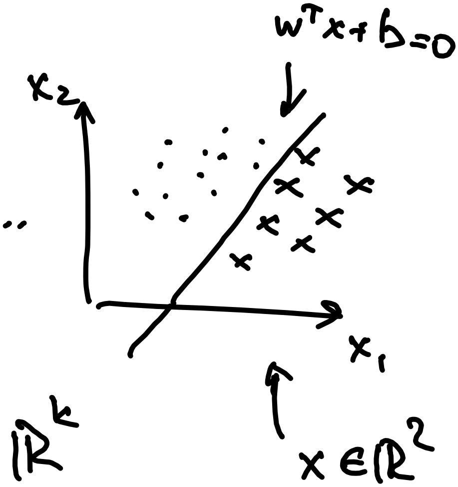
Линейные методы классификации.

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

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



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

$$\downarrow$$

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

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

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

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

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

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

w, b?

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

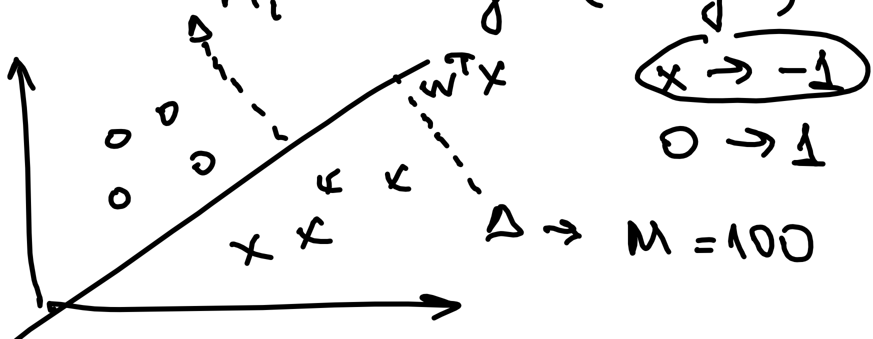
$$y_i \in \{-1, 1\}$$

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

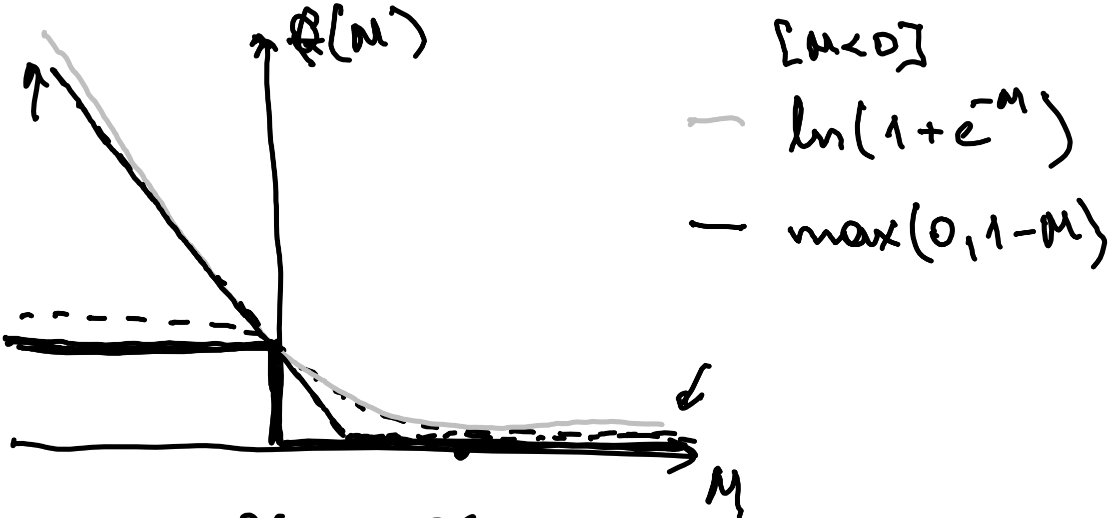
$$= \frac{1}{n} \sum_{i=1}^{n} [y_i \cdot \operatorname{sign}(\langle w, x_i \rangle) \neq \underbrace{y_i^2}_{1}] =$$

