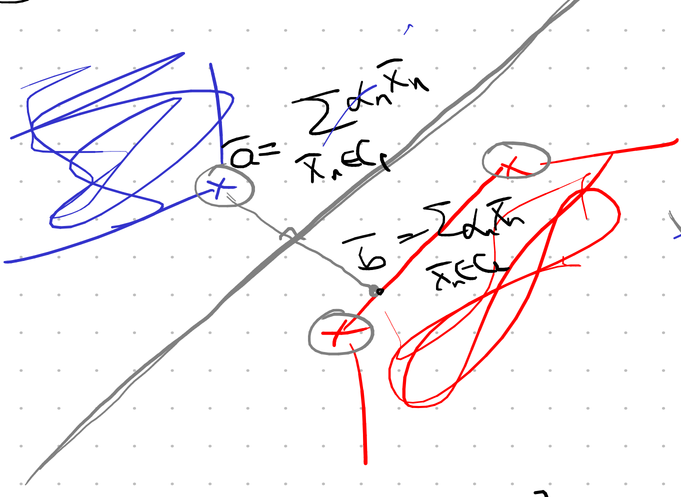
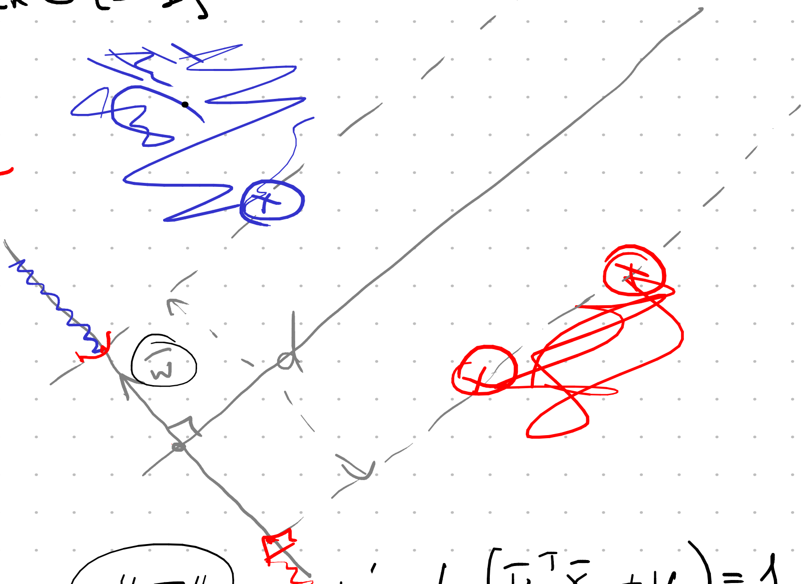
① SVM $D=\{(\bar{x}_n,t_n)\}$, $t_n\in\{\pm 1\}$





$$\min_{\bar{\alpha}}\left\|\sum_{\bar{x}_n\in C_1}\alpha_n\bar{x}_n-\sum_{\bar{x}_n\in C_2}\alpha_n\bar{x}_n\right\|^2$$

$$=\min_{\bar{\alpha}}\left\|\sum_n t_n\alpha_n\bar{x}_n\right\|^2$$

$$\forall n\ \ \alpha_n\geq 0,\quad \sum_{t_n=1}\alpha_n=\sum_{t_n=-1}\alpha_n=1$$

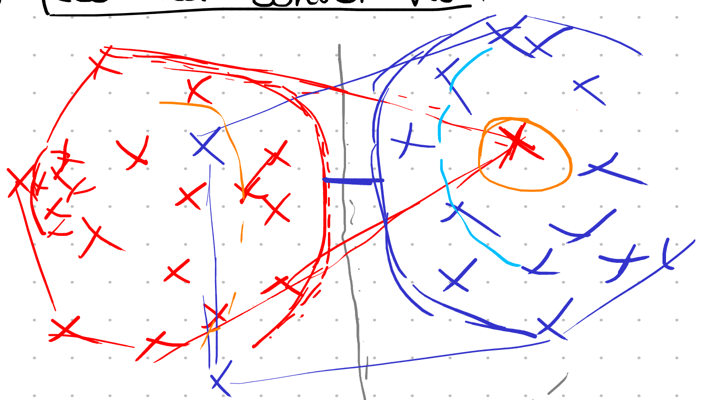
$\|\bar{w}\|$ $\quad \min_n t_n(\bar{w}^T\bar{x}_n+w_0)=1$

