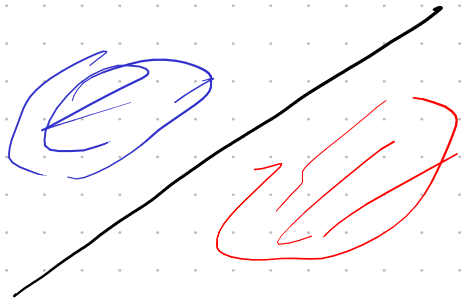
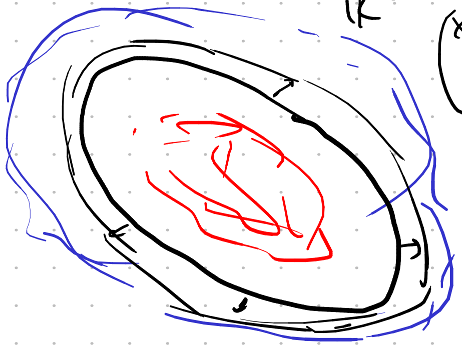




$$\mathbb{R}^2 \quad \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mapsto \begin{pmatrix} x_1 \\ x_2 \\ x_1^2 \\ x_2^2 \\ x_1 x_2 \end{pmatrix} \quad \mathbb{R}^5$$

$$\omega_0 + \omega_1 x_1 + \omega_2 x_2 + \omega_3 x_1^2 + \omega_4 x_2^2 + \omega_5 x_1 x_2 = 0$$

Generative models

$p(C_1|\bar{x}) = ?$ → discriminative

$p(C_2|\bar{x}) = 1 - p(C_1|\bar{x})$

prior

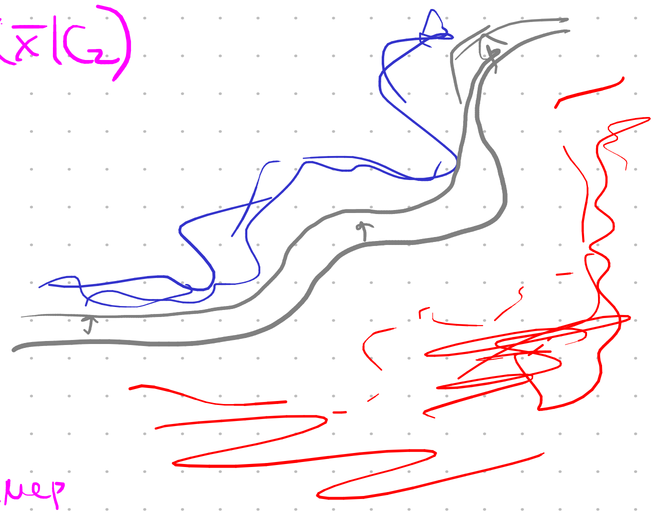
$$p(C_1|\bar{x}) = \frac{p(C_1)\,p(\bar{x}|C_1)}{p(C_1)\,p(\bar{x}|C_1) + p(C_2)\,p(\bar{x}|C_2)}$$

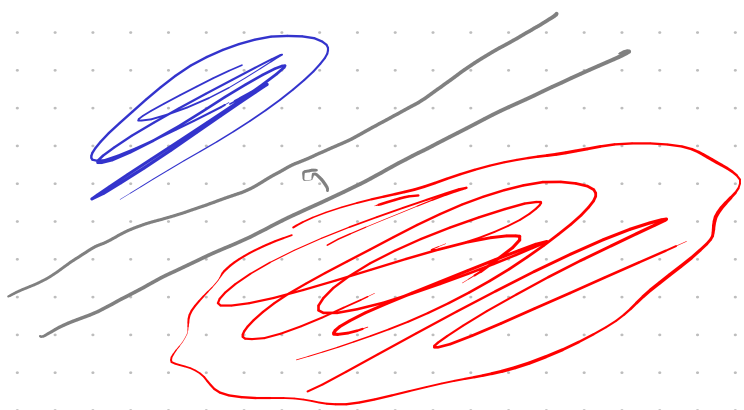
optimal Bayesian classifier

Decision surface: $p(C_1|\bar{x}) = \frac{1}{2} \iff p(C_1|\bar{x}) = p(C_2|\bar{x})$

