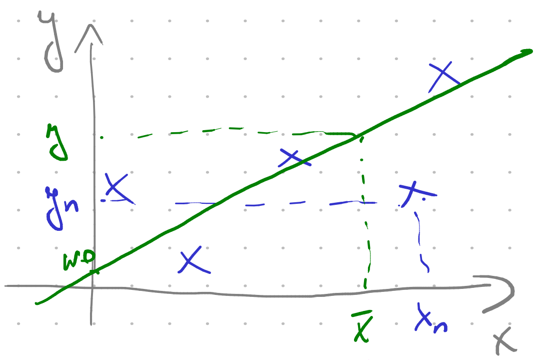
1 Linear regression



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

$$\hat{y} = w_1 x_1 + w_2 x_2 + \ldots + w_d x_d = \bar{w}^T \bar{x}$$

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

$$X\bar{w}$$

– МНК

least squares

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

$$\nabla_{\bar{w}} L \qquad \bar{w} = \begin{pmatrix} w_1 \\ \vdots \\ w_d \end{pmatrix}, \quad \bar{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix}, \quad X = N\begin{pmatrix} \text{---} \\ \text{---}\ \bar{x}_n\ \text{---} \\ \text{---} \end{pmatrix}_{d}$$

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

$$\nabla_{\bar{w}} (\bar{w}^T \bar{a}) = \left(\ldots \frac{\partial \bar{w}^T \bar{a}}{\partial w_i} \ldots \right) = \bar{a} \qquad \text{“}a_i\text{”}$$

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

$$\nabla_{\bar{w}} (\bar{w}^T A \bar{w}) = A\bar{w} + A^T \bar{w} \quad / \quad 2A\bar{w}$$

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

$$= 2a_{kk} w_k + \sum_{j \neq k} a_{kj} w_j + \sum_{i \neq k} a_{ik} w_i = \underbrace{\sum_j a_{kj} w_j}_{(A^T \bar{w})_k} + \underbrace{\sum_i a_{ik} w_i}_{(A \cdot \bar{w})_k}$$

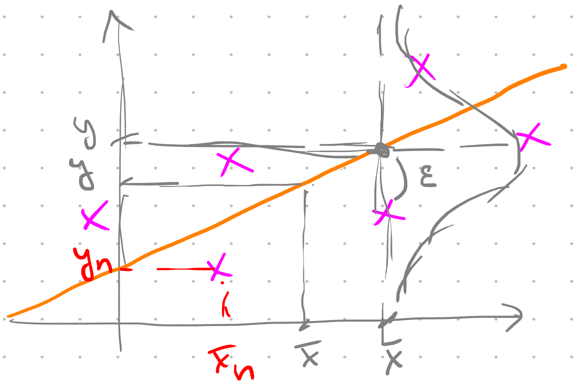
$$L = \bar{y}^T \bar{y} - (X\bar{w})^T \bar{y} - \bar{y}^T (X\bar{w}) + \bar{w}^T X^T X \bar{w} =$$

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

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

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

② Предположения лин. регрессии



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

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

$$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(y_n - \bar w^T \bar x_n)^2}$$

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

$$\log p(\bar y|X, \bar w) = \sum_{n=1}^{N}\left[-\frac{1}{2}\log(2\pi\sigma^2) - \overset{④}{\frac{1}{2\sigma^2}}(y_n - \bar w^T \bar x_n)^2\right] =$$

$\max_{\bar w}$

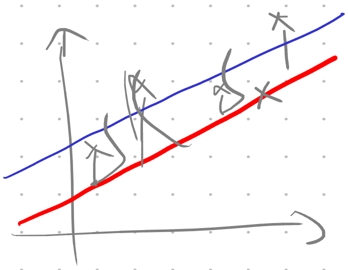
$$= -\frac{N}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{n=1}^{N}(y_n - \bar w^T \bar x_n)^2 \xrightarrow[\bar w]{} \min$$

