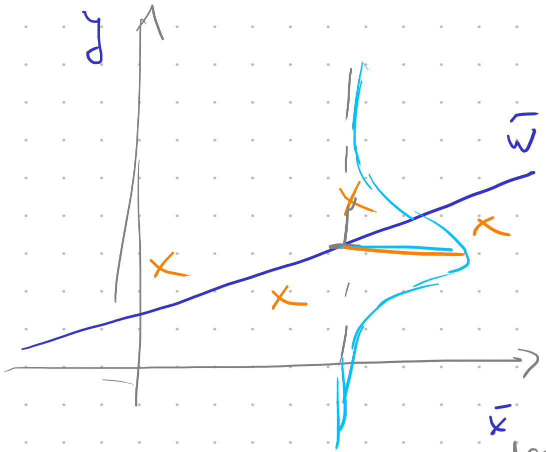


$$y \approx \bar{w}^T \bar{x}$$

$$p(y|\bar{w}, \bar{x}) = \mathcal{N}(y|\bar{w}^T \bar{x}, \sigma^2)$$

$$p(\bar{y}|\bar{w}, X) = \prod_n \mathcal{N}(y_n|\bar{w}^T \bar{x}_n, \sigma^2)$$

$$\log p(\bar{y}|\bar{w}, X) = \text{Const} - \frac{1}{2\sigma^2} \sum_n (y_n - \bar{w}^T \bar{x}_n)^2$$

regularizer

$$L = \sum_n (y_n - \bar{w}^T \bar{x}_n)^2 + \alpha \cdot \|\bar{w}\|^2 \xrightarrow{\bar{w}} \min$$

L_2-reg, Ridge regression | L_F-reg

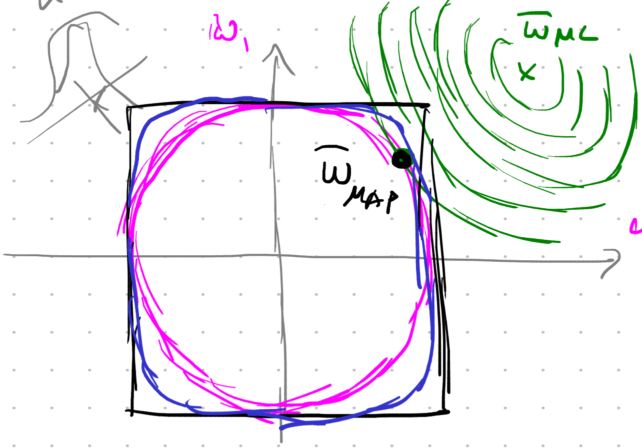
$$L = \sum_n (y_n - \bar{w}^T \bar{x}_n)^2 + \alpha \cdot \sum_{i=1}^{d} |w_i| \xrightarrow{\bar{w}} \min$$

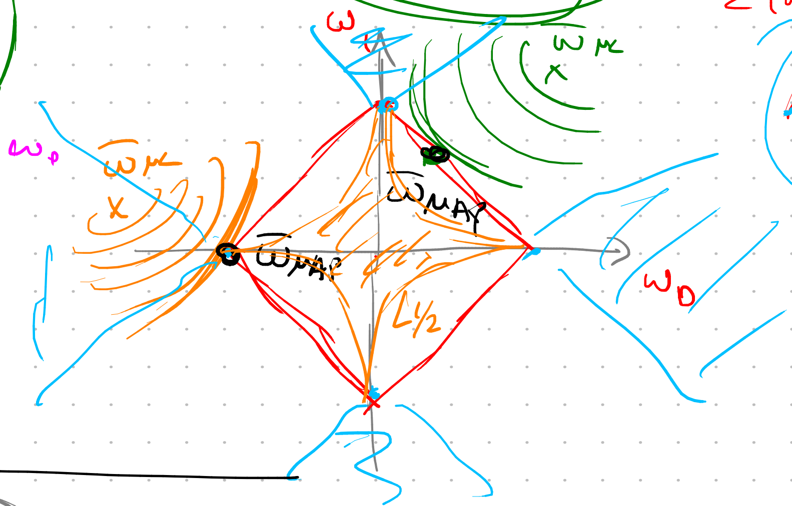
L_1-reg, Lasso regression

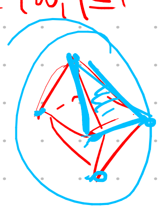
$$\sum_n (\dots)^2 + \alpha \|\bar{w}\|^2 \xrightarrow{\bar{w}} \min \iff \sum_n (\dots)^2 \to \min, \quad \|\bar{w}\|^2 \le \alpha$$

