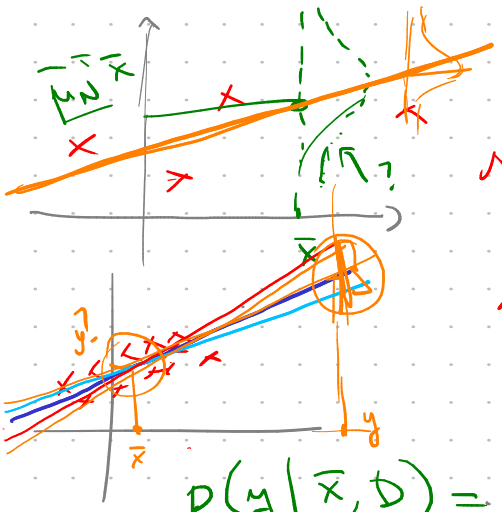
(1) Predictive distribution in LR



$$p(\bar{w}|D) \propto p(\bar{w})\,p(D|\bar{w})$$

$$\mathcal{N}(\bar{w}|\bar{\mu}_N,\Sigma_N) \propto \mathcal{N}(\bar{w}|\bar{\mu}_0,\Sigma_0)\cdot \prod_n \mathcal{N}(y_n|\bar{w}^T\bar{x}_n,\sigma^2)$$

$$\begin{cases} \Sigma_N^{-1} = \Sigma_0^{-1} + \frac{1}{\sigma^2}X^TX \\ \bar{\mu}_N = \Sigma_N\left(\Sigma_0^{-1}\bar{\mu}_0 + \frac{1}{\sigma^2}X^T\bar{y}\right) \end{cases}$$

$$p(y|\bar{x},D) = \int p(y|\bar{x},\bar{w})\,p(\bar{w}|D)\,d\bar{w} \approx \frac{1}{R}\sum_r p(y|\bar{x},\bar{w}^{(r)}), \quad \bar{w}^{(r)} \sim p(\bar{w}|D)$$

$$= \int \mathcal{N}(y|\bar{w}^T\bar{x},\sigma^2)\,\mathcal{N}(\bar{w}|\bar{\mu}_N,\Sigma_N)\,d\bar{w}$$

$$= \text{const}(y)\cdot \mathcal{N}(\bar{w}|\bar{\mu}',\Sigma')$$

$$\log p(y|\bar{x},\bar{w}) + \log p(\bar{w}|D) =$$

$$= -\frac{1}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}(y-\bar{w}^T\bar{x})^2 - \frac{d}{2}\log 2\pi - \frac{1}{2}\log\det\Sigma_N -$$

$$- \frac{1}{2}(\bar{w}-\bar{\mu}_N)^T\Sigma_N^{-1}(\bar{w}-\bar{\mu}_N) =$$

$$= \text{Const} - \frac{1}{2\sigma^2}\left(y^2 - 2y\,\bar{w}^T\bar{x} + \bar{w}^T\bar{x}\bar{x}^T\bar{w}\right) -$$

$$- \frac{1}{2}\left(\bar{w}^T\Sigma_N^{-1}\bar{w} - 2\bar{w}^T\left(\Sigma_N^{-1}\bar{\mu}_N\right) + \bar{\mu}_N^T\Sigma_N^{-1}\bar{\mu}_N\right) =$$

$$= \text{Const} - \frac{y^2}{2\sigma^2} - \frac{1}{2}\left(\bar{w}^T\left(\Sigma_N^{-1} + \frac{1}{\sigma^2}\bar{x}\bar{x}^T\right)\bar{w} - 2\bar{w}^T\left(\Sigma_N^{-1}\bar{\mu}_N + \frac{y}{\sigma^2}\bar{x}\right)\right)$$

$$\pm \frac{1}{2}\bar{\mu}'^T\Sigma'^{-1}\bar{\mu}' \quad -$$

$$p(y|\bar{x},D) = \mathcal{N}(y|\bar{\mu}_{pred},\sigma^2_{pred})$$

$$\bar{\mu}_{pred} = \bar{\mu}_N^T\bar{x} = \bar{w}_{MAP}^T\bar{x}, \qquad \boxed{\sigma^2_{pred} = \sigma^2 + \bar{x}^T\Sigma_N\bar{x}}$$

