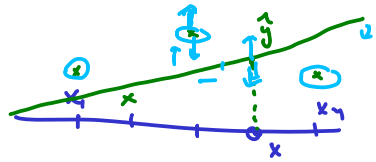
$$X^T \bar y = \sum (\bar x_n) \cdot y_n$$

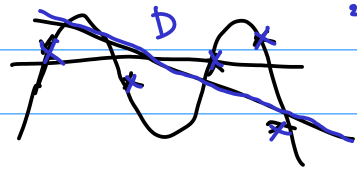


$$\mu_N = \beta \Sigma_N X^T \bar y$$

$$\hat y(\bar x) = \bar x^T \cdot \beta \Sigma_N X^T \bar y = \beta \sum_n \bar x^T \cdot \Sigma_N y_n \cdot \bar x_n = \sum_n \Big(\overbrace{\beta \cdot \bar x^T \Sigma_N \bar x_n}^{k(\bar x, \bar x_n)}\Big) \boxed{y_n}$$

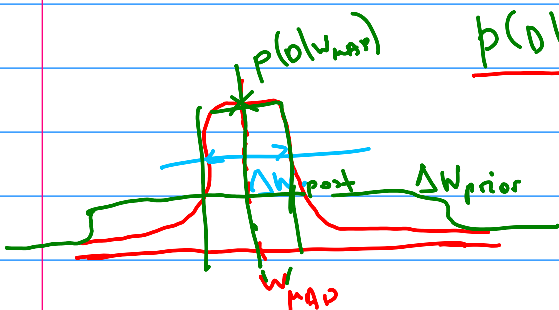
$$y(\bar x) = \sum_n \underbrace{k(\bar x, \bar x_n)}_{\text{kernel}} \cdot y_n$$

$$p(M_i|D) \propto p(M_i)\, p(D|M_i)$$



2/3 $\underset{\theta}{\boxed{M_1}}, \boxed{M_2}, \boxed{M_3}, \ldots, M_k$

$$p(\theta|D, M) = \frac{p(\theta|M) \cdot p(D|\theta, M)}{p(D|M)}$$

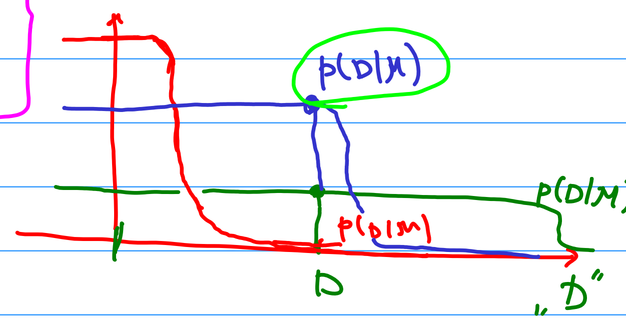


$$p(D|M) = \int \underbrace{p(\bar\theta|M)}_{\approx \frac{1}{\Delta w_{pri}}} p(D|\bar\theta, M)\, d\bar\theta \approx$$

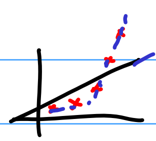
$$\approx \frac{1}{\Delta w_{pri}} \cdot p(D|w_{MAP}) \cdot \Delta w_{post}$$

$$p(D|M) \approx p(D|w_{MAP}) \cdot \frac{\Delta w_{post}}{\Delta w_{pri}}$$

$$\ln p(D|M) \approx \underbrace{\ln p(D|w_{MAP})}_{\text{likelihood}} \underbrace{- \ln \frac{\Delta w_{pri}}{\Delta w_{post}}}_{\text{penalty}}$$

