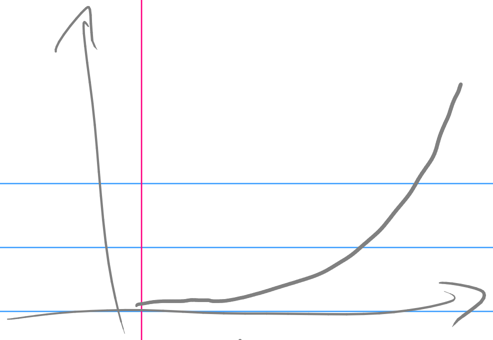
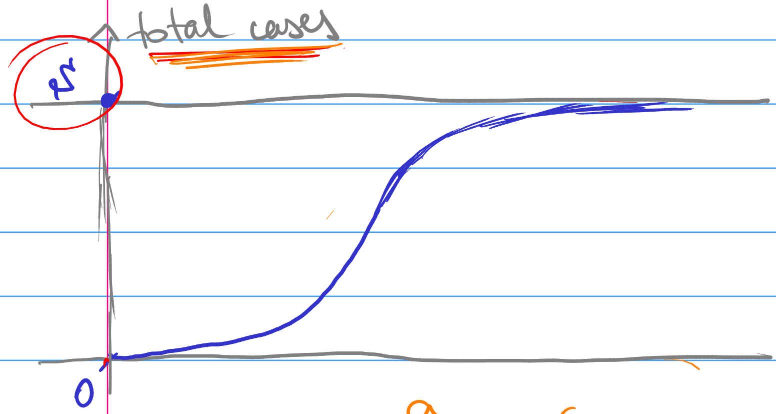


$$y \sim e^{w_1 x + w_0}$$

$$\log y \sim w_1 x + w_0$$



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

$$y \sim \sigma(w_1 x + w_0)$$

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

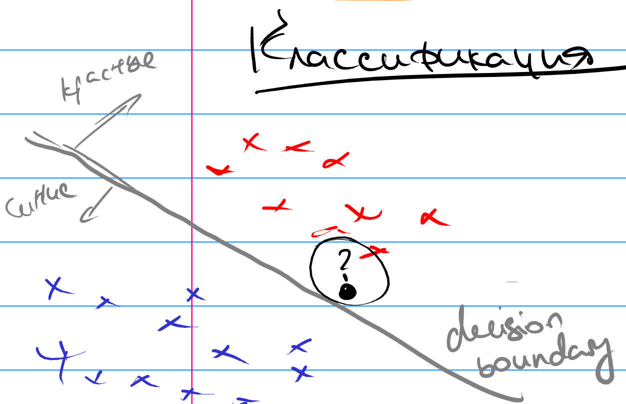
$$a = \ln\left(\frac{\sigma(a)}{1-\sigma(a)}\right)$$

2) $$\Phi(a) = \int_{-\infty}^{a} e^{-(w_0 + w_1 x + w_2 x^2)}\, dx$$

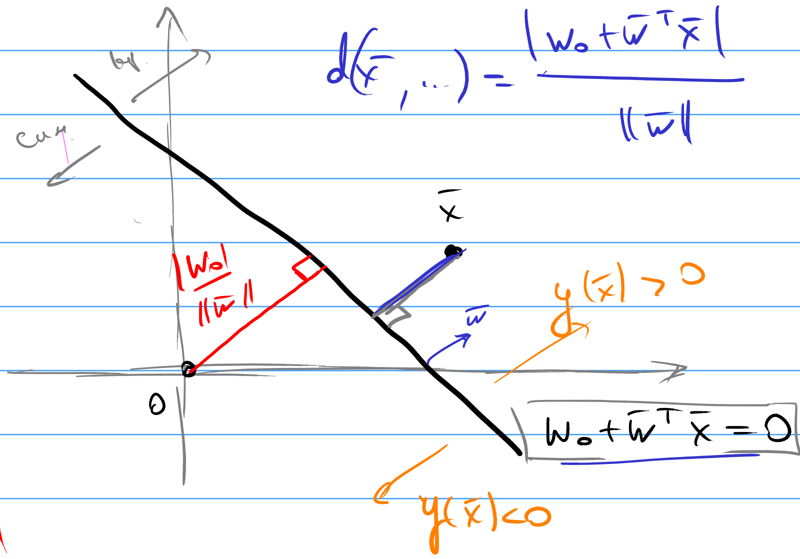
$y \sim e^{w_0 + w_1 x + w_2 x^2}$, где y = „new cases"

