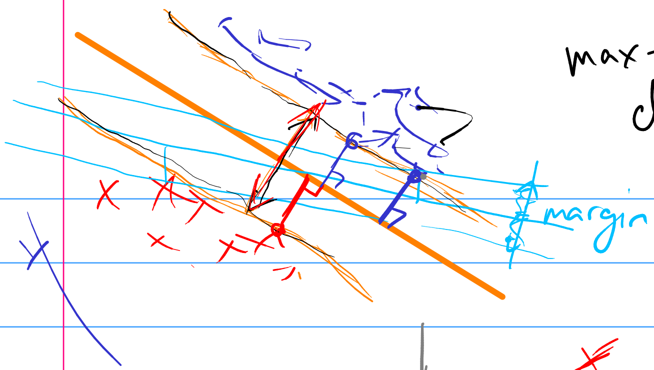
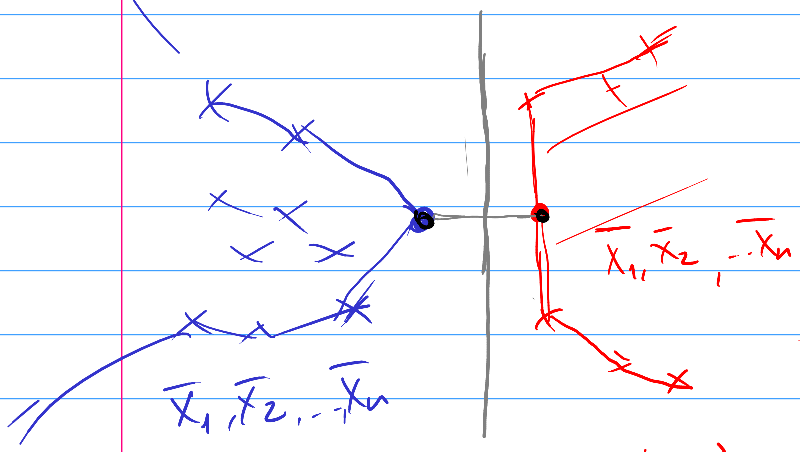
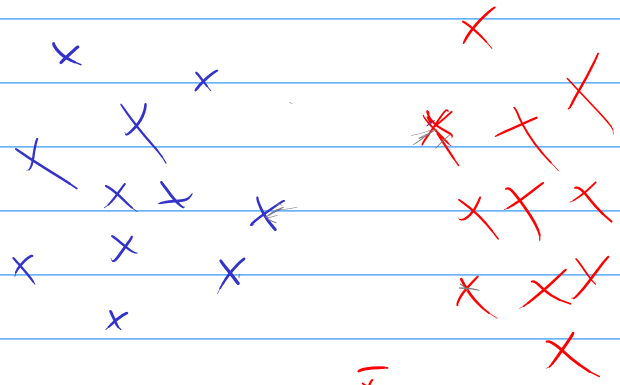
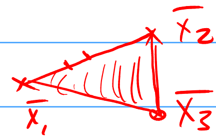


max-margin classifiers







$\{\alpha \bar{x}_1 + (1-\alpha)\bar{x}_2 \mid \alpha \in [0,1]\}$

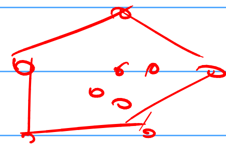
$\{\alpha_1 \bar{x}_1 + \alpha_2 \bar{x}_2 + \alpha_3 \bar{x}_3 \mid \alpha_i \geq 0,\ \sum \alpha_i = 1\}$

$$\mathrm{Conv}(C_1) = \Big\{ \sum_{\bar{x}_n \in C_1} \alpha_n \bar{x}_n \ \Big|\ \alpha_n \geq 0,\ \sum \alpha_n = 1 \Big\}$$

