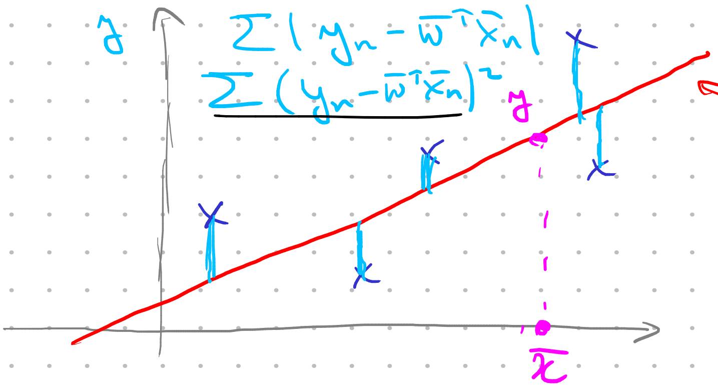


$D = \{(\bar x, y)\}$ $\quad \bar x \in \mathbb{R}^d$, $y \in \mathbb{R}$

$$\hat y = \bar w^T \bar x + \cancel{w_0} = w_0 + w_1 x_1 + \dots + w_d x_d$$

$$\bar x \mapsto \begin{pmatrix} \bar x \\ 1 \end{pmatrix}$$

Метод наименьших квадратов

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

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

$N \times d$

$$\bar y = \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix} \quad \bar w = \begin{pmatrix} w_1 \\ \vdots \\ w_d \end{pmatrix} \quad X = \begin{pmatrix} - \bar x_1 - \\ \vdots \\ - \bar x_N - \end{pmatrix}$$

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

$$\boxed{\bar a = \nabla_{\bar w} (\bar w^T \bar a)} = \begin{pmatrix} \partial(\bar w^T \bar a)/\partial w_1 \\ \dots \\ \partial \dots / \partial w_d \end{pmatrix} = \begin{pmatrix} a_1 \\ \vdots \\ a_d \end{pmatrix} = \bar a$$

$w_1 a_1 + \dots + w_d a_d$

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

$w_1^2 + \dots + w_d^2$

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

$$\nabla_w (\bar w^T A \bar w) = 2A\bar w$$

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

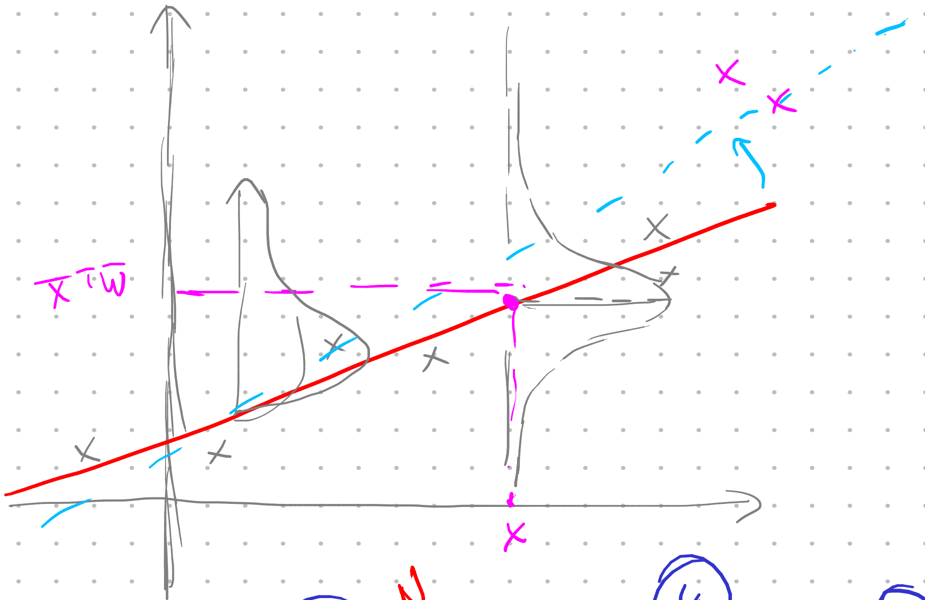
$$= 2 a_{kk} w_k + \sum_{i \neq k} a_{ik} w_i + \sum_{j \neq k} a_{kj} w_j = \sum_i a_{ik} w_i + \sum_j a_{kj} w_j$$