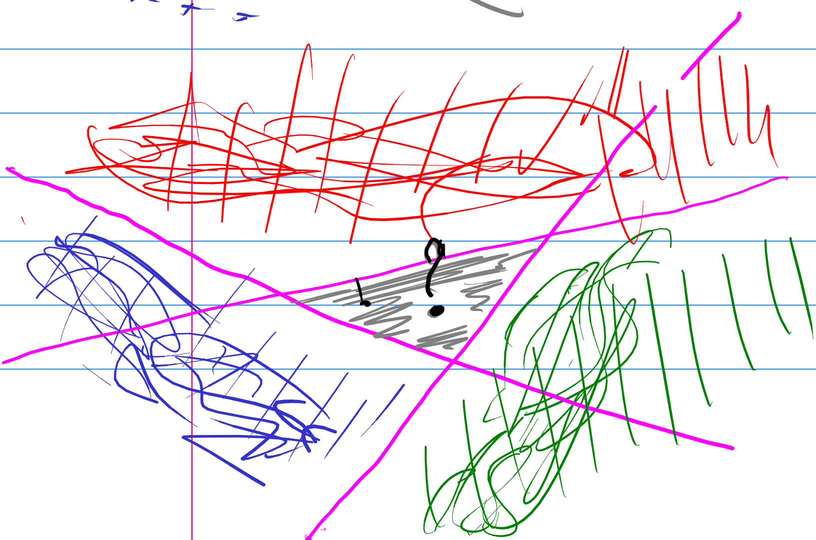
$$\ln y \sim w_0 + w_1 x + w_2 x^2$$

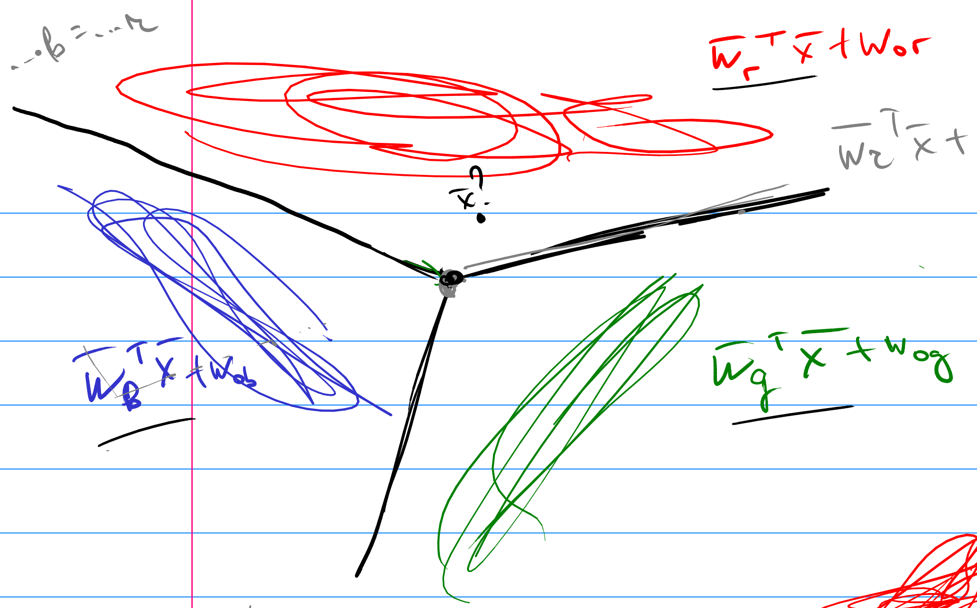
Классификация



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



