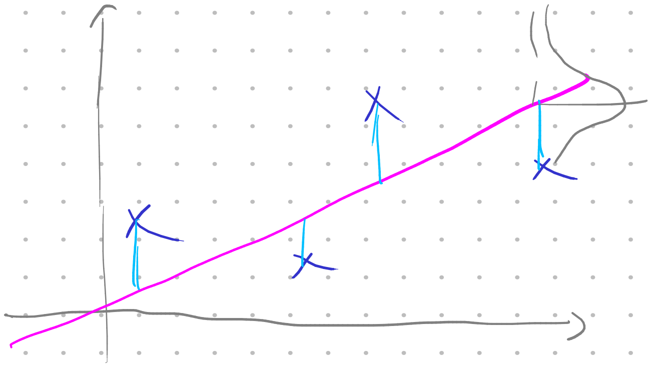


$$D = \{(\bar{x}_n, y_n)\}_{n=1}^{N}$$

$$\hat{y}_n = \bar{w}^T \bar{x}_n \approx y_n$$

$$L = \sum_n (y_n - \bar{w}^T \bar{x}_n)^2 \longrightarrow \min$$

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

$$p(D|\bar{w}) = \prod_{n=1}^{N} \mathcal{N}(y_n|\bar{w}^T\bar{x}_n, \sigma^2) \longrightarrow \max$$

L_2 – ridge regression

$$L = \sum_n (y_n - \bar{w}^T\bar{x}_n)^2 + \alpha \cdot \boxed{\|\bar{w}\|_2^2} \longrightarrow \min$$

$+ \alpha \boxed{\sum |w|}$ L_1 – lasso regression

$$L = \underline{\sum(\_\_)^2} + \underline{\alpha \|\bar{w}\|^2} \to \min \qquad + \alpha \cdot \sum \|w\|^q$$

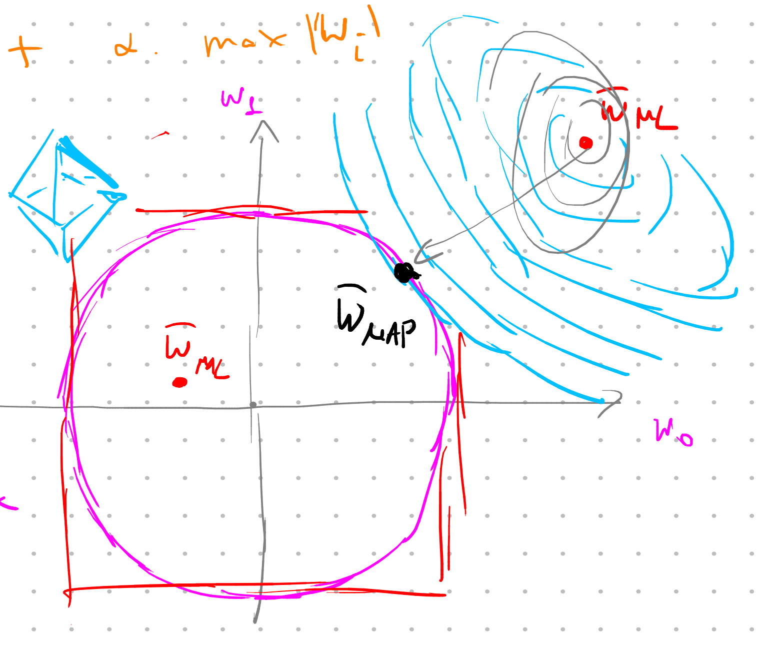
$+ \alpha \cdot \max |w_i|$

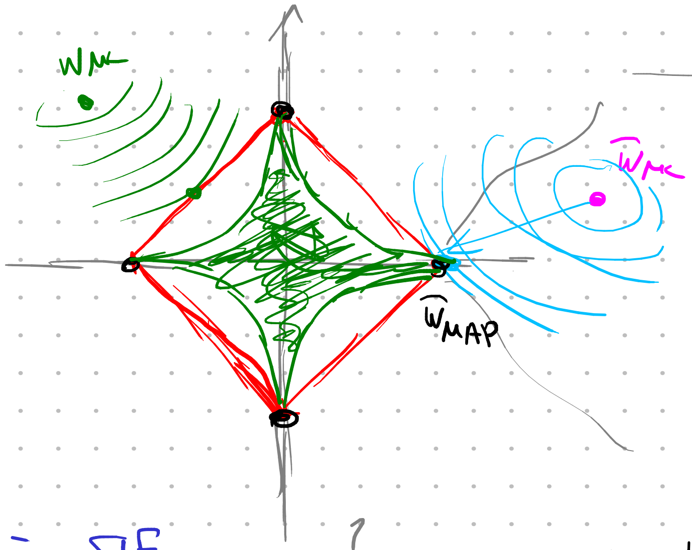
$\Updownarrow$

$$L = \sum(\_)^2 \longrightarrow \min$$

s.t. $\|\bar{w}\|^2 \le a$

$\sum |w_i| < a$





$$\sum_n (\_\_)^2$$

$$F(\bar{x} + \alpha \nabla_{\bar{x}} F) \xrightarrow[\alpha]{} \max$$

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

$$\log p(\bar{w}|D) = \text{Const} + \log p(\bar{w}) + \log p(D|\bar{w})$$

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

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

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

$$= \text{Const} - \frac{1}{2\sigma^2}\sum_{n=1}^{N}(y_n - \bar w^T \bar x_n)^2 - \frac{1}{2\sigma_0^2}\bar w^T \bar w$$

