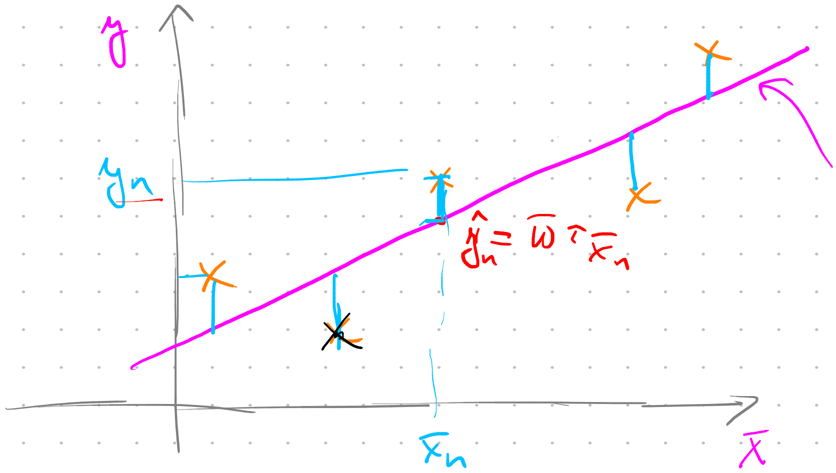


$$D = \{(\bar{x}_n, y_n)\}_{n=1}^{N}$$

$$\hat{y} = \bar{w}^T \bar{x} + \cancel{w_0} =$$

$$= \cancel{w_0 \cdot 1} + w_1 x_1 + \ldots + w_d x_d$$

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

Метод наименьших квадратов

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

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

$$L = (\bar{y} - \underset{N \times d}{X}\bar{w})^T (\bar{y} - X\bar{w}) \qquad \nabla_{\bar{w}} L = ?$$

$$\sum w_i a_i$$

$$\nabla_{\bar{w}} (\bar{w}^T \bar{a}) = \begin{pmatrix} \vdots \\ \partial(\bar{w}^T \bar{a}) / \partial w_i \\ \vdots \end{pmatrix} = \begin{pmatrix} \vdots \\ a_i \\ \vdots \end{pmatrix} = \bar{a}$$

$$\nabla_{\bar{w}} (\bar{w}^T \bar{w}) = \nabla_{\bar{w}} (w_1^2 + \ldots + w_d^2) = 2\bar{w}$$

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

A — симм.

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

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

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

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

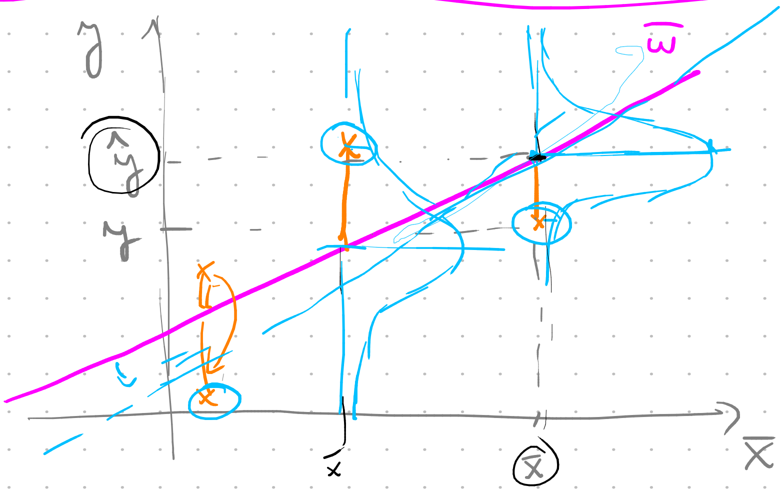
$$= \bar{y}^T \bar{y} - 2\bar{w}^T (X^T \bar{y}) + \bar{w}^T (X^T X) \bar{w}$$