$$\min\|\bar{w}\|^2$$

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

$$\max\frac{1}{\|\bar{w}\|}=\max\frac{1}{\sqrt{w_1^2+\dots+w_d^2}}\to\min$$

② Reduced convex hull



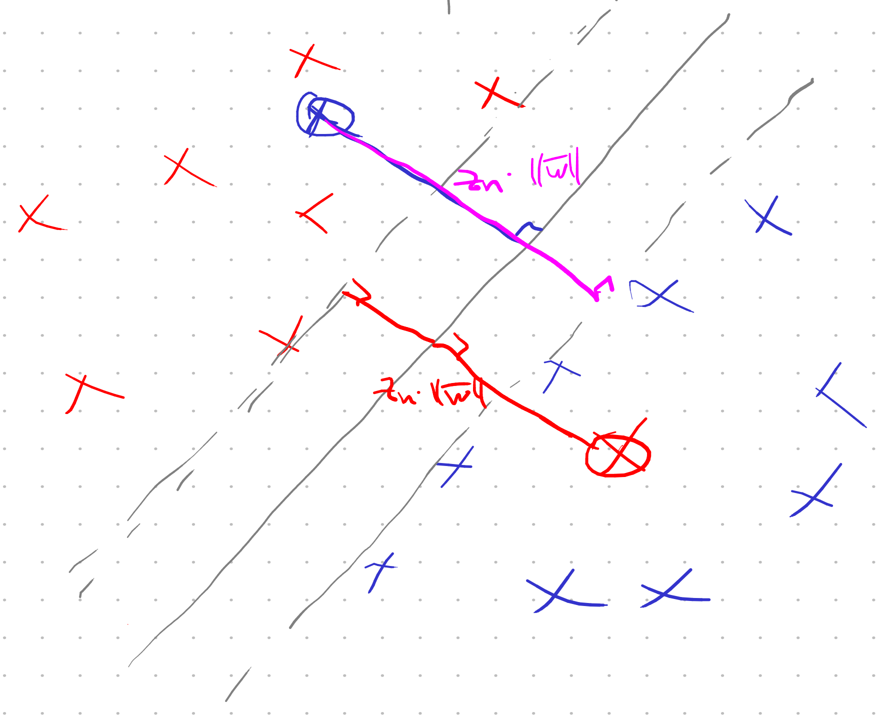
$$\mathrm{Conv}(X)=\left\{\sum\alpha_n\bar{x}_n\ \middle|\ \begin{array}{l}\alpha_n\geq 0\\ \sum\alpha_n=1\end{array}\right\}$$

$$\mathrm{Conv}_A(X)=\left\{\sum\alpha_n\bar{x}_n\ \middle|\ \begin{array}{l}\sum\alpha_n=1\\ 0\leq\alpha_n\leq \boxed{A}\end{array}\right\}$$

$$\min_{\alpha}\|\bar{a}-\bar{b}\|^2=$$

$$=\min_{\alpha}\left\|\sum t_n\alpha_n\bar{x}_n\right\|^2$$

$$\sum\alpha_n=\sum\alpha_n=1\qquad\forall n\ \ 0\leq\alpha_n\leq A$$



slack z_n

$$\min_{\bar{w},w_0,z_n}\left\{\|\bar{w}\|^2+C\cdot\sum_n z_n\right\}$$

$$t_n(\bar{w}^T\bar{x}_n+w_0)+z_n\geq 1$$

$$z_n\geq 0$$

(3) $\min_{\bar w, w_0} \frac{1}{2}\|\bar w\|^2 + C\sum z_n \quad \forall n \quad t_n(\bar w^T \bar x_n + w_0) \geq 1 - z_n$

$$L(\bar w, w_0, \bar\alpha) = \frac{1}{2}\|\bar w\|^2 + C z_n - \sum_n \alpha_n \left(t_n (\bar w^T \bar x_n + w_0) - 1 + z_n \right)$$