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

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



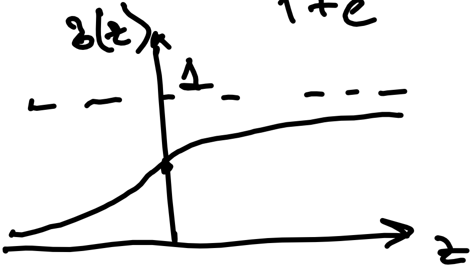
$$= \frac{1}{n} \sum_{i=1}^{n} [M_i < 0] \to \min_w$$



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

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

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



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

....

$$\langle w, x \rangle = \text{Const}_2$$

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

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

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

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

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

$$\frac{1}{1 + \exp(\langle w, x_i \rangle)}$$

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

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

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

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

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

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

$$\nabla_w \tilde{Q} = 0 \nRightarrow \text{аналитич. реш}$$

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

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

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

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

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

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

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

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

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

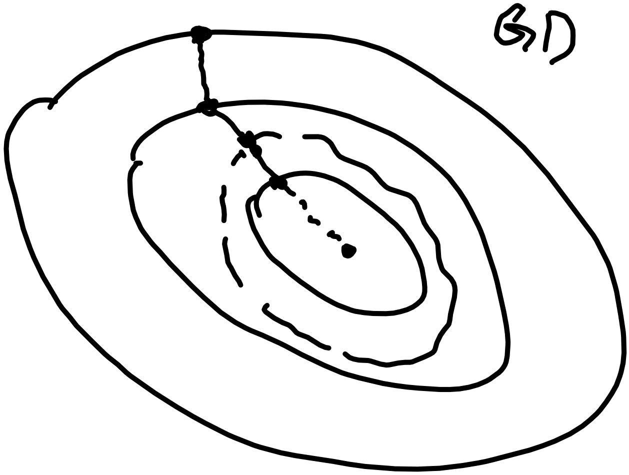
$$+ \underbrace{\frac{\lambda}{2}\|w\|_2^2}_{} \to \min_w \qquad \left(\|w\|_2^2 = w^T w = \sum_{j=1}^{d} w_j^2\right)$$

l_2 – регуляризация.

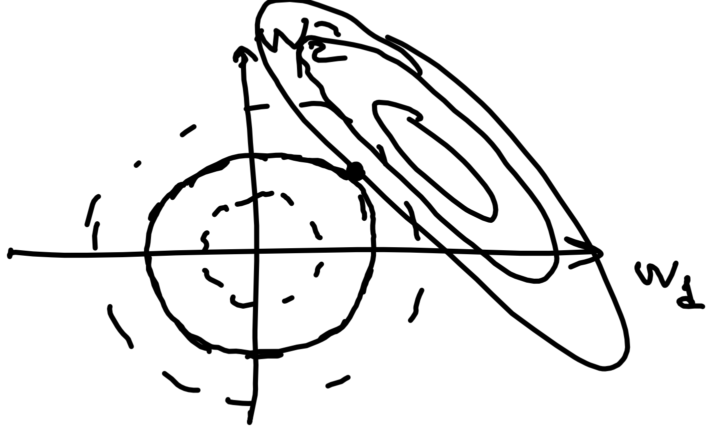
$$\nabla_w \tilde{Q} = \frac{1}{n}\sum_{i=1}^{n}\Big(\underset{\in \mathbb{R}}{-y_i}\, \underset{\in \mathbb{R}^d}{x_i} \cdot \sigma(\underbrace{-y_i\langle w, x_i\rangle}_{\in \mathbb{R}})\Big) + \frac{\lambda}{\cancel{2}}\underset{\in \mathbb{R}^d}{\cancel{2}w}$$

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

$|N|$ = Batch size.