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

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

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

$$p(D|\bar\theta) = \prod_n p(d_n|\bar\theta)$$



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

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

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

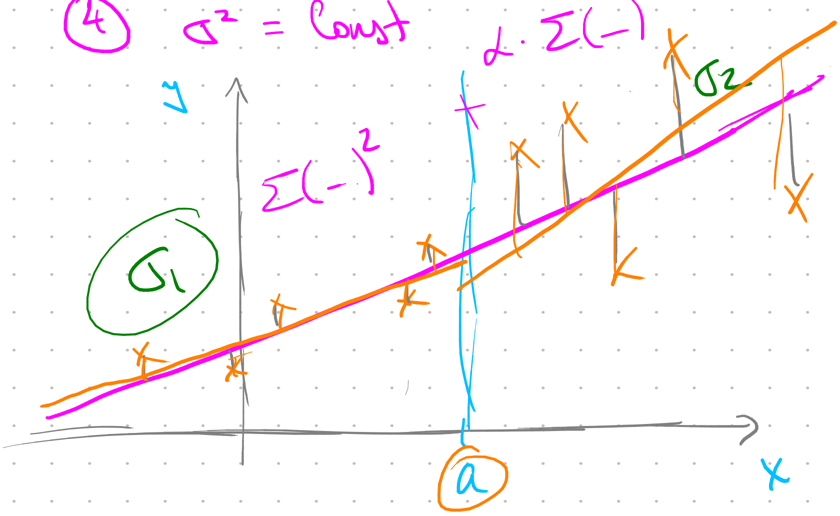
$$\log p(\bar y|X,\bar w) = \sum_{n=1}^N\left(\overbrace{-\frac{1}{2}\log(2\pi\sigma^2)}^{const} - \frac{1}{2\sigma^2}(y_n-\bar w^T\bar x_n)^2\right) =$$

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

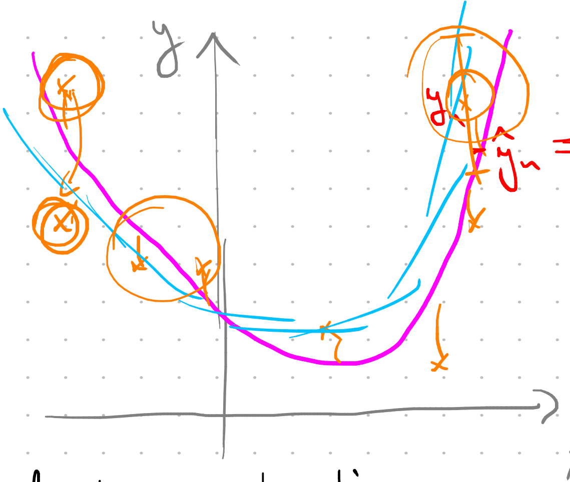
$$\xrightarrow[\bar w]{} \min$$

1. y_n усл.-нез. при усл. $\bar w$
2. y лин. зав. от $\bar x$ $\quad y \approx \bar w^T\bar x$
3. Гауссовский шум
4. $\sigma^2 = const$

$$\frac{1}{\sigma_1^2}\sum_n(-)^2 + \frac{1}{\sigma_2^2}\sum_n(-)^2 \xrightarrow[\bar w]{} \min$$