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

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





Пример

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

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

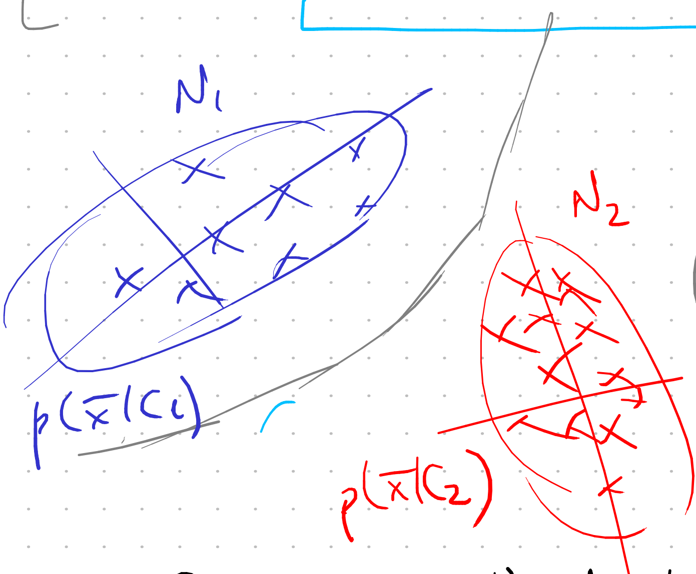
$$p(C_1|\bar{x}) = p(C_2|\bar{x})$$

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

$$\log p(C_1) - \cancel{\frac{d}{2}\log 2\pi} - \frac{1}{2}\log\det\Sigma_1 - \frac{1}{2}(\bar{x}-\bar{\mu}_1)^T\,\Sigma_1^{-1}\,(\bar{x}-\bar{\mu}_1)$$

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

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



$$\hat{\bar{\mu}}_1 = \frac{1}{N_1}\sum_{n\in C_1}\bar{x}_n \qquad \hat{\bar{\mu}}_2$$

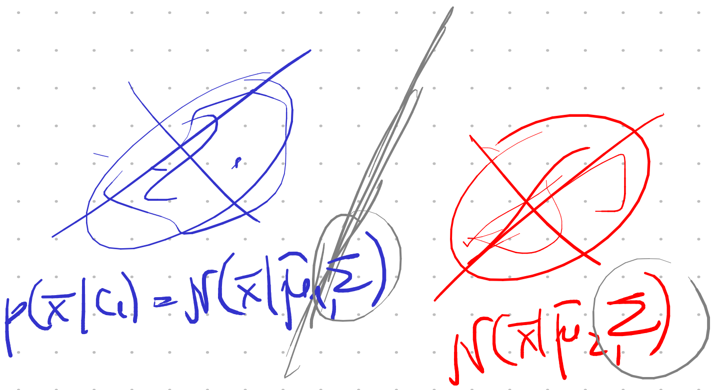
$$\boxed{\hat{\Sigma}_1 = \frac{1}{N_1}\sum_{n\in C_1}(\bar{x}-\hat{\bar{\mu}}_1)(\bar{x}-\hat{\bar{\mu}}_1)^T} \qquad \hat{\Sigma}_2$$