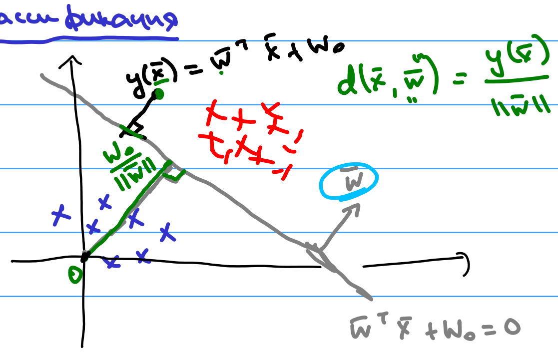


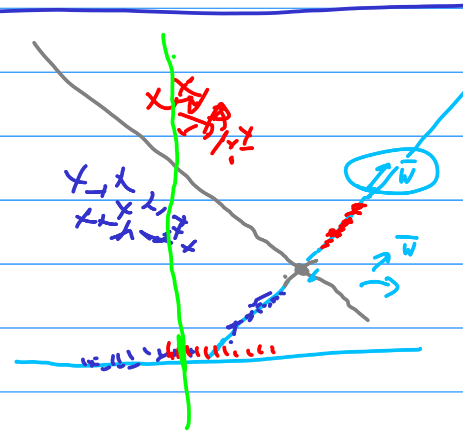
$$D \sim p(D|M_{true})$$

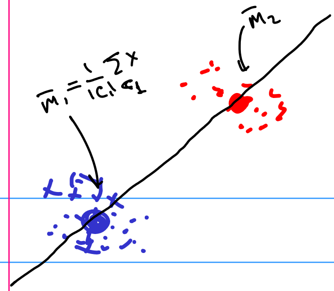


Классификация

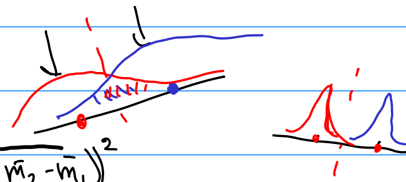
$$y(\bar x) = \bar w^T \bar x + w_0 \qquad d(\bar x, \bar w) = \frac{y(\bar x)}{\|\bar w\|}$$

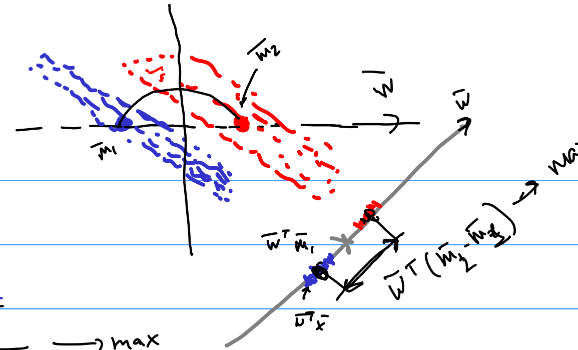






$\bar{w} = \bar{m}_2 - \bar{m}_1$





$$J(\bar{w}) = \frac{\left(\bar{w}^T(\bar{m}_2 - \bar{m}_1)\right)^2}{\sum\limits_{\bar{x}\in C_1}\underbrace{\left(\bar{w}^T\bar{m}_1 - \bar{w}^T\bar{x}\right)}_{\bar{w}^T(\bar{m}_1-\bar{x})}^2 + \sum\limits_{\bar{x}\in C_2}\underbrace{\left(\bar{w}^T\bar{m}_2 - \bar{w}^T\bar{x}\right)}_{\bar{w}^T(\bar{m}_2-\bar{x})}^2} \longrightarrow \max$$

$\cancel{A\bar{w} \propto B\bar{w}}$

$\cancel{\bar{w} \propto A^{-1}B\bar{w}}$

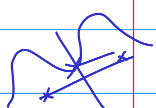
$$= \frac{\bar{w}^T \overbrace{(\bar{m}_2 - \bar{m}_1)(\bar{m}_2 - \bar{m}_1)^T}^{S_B \text{ - between-class}} \bar{w}}{\bar{w}^T \underbrace{\left(\sum_{x\in C_1}(\bar{m}_1 - \bar{x})(\bar{m}_1 - \bar{x})^T + \sum_{x \in C_2}(\bar{m}_2 - \bar{x})(\bar{m}_2 - \bar{x})^T\right)}_{S_W}\bar{w}} = \frac{\bar{w}^T S_B \bar{w}}{\bar{w}^T S_W \bar{w}} \longrightarrow \max$$

$$\nabla_{\bar{w}} J(\bar{w}) = \frac{2S_B\bar{w}\cdot(\bar{w}^T S_W \bar{w}) - 2S_W\bar{w}\,(\bar{w}^T S_B \bar{w})}{(\bar{w}^T S_W \bar{w})^2} = 0$$

$$\underbrace{(\bar{w}^T S_W \bar{w})}_{1\times 1}\cdot \underbrace{S_B\bar{w}}_{d\times 1} - \underbrace{(\bar{w}^T S_B \bar{w})}_{1\times 1}\cdot \underbrace{S_W\bar{w}}_{d\times 1} = 0$$