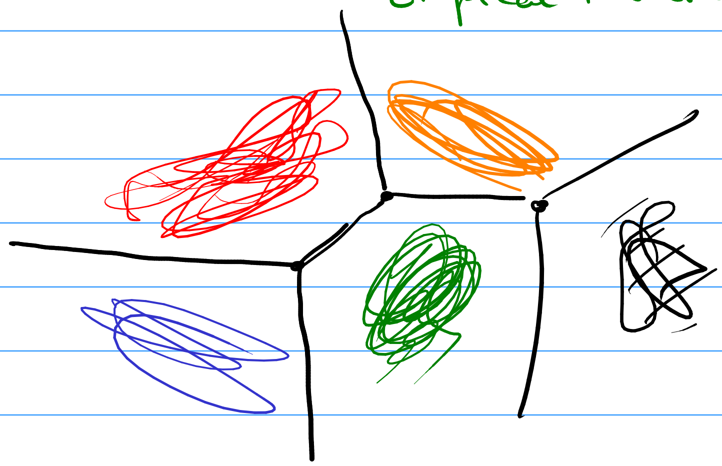




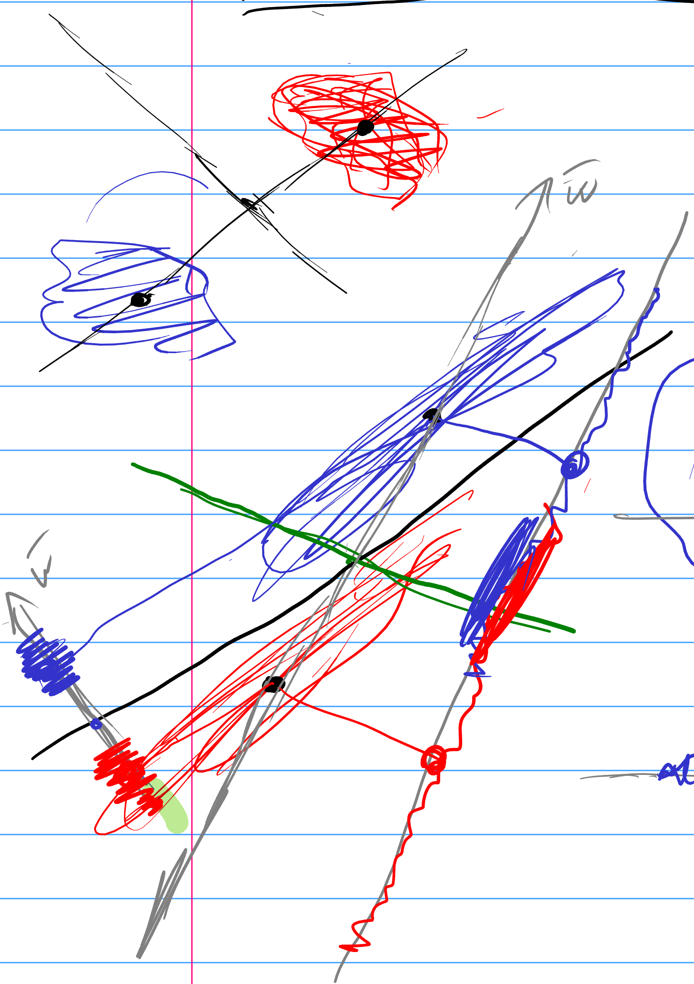
$\max(\bar{w}_r^T \bar{x} + w_{0r}, \ldots b, \ldots g)$