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

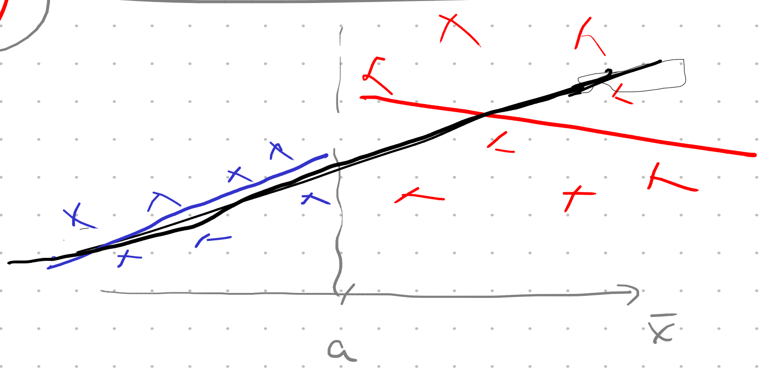
QDA – quadratic discriminant analysis

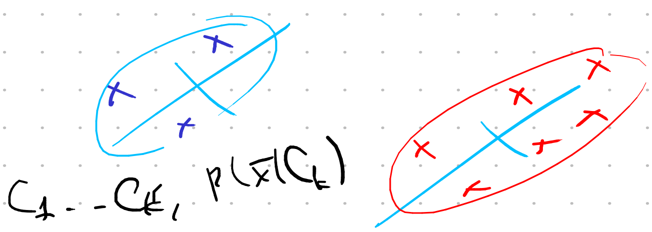
LDA – linear discriminant analysis

$$\Sigma_1 = \Sigma_2 = \Sigma$$



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





$$p(C_k|\bar{x}) = \frac{p(C_k)\,p(\bar{x}|C_k)}{\sum_\ell p(C_\ell)\,p(\bar{x}|C_\ell)}$$

Logistic Regression - discriminative model

$\bar{x}$

$$D = \{(\bar{x}_n, t_n)\}_{n=1}^{N}, \quad t_n = \begin{cases} 1, & \bar{x}_n \in C_1 \\ 0, & \bar{x}_n \in C_2 \end{cases}$$

$\cancel{f(\bar{w}^T\bar{x})}$ $\cancel{\approx \bar{w}^T\bar{x}}$?

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

odds $\cancel{\approx \bar{w}^T\bar{x}}$?

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

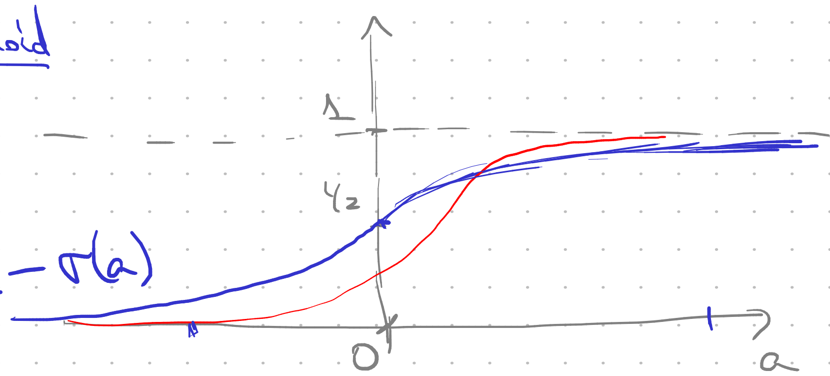
log-odds $\approx \bar{w}^T\bar{x}$

$$= \frac{1}{1 + e^{-\underbrace{\log\frac{p(C_1)\,p(\bar{x}|C_1)}{p(C_2)\,p(\bar{x}|C_2)}}_{=a}}}$$

$a \approx \bar{w}^T\bar{x}$ logistic sigmoid

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

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



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

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

$$\sigma(\bar{w}^T\bar{x}) = \frac{1}{2}$$

$$\boxed{\bar{w}^T\bar{x} = 0}$$

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

$$D = \{(\bar{x}_n, t_n)\}_{n=1}^{N} \qquad p(D|\bar{w}) = \prod p(t_n|\bar{w},\bar{x}_n) \xrightarrow{\bar{w}} \max$$

$$p(D|\bar w) = \prod_n p(t_n|\bar w, \bar x_n) = \prod_n \begin{cases} \sigma(\bar w^T \bar x_n), & \text{если } t_n = 1 \\ 1-\sigma(\bar w^T \bar x_n), & \text{если } t_n = 0 \end{cases}$$