$$S_B\bar{w} \propto S_W\bar{w}$$

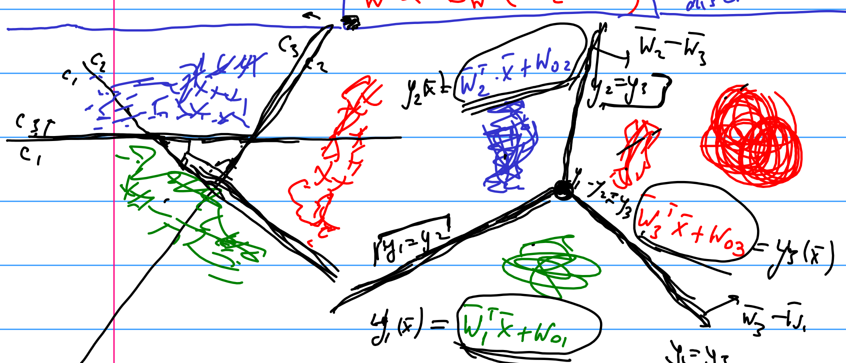
$$(\bar{m}_2 - \bar{m}_1)\underbrace{(\bar{m}_2 - \bar{m}_1)^T\bar{w}}_{"} \propto \bar{m}_2 - \bar{m}_1$$



$$S_W\bar{w} \propto \bar{m}_2 - \bar{m}_1$$

$$\boxed{\bar{w} \propto S_W^{-1}(\bar{m}_2 - \bar{m}_1)}$$

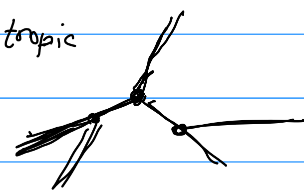
Fischer's linear discriminant



$$\bar{w}_1^T\bar{x} + w_{01} = \bar{w}_2^T\bar{x} + w_{02}$$

$$(\bar{w}_1 - \bar{w}_2)^T\bar{x} + (w_{01} - w_{02}) = 0$$

tropic



$$p(C_1|\bar{x}) = \frac{\boxed{p(C_1)}\,p(\bar{x}|C_1)}{\boxed{p(C_1)}\,p(\bar{x}|C_1) + \boxed{p(C_2)}\,p(\bar{x}|C_2)}$$

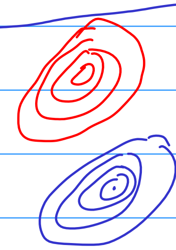
optimal Bayes classifier

$$p(C_1) \approx \frac{N_1}{N_1 + N_2}$$

$p(C_1) = p(C_2) = \frac{1}{2}$

$$p(\bar{x}|C_1) = \mathcal{N}(\bar{x}\,|\,\bar{\mu}_1, \Sigma)$$

$$p(\bar{x}|C_2) = \mathcal{N}(\bar{x}\,|\,\bar{\mu}_2, \Sigma)$$



$p(C_1|\bar{x}) \propto p(\bar{x}|C_1)\,p(C_1)$

$$\frac{1}{\mathrm{Area}(C_1)} \overset{?}{\underset{}{}} \frac{1}{\mathrm{Area}(C_2)}$$

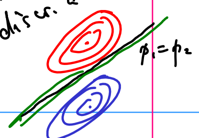
$$\ln p(\bar{x}|C_1) = \ln p(\bar{x}|C_2)$$

$$\cancel{-\frac{d}{2}\ln 2\pi} - \cancel{\frac{1}{2}\ln\det\Sigma} - \cancel{\frac{1}{2}}(\bar{x} - \bar{\mu}_1)^T\Sigma^{-1}(\bar{x} - \bar{\mu}_1) =$$

$$= \cancel{-\frac{d}{2}\ln 2\pi} - \cancel{\frac{1}{2}\ln\det\Sigma} - \cancel{\frac{1}{2}}(\bar{x} - \bar{\mu}_2)^T\Sigma^{-1}(\bar{x} - \bar{\mu}_2)$$

$$\bar{x}^T\Sigma^{-1}\bar{x} - 2\bar{x}^T\Sigma^{-1}\bar{\mu}_1 + \bar{\mu}_1^T\Sigma^{-1}\bar{\mu}_1 = \bar{x}^T\Sigma^{-1}$$