$\mathrm{Conv}(C_2)$

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

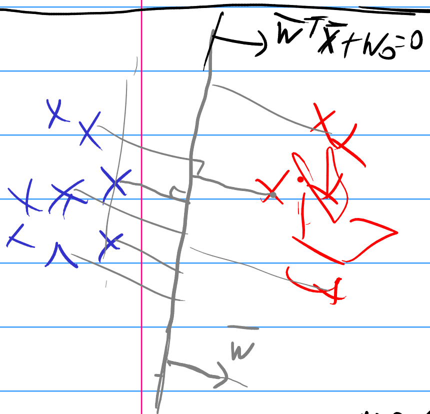
$t_n \in \{\pm 1\}$



$$\min_{\substack{\bar{c} \in \mathrm{Conv}(C_1) \\ \bar{d} \in \mathrm{Conv}(C_2)}} \|\bar{c} - \bar{d}\|^2 = \min_{\alpha_n} \Big\| \sum_{n:t_n=1} \alpha_n \bar{x}_n - \sum_{n:t_n=-1} \alpha_n \bar{x}_n \Big\|^2$$

$$\sum_{n:t_n=1} \alpha_n = \sum_{n:t_n=-1} \alpha_n = 1$$

$$\forall n\ \ \alpha_n \geq 0$$



$$d(\bar{x}) = \frac{|\bar{w}^T\bar{x} + w_0|}{\|\bar{w}\|}$$

$\forall n: t_n = 1 \quad \bar{w}^T\bar{x}_n + w_0 > 0$

$\forall n: t_n = -1 \quad \bar{w}^T\bar{x}_n + w_0 < 0$

$$\max_{\bar{w}} \min_{n} \frac{|\bar{w}^T\bar{x}_n + w_0|}{\|\bar{w}\|} = \max_{\bar{w}} \frac{1}{\|\bar{w}\|} \cdot \underbrace{\min_n |\bar{w}^T\bar{x}_n + w_0|}_{1}$$

$\forall n: t_n = 1 \quad \bar{w}^T\bar{x}_n + w_0 \geq 1$

$\forall n: t_n = -1 \quad \bar{w}^T\bar{x}_n + w_0 \leq -1$

$$\max_{\bar{w}} \frac{1}{\|\bar{w}\|} = \min_{\bar{w}} \|\bar{w}\|^2$$

$$\forall n \quad t_n(\bar{w}^T\bar{x}_n + w_0) - 1 \geq 0$$

$$L(\bar w, \bar\alpha) = \frac{1}{2}\|\bar w\|^2 - \sum_{n=1}^{N} \alpha_n \left(t_n(\bar w^T \bar x_n + w_0) - 1\right) \xrightarrow[\bar w, \bar\alpha]{} \min$$

$\alpha_n \geq 0$

$$\nabla_{\bar w} L = \bar w - \sum_{n=1}^{N} \alpha_n t_n \bar x_n = 0 \Rightarrow \boxed{\bar w = \sum_n \alpha_n t_n \bar x_n}$$

$$\partial L / \partial w_0 = -\sum_n \alpha_n t_n \Rightarrow \sum_n \alpha_n t_n = 0$$

$$L(\bar\alpha) = \frac{1}{2}\Big(\sum_n \alpha_n t_n \bar x_n\Big)^T \Big(\sum_m \alpha_m t_m \bar x_m\Big) - \sum_n \alpha_n t_n \bar x_n^T \Big(\sum_m \alpha_m t_m \bar x_m\Big) +$$

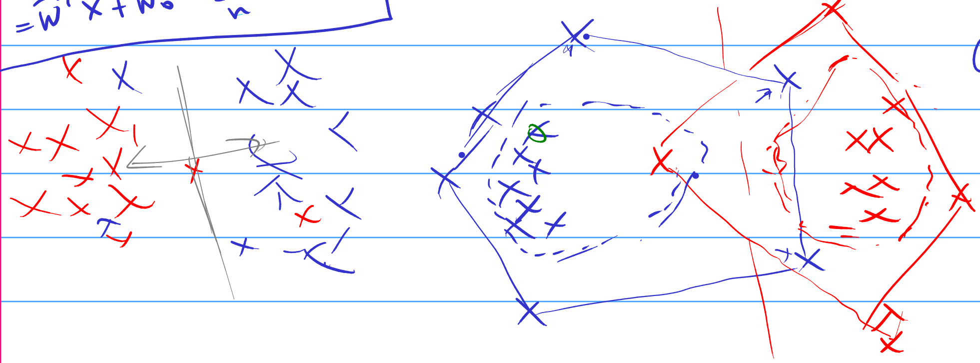
$$\sum_{n,m} \alpha_n \alpha_m t_n t_m \bar x_n^T \bar x_m \qquad \sum \ — \qquad + \sum_n \alpha_n$$

