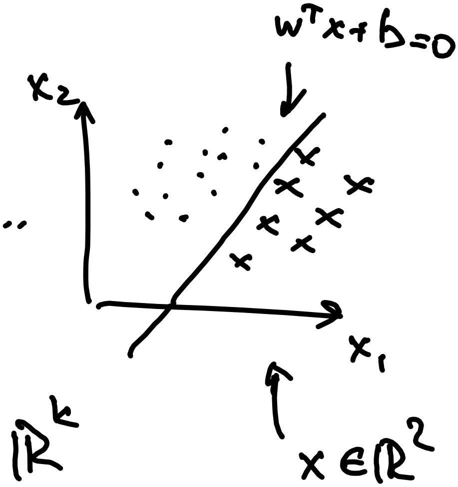
Линейные методы классификации.

$a(x) = \mathrm{sign}(\langle w, x \rangle + b)$

$w, x \in \mathbb{R}^d, \; b \in \mathbb{R}$



$y_i = \beta_0 + \beta_1 x_i^{(1)} + \beta_2 x_i^{(2)} + \ldots + \beta_k x_i^{(k)}$

$\downarrow$

$\langle \vec{\beta}, x_i \rangle + \beta_0, \quad \vec{\beta} \in \mathbb{R}^k$

$x_i = (x_i^{(0)}, x_i^{(1)}, \ldots, x_i^{(k)}) \in \mathbb{R}^{k+1}$

$\rightarrow \underline{\langle \vec{\beta}, \vec{x} \rangle}$

$\underbrace{w^T x + b}_{} = 0 \rightarrow \mathbb{R}^{d-1}$

$(w_1 x_1 + w_2 x_2 + \ldots + w_d x_d + b = 0$

$x_1 = -\frac{w_2}{w_1} x_2 - \ldots - \frac{w_d}{w_1} x_d - b$

w, b?

$$Q = \frac{1}{n}\sum_{i=1}^{n}[a(x_i) \neq y_i] =$$

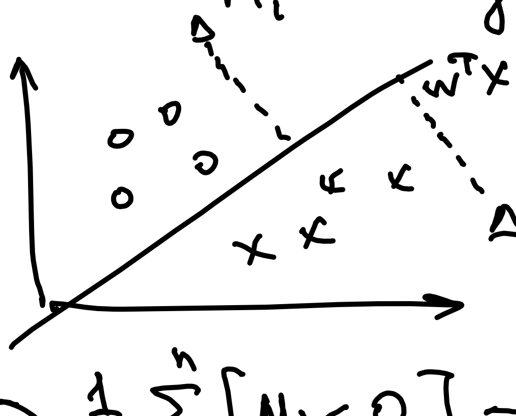
$y_i \in \{-1, 1\}$

$$= \frac{1}{n}\sum_{i=1}^{n}[\text{sign}(\langle w, x_i\rangle) \neq y_i] =$$

$$= \frac{1}{n}\sum_{i=1}^{n}[y_i \cdot \text{sign}(\langle w, x_i\rangle) \neq \underbrace{y_i^2}_{1}] =$$

$$= \frac{1}{n}\sum_{i=1}^{n}[\underbrace{y_i \cdot \text{sign}(\langle w, x_i\rangle) < 0}_{y_i \langle w, x_i\rangle < 0}] =$$

$$= \frac{1}{n}\sum_{i=1}^{n}[\underbrace{y_i \cdot \langle w, x_i\rangle}_{M_i \text{ — margin (отступ)}} < 0] =$$