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

$z_n \geq 0$

$$\nabla_{\bar w} L = \bar w - \sum_n \alpha_n t_n \bar x_n = 0$$

$$\boxed{\bar w = \sum_n \alpha_n t_n \bar x_n}$$

$$\frac{\partial L}{\partial w_0} = -\sum_n \alpha_n t_n = 0$$

$$\boxed{\sum_n \alpha_n t_n = 0}$$

$$L(\bar\alpha) = \cancel{\frac{1}{2}\|\sum_n \alpha_n t_n \bar x_n\|^2} - \frac{1}{2}\left(\sum_n \alpha_n t_n \bar x_n\right)^T \left(\sum_n \alpha_n t_n \bar x_n\right) + \sum_n \alpha_n$$

$$L(\bar\alpha, \bar z) = \sum_n \alpha_n - \sum_n \sum_m \alpha_n \alpha_m t_n t_m \boxed{(\bar x_n^T \bar x_m)} \xrightarrow[\bar\alpha]{} \min$$

$n, m: \quad \bar x_n^T \bar x_m$

$+ \sum_n z_n (C - \alpha_n)$ при условии $\forall n \quad \alpha_n \geq 0, \quad \sum_n \alpha_n t_n = 0$

$z_n \geq 0$

Karush – Kuhn – Tucker:

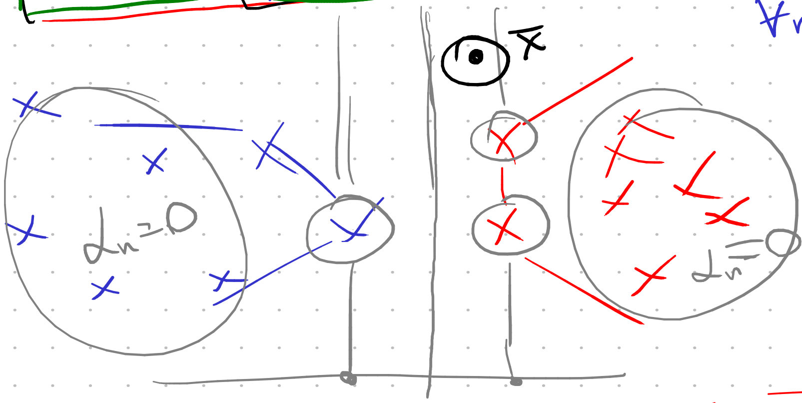
$\forall n \quad \alpha_n \geq 0$

$\forall n \quad t_n y(\bar x_n) - 1 \geq 0$

$\forall n \quad \alpha_n \left(t_n y(\bar x_n) - 1 \right) = 0$

$0 \quad\quad 0$

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

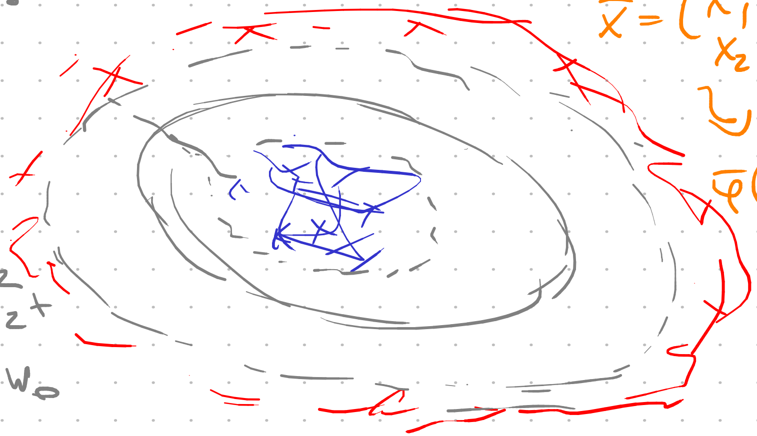


$$\boxed{\bar w = \sum_n \alpha_n t_n \bar x_n}$$

(4) Kernel trick

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

$$y(\bar x) = w_{11} x_1^2 + w_{12} x_1 x_2 + w_{22} x_2^2 + v_1 x_1 + w_2 x_2 + w_0$$



$\bar x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \quad \mathbb{R}^2$

$\bar\varphi(\bar x) = \begin{pmatrix} x_1^2 \\ x_1 x_2 \\ x_2^2 \\ x_1 \\ x_2 \end{pmatrix} \quad \mathbb{R}^5$

$$\begin{pmatrix} x_1 \\ \vdots \\ x_d \end{pmatrix} \mapsto \bar\varphi(\bar x) = \begin{pmatrix} x_1^2 \\ \vdots \\ x_d^2 \\ \boxed{\begin{matrix} x_1 \\ \vdots \\ x_d \end{matrix}} \end{pmatrix} \in \mathbb{R}^{\frac{d(d+1)}{2}}$$

