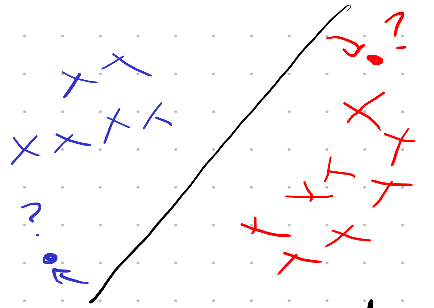
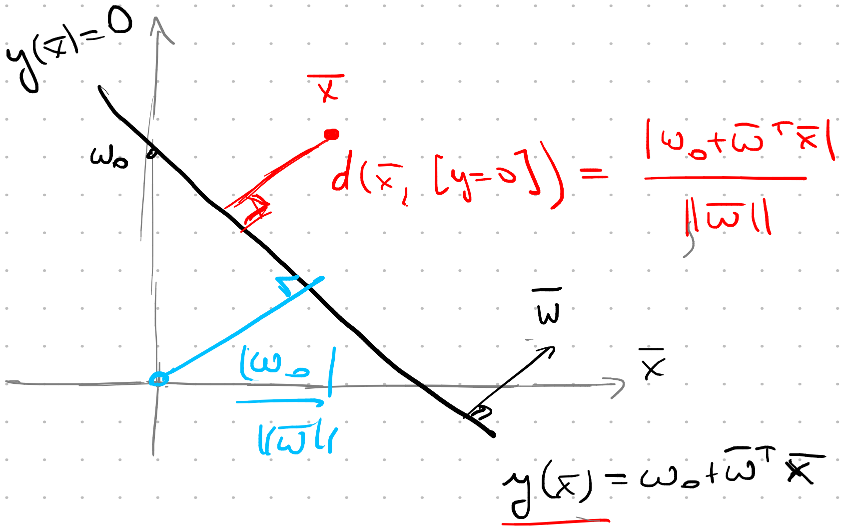
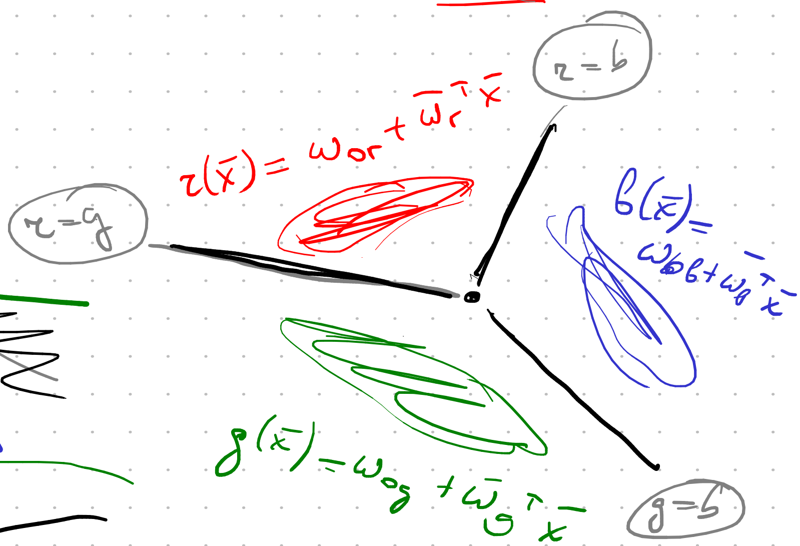
① Classification