GD

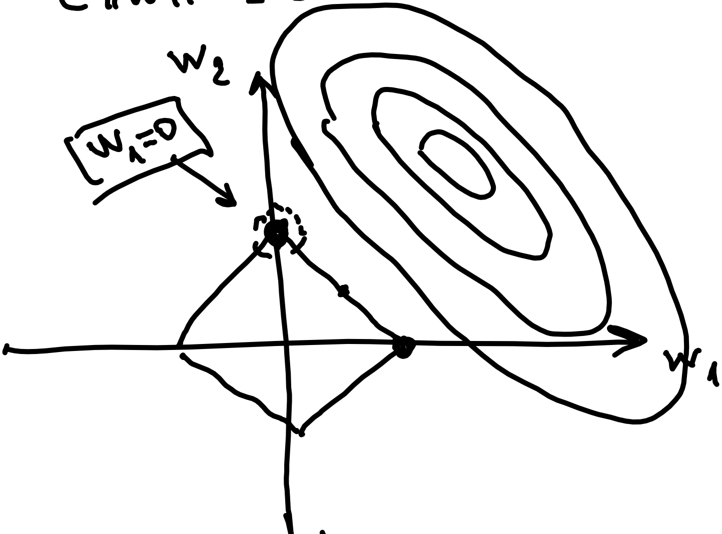

SGD



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

$\|w\|^2$

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



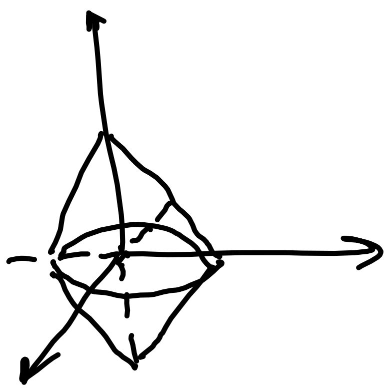
$\|w\|_1$

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

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

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

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

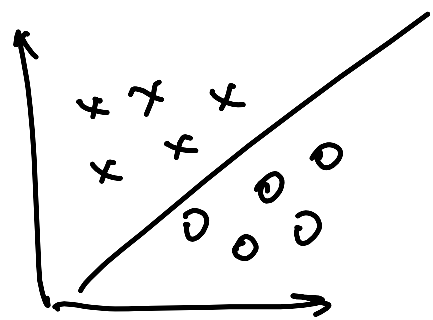


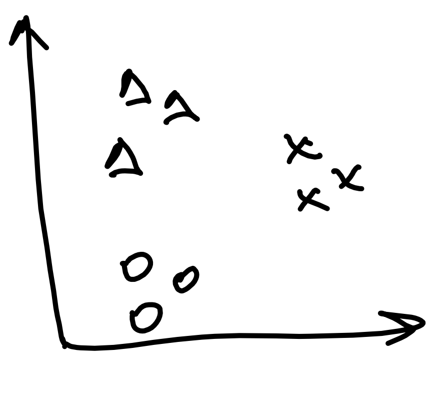
$\boxed{\text{SVM}}$

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

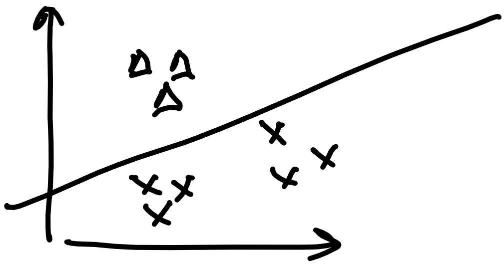
$$\to \frac{1}{n}\sum_{i=1}^{n}[\underbrace{y_i\langle w, x_i\rangle}_{M_i} < 0]$$

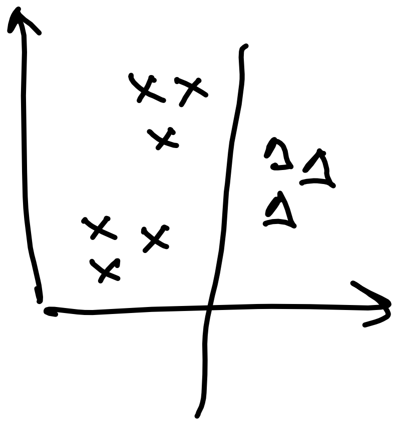
$$a(x) = \mathrm{sign}(\langle w, x\rangle + b)$$