$\mathbb{R}^d$

$$\begin{pmatrix} x_1^2 \\ \sqrt{2}x_1x_2 \\ x_2^2 \\ \vdots \\ x_d^2 \end{pmatrix}^T \begin{pmatrix} y_1^2 \\ \sqrt{2}y_1y_2 \\ y_2^2 \\ \vdots \\ y_d^2 \end{pmatrix} = x_1^2y_1^2 + 2x_1x_2y_1y_2 + x_2^2y_2^2 =$$

$$= (x_1y_1 + x_2y_2)^2 = \left(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}^T \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}\right)^2$$

$$(x_1y_1 + \ldots + x_dy_d)^2$$

$$\bar\varphi(\bar x)^T \bar\varphi(\bar y) = (\bar x^T \bar y)^2$$

$\underline{k(\bar x, \bar y)}$ – kernel $\qquad k(\bar x, \bar y) = \bar\varphi(\bar x)^T \bar\varphi(\bar y)$

$$L(\bar\alpha) = \sum_n \alpha_n - \frac{1}{2} \sum_n \sum_m \alpha_n \alpha_m t_n t_m k(\bar x_n, \bar x_m) \xrightarrow[\bar\alpha]{} \min$$

$$\bar w = \sum_n \alpha_n t_n \bar x_n \qquad y(\bar x) = \sum_n \alpha_n t_n k(\bar x_n, \bar x) + w_0$$

Теорема Мерсера:

1) $k(\bar x, \bar y) = k(\bar y, \bar x)$

2) $\iint k(\bar x, \bar y) g(\bar x) g(\bar y)\, d\bar x\, d\bar y > 0$

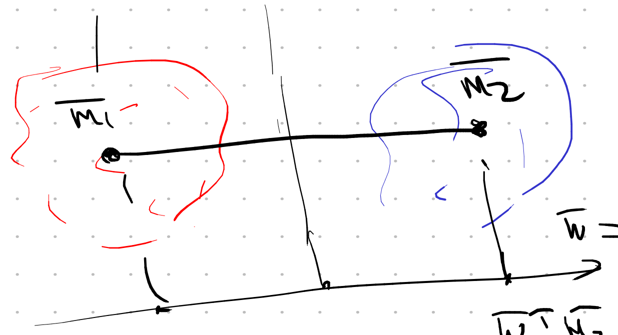
$\forall g: \int g^2(\bar x)\, d\bar x < \infty$

$$\begin{pmatrix} x_1^2 \\ \vdots \\ x_d^2 \\ \sqrt{2}x_1 \\ \vdots \\ x_d \end{pmatrix}^T \begin{pmatrix} y_1^2 \\ \vdots \\ y_d^2 \\ \sqrt{2}y_1 \\ \vdots \\ y_d \end{pmatrix} = (\bar x^T \bar y)^2 + (\bar x^T \bar y)$$

$$= (\bar x^T \bar y + 1)^2 - 1$$

$$(\bar x^T \bar y + 1)^k - 1$$

(5) Kernel methods



$$y(\bar{x}) = \bar{w}^T \bar{x} + w_0 = (\bar{m}_2 - \bar{m}_1)^T \bar{x} + \frac{\bar{m}_2^T \bar{m}_2 - \bar{m}_1^T \bar{m}_1}{2}$$

$$\bar{w} = \bar{m}_2 - \bar{m}_1$$

$$w_0 = \frac{\bar{w}^T \bar{m}_2 + \bar{w}^T \bar{m}_1}{2} = \frac{(\bar{m}_2 - \bar{m}_1)^T (\bar{m}_2 + \bar{m}_1)}{2} = \frac{\bar{m}_2^T \bar{m}_2 - \bar{m}_1^T \bar{m}_1}{2}$$

$$\bar{m}_1 = \frac{1}{N_1} \sum_{t_n = 1} \bar{\varphi}(\bar{x}_n) \qquad \bar{m}_2 = \frac{1}{N_2} \sum_{t_n = -1} \bar{\varphi}(\bar{x}_n)$$