sparsity

$\sum |w_i| \le \alpha$







$$\mathcal{N}(x|\mu, \sigma^2)$$

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

$$p(\bar{w}) = \mathcal{N}(\bar{w}|\bar{0}, \sigma_0^2 \cdot I) = \frac{1}{\sqrt{(2\pi)^d \cdot \sigma_0^{2d}}} e^{-\frac{1}{2\sigma_0^2} \bar{w}^T \bar{w}}$$

$$\mathcal{N}(\underbrace{\bar{x}}_{\in \mathbb{R}^d}|\bar{\mu}, \Sigma) = \frac{1}{\sqrt{(2\pi)^d \det \Sigma}} e^{-\frac{1}{2}(\bar{x} - \bar{\mu})^T \Sigma^{-1} (\bar{x} - \bar{\mu})}$$

$$p(w_i) = \mathcal{N}(w_i|0, \sigma_0^2), \quad p(\bar{w}) = \prod p(w_i) \qquad \mathcal{N}\left(\bar{w}\middle|\begin{pmatrix}0\\ \vdots \\ 0\end{pmatrix}, \begin{pmatrix}\sigma_0^2 & & \\ & \ddots & \\ & & \sigma_0^2\end{pmatrix}\right)$$

$$\log p(\bar{w} \mid D) = \text{const} + \log p(\bar{w}) + \log p(D \mid \bar{w}) =$$

$$= \text{const} - \frac{1}{2\sigma_0^2} \bar{w}^T \bar{w} - \frac{1}{2\sigma^2} \sum_n (y_n - \bar{w}^T \bar{x}_n) \xrightarrow{\bar{w}} \max$$

$$\sum_n (y_n - \bar{w}^T \bar{x}_n)^2 + \underbrace{\boxed{\frac{\sigma^2}{\sigma_0^2}}}_{= \alpha} \bar{w}^T \bar{w} \xrightarrow{\bar{w}} \min$$

L_1-reg. $\rightsquigarrow$ $p(\bar{w}) = \text{const} \cdot e^{-\text{const} \cdot \sum |w_i|}$

$\theta^{\alpha-1}(1-\theta)^{\beta-1}$ Conjugate prior $\theta^n (1-\theta)^m$ $\propto \theta^{n+\alpha-1}(1-\theta)^{m+\beta-1}$

$$\underline{p(\bar{w} \mid \bar{\alpha})} \times p(D \mid \bar{w}) \propto p(\bar{w} \mid \bar{\alpha}')$$

$$\prod_{n=1}^{N} \mathcal{N}(y_n \mid \bar{w}^T \bar{x}_n, \sigma^2)$$

$\propto \mathcal{N}(\bar{w} \mid \mu_N, \Sigma_N)$? $\times$ $c \cdot e^{-\frac{1}{\sigma^2} \sum (y_n - \bar{w}^T \bar{x}_n)^2}$ $\propto$!

$$p(\bar{w} \mid \bar{\mu}_0, \Sigma_0) = \mathcal{N}(\bar{w} \mid \bar{\mu}_0, \Sigma_0) = \frac{1}{(2\pi)^{d/2} \sqrt{\det \Sigma_0}} e^{-\frac{1}{2}(\bar{w} - \bar{\mu}_0)^T \Sigma_0^{-1} (\bar{w} - \bar{\mu}_0)}$$

$\times$

$$p(\bar{y} \mid X, \bar{w}) = \prod_n \mathcal{N}(y_n \mid \bar{w}^T \bar{x}_n, \sigma^2)$$

$$\log p(\bar{w} \mid \bar{y}, \bar{\mu}_0, \Sigma_0, X) = \text{const} - \frac{1}{2} (\bar{w} - \bar{\mu}_0)^T \Sigma_0^{-1} (\bar{w} - \bar{\mu}_0)$$

$$- \frac{1}{2\sigma^2} \cdot \sum_n (\bar{w}^T \bar{x}_n - y_n)^2 = (\bar{y} - X\bar{w})^T (\bar{y} - X\bar{w})$$

$$= \text{const} - \frac{1}{2} \left(\bar{w}^T \Sigma_0^{-1} \bar{w} - 2 \bar{w}^T \Sigma_0^{-1} \bar{\mu}_0 + \cancel{\bar{\mu}_0^T \Sigma_0^{-1} \bar{\mu}_0} \right)$$

$$- \frac{1}{2\sigma^2} \left(\cancel{\bar{y}^T \bar{y}} - 2 \bar{w}^T X^T \bar{y} + \bar{w}^T X^T X \bar{w} \right) =$$