$\bar\alpha$

$$L(\bar\alpha) = \sum_n \alpha_n - \frac{1}{2} \sum_{n,m} \alpha_n \alpha_m t_n t_m \bar x_n^T \bar x_m \xrightarrow[\bar\alpha]{} \min$$

$$\bar w = \sum_n \alpha_n t_n \bar x_n \qquad \bar x_n^T \bar x + w_0 \qquad \sum_n \alpha_n t_n = 0 \qquad \forall n \ \alpha_n \geq 0, \ \alpha_n \leq C$$

$$y(\bar x) = \bar w^T \bar x + w_0 = \sum_n \alpha_n t_n \bar x_n^T \bar x + w_0$$



$$\mathrm{Conv}(C) = \Big\{ \sum \alpha_n \bar x_n \ \Big| \ \sum \alpha_n = 1, \ \alpha_n \geq 0 \Big\}$$

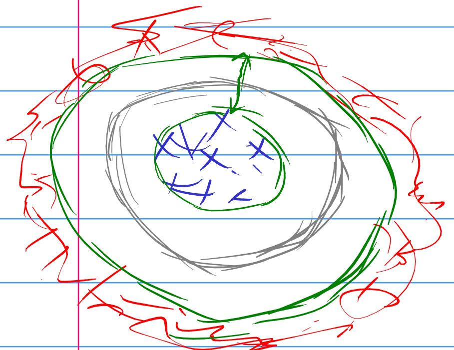
C_1

$$\sum_{n: t_n = 1} \alpha_n \bar x_n = \bar c$$

$$\sum \alpha_n = 1, \quad 0 \leq \alpha_n \leq D$$

$$\min_{\bar w, \bar z} \Big[\|\bar w\|^2 + C \cdot \sum z_n\Big] \qquad t_n(\bar w^T \bar x_n + w_0) + z_n \geq 1 \quad \forall n$$

$$z_n \geq 0$$



$$\bar x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \longmapsto \bar x = \begin{pmatrix} x_1^2 \\ x_2^2 \\ x_1 x_2 \\ x_1 \\ x_2 \end{pmatrix} \begin{matrix} \}\ 10 \cdot 11/2 \\ \}\ 10 \end{matrix}$$

$$\mathbb{R}^2 \longrightarrow \mathbb{R}^5$$

$$\bar x = \begin{pmatrix} x_1 \\ \vdots \\ x_{10} \end{pmatrix} \longmapsto \mathbb{R}^{55+10}$$

$$O(d^k) \approx$$

$d \sim 100$s

$$\bar x = \begin{pmatrix} x_1^2 \\ x_2^2 \\ \sqrt{2} x_1 x_2 \\ x_1 \\ x_2 \end{pmatrix} \qquad \bar y = \begin{pmatrix} y_1^2 \\ y_2^2 \\ \sqrt{2} y_1 y_2 \\ y_1 \\ y_2 \end{pmatrix}$$

kernel trick

$$\bar x^T \bar y = x_1^2 y_1^2 + x_2^2 y_2^2 + 2 x_1 x_2 y_1 y_2 + x_1 y_1 + x_2 y_2 = (x_1 y_1 + x_2 y_2)^2 + \bar x^T \bar y = (\bar x^T \bar y)^2 + \bar x^T \bar y$$

$$\varphi(\bar{x}) = \begin{pmatrix} x_1^d \\ \vdots \\ \vdots \end{pmatrix} \qquad \varphi(\bar{y}) = \begin{pmatrix} y_1^d \\ \vdots \\ \vdots \end{pmatrix}$$

$$\bar{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \qquad \bar{y} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$$

$$\underline{(\bar{x}^T \bar{y} + 1)^d - 1} = \varphi(\bar{x})^T \varphi(\bar{y})$$

$$\bar{x} \in \mathbb{R}^d \longrightarrow \varphi(\bar{x}) \in \mathbb{R}^{D}$$

$$\varphi(\bar{x}_n)^T \varphi(\bar{x}_m) = k(\bar{x}_n, \bar{x}_m)$$

ядро / kernel

$$L(\alpha) = \sum_n \alpha_n - \frac{1}{2} \sum_{n,m} \alpha_n \alpha_m t_n t_m k(\bar{x}_n, \bar{x}_m)$$