$$y(\bar{x}) = (\bar{m}_2 - \bar{m}_1)^T \bar{\varphi}(\bar{x}) + \frac{\|\bar{m}_2\|^2 - \|\bar{m}_1\|^2}{2} =$$

$$= \frac{1}{N_2} \sum_{t_n = -1} k(\bar{x}_n, \bar{x}) - \frac{1}{N_1} \sum_{t_n = 1} k(\bar{x}_n, \bar{x}) + \frac{1}{2} \left(\frac{1}{N_2^2} \sum_{\substack{n,m \\ t_n = t_m = -1}} k(\bar{x}_n, \bar{x}_m) - \frac{1}{N_1^2} \sum_{t_n = 1} \dots \right)$$

(6) ν-SVM

$$L(\bar{w}, w_0, \bar{z}) = \frac{1}{2} \bar{w}^T \bar{w} + C \sum z_n$$
$$t_n(\bar{w}^T \bar{x}_n + w_0) + z_n \geq 1$$

$$L(\bar{w}, w_0, \bar{z}, \rho) = \frac{1}{2} \bar{w}^T \bar{w} + \frac{1}{N} \sum z_n - \nu \cdot \rho$$

$$t_n(\bar{w}^T \bar{x}_n + w_0) + z_n \geq \rho, \quad z_n \geq 0, \quad \rho \geq 0$$

$$L(\bar{w}, w_0, \bar{z}, \rho, \bar{\alpha}, \bar{\beta}, \delta) = \frac{1}{2} \bar{w}^T \bar{w} + \frac{1}{N} \sum z_n - \nu \rho$$
$$- \sum_n \alpha_n \left(t_n (\bar{w}^T \bar{x}_n + w_0) + z_n - \rho \right) - \sum_n \beta_n z_n + \delta \rho =$$

$$\nabla_{\bar{w}} L = \bar{w} - \sum_n \alpha_n t_n \bar{x}_n = 0 \qquad \sum_n \alpha_n t_n = 0$$

$$z_n: \quad \alpha_n + \beta_n = \frac{1}{N}$$

$$\rho: \quad -\sum_n \alpha_n + \nu = \delta$$

$$\forall n \quad \alpha_n \left(t_n (\bar{w}^T \bar{x}_n + w_0) + z_n - \rho \right) = 0$$

$$\delta \cdot \rho = 0$$

$$= -\frac{1}{2}\sum_{n,m} \alpha_n \alpha_m t_n t_m \bar{x}_n^T \bar{x}_m$$

$\forall n \quad \boxed{\alpha_n \le 1/N}$

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

$$\mu > 0 \Rightarrow \delta = 0, \quad \sum_n \alpha_n = \nu$$

если $z_n > 0$, то $\alpha_n = 1/N \Rightarrow$ векторов, для к-рых $z_n > 0$, не больше νN

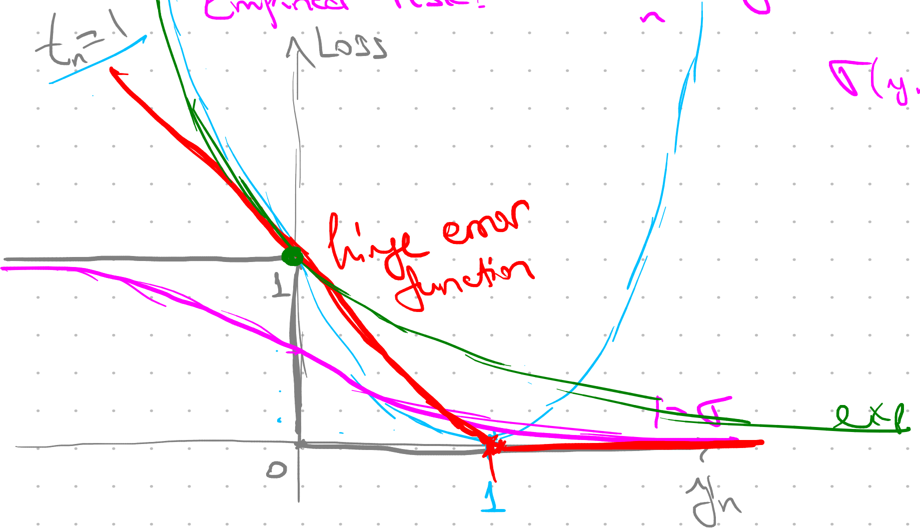
Опорн. в-ров, у к-рых $\alpha_n > 0$, не меньше νN