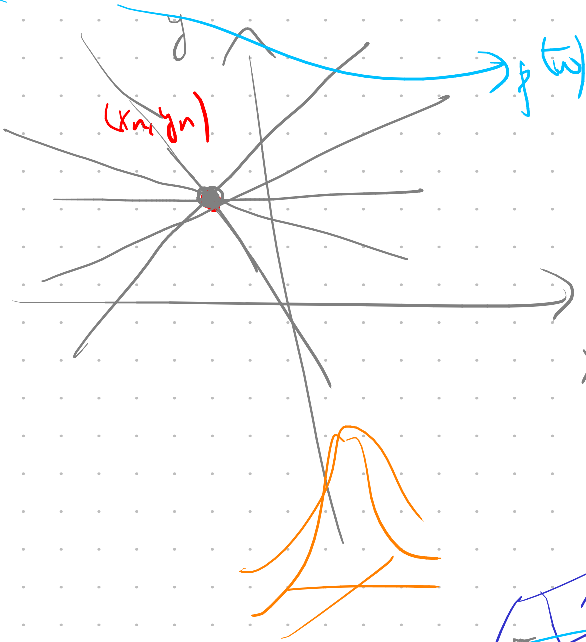
$$= \text{Const} - \frac{1}{2}\bar{\omega}^T \underbrace{\left(\Sigma_0^{-1} + \frac{1}{\sigma^2}X^TX\right)}_{=\Sigma_N^{-1}}\bar{\omega} + \bar{\omega}^T \underbrace{\left(\Sigma_0^{-1}\bar{\mu}_0 + \frac{1}{\sigma^2}X^T\bar{y}\right)}_{=\Sigma_N^{-1}\bar{\mu}_N}$$

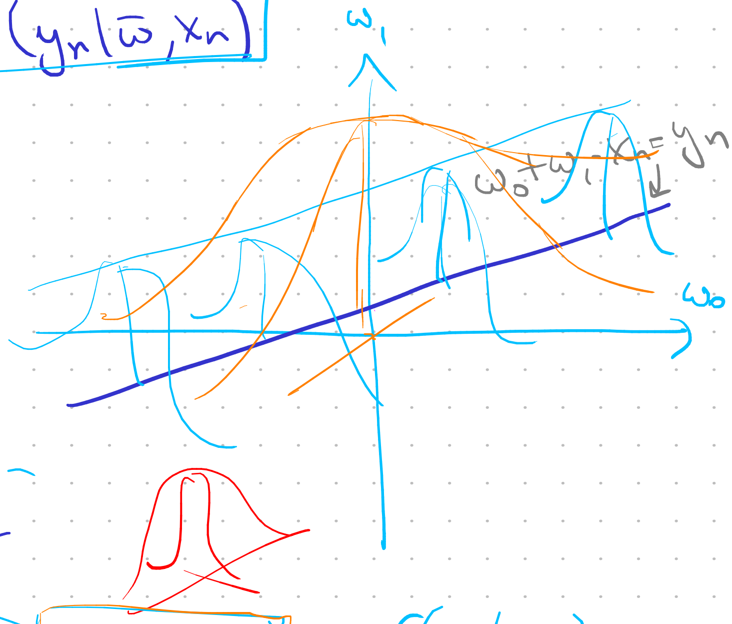
$$p(\bar\omega|\bar y, X) = \mathcal{N}(\bar\omega|\bar\mu_N, \Sigma_N),$$ где