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



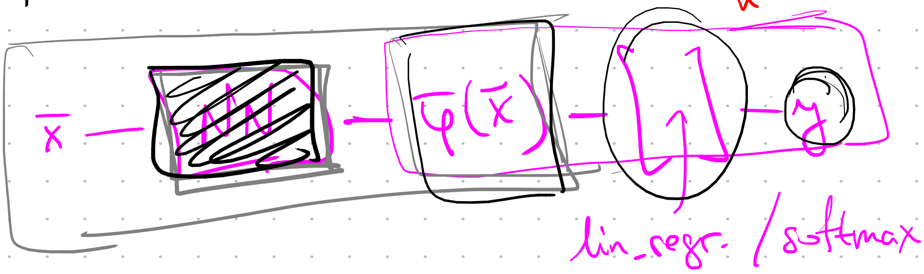
polynomial regression

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

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

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

feature extraction



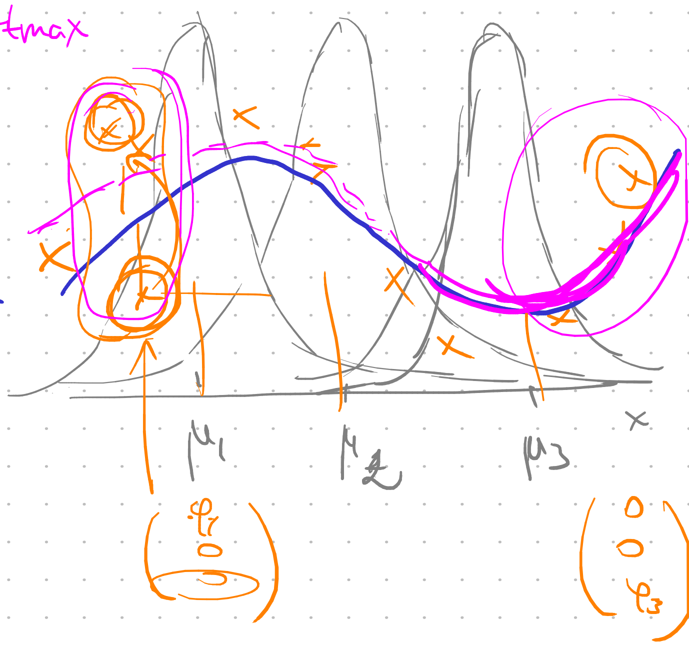
$$L = (y - \hat{y})^2$$

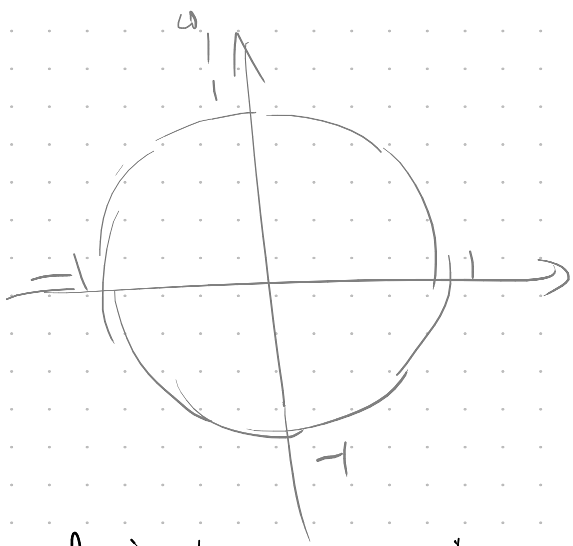
Local features

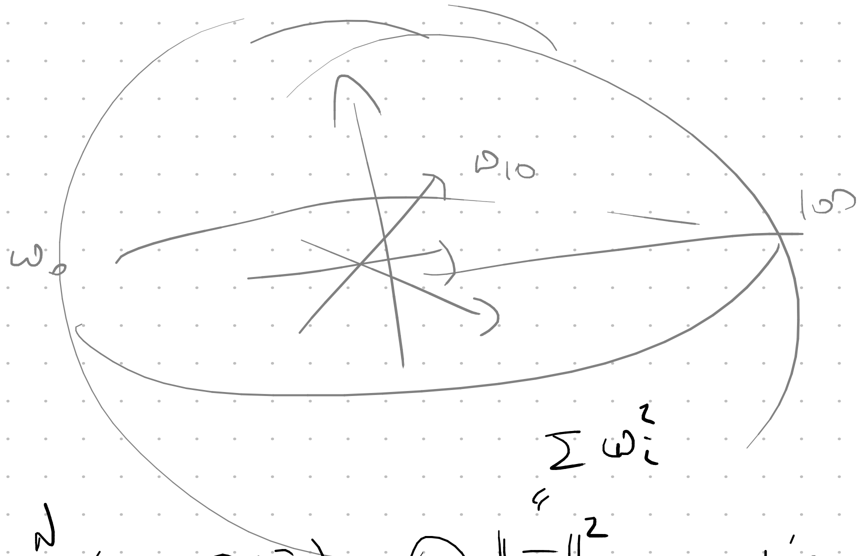
RBF – radial basis functions

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

$$x \longmapsto \begin{pmatrix} 1 \\ \varphi_1(x) \\ \dots \\ \varphi_m(x) \end{pmatrix}^T \begin{pmatrix} w_0 \\ w_1 \\ \vdots \\ w_m \end{pmatrix}$$







$$\sum w_i^2$$

Regularization

$$L = \sum_{n=1}^{N} (y_n - \bar{w}^T \bar{x}_n) + \alpha \cdot \|\bar{w}\|^2 \xrightarrow[\bar{w}]{} \min$$

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

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