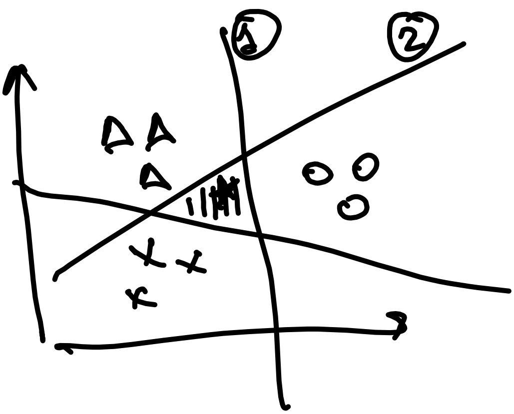




$\boxed{\text{One-vs-All}}$







$$
\begin{array}{l}
(A) \rightarrow a_1 \rightarrow p_1 \\
\quad \rightarrow a_2 \rightarrow p_2 \quad \rightarrow \arg\max_k \\
\quad \rightarrow \quad \vdots \\
\qquad a_K \rightarrow p_K
\end{array}
$$

$\sum p_k = 1$

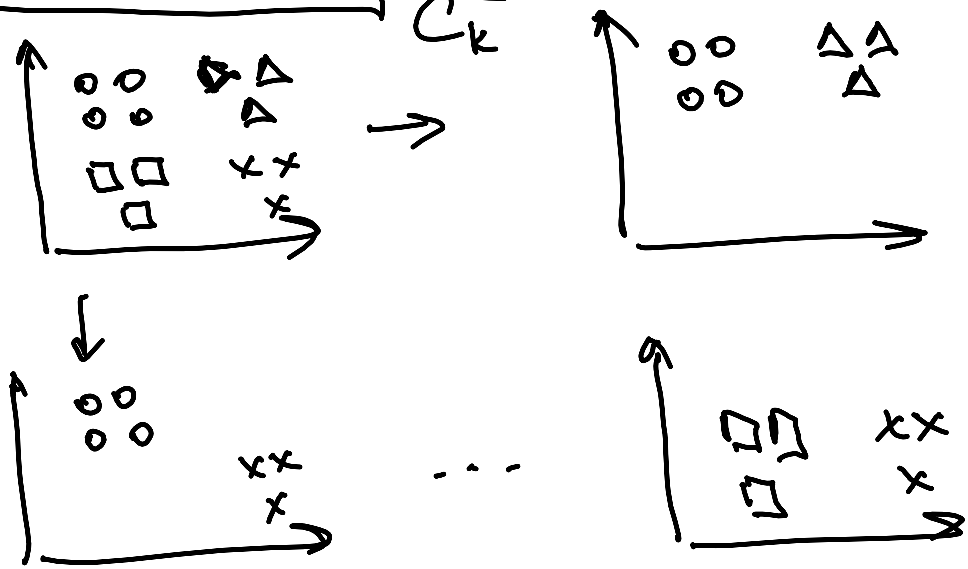
softmax

$$\mathrm{softmax}(b_1, b_2, \dots, b_n) =$$

$$= \left(\frac{e^{b_1}}{e^{b_1} + \dots + e^{b_n}}, \frac{e^{b_2}}{e^{b_1} + \dots + e^{b_n}}, \dots, \frac{e^{b_n}}{e^{b_1} + \dots + e^{b_n}} \right)$$