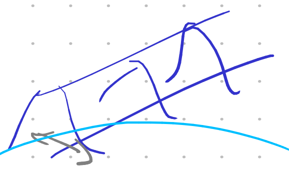
$$\begin{cases} \Sigma_N^{-1} = \Sigma_0^{-1} + \frac{1}{\sigma^2}X^TX, \\ \bar\mu_N = \Sigma_N\cdot\left(\Sigma_0^{-1}\bar\mu_0 + \frac{1}{\sigma^2}X^T\bar y\right) \end{cases}$$

$$p(\bar\omega) \xrightarrow{D_1=(\bar x_1, y_1)} p(\bar\omega|y_1) \xrightarrow{D_2=(\bar x_2, y_2)} p(\bar\omega|y_1, y_2) \longrightarrow$$

$$p(\bar\omega) \times \boxed{p(y_n|\bar\omega, x_n)}$$







Predictive distribution

$$= f(y, \bar x, D)\cdot\mathcal{N}(\bar\omega|\ldots)$$

$$p(y|\bar x, D) = \int p(y|\bar\omega, \bar x)\cdot p(\bar\omega|D)\,d\bar\omega =$$

$$= E_{p(\bar\omega|D)}[p(y|\bar\omega,\bar x)] \approx \frac{1}{R}\sum_r p(y|\bar\omega^{(r)}, \bar x)$$

$$\log p(y|\bar\omega,\bar x) + \log p(\bar\omega|D) = -\frac{1}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}(y - \bar\omega^T\bar x)^2$$

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

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

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

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

$$= \text{Const} - \frac{y^2}{2\sigma^2} - \frac{1}{2}(\bar{w}-\bar{\mu}')^T\Sigma'^{-1}(\bar{w}-\bar{\mu}') + \frac{1}{2}\bar{\mu}'^T\Sigma'^{-1}\bar{\mu}'$$

где $\Sigma'^{-1} = \Sigma_N^{-1} + \frac{1}{\sigma^2}\bar{x}\bar{x}^T$

$\bar{\mu}' = \Sigma'\left(\Sigma_N^{-1}\bar{\mu}_N + \frac{y}{\sigma^2}\bar{x}\right)$

$$+\frac{1}{2}(\_\_)^T\Sigma'\cancel{\Sigma'^{-1}\Sigma'}(\_\_)$$

$$= \frac{1}{2}(\_\_)^T\Sigma'(\_\_)$$

$$\log p(y|\bar{x},D) = \text{Const} - \frac{y^2}{2\sigma^2} + \frac{1}{2}\bar{\mu}'^T\Sigma'^{-1}\bar{\mu}' =$$

$$\Sigma_N^{-1}\bar{\mu}_N + \frac{y}{\sigma^2}\bar{x}$$

$$= \text{Const} - \frac{1}{2\sigma_{pred}^2}(y-\mu_{pred})^2$$

$\Sigma'\bar{x} \;|\; \Sigma_N\bar{x} \;|\; \Sigma'^{-1}\Sigma_N\bar{x}$

$$\Sigma_N\bar{x} = \Sigma'(\Sigma'^{-1}\Sigma_N\bar{x})$$

$$(\Sigma'^{-1}\Sigma_N)\bar{x} = \bar{x} + \frac{1}{\sigma^2}\bar{x}(\bar{x}^T\Sigma_N\bar{x}) = \left(1 + \frac{1}{\sigma^2}\bar{x}^T\Sigma_N\bar{x}\right)\bar{x}$$

$$\Sigma_N\bar{x} = \left(1 + \frac{1}{\sigma^2}\bar{x}^T\Sigma_N\bar{x}\right)\cdot\Sigma'\bar{x}$$

$$\Sigma'\bar{x} = \frac{\Sigma_N\bar{x}}{1 + \frac{1}{\sigma^2}\bar{x}^T\Sigma_N\bar{x}}$$

$$y^2: \; \frac{1}{2\sigma_{pred}} - \frac{1}{2}\left(\frac{1}{\sigma^2} - \frac{1}{\sigma^4}\bar{x}^T\Sigma'\bar{x}\right) =$$

$$= -\frac{1}{2}\left(\frac{1}{\sigma^2} - \frac{\bar{x}^T\Sigma_N\bar{x}}{\sigma^4 + \sigma^2\bar{x}^T\Sigma_N\bar{x}}\right) = -\frac{1}{2\sigma^2}\left(1 - \frac{(-)}{\sigma^2+(-)}\right) =$$

$$= -\frac{1}{2\sigma^2}\cdot\frac{\sigma^2}{\sigma^2+(-)} = -\frac{1}{2(\sigma^2+\bar{x}^T\Sigma_N\bar{x})}$$