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

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

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

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



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

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

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

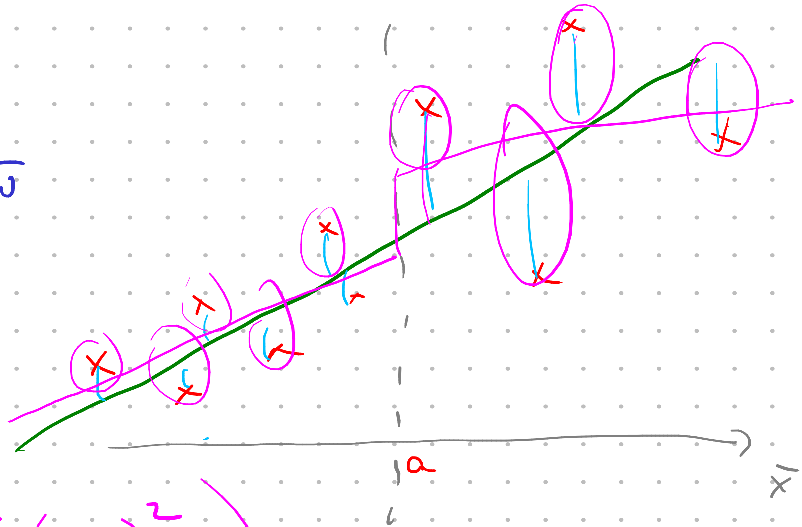
$$= \prod_{n=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{1}{2\sigma^2}(y_n - \bar{w}^T\bar{x}_n)^2} \xrightarrow{\bar{w}} \max$$

$$\log p(D \mid \bar{w}, X) = \sum_{n=1}^{N} \left[-\frac{1}{2}\cancel{(\log 2\pi\sigma^2)} - \frac{1}{2\sigma^2}(y_n - \bar{w}^T\bar{x}_n)^2 \right]$$

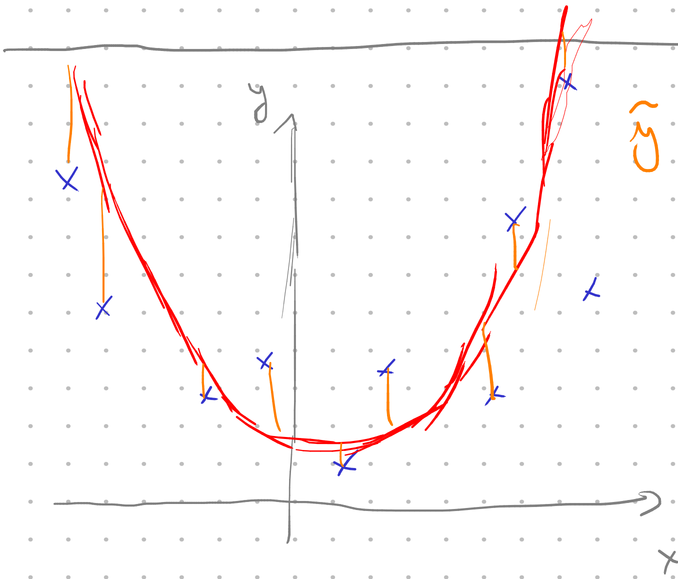
$$= \text{Const} - \frac{1}{2\sigma^2} \boxed{\sum_{n=1}^{N} (y_n - \bar{w}^T\bar{x}_n)^2} \xrightarrow{\bar{w}} \max$$

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

1. Гауссовский шум
2. Постоянная σ^2
3. y_n усл. незав. при усл. $\bar{w}$
4. линейность $\hat{y} = \bar{w}^T \bar{x}$



$$\log p(D|\bar{w}) = \text{Const} - \left(\frac{1}{2\sigma_1^2} \sum_{x<a} (\ -\ )^2 + \boxed{\frac{\sigma_1^2}{\sigma_2^2}} \sum_{x \geq a} (\ -\ )^2 \right) \to \min$$