LDA
linear discr. analysis



$$\bar x^T \Sigma_1^{-1}\bar x - 2\bar x^T\Sigma^{-1}\bar\mu_1 + \bar\mu_1^T\Sigma^{-1}\bar\mu_1 = \bar x^T\Sigma_2^{-1}\bar x - 2\bar x^T\Sigma^{-1}\bar\mu_2+\bar\mu_2^T\Sigma^{-1}\bar\mu_2$$

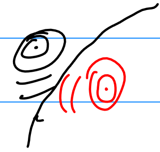
$$\bar x^T \cdot \underbrace{\Sigma^{-1}(\bar\mu_1-\bar\mu_2)}_{\boxed{\Sigma_W^{-1}}(\bar m_1-\bar m_2)} - \frac12\bar\mu_1^T\Sigma^{-1}\mu_1+\frac12\bar\mu_2^T\Sigma^{-1}\bar\mu_2 = 0$$

$p(\bar x|C_1) = \mathcal N(\bar x|\bar\mu_1,\Sigma)$

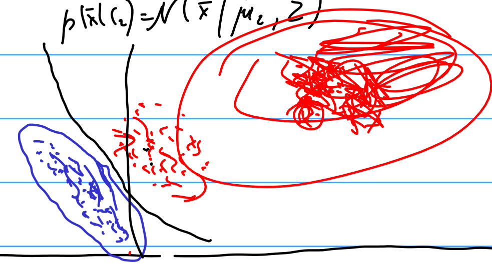
$p(\bar x|C_2) = \mathcal N(\bar x|\bar\mu_2,\Sigma)$

$p(\bar x|C_1) = \mathcal N(\bar x|\bar\mu_1,\Sigma_1)$

$p(\bar x|C_2) = \mathcal N(\bar x|\bar\mu_2,\Sigma_2)$



QDA
quadratic discr. analysis

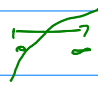


логистическая регрессия

$$p(C_1|\bar x) = \frac{p(C_1)p(\bar x|C_1)}{p(C_1)p(\bar x|C_1)+p(C_2)p(\bar x|C_2)} = \frac{1}{1+\frac{p(C_2)p(\bar x|C_2)}{p(C_1)p(\bar x|C_1)}} = \frac{1}{1+e^{-\boxed{\ln\frac{p(C_1)p(\bar x|C_1)}{p(C_2)p(\bar x|C_2)}}}}$$

log odds $\approx \bar w^T\bar x$ $\to +\infty$ C_1, $\to -\infty$ C_2

5:1 odds



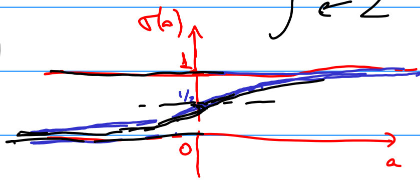
$$p(C_1|\bar x) = \frac{1}{1+e^{-\bar w^T\bar x}}$$

$$\prod_{x_n,y_n\in D} p(\bar x_n|y_n) \to \max$$

$\int e \leftarrow \Sigma$

$$\sigma(a) = \left(\frac{1}{1+e^{-a}}\right)$$

логист. сигмоид



$$1-\sigma(a) = 1-\frac{1}{1+e^{-a}} = \frac{e^{-a}}{1+e^{-a}} = \frac{1}{1+e^{a}} = \sigma(-a)$$

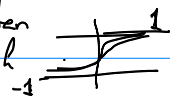
$$\sigma'(a) = \frac{e^{-a}}{(1+e^{-a})^2} = \frac{1}{1+e^{-a}}\cdot\frac{e^{-a}}{1+e^{-a}} =$$

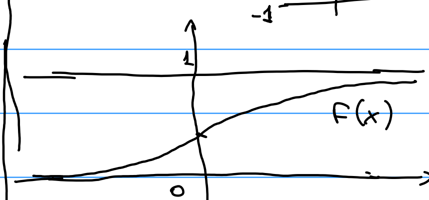
$$\sigma'(a) = \sigma(a)\cdot(1-\sigma(a))$$

$$(\ln\sigma(a))' = \frac{\sigma(a)(1-\sigma(a))}{\sigma(a)} = 1-\sigma(a)$$

$$(\ln(1-\sigma(a)))' = \sigma(a)$$

Сигмоиды: arctan, tanh





Probit:

$$\Phi(a) = \int_{-\infty}^{a} \mathcal N(x|0,\sigma^2)\,dx$$