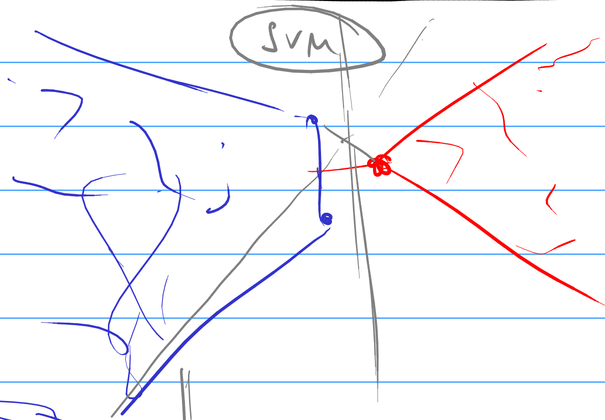
не единственный

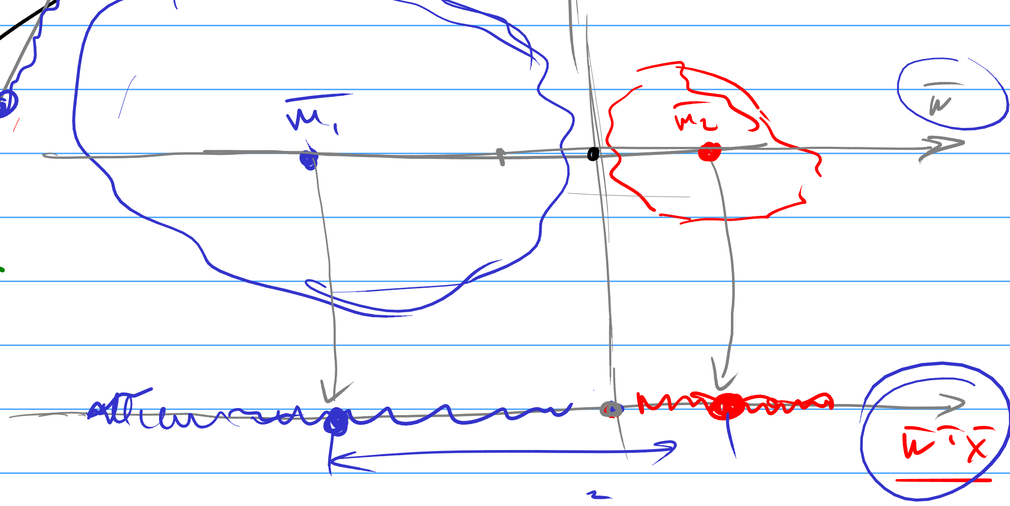
tropical mathematics



Fischer's linear discriminant







$$\left(\bar{w}^T(\bar{m}_1 - \bar{m}_2)\right)^2 \xrightarrow{\bar{w}} \max$$

$$\bar{w}^T(\bar{x} - \bar{m}_1)$$

$$\sum_{\bar{x} \in C_1} \left(\bar{w}^T \bar{x} - \bar{w}^T \bar{m}_1\right)^2 + \sum_{\bar{x} \in C_2} \left(\bar{w}^T \bar{x} - \bar{w}^T \bar{m}_2\right)^2 \xrightarrow[\bar{w}]{} \min$$

$$J(\bar{w}) = \frac{\left(\bar{w}^T(\bar{m}_1 - \bar{m}_2)\right)^2}{\sum(\ )^2 + \sum(\ )^2} = \xrightarrow[\bar{w}]{} \max$$

$$= \frac{\bar{w}^T \boxed{(\bar{m}_1 - \bar{m}_2)(\bar{m}_1 - \bar{m}_2)^T} \bar{w}}{\bar{w}^T \left(\boxed{\sum_{\bar{x} \in C_1} (\bar{x} - \bar{m}_1)(\bar{x} - \bar{m}_1)^T + \sum_{\bar{x} \in C_2} (\bar{x} - \bar{m}_2)(\bar{x} - \bar{m}_2)^T} \right) \bar{w}} \xrightarrow[\bar{w}]{} \max$$