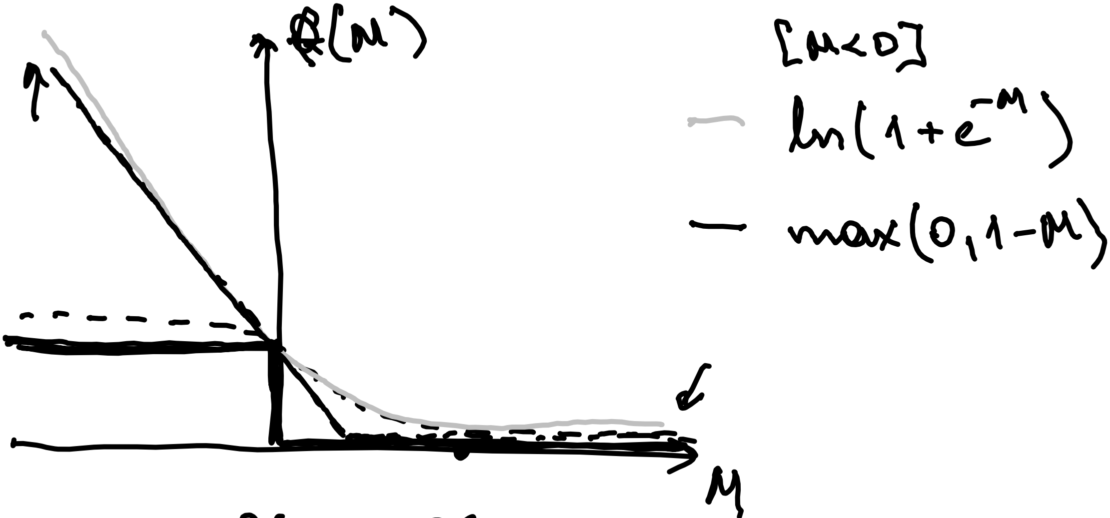


$x \to -1$

$o \to 1$

$\Delta \to M = 100$

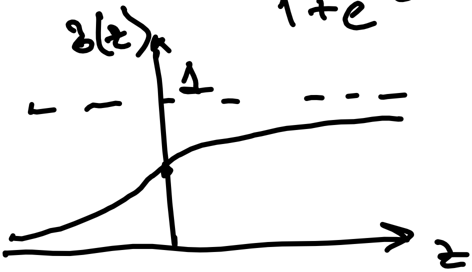
$$= \frac{1}{n}\sum_{i=1}^{n}[M_i < 0] \to \min_w$$



$$Q \le \tilde{Q} \qquad \tilde{Q} \to \min \Rightarrow Q \to \min$$

$$a(x) = \sigma(\langle w, x \rangle) = \frac{1}{1+e^{-\langle w, x \rangle}}$$

$$\sigma(z) = \frac{1}{1+e^{-z}}, \quad P(y=+1|x) = a(x)$$



$$\frac{1}{1+e^{-\langle w, x \rangle}} = Const$$

....

$$\langle w, x \rangle = Const_2$$

$$L = \prod_{i=1}^{n} a(x_i)^{[y_i=1]} \cdot (1 - a(x_i))^{[y_i=-1]} \Rightarrow$$

$$\Rightarrow \log L = \sum_{i=1}^{n} \Big([y_i = 1] \cdot \log a(x_i) +$$

$$+ [y_i = -1] \cdot \log (1 - a(x_i)) \Big) =$$

$$= \sum_{i=1}^{n} \Big(-[y_i = 1] \cdot \log (1 + \exp(-\langle w, x_i \rangle)) +$$

$$+ [y_i = -1] \cdot \log \Big(\underbrace{\frac{\exp(-\langle w, x_i \rangle)}{1 + \exp(-\langle w, x_i \rangle)}}_{\frac{1}{1 + \exp(\langle w, x_i \rangle)}} \Big) \Big) =$$

$$= - \sum_{i=1}^{n} \Big([y_i = 1] \cdot \log (1 + \exp(-\langle w, x_i \rangle)) +$$

$$+ [y_i = -1] \cdot \log (1 + \exp(\langle w, x_i \rangle)) \Big) =$$

$$= - \sum_{i=1}^{n} \log (1 + \exp(-y_i \langle w, x_i \rangle)) \to \max_{w}$$

$$\tilde{Q} = \frac{1}{n}\sum_{i=1}^{n} \log\left(1+\exp\left(-\underbrace{y_i \langle w, x_i \rangle}_{\text{Margin}_i}\right)\right) \to \min_w$$

$$\log\left(1+e^{-M}\right)$$

$$a(x) = \sigma(\langle w, x \rangle)$$

$$\nabla_w \tilde{Q} = 0 \not\Rightarrow \text{аналитич. реш.}$$

$$w_{k+1} = w_k - \alpha \cdot \nabla_w \tilde{Q}(w_k) \quad (GD)$$

$$w_{k+1} = w_k - \alpha \cdot H^{-1} \cdot \nabla_w \tilde{Q}(w_k) \quad (NW)$$

$$H = \left(\frac{\partial^2 \tilde{Q}}{\partial w_k^{(i)} \partial w_k^{(j)}} \right)$$

$$\nabla_w \langle w, x_i \rangle = x_i \qquad \nabla_a \left(\underset{1\times n}{a^T} \underset{n\times 1}{x} \right) = x$$

$$\nabla_a \left(a^T \underset{n\times n}{X} \underset{n\times 1}{a} \right) = (X + X^T)\, a$$

$$\nabla_w \tilde{Q} = \frac{1}{n}\sum_{i=1}^{n} \frac{-\exp(-y_i \langle w, x_i \rangle)\, y_i x_i}{\underbrace{1 + \exp(-y_i \langle w, x_i \rangle)}} =$$

$$= \frac{1}{n}\sum_{i=1}^{n} -y_i x_i \cdot \frac{1}{1 + \exp(y_i \langle w, x_i \rangle)} =$$

$$= \boxed{\frac{1}{n}\sum_{i=1}^{n} -y_i x_i \cdot \sigma(-y_i \langle w, x_i \rangle)}$$