μ_{pred}, σ^2_{pred}

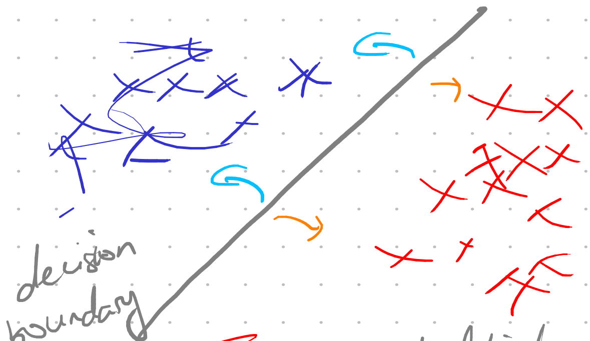
$$\boxed{\sigma^2_{pred} = \sigma^2 + \bar{x}^T\Sigma_N\bar{x}} \qquad \boxed{\mu_{pred} = \bar{\mu}_N^T\bar{x}}$$

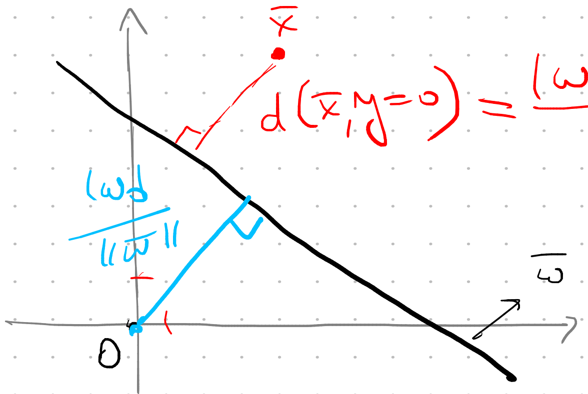
$\bar{w}_{MAP}$

$$y^1: \; \frac{1}{\sigma^2}\bar{\mu}_N^T\Sigma_N^{-1}\Sigma'\bar{x} = \frac{1}{\sigma^2}\,\frac{\bar{\mu}_N^T\bar{x}}{1+\frac{1}{\sigma^2}\bar{x}^T\Sigma_N\bar{x}} =$$

$$= \frac{\bar{\mu}_N^T\bar{x}}{\sigma^2+\bar{x}^T\Sigma_N\bar{x}} = \frac{\bar{\mu}_N^T\bar{x}}{\sigma^2_{pred}}$$