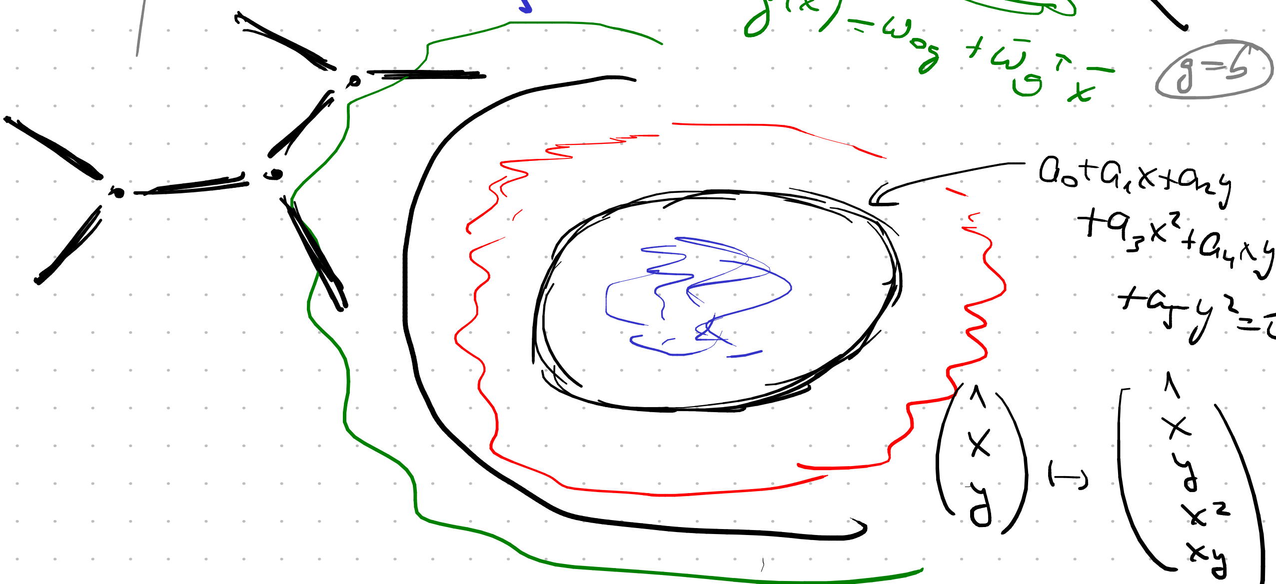


decision boundary/surface

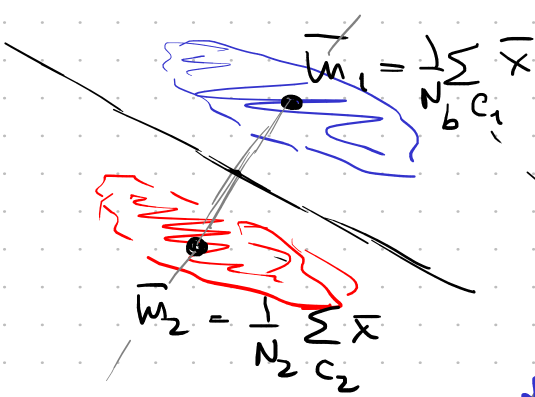


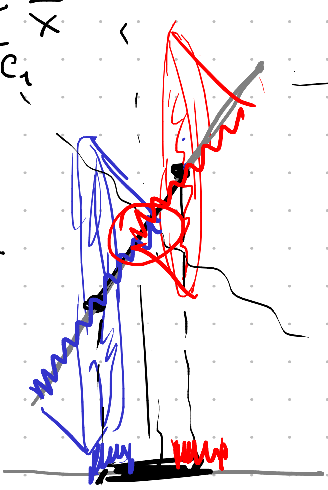


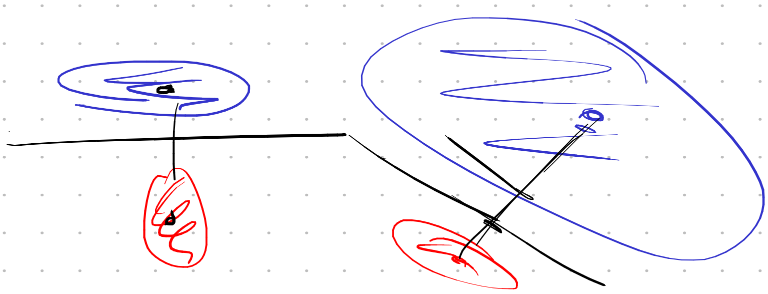
tropical mathematics

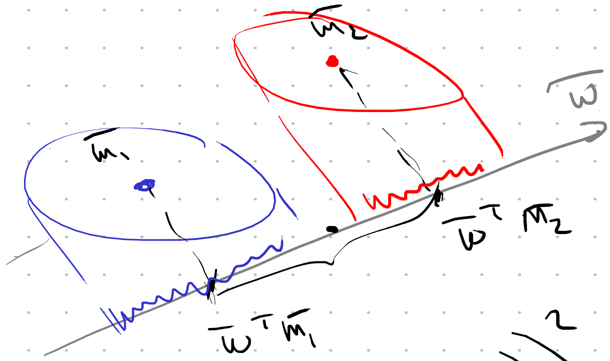


② Fisher's linear discriminant









$$\bar\omega^T(\bar x-\bar m_1)(\bar x-\bar m_1)^T\bar\omega$$

$$\sum_{\bar x\in C_1}\left(\bar w^T\bar x-\bar w^T\bar m_1\right)^2 + \sum_{\bar x\in C_2}\left(\bar w^T\bar x-\bar w^T\bar m_2\right)^2 \xrightarrow[\bar w]{} \min$$

Fisher's linear discriminator

$$\left(\bar w^T(\bar m_1-\bar m_2)\right)^2 \xrightarrow[\bar w]{} \max$$

$$J(\bar w)=\frac{\bar\omega^T\underbrace{(\bar m_1-\bar m_2)(\bar m_1-\bar m_2)^T}_{S_B\ \text{between-class covariance}}\bar\omega}{\bar\omega^T\underbrace{\left(\sum_{C_1}(\bar x-\bar m_1)(\bar x-\bar m_1)^T+\sum_{C_2}(\bar x-\bar m_2)(\bar x-\bar m_2)^T\right)}_{S_W\ \text{within-class covariance}}\bar\omega}\xrightarrow[\bar\omega]{}\max$$