$= S_B$ – between-class covariance

S_W – within-class covariance

$$S_W + S_B \stackrel{?}{=} S = \sum_{\bar{x}} (\bar{x} - \bar{m})(\bar{x} - \bar{m})^T$$

$$J(\bar{w}) = \frac{\bar{w}^T S_B \bar{w}}{\bar{w}^T S_W \bar{w}} \xrightarrow[\bar{w}]{} \max$$

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

$$(\bar{w}^T S_W \bar{w}) \cdot S_B \bar{w} = (\bar{w}^T S_B \bar{w}) \cdot S_W \bar{w}$$

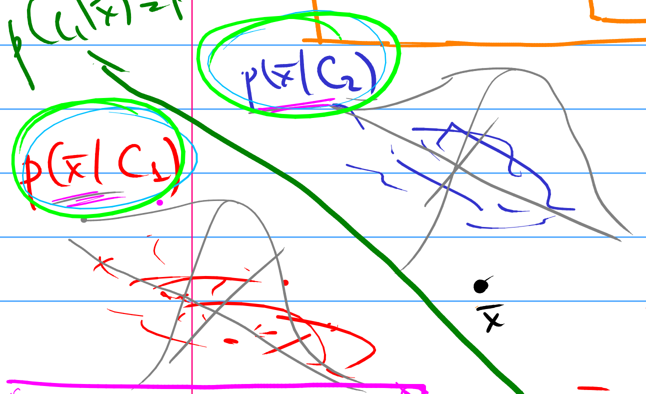
$$\boxed{S_B \bar{w} \propto S_W \bar{w}}$$

$$(\bar{m}_1 - \bar{m}_2) \underbrace{\left[(\bar{m}_1 - \bar{m}_2)^T \bar{w}\right]}_{} \propto \bar{m}_1 - \bar{m}_2$$

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

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

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



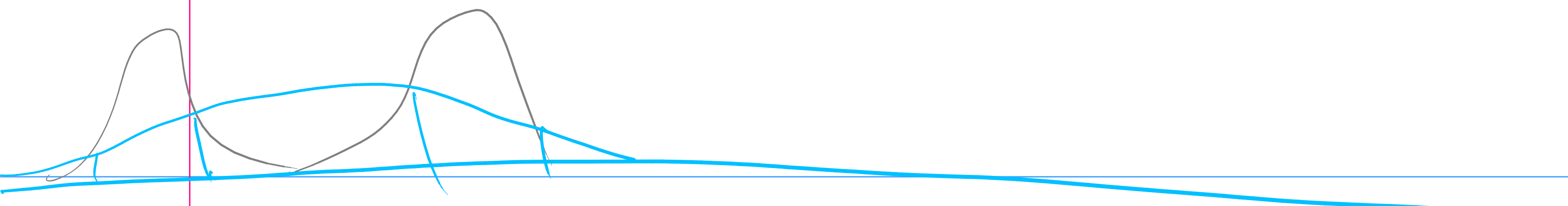
Generative models

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

optimal Bayesian classifier

$$0 \quad \bullet\, p(C_1) \quad 1$$

$$\bar{x} \sim p(\bar{x}|C_2) \qquad \bar{x} \sim p(\bar{x}|C_1)$$



$$p(\bar{x}|C_1) = \mathcal{N}(\bar{x} \,|\, \bar{\mu}_1, \Sigma_1) \qquad p(C_1) \quad p(C_2)$$

$$p(\bar{x}|C_2) = \mathcal{N}(\bar{x} \,|\, \bar{\mu}_2, \Sigma_2)$$

$$\boxed{\ln p(\bar{x}|C_1) - \ln p(\bar{x}|C_2) + \ln \frac{p(C_1)}{p(C_2)} = 0}$$

