log p(\bar{y} \mid \bar{w}, X) =$$

$$= \text{Const} - \frac{1}{\sigma_1^2} \sum_{x_i < a} (y_i - \bar{x}_i^T \bar{w})^2 - \frac{1}{\sigma_2^2} \sum_{x_i > a} (y_i - \bar{x}_i^T \bar{w})^2 \xrightarrow[\bar{w}]{} \max$$

$$\frac{\sigma_1^2}{\sigma_2^2}$$

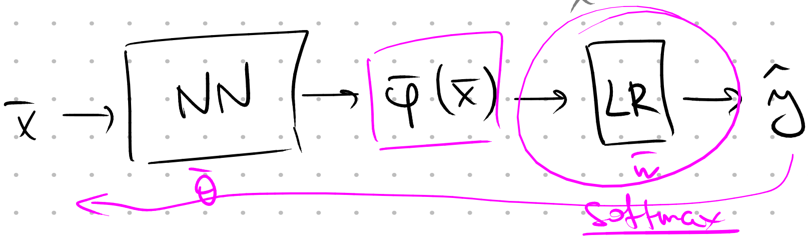
$$\hat{y} = w_0 + w_1 x + w_2 x^2 = \begin{pmatrix} w_0 \\ w_1 \\ w_2 \end{pmatrix}^T \begin{pmatrix} 1 \\ x \\ x^2 \end{pmatrix}$$

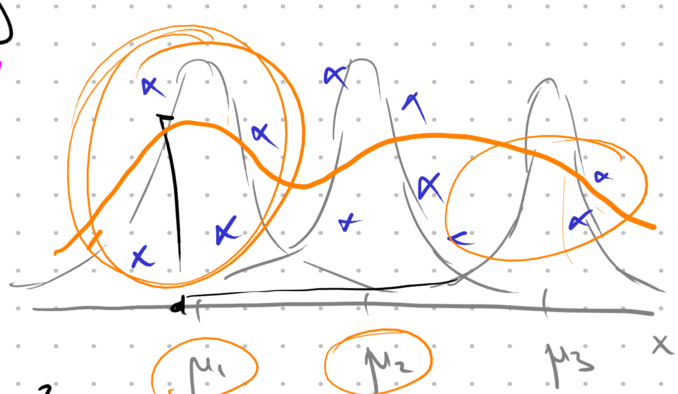


features

$$p(y \mid \bar{x}, \bar{w}) = \mathcal{N}(y \mid \bar{w}^T \bar{\varphi}(\bar{x}), \sigma^2)$$

$$L = \sum_i (y_i - \bar{w}^T \bar{\varphi}(\bar{x}_i))^2 \xrightarrow[\bar{w}]{} \min$$



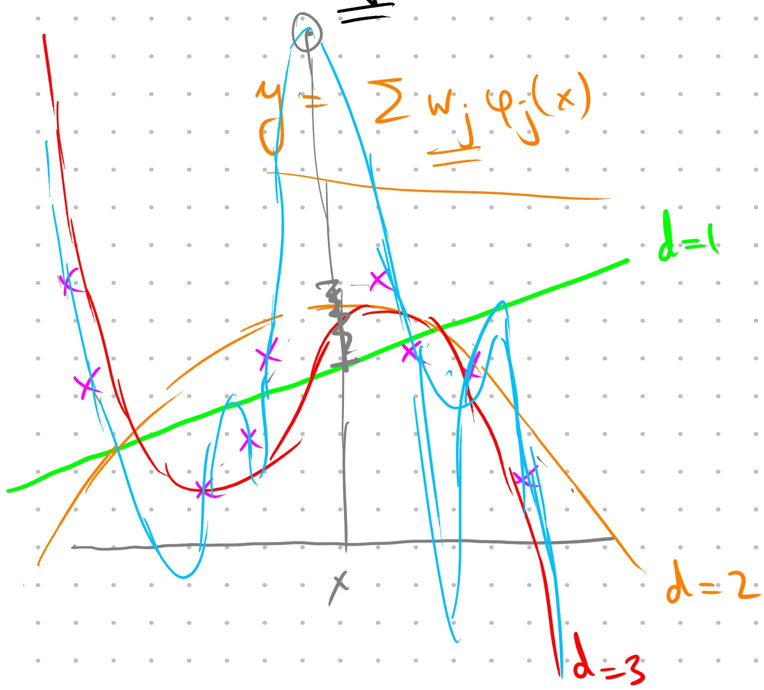


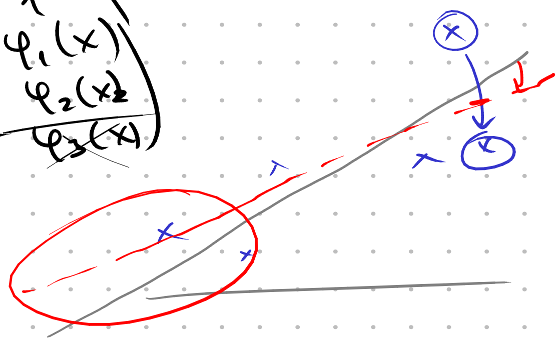
Local features

RBF – radial basis functions

$$\varphi_j(x) = c \cdot e^{-c' \cdot (x - \mu_j)^2}$$

$$x \longmapsto \begin{pmatrix} 1 \\ \varphi_1(x) \\ \varphi_2(x) \\ \varphi_3(x) \end{pmatrix}$$





$d=1$ $\quad \hat{y} = w_0 + w_1 x$

$$\hat{y} = w_0 + w_1 x + w_2 x^2$$

$$\hat{y} = \ldots + w_3 x^3$$

$$\hat{y} = \ldots + w_4 x^4$$

$$\vdots$$

$$\hat{y} = \ldots + w_7 x^7$$

overfitting