$$S_B+S_W=S=\sum(\bar x-\bar m)(\bar x-\bar m)^T$$

$$J(\bar w)=\frac{\bar\omega^T S_B\bar\omega}{\bar\omega^T S_W\bar w}\to\max$$

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

$$\underbrace{(\bar w^T S_W\bar w)}_{\in\mathbb{R}}\cdot S_B\bar w=\underbrace{(\bar w^T S_B\bar w)}_{\in\mathbb{R}}\cdot S_W\bar w$$

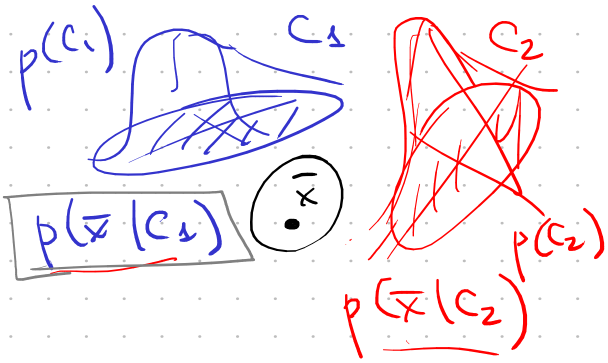
$$S_B\bar w\propto S_W\bar w$$

$$(\bar m_1-\bar m_2)\underbrace{\left[(\bar m_1-\bar m_2)^T\bar w\right]}_{\in\mathbb{R}}\propto\bar m_1-\bar m_2$$

$$S_W\bar\omega\propto\bar m_1-\bar m_2$$

$$\boxed{\bar\omega\propto S_W^{-1}(\bar m_1-\bar m_2)}$$

3 Generative models



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

optimal Bayes classifier

discriminative models: $p(C_1|\bar{x}) = \dots \bar{w}$ ___

generative models: $p(\bar{x}|C_1) =$ _ _ $\bar{w}$ _

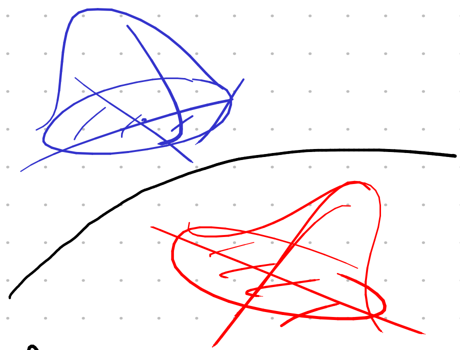
$$p(C_k|\bar{x}) = \frac{p(\bar{x}|C_k)\,p(C_k)}{\sum p(\bar{x}|C_j)p(C_j)}$$