$\boxed{\text{One-vs-One}}$

C_k^2



$2 \leftarrow a_1, a_2$

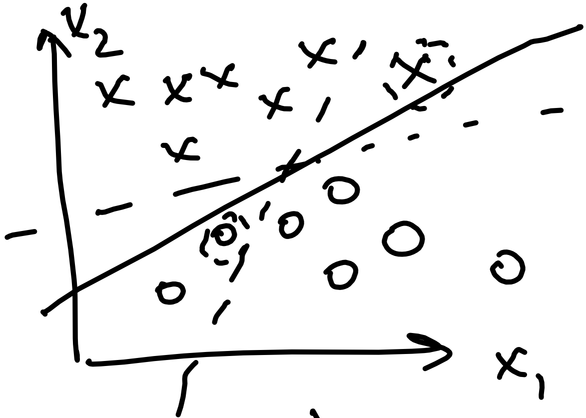
$1 \leftarrow a_1, a_3$

$\dots$

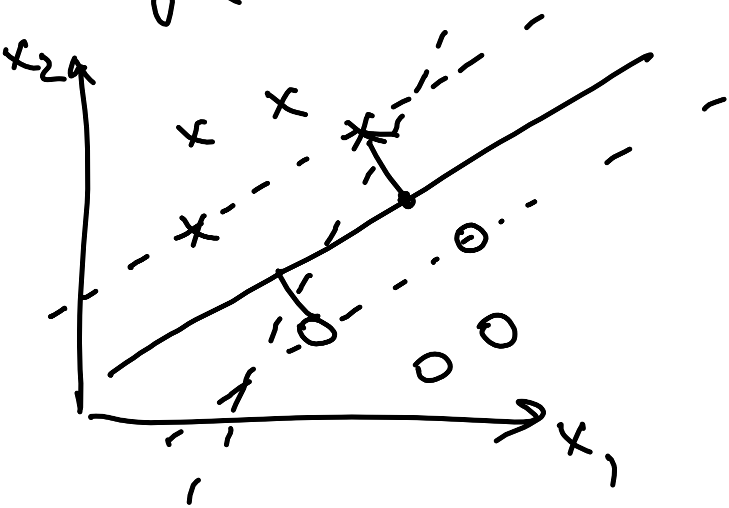
a_{k-1}, a_k

argmax

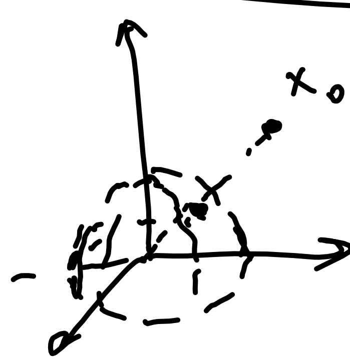
SVM



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



$$\begin{cases} \frac{1}{2}\|x - x_0\|^2 \to \min\limits_{x} \\ \|x\|_2^2 \le 1 \end{cases}$$



$$L = \frac{1}{2}\left(x^T x - 2x_0^T x + x_0^T x_0\right) +$$

$$\nabla_x (a^T x) = a$$

$$+ \lambda (x^T x - 1)$$

K.K.T:

$$\nabla_x L = x - x_0 + 2\lambda x = 0$$

$$(1 + 2\lambda) x = x_0 \Rightarrow \boxed{x^* = \frac{x_0}{1+2\lambda}}$$

$\lambda = 0$ или $x^T x = 1$ $\quad (1+2\lambda)^2 = x_0^T x_0$

$$1 + 2\lambda = \pm\sqrt{x_0^T x_0}$$

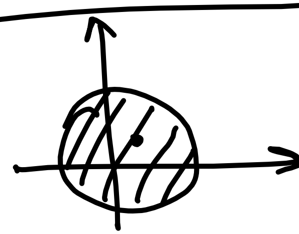
$$2\lambda = \pm\sqrt{x_0^T x_0} - 1$$

$$\boxed{\lambda^* = \frac{1}{2}\left(\pm\sqrt{x_0^T x_0} - 1\right)}$$

$\lambda \geq 0$

$$\lambda \cdot \left((x^*)^T x^* - 1\right) = 0$$

$$\lambda \left(\frac{x_0^T x_0}{(1+2\lambda)^2} - 1\right) = 0.$$

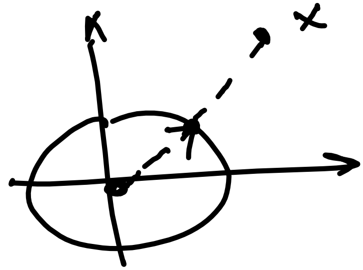


1. $\lambda = 0 \Rightarrow \boxed{x^* = x_0}$

2. $\lambda \neq 0 \quad \frac{x_0^T x_0}{(1+2\lambda)^2} = 1 \Rightarrow \lambda^* = \pm\frac{1}{2}\sqrt{x_0^T x} - \frac{1}{2}$

$$\lambda^* = \frac{1}{2}\left(\sqrt{x_0^T x_0} - 1\right) = \frac{1}{2}\left(\|x_0\| - 1\right)$$

$$x^* = \frac{x_0}{\cancel{1} + 2 \cdot \frac{1}{2}\|x_0\| - \cancel{1}} = \boxed{\frac{x_0}{\|x_0\|}}$$