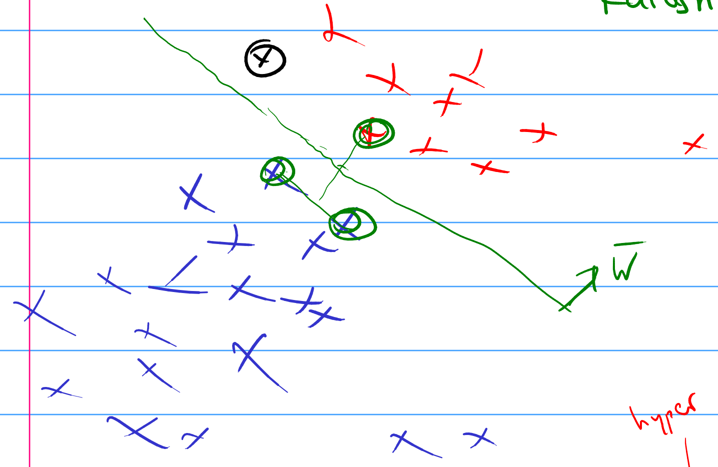
$$\sum \alpha_n t_n = 0, \quad \alpha_n \geq 0$$

$$\mathbb{R}^D \ni \bar{w} = \sum_n \alpha_n t_n \varphi(\bar{x}_n)$$

$$\underline{y(\bar{x})} = \bar{w}^T \underbrace{\varphi(\bar{x})}_{\in \mathbb{R}^D} + w_0 = \sum_n \alpha_n t_n \underline{k(\bar{x}_n, \bar{x})} + w_0$$

Karush-Kuhn-Tucker

$$\alpha_n \geq 0$$
$$t_n y(\bar{x}_n) - 1 \geq 0$$
$$\alpha_n (t_n y(\bar{x}_n) - 1) = 0$$



hyper перем.

$$\min \left(\frac{1}{2} \|\bar{w}\|^2 + \frac{1}{2} \sum_n z_n - \nu \cdot \rho \right) \qquad t_n(\bar{w}^T \bar{x}_n + w_0) + z_n \geq \rho$$

$$z_n \geq 0, \quad \rho \geq 0$$

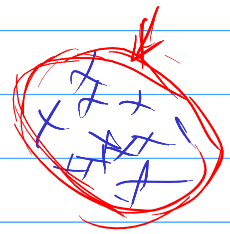
$\rho > 0 \Rightarrow$

(1) $\nu \geq \frac{1}{N} \cdot \#\{n : z_n > 0\}$

(2) $\nu \leq \frac{1}{N} \#\{\bar{x}_n - \text{support vectors}\}$

$\boxed{\nu} \approx$ доля выбросов

one-class SVM



$t_n = 1$

$$C\sum z_n + \frac{1}{2}\|\bar{w}\|^2 \qquad z_n = \begin{cases} 0, & t_n y(\bar{x}_n) \geq 1 \\ 1 - t_n y(\bar{x}_n), & t_n y(\bar{x}_n) < 1 \end{cases}$$

$$E_{SV}(t_n y(\bar{x}_n)) = \mathrm{ReLU}(1 - t_n y(\bar{x}_n))$$

$$L(\bar{w}) = \sum_n E_{SV}(\underbrace{t_n(\bar{w}^T\bar{x}_n + w_0)}_{t_n y(x_n)}) + \underbrace{\lambda \cdot \|\bar{w}\|^2}_{L_2\text{-regularizer}}$$