⑦ Error functions for classification

Empirical risk: $\sum_n [y_n \ne t_n] \xrightarrow[w]{} \min$

$\sigma(y_n)$

Exp: $e^{-t_n y_n}$



SVM:

$t_n \cdot y(\bar{x}_n) \ge 1$

$z_n + t_n y_n = 1$

⑧ Equivalent kernel

$y \sim \bar{w}^T \bar{x} \qquad p(\bar{w}) = N(\bar{0}, \Sigma_0) \qquad p(y \mid \bar{x}, \bar{w}) = N(y \mid \bar{w}^T\bar{x}, \sigma^2)$

$$p(\bar{w} \mid D) = N(\bar{w} \mid \bar{\mu}_N, \Sigma_N)$$

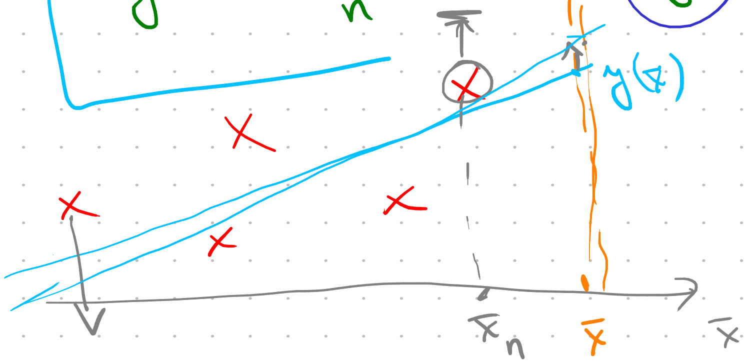
$$\Sigma_N^{-1} = \Sigma_0^{-1} + \frac{1}{\sigma^2} X^T X$$

$$\bar{\mu}_N = \frac{1}{\sigma^2}\Sigma_N X^T \bar{y}$$

$$y(\bar{x}) = \bar{\mu}_N^T \bar{x} = \frac{1}{\sigma^2}\left(\Sigma_N X^T \bar{y}\right)^T \bar{x} = \frac{1}{\sigma^2}\bar{x}^T\left(\Sigma_N X^T \boxed{\bar{y}}\right) = \sum_n \left(\Sigma_N X^T\right)_n y_n$$

$$= \frac{1}{\sigma^2} \sum_n \bar{x}^T \Sigma_N \bar{x}_n \boxed{y_n}$$

$$\boxed{y(\bar{x}) = \sum_n k(\bar{x}, \bar{x}_n)\, y_n,}$$ где $$k(\bar{x}, \bar{x}_n) = \frac{1}{\sigma^2} \bar{x}^T \Sigma_N \bar{x}_n$$



9 Relevant vector machines (RVM)

$$p(y \mid \bar{w}, \bar{x}) = \mathcal{N}\left(y \mid \bar{w}^T \bar{x},\ \beta^{-1} = \frac{1}{\sigma^2}\right)$$

$$p(\bar{y} \mid X, \bar{w}) = \prod_n \mathcal{N}(y_n \mid \bar{w}^T \bar{x}_n, \beta^{-1})$$

$$p(\bar{w} \mid \bar{\alpha}) = \prod_{i=1}^{d} \mathcal{N}(w_i \mid 0, \alpha_i^{-1})$$

$$\Sigma_0 = \begin{pmatrix} \alpha_1^{-1} & & 0 \\ & \ddots & \\ 0 & & \alpha_d^{-1} \end{pmatrix}$$

$$\begin{cases} \Sigma_N^{-1} = \Sigma_0^{-1} + \beta \cdot X^T X \\ \bar{\mu}_N = \beta \cdot \Sigma_N X^T \bar{y} \end{cases}$$