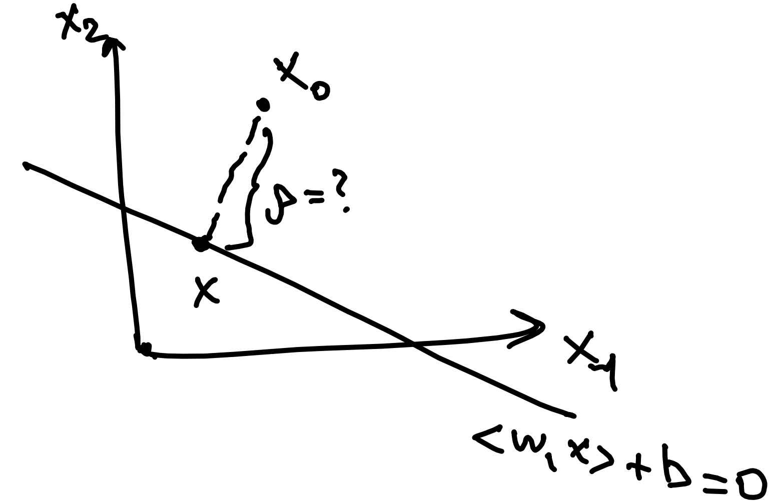
$$g(\lambda, \mu) = \inf_x L(x, \lambda, \mu)$$

$$(P): \begin{cases} f_0(x) \to \min \\ \dots \end{cases} \qquad (D): \begin{cases} g(\lambda, \mu) \to \max\limits_{\lambda, \mu} \\ \dots \end{cases}$$

$$g(\lambda, \mu) \le f(x) \quad \forall x, \lambda, \mu.$$

Сильная двойств:

$$g(\lambda^*, \mu^*) = f(x^*)$$



$$\begin{cases} \frac{1}{2}\|x - x_0\|_2^2 \to \min\limits_x \\ \langle w, x \rangle + b = 0 \end{cases}$$

$$L = \frac{1}{2}\left(x^T x - 2x_0^T x + x_0^T x_0\right) + \mu\left(w^T x + b\right)$$

K.K.T:

$$\begin{cases} \nabla_x L = x - x_0 + \mu w = 0. \\ \boxed{x^* = x_0 - \mu w} \end{cases}$$

$$q(\mu) = \frac{1}{2}\|\cancel{x_0} - \mu w - \cancel{x_0}\|^2 + \mu\left(w^T(x_0 - \mu w) + b\right) = \frac{1}{2}\mu^2 \cdot w^T w + \mu \cdot w^T x_0 - \mu^2 w^T w + \mu b \to \max_\mu$$

$$w^T x^* + b = 0 \Rightarrow w^T(x_0 - \mu w) + b = 0.$$

$$w^T x_0 - \mu w^T w + b = 0.$$

$$\boxed{\mu^* = \frac{w^T x_0 + b}{w^T w}}$$

$$\boxed{x^* = x_0 - \frac{w^T x_0 + b}{w^T w} w}$$

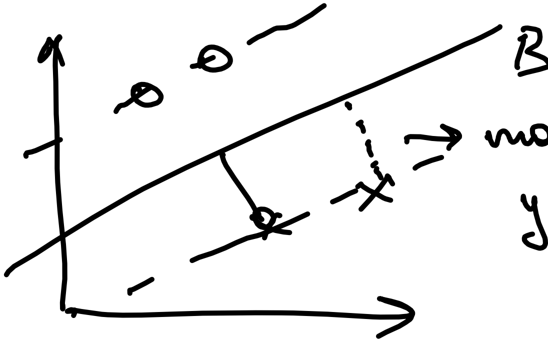
$$\rho(x, \langle w, x\rangle + b) = \|x - x_0\| =$$