$$\mathcal{N}(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

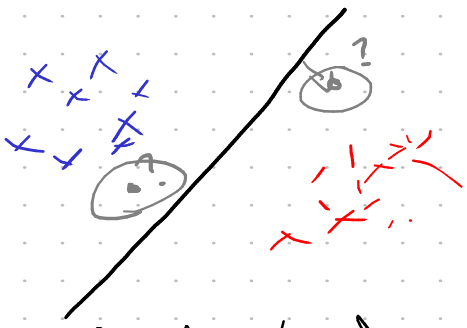
precision $\tau = \frac{1}{\sigma^2}$

$$\mathcal{N}(x \mid \mu, \tau) = \sqrt{\frac{\tau}{2\pi}} e^{-\frac{\tau}{2}(x-\mu)^2}$$

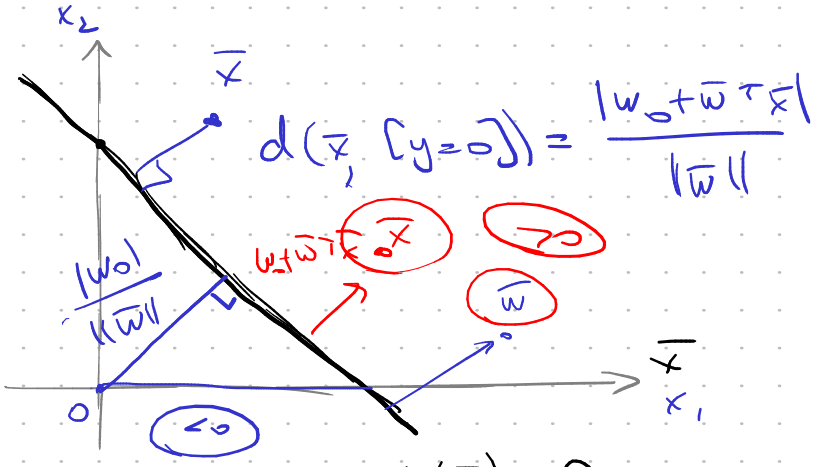
matrix $\Lambda = \Sigma^{-1}$

$$\mathcal{N}(\bar{x} \mid \mu, \Lambda) = \frac{\sqrt{\det \Lambda}}{(2\pi)^{d/2}} \cdot e^{-\frac{1}{2}(\bar{x}-\bar{\mu})^T \Lambda (\bar{x}-\bar{\mu})}$$