Classification

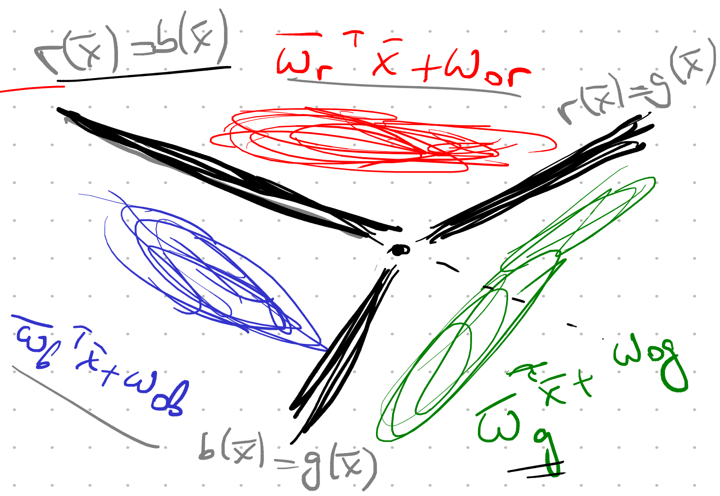


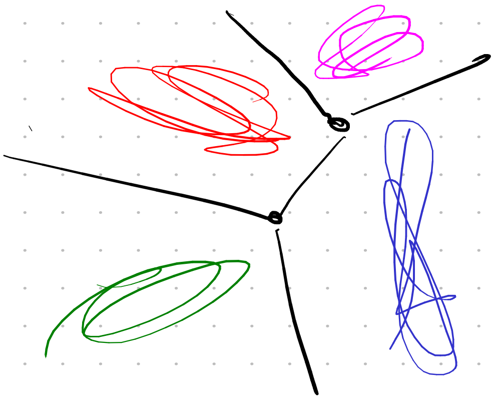


$$d(\bar{x}, y=0) = \frac{|w_0 + \bar{w}^T\bar{x}|}{\|\bar{w}\|}$$

$$y(\bar{x}) = w_0 + \bar{w}^T\bar{x}$$

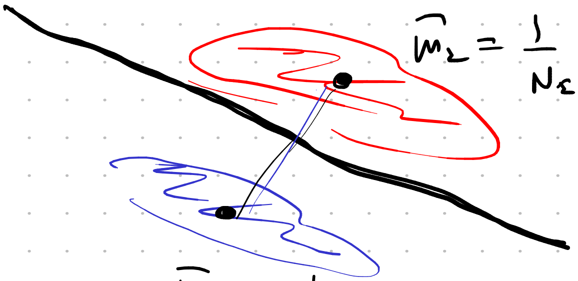
Multiclass classification



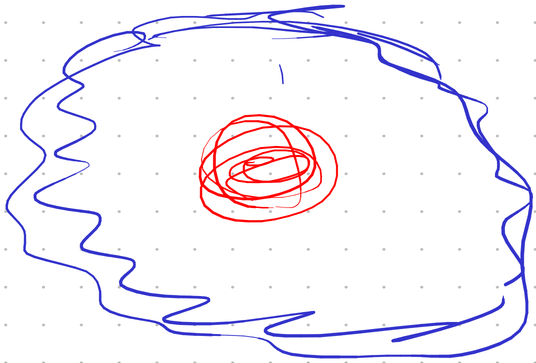


Tropical mathematics

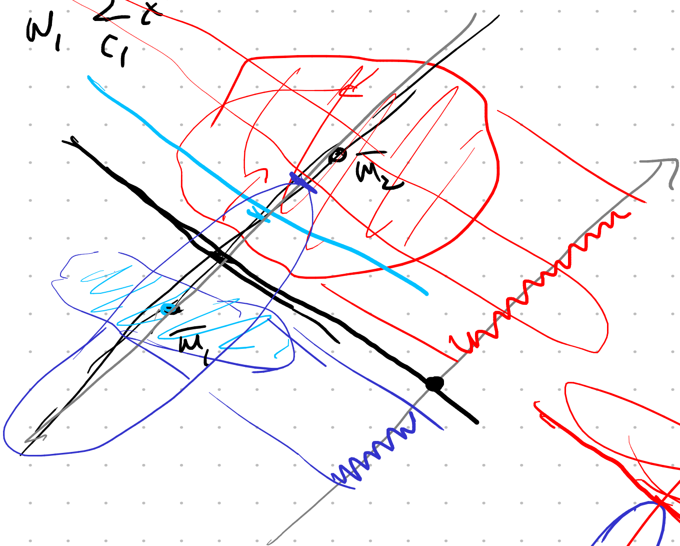
Fischer's linear discriminant

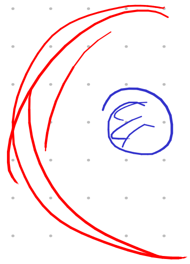


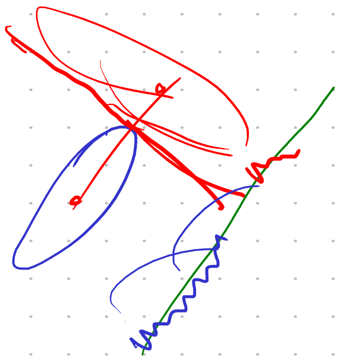
$$\bar m_2 = \frac{1}{N_2}\sum_{C_2} \bar x$$