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

Empirical Bayes: $$p(\bar{y} \mid X, \bar{\alpha}, \beta) = \int \underbrace{p(\bar{w} \mid \bar{\alpha})\, p(\bar{y} \mid X, \bar{w}, \beta)}\, d\bar{w} \xrightarrow[\bar{\alpha}, \beta]{} \max$$

$$p(\bar{y} \mid X, \bar{\alpha}, \beta) = \mathcal{N}(\bar{y} \mid \bar{0}, C) \xrightarrow{\bar{\alpha}, \beta} \max,$$ где $$C = \beta^{-1} I + X \Sigma_0^{-1} X^T$$

$$p(y_n \mid X, \bar{\alpha}, \beta) = \mathcal{N}(y_n \mid \bar{\mu}_0^T \bar{x}_n,\ \bar{x}_n^T \Sigma_N \bar{x} + \sigma^2)$$ Done

$$\log p(\bar{y} \mid X, \bar{\alpha}, \beta) = -\frac{N}{2} \log 2\pi - \frac{1}{2} \log \det C - \frac{1}{2} \bar{y}^T C^{-1} \bar{y} \xrightarrow{\bar{\alpha}, \beta} \max$$

$\gamma_n = 1 - \alpha_n \Sigma_{nn}$

$$\begin{cases} \alpha_n = \dfrac{\gamma_n}{\mu_N^2} \to \infty \\ \beta^{-1} = \dfrac{\|\bar{y} - X\bar{\mu}_N\|^2}{N - \sum_n \gamma_n} \end{cases}$$

$$\hat{y}(\bar{x}) = \sum_n \boxed{k(\bar{x}, \bar{x}_n)} \cdot y_n$$

$\to 0$ если $\alpha_n \to \infty$

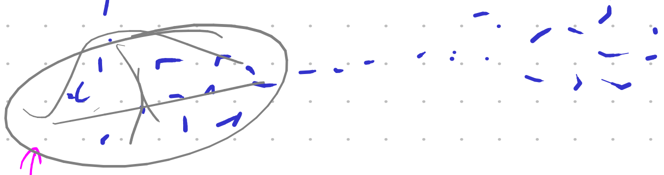
$$y(\bar{x}, \bar{w}) = \sigma(\bar{w}^T \bar{x}) \qquad A = \begin{pmatrix} \alpha_1 & & \\ & \ddots & \\ & & \alpha_d \end{pmatrix}$$

$$p(\bar{w}|\bar{\alpha}) = \prod N(w_i | 0, \alpha_i^{-1})$$

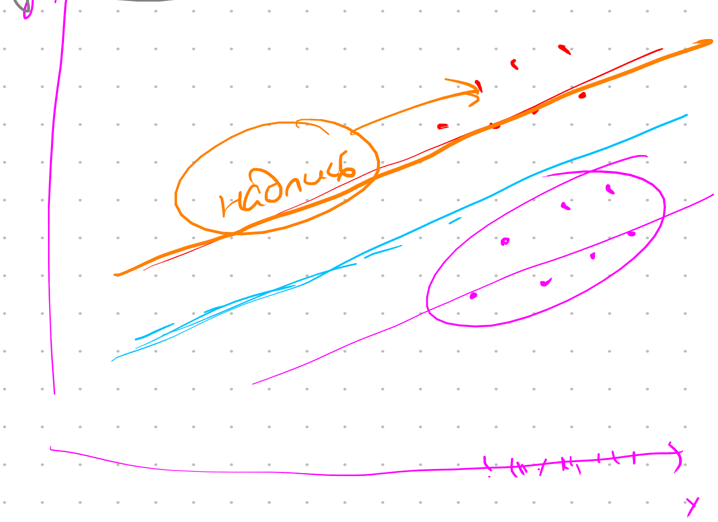
$$\log p(\bar{t} | \ldots) = \sum_n \left[t_n \log y_n + (1 - t_n)\log(1 - y_n) \right] - \frac{1}{2}\bar{w}^T A \bar{w} + \text{Const}$$

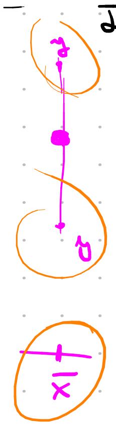
IRLS

$$p(\bar{t}|\bar{\alpha}) = \int p(\bar{t}|\bar{w}) \, p(\bar{w}|\bar{\alpha}) \, d\bar{w} = \ldots$$



$\bar{\alpha} \to \max$





$D($

$D = \{ (\bar{x}, y_1), (\bar{x}, y_2) \}$