(2) Classification



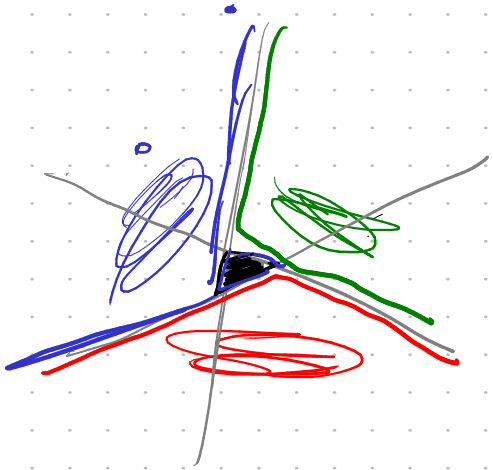
decision boundary / surface



$$y(\bar{x}) = 0$$

$$w_0 + \bar{w}^T \bar{x} = 0$$





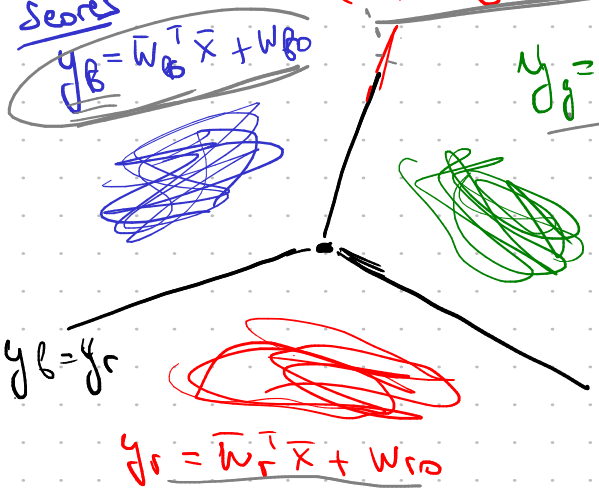
affinity scores

$y_B = \bar{w}_B^T \bar{x} + w_{B0}$

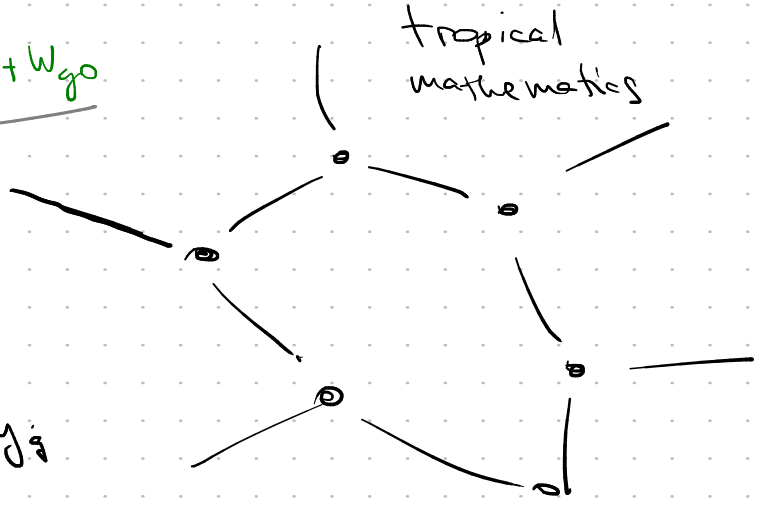
$y_g = \bar{w}_g^T \bar{x} + w_{g0}$

$y_B = y_g$

$(\bar{w}_B^T - \bar{w}_g^T)\bar{x} + w_{B0} - w_{g0} = 0$

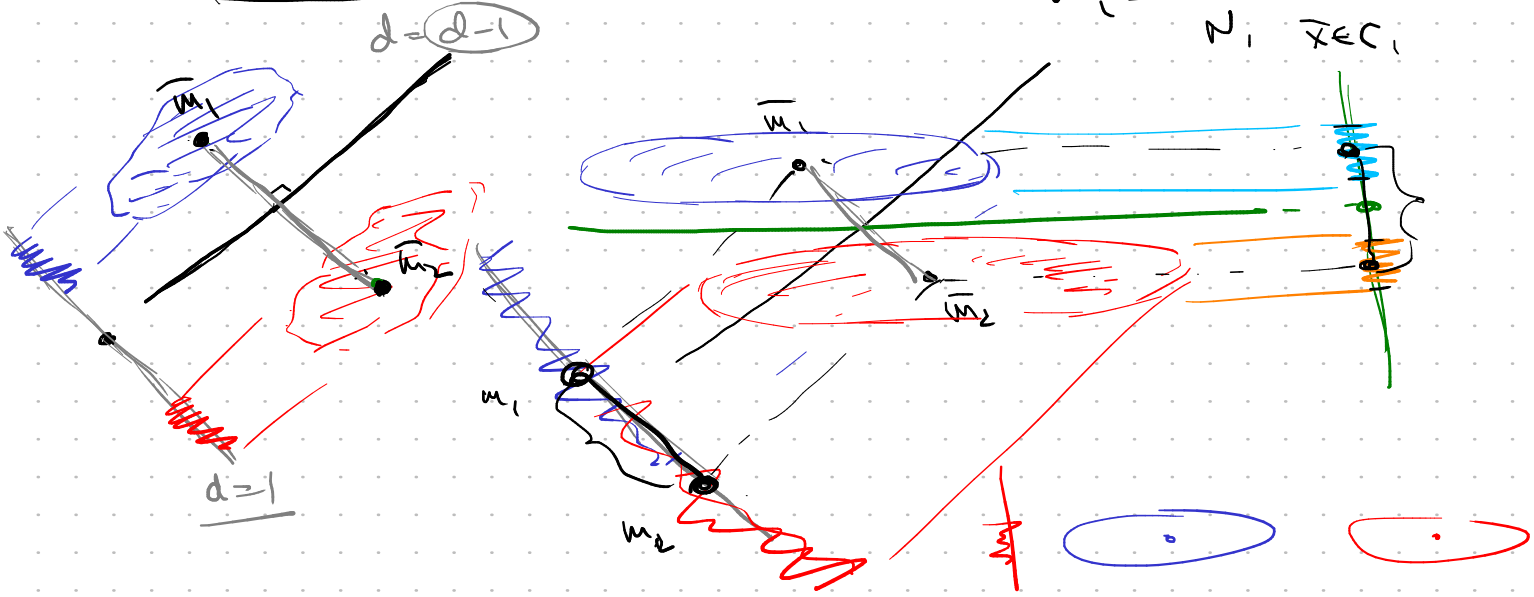


tropical mathematics



(3) Fisher's linear discriminant

$$\bar{m}_i = \frac{1}{N_i} \sum_{\bar{x} \in C_i} \bar{x}$$



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

S_B – between-class covariance

$$J(\bar{w}) = \frac{\bar{w}^T \underbrace{(\bar{m}_1 - \bar{m}_2)(\bar{m}_1 - \bar{m}_2)^T}_{S_B} \bar{w}}{\bar{w}^T \underbrace{\left(\sum_{C_1} (\bar{x} - \bar{m}_1)(\bar{x} - \bar{m}_1)^T + \sum_{C_2} (\bar{x} - \bar{m}_2)(\bar{x} - \bar{m}_2)^T\right)}_{S_w} \bar{w}} \xrightarrow{\bar{w}} \max$$

S_w – within-class cov.

$$S_B + S_w = S = \sum (\bar{x} - \bar{m})(\bar{x} - \bar{m})^T$$

$$J(\bar{w}) = \frac{\bar{w}^T S_B \bar{w}}{\bar{w}^T S_w \bar{w}}$$

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

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

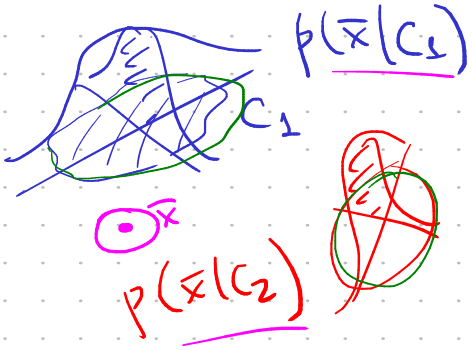
$$(\bar{w}^T S_w \bar{w}) \cdot S_B \bar{w} = (\bar{w}^T S_B \bar{w}) S_w \bar{w}$$

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