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

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

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

4 LDA, QDA



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

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

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

$$\log\frac{p(C_1)}{p(C_2)} - \frac{d}{2}\log 2\pi - \frac{1}{2}\log\det\Sigma_1 - \frac{1}{2}(\bar{x}-\bar{\mu}_1)\Sigma_1^{-1}(\bar{x}-\bar{\mu}_1) =$$

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

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

optimal Bayes classifier

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

$$\begin{cases}\hat{\mu}_1=\frac{1}{N}\sum\limits_{C_1}\bar{x}\\ \hat{\Sigma}_1=\frac{1}{N_1}\sum\limits_{C_1}(\bar{x}-\hat{\mu}_1)(\bar{x}-\hat{\mu}_1)^T\end{cases}$$

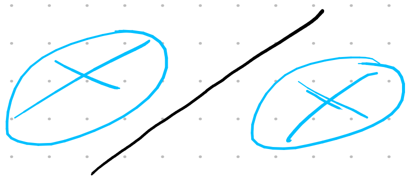
QDA

quadratic discriminant analysis

LDA - linear discriminant analysis

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

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



$\bar{\mu}_1, \bar{\mu}_2, \Sigma \to \max$

$$p(D|\bar{\mu}_1,\bar{\mu}_2,\Sigma)=\underbrace{\prod_{C_1}p(\bar{x}|\bar{\mu}_1,\Sigma_1)}\underbrace{\prod_{C_2}p(\bar{x}|\bar{\mu}_2,\Sigma_2)}$$

$$=\underbrace{\left(\prod_{C_1}\ldots\bar{\mu}_1,\hat{\Sigma}_1\right)}_{\bar{\mu}_1,\Sigma_1\to\max}\times\underbrace{\left(\prod_{C_2}\ldots\bar{\mu}_2,\Sigma_2\right)}_{\bar{\mu}_2,\Sigma_2\to\max}$$

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

5 Logistic regression / discriminative model

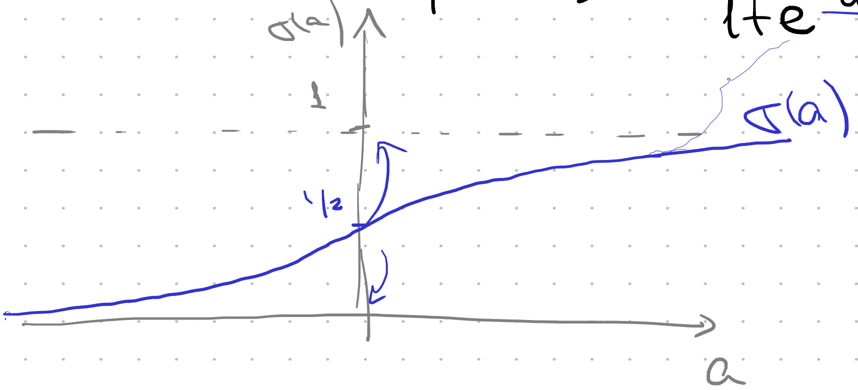
$$\overbrace{p(C_1|\bar{x})}^{\approx\bar{w}^T\bar{x}}=\frac{p(C_1)p(\bar{x}|C_1)}{p(C_1)p(\bar{x}|C_1)+p(C_2)p(\bar{x}|C_2)}=$$

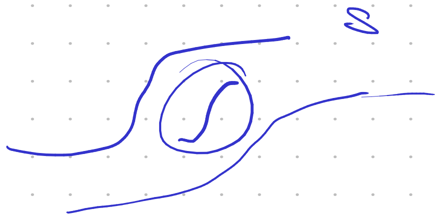
odds — $\frac{p(C_2|\bar{x})}{p(C_1|\bar{x})}$

log-odds

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

$\bar{w}^T\bar{x} \approx a$ $\quad p(C_1|\bar{x}) = \frac{1}{1+e^{-a}} = \sigma(a)$ — logistic sigmoid





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

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

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

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

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

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

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