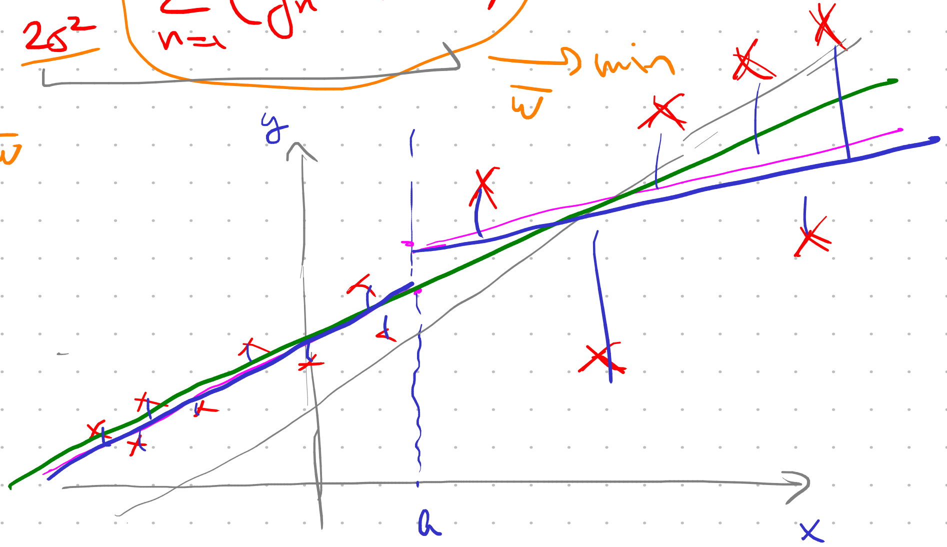
① y_n усл. нез при усл. $\bar w$

② $y \approx \bar w^T \bar x$

③ Гауссовский шум

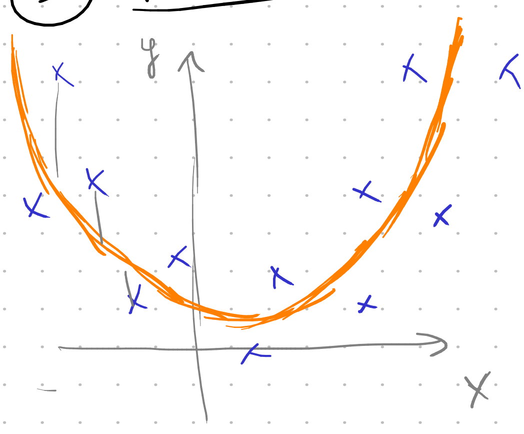
④ $\sigma_n = \sigma$





$$\frac{1}{\sigma_1^2}\sum_{n: x_n < a}(\ldots)^2 + \frac{1}{\sigma_2^2}\sum_{n: x_n \geq a}(\ldots)^2 \xrightarrow[\bar w]{} \min$$

③ Feature extraction



$y \sim w_0 + w_1 x \qquad x \mapsto \begin{pmatrix} 1 \\ x \end{pmatrix}$

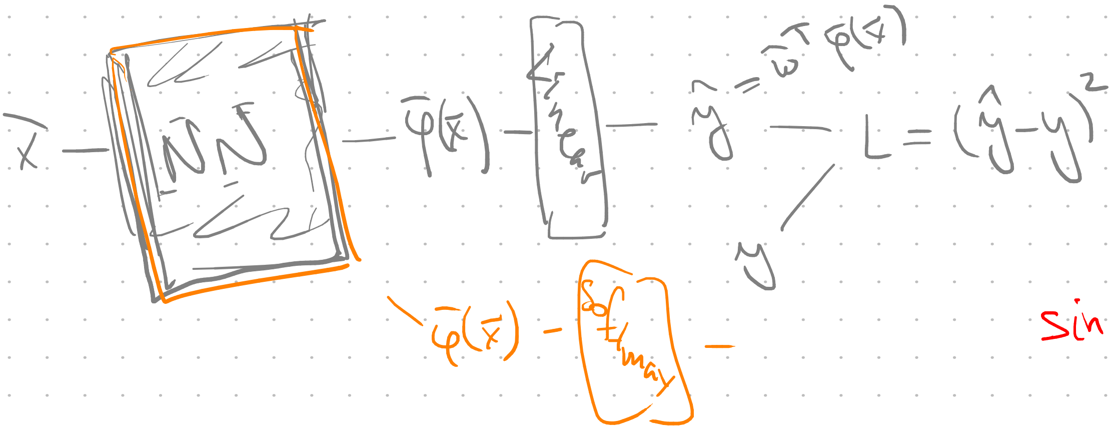
$y \sim w_0 + w_1 x + w_2 x^2$

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

Polynomial regression

$$\sum_n \left(y_n - (w_0 + w_1 x_n + w_2 x_n^2)\right)^2 \xrightarrow[\bar w]{} \min$$