$$= \|\cancel{x_0} - \frac{w^T x_0 + b}{w^T w} \cdot w - \cancel{x_0}\| =$$

$$= \frac{|w^T x_0 + b|}{\cancel{w^T w}} \cdot \underbrace{\|w\|}_{\cancel{\sqrt{w^T w}}} = \frac{|w^T x_0 + b|}{\sqrt{w^T w}} =$$

$$= \boxed{\frac{|w^T x_0 + b|}{\|w\|}}$$

SVM



$\to$ margin

$y_i(\langle w, x_i \rangle + b)$

$$\rho(B, x) = \frac{|w^T x + b|}{\|w\|} \to \max$$

$$2x + 5 = 0 \quad | \cdot 100$$

$$\min_x |w^T x + b| = 1 \Rightarrow \rho(B, x) = \frac{1}{\|w\|} \to \max_w$$

$$\begin{cases} \frac{1}{2}\|w\|^2 \to \min\limits_{w,b} \\ y_i(\langle w, x_i \rangle + b) \geq 1 \end{cases}$$

$$L = \frac{1}{2} w^T w - \sum_{i=1}^{n} \lambda_i \cdot \left[y_i(\langle w, x_i \rangle + b) - 1 \right]$$

$\boxed{K.K.T:}$

$$\begin{cases} \nabla_w L = w - \sum_{i=1}^n \lambda_i y_i x_i = 0 \\ L'_b = -\sum_{i=1}^n \lambda_i y_i = 0 \\ \lambda_i = 0 \text{ unu } y_i(\langle w, x_i \rangle + b) = 1 \\ \lambda_i \geq 0. \end{cases}$$

$$\langle w, \sum_i \lambda_i y_i x_i \rangle$$

$$\left(\sum_i \lambda_i y_i \langle w, x_i \rangle = \lambda_1 y_1 \langle w, x_1 \rangle + \lambda_2 y_2 \langle w, x_2 \rangle \ldots = \langle w, \lambda_1 x_1 y_1 \rangle + {} + \langle w, \lambda_2 y_2 x_2 \rangle + \ldots \right)$$

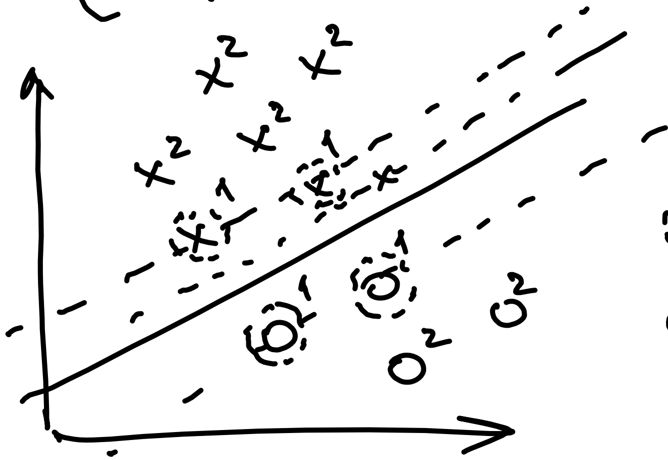
$$\boxed{w^* = \sum_{i=1}^n \lambda_i y_i x_i}$$

$$g(\lambda) = \frac{1}{2} \underbrace{\left\langle \sum_{i=1}^n \lambda_i y_i x_i, \sum_{j=1}^n \lambda_j y_j x_j \right\rangle} -$$

$$- \underbrace{\sum_j \sum_i \lambda_i \lambda_j y_i y_j \langle x_i, x_j \rangle} - b \underbrace{\sum_i \lambda_i y_i}_{0} + \sum_i \lambda_i =$$

$$= -\frac{1}{2} \sum_i \sum_j \lambda_i \lambda_j y_i y_j \langle x_i, x_j \rangle + \sum_i \lambda_i$$