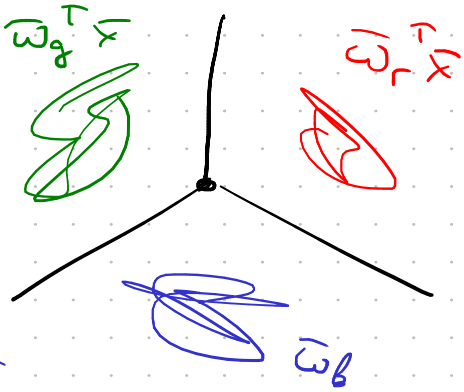
$$= \prod_n \cancel{\left(t_n \cdot \sigma(\bar w^T \bar x_n) + (1-t_n)(1-\sigma(\bar w^T \bar x_n))\right)}$$

$$= \prod_{n=1}^{N} \sigma(\bar w^T \bar x_n)^{t_n} \left(1-\sigma(\bar w^T \bar x_n)\right)^{1-t_n}$$

$$\nabla_w(\cdot) = t_n \cdot \frac{1}{\sigma(\bar w^T \bar x_n)} \cdot \sigma \cdot (1-\sigma) \cdot \bar x_n$$

$$\log p(D|\bar w) = \sum_{n=1}^{N} \left(t_n \log \sigma(\bar w^T \bar x_n) + (1-t_n) \log(1-\sigma(\bar w^T \bar x_n)) \right) \xrightarrow[\bar w]{} \max$$

$$p(C_k|\bar x) = \frac{p(C_k)\, p(\bar x|C_k)}{\sum_{s=1}^{K} p(C_s)\, p(\bar x|C_s)} = (*)$$



$$\bar w_k^T \bar x \approx \log p(C_k)\, p(\bar x|C_k) + const$$

$$(*) = \frac{e^{\bar w_k^T \bar x \cancel{+const}}}{\sum_{s=1}^{K} e^{\bar w_s^T \bar x \cancel{+const}}}$$

$$\text{softmax}(a_1, \dots, a_K) = \left(\dots, \frac{e^{a_k}}{\sum_s e^{a_s}}, \dots \right)$$

$\bar w_1, \dots, \bar w_K$ $\qquad$ $\bar w_k^T \bar x$

$$\bar t_n = (0 \dots \underset{k}{1} \dots 0)$$ one-hot encoding

$$t_{nk} = 1 \iff \bar x_n \in C_k$$

$$p(D|W) = \prod_n p(\bar t_n | W, \bar x_n) = \prod_{n=1}^{N} \prod_{k=1}^{K} p(C_k|\bar w_k, \bar x_n)^{t_{nk}}$$

$$\log p(D|W) = \sum_n \sum_k t_{nk} \log y_{nk},$$

где $\bar y_n = \text{softmax}(\bar w_1^T \bar x_n, \dots, \bar w_K^T \bar x_n)$, т.е. $y_{nk} = \dfrac{e^{\bar w_k^T \bar x_n}}{\sum_s e^{\bar w_s^T \bar x_n}}$

$$\nabla_{\bar{w}} \log p(D|\bar{w}) = \sum_{n=1}^{N} \Big[t_n \cdot \frac{1}{\cancel{\sigma(\bar{w}^T \bar{x}_n)}} \cdot \cancel{\sigma(\bar{w}^T \bar{x}_n)} \big(1-\sigma(\bar{w}^T \bar{x}_n)\big) \cdot \bar{x}_n \; -$$

$$- (1-t_n) \cdot \frac{1}{\cancel{1-\sigma(\bar{w}^T \bar{x}_n)}} \cdot \sigma(\bar{w}^T \bar{x}_n) \cancel{\big(1-\sigma(\bar{w}^T \bar{x}_n)\big)} \cdot \bar{x}_n \Big]$$

$$= \sum_n \Big[t_n \big(1-\sigma(\bar{w}^T \bar{x}_n)\big) - (1-t_n)\, \sigma(\bar{w}^T \bar{x}_n) \Big] \bar{x}_n$$

$$t_n - \cancel{\sigma_n t_n} - \sigma + \cancel{\sigma_n t_n}$$

$$\boxed{\nabla_{\bar{w}} \log p(D|\bar{w}) = \sum_n \big[t_n - \sigma(\bar{w}^T \bar{x}_n) \big] \cdot \bar{x}_n}$$