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

$$p(C_1) \cdot \frac{1}{\cancel{(\sqrt{2\pi})^d}\sqrt{\det \Sigma_1}} e^{-\frac{1}{2}(\bar{x}-\bar{\mu}_1)^T \Sigma_1^{-1}(\bar{x}-\bar{\mu}_1)} =$$

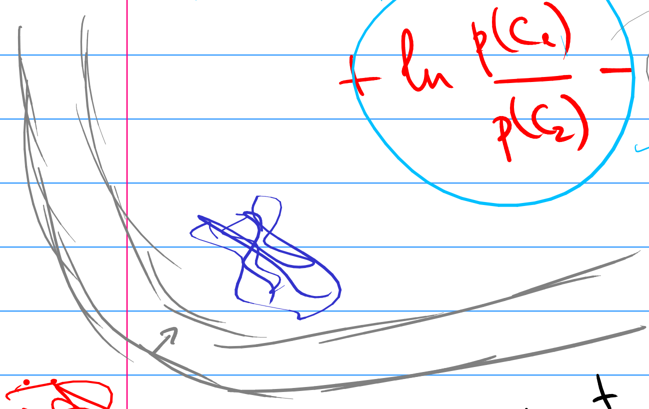
$$= p(C_2) \cdot \frac{1}{\cancel{(2\pi)^{d/2}}\sqrt{\det \Sigma_2}} e^{-\frac{1}{2}(\bar{x}-\bar{\mu}_2)^T \Sigma_2^{-1}(\bar{x}-\bar{\mu}_2)}$$

$$\ln p(C_1) - \frac{1}{2}\ln\det\Sigma_1 - \frac{1}{2}(\bar{x}-\bar{\mu}_1)^T \Sigma_1^{-1}(\bar{x}-\bar{\mu}_1) =$$

$$= \ln p(C_2) - \frac{1}{2}\ln\det\Sigma_2 - \frac{1}{2}(\bar{x}-\bar{\mu}_2)^T \Sigma_2^{-1}(\bar{x}-\bar{\mu}_2)$$

$$-\frac{1}{2}\bar{x}^T\left(\Sigma_1^{-1} - \Sigma_2^{-1}\right)\bar{x} + \bar{x}^T\left(\Sigma_1^{-1}\bar{\mu}_1 - \Sigma_2^{-1}\bar{\mu}_2\right) +$$

$$+ \ln\frac{p(C_1)}{p(C_2)} - \frac{1}{2}\ln\frac{\det\Sigma_1}{\det\Sigma_2} - \frac{1}{2}\bar{\mu}_1^T\Sigma_1^{-1}\bar{\mu}_1 + \frac{1}{2}\bar{\mu}_2^T\Sigma_2^{-1}\bar{\mu}_2 = 0$$



$$\hat{\mu}_1 = \frac{1}{N_1}\sum_{\bar{x}\in C_1} \bar{x} \qquad - , -$$

$$\hat{\Sigma}_1 = \frac{1}{N_1}\sum_{\bar{x}\in C_1} (\bar{x}-\hat{\mu}_1)(\bar{x}-\hat{\mu}_1)^T$$

QDA

quadratic discriminant analysis

$$p(C_k|\bar{x}) = \frac{p(\bar{x}|C_k)\,p(C_k)}{\sum_i p(\bar{x}|C_i)\,p(C_i)} \propto p(\bar{x}|C_k)\,p(C_k)$$

LDA – linear discriminant analysis

$$p(\bar{x}|C_1) = \mathcal{N}(\bar{x}|\bar{\mu}_1, \Sigma)$$
$$p(\bar{x}|C_2) = \mathcal{N}(\bar{x}|\bar{\mu}_2, \Sigma)$$

$$\bar{x}^T \Sigma^{-1}(\bar{\mu}_1 - \bar{\mu}_2) + \ln\frac{p(C_1)}{p(C_2)} - \frac{1}{2}\bar{\mu}_1^T \Sigma^{-1}\bar{\mu}_1 + \frac{1}{2}\bar{\mu}_2^T \Sigma^{-1} \bar{\mu}_2 = 0$$