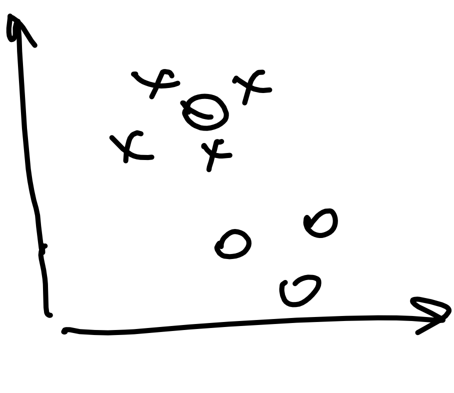
$$(D): \begin{cases} q(\lambda) = -\dfrac{1}{2} \sum\limits_i \sum\limits_j \lambda_i \lambda_j y_i y_j \langle x_i, x_j \rangle + \sum\limits_i \lambda_i \to \max\limits_{\lambda} \\ \sum\limits_i \lambda_i y_i = 0 \\ \lambda_i \geq 0 \end{cases}$$



$(\lambda > 0)$

① → опорные

② – периф. (эталонные) $(\lambda = 0)$



$$\begin{cases} \frac{1}{2}\|w\|^2 + C \cdot \sum_i \xi_i \to \min\limits_{w, b, \xi_i} \\ y_i(\langle w, x_i \rangle + b) \geq 1 - \xi_i \\ \xi_i \geq 0. \end{cases}$$

X - np.array (n×d)

y - np.array.

```
def gini(X, y, feat_num, thr):

    ...

def find_best_split(X, y):
    best_gini, best_thr, best_fn = [None]*3
for feat_num in range(X.shape[1]):
    unq_vals = np.unique(X[:, feat_num])
    for split in unq_vals:
        curr_gini = gini(X, y, feat_num,
                         thr = split)

        if best_gini is None:
            best_gini = curr_gini
        ...
```

```
if curr_gini < best_gini:
    best_gini = curr_gini
    best_thr = split
    ...
```

$\nabla_x (a^T x)$

---

x_1	x_2	y	
0	2	1	✓
0	3	1	
2	1	-1	✓
3	1	-1	

$$\begin{cases} \frac{1}{2}(w_1^2 + w_2^2) \to \min\limits_{w,b}. \\ 1 \cdot (w_1 \cdot 0 + w_2 2 + b) \geq 1 \\ 1 \cdot (w_1 \cdot 0 + w_2 \cdot 3 + b) \geq 1 \\ -1 (w_1 \cdot 2 + w_2 1 + b) \geq 1 \\ -1 (w_1 \cdot 3 + w_2 \cdot 1 + b) \geq 1. \end{cases}$$

$$\begin{cases} \frac{1}{2}(w_1^2 + w_2^2) \to \min\limits_{w,b}. \\ 2w_2 + b - 1 \geq 0. \\ 3w_2 + b - 1 \geq 0. \\ +2w_1 + w_2 + b + 1 \leq 0 \\ 3w_1 + w_2 + b + 1 \leq 0 \end{cases}$$

$$L = \frac{1}{2}(w_1^2 + w_2^2) - \lambda_1 (2w_2 + b - 1) - \lambda_2 (3w_2 + b - 1) + \lambda_3 (2w_1 + w_2 + b + 1) + \lambda_4 (3w_1 + w_2 + b + 1)$$

$\boxed{K.K.T:}$

$$\begin{cases} L'_{w_1} = w_1 + 2\lambda_3 + 3\lambda_4 = 0 & (1) \\ L'_{w_2} = w_2 - 2\lambda_1 - 3\lambda_2 + \lambda_3 + \lambda_4 = 0 & (2) \\ L'_b = -\lambda_1 - \lambda_2 + \lambda_3 + \lambda_4 = 0. & (3) \\ \lambda_1 = 0 \text{ или } 2w_2 + b = 1 & (4) \\ \lambda_2 = 0 \text{ или } 3w_2 + b = 1 & \lambda_2 = 0 \\ \lambda_3 = 0 \text{ или } 2w_1 + w_2 + b = -1 & (5) \\ \lambda_4 = 0 \text{ или } 3w_1 + w_2 + b = -1 & \lambda_4 = 0. \end{cases}$$

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 2 \\ 0 & 1 & 0 & -2 & 1 \\ 0 & 0 & 0 & -1 & 1 \\ 0 & 2 & 1 & 0 & 0 \\ 2 & 1 & 1 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} w_1 \\ w_2 \\ b \\ \lambda_1 \\ \lambda_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ -1 \end{bmatrix}$$