$\hat{y} \approx y$

$$\sum_n (y_n - \hat{y}_n)^2 \to \min$$

$$\sum_n \left(y_n - \underline{(w_0 + w_1 x_n + w_2 x_n^2)} \right)^2 \to \min$$

$$x \mapsto \begin{pmatrix} 1 \\ x \\ x^2 \end{pmatrix}^T \begin{pmatrix} w_0 \\ w_1 \\ w_2 \end{pmatrix}$$

lin. regr.
softmax

$\bar{x}$ — NN — $\bar{\varphi}(\bar{x})$ — [ ] — y

features

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

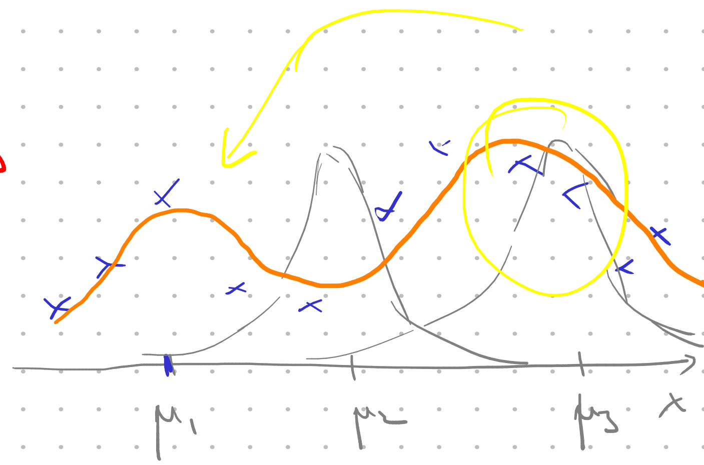
$$p(y|\bar{x}, \bar{w}) = \mathcal{N}(y \,|\, \bar{w}^T \underline{\bar{\varphi}(\bar{x})}, \sigma^2)$$

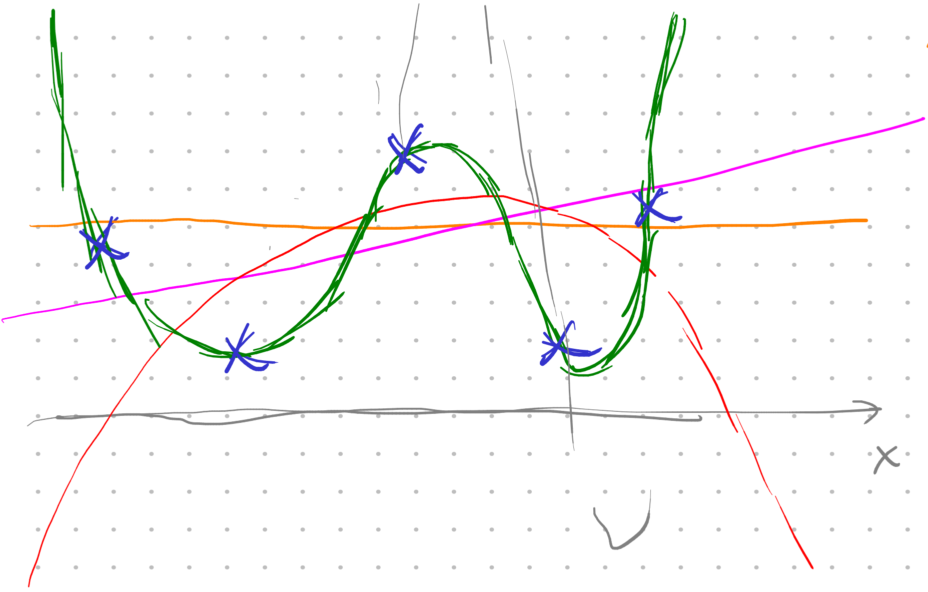
Local features

RBF – radial basis functions

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

$$x \mapsto \begin{pmatrix} \varphi_1(x) \\ \vdots \\ \varphi_m(x) \end{pmatrix}$$





$\hat{y} = w_0$

$\hat{y} = w_0 + w_1 x$

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

$\hat{y} = w_0 + w_1 x + \dots + w_4 x^4$

Overfitting

Regularization

regularizer

$$L = (\bar{y} - X\bar{w})^T (\bar{y} - X\bar{w}) + \lambda \cdot \sum_{i=1}^{d} w_i^2 \xrightarrow{\bar{w}} \min$$

$$+ \lambda \sum_i |w_i|$$

$$+ \lambda \sum_{i=1}^{d} w_i^4$$

param

$$p(D|\theta) = \prod_{n=1}^{N} p(d_n|\theta)$$