$$\hat{\Sigma} = \frac{N_1}{N_1+N_2}\hat{\Sigma}_1 + \frac{N_2}{N_1+N_2}\hat{\Sigma}_2$$

$$\left(\prod_{\bar{x}\in C_1} p(\bar{x}|C_1)p(C_1)\right)\cdot\left(\prod_{\bar{x}\in C_2} p(C_2)\,p(\bar{x}|C_2)\right) \xrightarrow[\bar{\mu}_1, \bar{\mu}_2, \Sigma]{} \max$$

$$\Sigma \sim \quad + \quad \underbrace{N_1 \ln p(C_1)} + \underbrace{N_2 \ln p(C_2)} + \Sigma \; - - -$$

$$N_1 \ln p(C_1) + N_2 \ln(1 - p(C_1)) \xrightarrow[p(C_1)]{} \max$$

Generative models:

– обучить $p(\bar{x}|C_k)$, $p(C_k)$

– $p(C_k|\bar{x}) = \dfrac{p(\bar{x}|C_k)p(C_k)}{\sum_i p(\bar{x}|C_i)\,p(C_i)}$

Discriminative models:

$\underline{p(C_k|\bar{x})} \approx \text{model}(\bar{x}, \bar{\theta})$

$\dfrac{p(C_2|\bar{x})}{p(C_1|\bar{x})}$ – odds

$[0, \infty)$

$\cancel{p(C_1|\bar{x}) \approx \bar{w}^T\bar{x}}$

$$p(C_1|\bar{x}) = \frac{p(\bar{x}|C_1)p(C_1)}{p(\bar{x}|C_1)p(C_1) + p(\bar{x}|C_2)p(C_2)} = \frac{1}{1 + \dfrac{p(\bar{x}|C_2)p(C_2)}{p(\bar{x}|C_1)p(C_1)}}$$

$\overset{?}{\approx} \bar{w}^T\bar{x}$

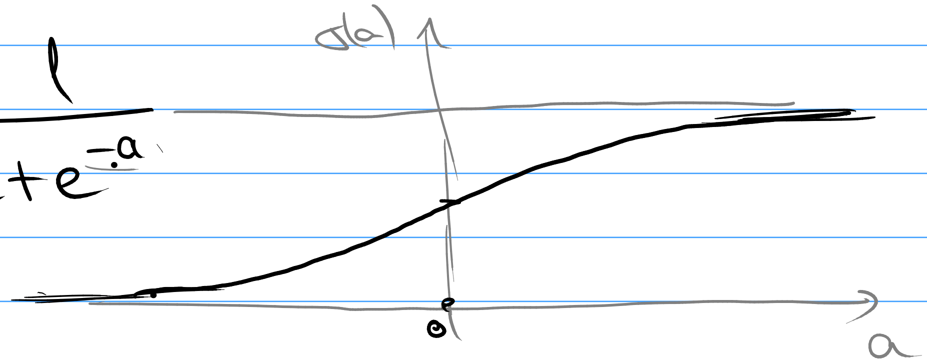
$$p(C_1|\bar{x}) = \frac{1}{1+e^{-\ln\frac{p(\bar{x}|C_1)p(C_1)}{p(\bar{x}|C_2)p(C_2)}}}$$

log odds

$\approx \bar{w}^T\bar{x}$

logistic regression

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



$$p(C_1|\bar{x}) = \sigma(\bar{w}^T\bar{x})$$

$$p(C_2|\bar{x}) = 1-\sigma(\bar{w}^T\bar{x})$$

$$\prod_{\bar{x}} \cdot \begin{cases} \sigma(\bar{w}^T\bar{x}), & \text{если } \bar{x}\in C_1 \\ 1-\sigma(\bar{w}^T\bar{x}), & \text{если } \bar{x}\in C_2 \end{cases} \xrightarrow[\bar{w}]{} \max$$