$\bar{x} \sim \boxed{\bar{\varphi}} \sim \bar{\varphi}(\bar{x}) \sim \boxed{LR}$



$\{1 \ldots K\}$

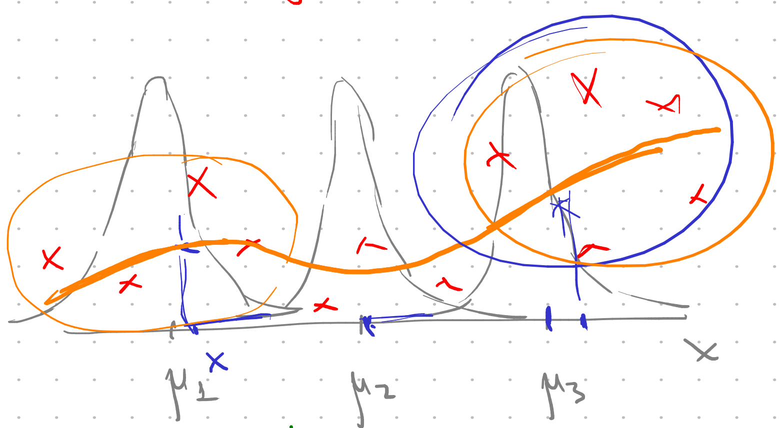
$\sin(\underline{k} \cdot x)$

Local features / RBF – radial basis functions

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

$$x \mapsto \begin{pmatrix} \varphi_1(x) \\ \cdots \\ \varphi_m(x) \end{pmatrix}^T \begin{pmatrix} w_1 \\ \vdots \\ w_m \end{pmatrix}$$

$$\hat{y} = \sum_{j=1}^{m} w_j \varphi_j(x)$$

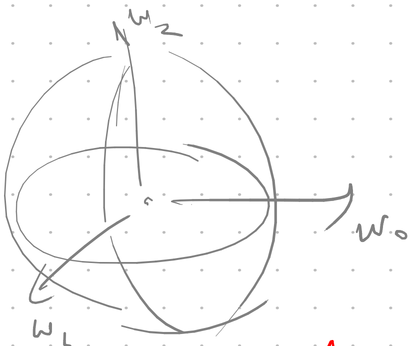


$$\begin{pmatrix} x_1 \\ \vdots \\ x_d \end{pmatrix} \mapsto \begin{pmatrix} 1 \\ x_1 \\ \vdots \\ x_d \\ x_1^2 \\ \vdots \\ x_d^2 \end{pmatrix}$$

$\frac{d(d+1)}{2}$

k $\quad O(d^k)$

④ Regularization



$$L = \sum_n (y_n - \bar{w}^T \bar{x}_n)^2 + \alpha \cdot \underline{\sum_i w_i^2}$$

$= \|\bar{w}\|_2^2$

L_2-regularization
ridge regression

$+ \alpha \cdot \sum_i |w_i|$

L_1-reg.
Lasso
regression

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

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

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

$p(\bar{w}) \propto \ldots e^{-\alpha \cdot \sum |w_i|}$

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

likelihood

$$p(\bar{w}) = \mathcal{N}(\bar{w} \mid \bar{0}, \sigma_0^2 I) = \frac{1}{(\sqrt{2\pi})^d \det(\sigma_0^2 I)} \cdot e^{-\frac{1}{2\sigma_0^2} \bar{w}^T \bar{w}}$$

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

$$= \text{Const} - \frac{1}{2\sigma^2} \sum_{n=1}^{N} (y_n - \bar{w}^T \bar{x}_n)^2 - \frac{1}{2\sigma_0^2} \bar{w}^T \bar{w}$$

$$= \text{Const} - \frac{1}{2} \Big(\underbrace{\sum_{n=1}^{N} (y_n - \bar{w}^T \bar{x}_n)^2 + \frac{\sigma^2}{\sigma_0^2} \bar{w}^T \bar{w}}_{\bar{w} \to \min} \Big)$$