$$S_w \bar{w} \propto (\bar{m}_1 - \bar{m}_2)$$

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

(4) Generative models for classification



prior

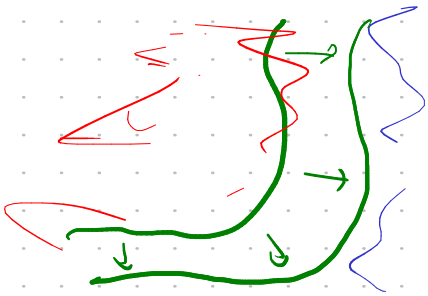
$$p(C_1|\bar{x}) = \frac{p(\bar{x}|C_1)\,p(C_1)}{p(\bar{x}|C_1)p(C_1) + p(\bar{x}|C_2)p(C_2)}$$

optimal Bayes classifier

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

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

$$p(C_k|\bar{x}) = \frac{p(\bar{x}|C_k)p(C_k)}{\sum_s p(\bar{x}|C_s)p(C_s)}$$



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

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

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

$w^T \bar{x}$ $\quad$ w_0

(5) LDA, QDA

$p(\bar{x}|C_1) = N(\bar{x}|\bar{\mu}_1, \Sigma_1)$ $\quad$ — " — C_2, μ_2, Σ_2

$\hat{\mu}_1 = \frac{1}{N_1}\sum_{C_1} x$ $\quad$ Σ_1

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

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

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

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

QDA - quadratic discriminant analysis

$\Sigma = \Sigma_1 = \Sigma_2 \Rightarrow$ LDA - linear discr. analysis

$$p(D \mid \bar{\mu}_1, \bar{\mu}_2, \Sigma) = \prod_{C_1} p(\bar{x} \mid \bar{\mu}_1, \Sigma) \cdot \prod_{C_2} p(\bar{x} \mid \bar{\mu}_2, \Sigma)$$

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

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