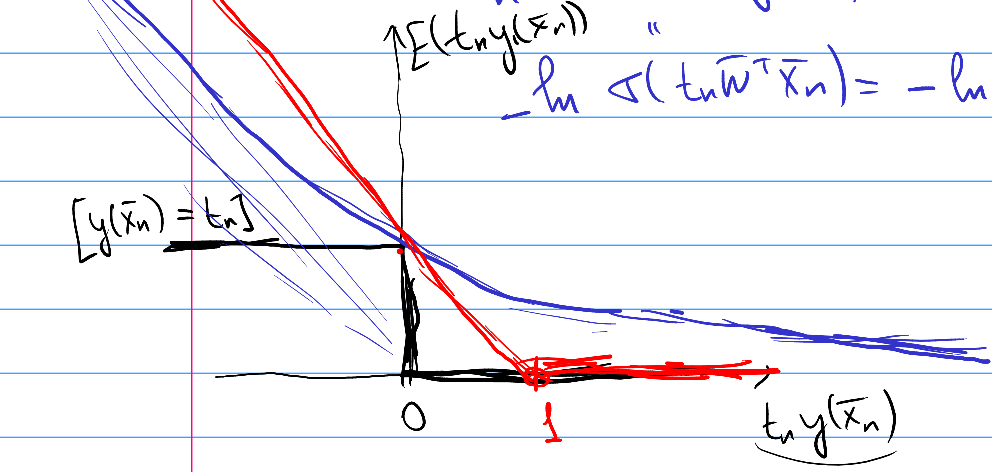
$$t_n \in \{\pm 1\} \qquad p(t_n = 1 \mid \bar{x}_n, \bar{w}) = \sigma(\bar{w}^T\bar{x}_n)$$

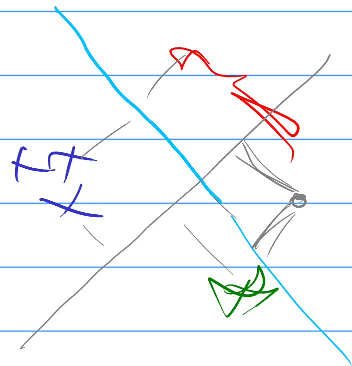
$$p(t_n = -1 \mid \bar{x}_n, \bar{w}) = \sigma(-\bar{w}^T\bar{x}_n) = 1 - \sigma(\bar{w}^T\bar{x}_n)$$

ESV hinge error function

$$L(\bar{w}) = \sum_n E_{LR}(t_n y(\bar{x}_n)) + \lambda \|\bar{w}^2\|$$

$$E_{LR} = -\ln \sigma(t_n \bar{w}^T\bar{x}_n) = -\ln \frac{1}{1+e^{\dots}} = \ln(1 + e^{-t_n y(\bar{x}_n)})$$





$$p(x \mid \mu, \tau) = \sqrt{\frac{\tau}{2\pi}}\, e^{-\frac{\tau}{2}(x-\mu)^2} \qquad \tau = 1/\sigma^2 \text{ (precision)}$$

$$\ln p(x \mid \mu, \tau) = \frac{1}{2}\ln\tau - \frac{1}{2}\ln 2\pi - \frac{\tau}{2}(x-\mu)^2$$

$$\ln p(\mu, \tau \mid x) = \text{const} + \ln p(\mu, \tau) + \ln p(x \mid \mu, \tau)$$

1) $\tau = \text{const}$

$$\ln p(x_1 \dots x_n \mid \mu) = \text{const} - \sum_n \frac{\tau}{2}(x_n - \mu)^2$$

$$+ \ln p(\mu \mid \mu_0, \tau_0) = -\frac{1}{2}\ln\tau_0 - \frac{1}{2}\ln 2\pi - \frac{\tau_0}{2}(\mu - \mu_0)^2$$

$$\ln p(\mu \mid x) = \text{const} - \frac{\tau_0}{2}(\mu - \mu_0)^2 - \sum_n \frac{\tau}{2}(x_n - \mu)^2 =$$

$$= \text{const} - \frac{\tau'}{2}(\mu - \mu')^2$$

$$\tau' = \tau_0 + n\tau$$

$$\mu' = \frac{\tau_0\mu_0 + \tau\sum x_n}{\tau_0 + n\tau}$$

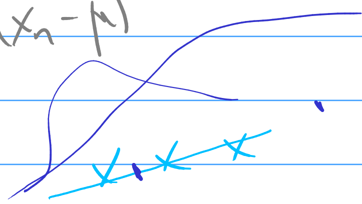
2) $\mu = const$ $\quad \ln p(x_1 - x_N|\tau) = N\frac{1}{2}\ln\tau - \frac{\tau}{2}(x-\mu)^2$ $\quad \frac{\tau}{2}\sum_n (x_n-\mu)^2$