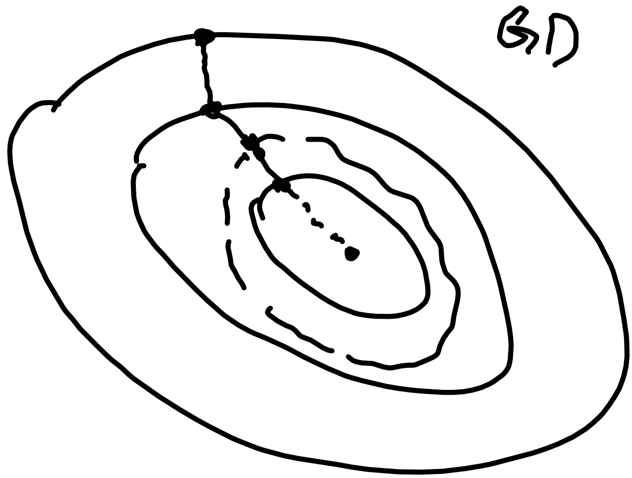
$$\tilde{Q} = \frac{1}{n}\sum_{i=1}^{n} \log(1 + \exp(-y_i \langle w, x_i \rangle)) +$$

$$+ \underbrace{\frac{\lambda}{2}\|w\|_2^2}_{l_2\text{ - регуляризация.}} \to \min_w \quad \left(\|w\|_2^2 = w^T w = \sum_{j=1}^{d} w_j^2\right)$$

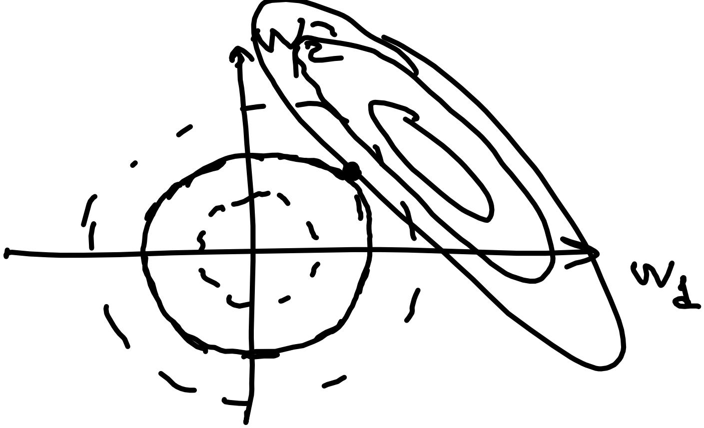
$$\nabla_w \tilde{Q} = \frac{1}{n}\sum_{i=1}^{n}\Big(\underset{\in \mathbb{R}}{-y_i}\, \underset{\in \mathbb{R}^d}{x_i} \cdot \sigma\underbrace{(-y_i \langle w, x_i \rangle)}_{\in \mathbb{R}}\Big) + \frac{\lambda}{\cancel{2}}\underset{\in \mathbb{R}^d}{\cancel{2}w}$$

$$\tilde{\nabla}_w \tilde{Q} = \frac{1}{N} \cdot \sum_{i \in N} \dots \qquad N = \{1, 3, 15, 105, \dots\}$$

$|N| =$ Batch size.


GD

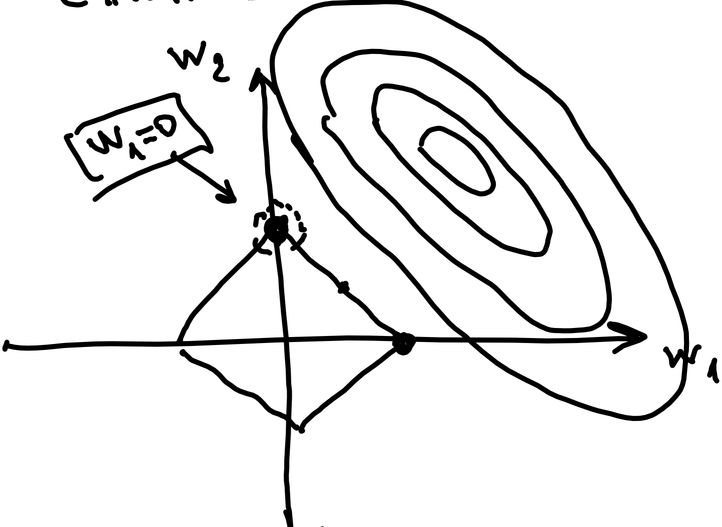

SGD



$w \in \mathbb{R}^2$

$\|w\|^2$

$$\begin{cases} f \to \min \\ \|w\|^2 \le Const \end{cases}$$



$\|w\|_1$

$$\|w\|_1 = \sum_{j=1}^{d} |w_j|$$

l_2-reg: $\frac{\lambda}{2}\|w\|^2$

l_1-reg: $\frac{\lambda}{2}\|w\|_1$

elastic-net: $\gamma_1\|w\|^2 + \gamma_2\|w_1\|$