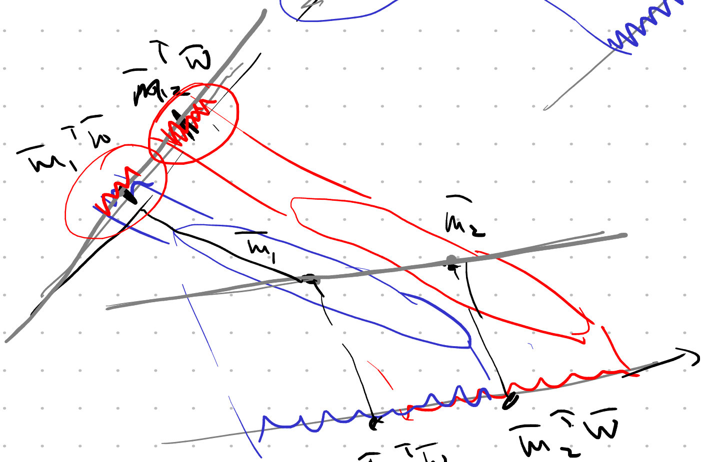


$$\bar m_1 = \frac{1}{N_1}\sum_{C_1} \bar x$$









$$\big((\bar m_1 - \bar m_2)^T \bar w\big)^2 \to max$$

$$\big(\bar w^T(\bar m_1 - \bar m_2)\big)^2 \to max$$

$$\sum_{x \in C_1}\Big(\underbrace{\bar w^T \bar x - \bar w^T \bar m_1}_{\bar w^T(\bar x - \bar m_1)}\Big)^2 +$$

$$+ \sum_{C_2}\big(\bar w^T \bar x - \bar w^T \bar m_2\big)^2 \to min$$

Fischer's lin. discr.

between-class covariance

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

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

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

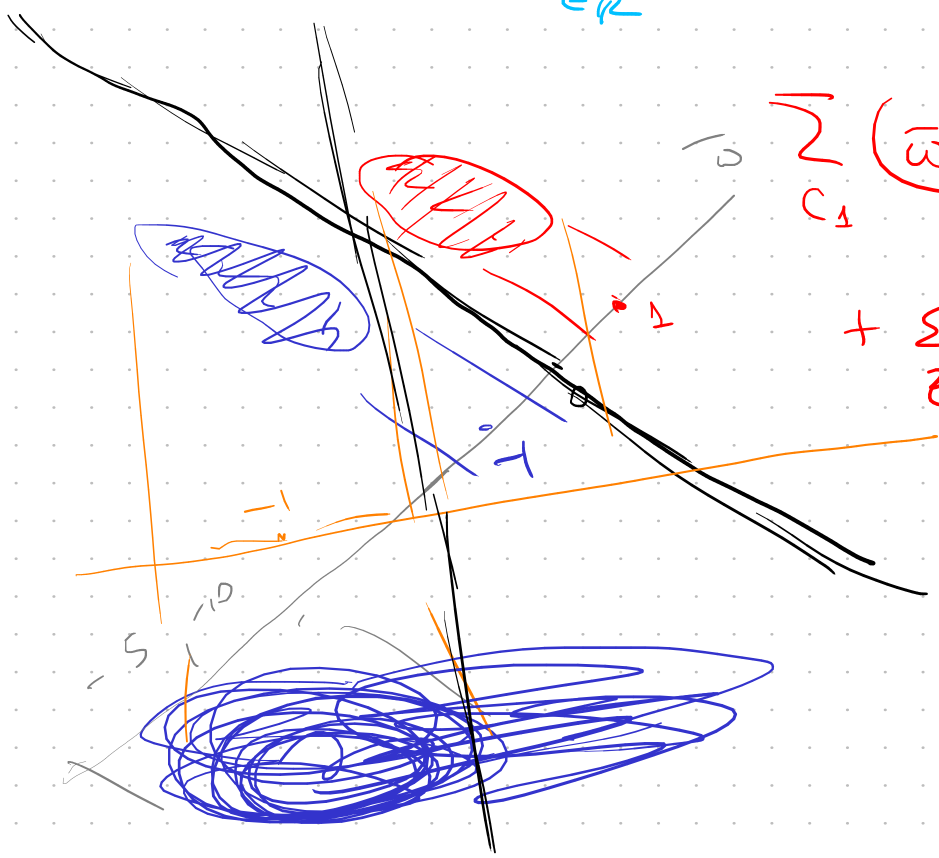
$$\nabla_{\bar w} J = \frac{2S_B\bar w(\bar w^T S_W\bar w) - 2S_W\bar w(\bar w^T S_B\bar w)}{(\bar w^T S_W\bar w)^2} = 0$$

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

$$S_B\bar w \propto S_W\cdot\bar w \qquad S_W\cdot\bar w \propto \bar m_1-\bar m_2$$

$$\| \qquad\qquad \boxed{\bar w \propto S_W^{-1}(\bar m_1-\bar m_2)}$$

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



$$\sum_{C_1}(\bar w^T\bar x_n - 1)^2 +$$

$$+ \sum_{C_2}(\bar w^T\bar x_n + 1)^2 \longrightarrow \min$$