$$p(\tau|a_0,b_0) = \frac{1}{z}\cdot\tau^{a_0-1}\cdot e^{-b_0\tau} \quad \Big| \quad \ln p(\tau|a_0,b_0) = (a_0-1)\ln\tau - b_0\tau + c$$

$$\ln p(\tau|x) = const + N\frac{1}{2}\ln\tau + (a_0-1)\ln\tau - \frac{\tau}{2}\sum_n (x_n-\mu)^2 - b_0\tau$$

$a' = a_0 + \frac{1}{2}$ $\qquad\qquad$ $a' = a_0 + \frac{N}{2}$

$b' = b_0 + \frac{1}{2}(x-\mu)^2$ $\qquad\qquad$ $b' = b_0 + \frac{1}{2}\sum_n (x_n-\mu)^2$



3) $$\ln p(x|\mu,\tau) = const + \frac{1}{2}\ln\tau - \frac{\tau}{2}(x-\mu)^2$$

$$p(\mu,\tau) = p(\mu|\mu_0,\tau_0)\cdot p(\tau|a_0,b_0)$$

$$p(\mu,\tau) = p(\tau)\,p(\mu|\tau) = p(\tau|a_0,b_0)\cdot p(\mu|\mu_0,\lambda_0\tau)$$

$$\ln p(\mu,\tau|x) = \ln p(\mu,\tau|..) + \ln p(x|\mu,\tau) + const =$$

$$= const + \frac{1}{2}\ln\tau - \frac{\tau}{2}(x-\mu)^2 + (a_0-1)\ln\tau - b_0\tau + \frac{1}{2}\ln\tau_0 - \frac{\tau_0}{2}(\mu-\mu_0)^2$$

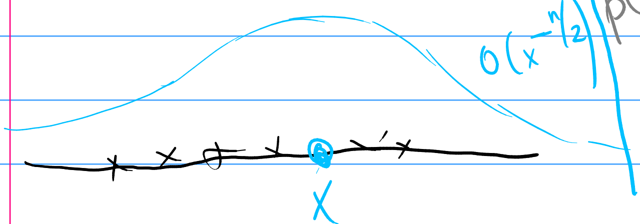
$$\ln p(\mu,\tau|x) = const + \frac{\ln\tau}{2} - \frac{\tau}{2}(x-\mu)^2 + (a_0-1)\ln\tau - b_0\tau + \boxed{\frac{\ln\tau}{2}} + \frac{\ln\lambda_0}{2} - \frac{\lambda_0\tau}{2}\sum_n(\mu-\mu_0)^2$$

$\lambda' = \lambda_0 + 1$ $\qquad$ $a' = a_0 + 1/2$ $\qquad$ $-\frac{1}{2}\mu^2(\tau + \lambda_0\tau) = -\frac{\tau}{2}\mu^2(\lambda_0+1)$

$\mu' = \dfrac{\lambda_0 + x}{\lambda_0 + 1}$ $\qquad$ $b' = b_0 - \frac{1}{2}(x^2 + \mu_0^2\lambda_0)$ $\qquad$ $\mu(\tau x + \lambda_0\tau\mu_0)$

$\lambda' = \lambda_0 + N$ $\qquad$ $a' = a_0 + N/2$

$\mu' = \dfrac{\lambda_0 + \sum x_n}{\lambda_0 + N}$ $\qquad$ $b' = b_0 - \frac{1}{2}\left(\mu_0^2\lambda_0 + \sum x_n^2\right)$



$O(x^{-N/2})$

$$p(x|x_1 - x_n) = \int p(x|\mu,\tau)\,p(\mu,\tau|x_1 - x_n)\,d\mu\,d\tau$$

$$= \int N(x|\mu,\tau)\cdot N(\mu|\mu',\lambda'\tau)\cdot Gam(\tau|a',b')\,d\mu\,d\tau$$

$$= \ldots = Student(x|\ldots)$$