$$= \text{Const} - c \cdot \underbrace{\left(\sum_{n=1}^{N}(y_n - \bar w^T \bar x_n)^2 + \frac{\sigma^2}{\sigma_0^2}\bar w^T \bar w\right)}_{\rightarrow \min} \rightarrow \max$$

$$(\bar y - X\bar w)^T(\bar y - X\bar w) + \underline{\lambda \bar w^T \bar w} \rightarrow \min$$

$$-2X^T\bar y + 2X^T X\bar w + 2\lambda\bar w = 0$$

$$\boxed{\bar w_{MAP} = (X^T X + \underset{>0}{\lambda I})^{-1} X^T \bar y}$$

$$\ldots + \lambda\sum_{i=1}^{d}|w_i| \quad = -\log p(\bar w) + \text{const} \quad \rightsquigarrow \quad p(\bar w) \propto e^{-\lambda \sum_i |w_i|}$$

Laplace distribution

Conjugate prior

$$e^{-w^2-} \qquad\qquad c \cdot e^{-\frac{1}{2\sigma^2}\sum(-)^2}$$

$$p(\bar w) \times p(D \mid \bar w) \propto p(\bar w \mid D)$$

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

$$\frac{1}{(2\pi)^{d/2}\sqrt{\det \Sigma_N}} e^{-\frac{1}{2} -}$$

$$\log p(\bar{w}|D) = \text{Const} + \log p(D|\bar{w}) + \log p(\bar{w}) =$$

$$= \text{Const} - \frac{1}{2\sigma^2}\sum_{n=1}^{N}(y_n - \bar{w}^T\bar{x}_n)^2 - \underline{\frac{d}{2}\log 2\pi - \frac{1}{2}\log\det\Sigma_0}$$

$$- \frac{1}{2}(\bar{w}-\bar{\mu}_0)^T \Sigma_0^{-1}(\bar{w}-\bar{\mu}_0) =$$

$$= \text{Const} - \overbrace{\frac{1}{2\sigma^2}(\bar{y}-X\bar{w})^T(\bar{y}-X\bar{w})} - \overbrace{\underline{\frac{1}{2}(\bar{w}-\bar{\mu}_0)^T\Sigma_0^{-1}(\bar{w}-\bar{\mu}_0)}}$$

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

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

$$= \text{Const} - \frac{1}{2\sigma^2}\left(\bar{w}^T X^T X\bar{w} - 2\bar{w}^T X^T\bar{y}\right) - \frac{1}{2}\left(\bar{w}^T\Sigma_0^{-1}\bar{w} - 2\bar{w}^T\Sigma_0^{-1}\bar{\mu}_0\right)$$

$$= \text{Const} - \frac{1}{2}\bar{w}^T\left(\Sigma_0^{-1} + \frac{1}{\sigma^2}X^T X\right)\bar{w} + \bar{w}^T\left(\Sigma_0^{-1}\bar{\mu}_0 + \frac{1}{\sigma^2}X^T\bar{y}\right)$$

$$= \text{Const} - \frac{1}{2}\bar{w}^T\Sigma_N^{-1}\bar{w} + \bar{w}^T\Sigma_N^{-1}\bar{\mu}_N$$

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

$$\boxed{\begin{cases}\Sigma_N = \left(\underset{\searrow 0}{\Sigma_0^{-1}} + \frac{1}{\sigma^2}X^T X\right)^{-1} \\ \bar{\mu}_N = \Sigma_N\left(\underset{\searrow 0}{\Sigma_0^{-1}}\bar{\mu}_0 + \frac{1}{\sigma^2}X^T\bar{y}\right)\end{cases}}$$

$$p(\bar{w}) = \mathcal{N}(\bar{w}|\bar{\mu}_0, \Sigma_0)$$

$$\times$$

$$p(D|\bar{w}) \quad \dots \quad (X, \bar{y})$$

$$\parallel$$

$$p(\bar{w}|D) = \mathcal{N}(\bar{w}|\bar{\mu}_N, \Sigma_N)$$

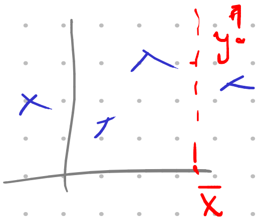
$$\overset{\in\mathbb{R}^d}{\underset{\propto 1}{\left(p(\overline{w})\right)}}\, p(D|\overline{w}) \to \max$$

improper priors

$$\swarrow f(y,\overline{x},D) \quad \mathcal{N}(\overline{w}|\ldots,\ldots)$$

Predictive distribution

$$\underline{p(y|\overline{x},D)} = \int p(y,\overline{w}|\overline{x},D)\,d\overline{w} = \int \underline{p(y|\overline{w},\overline{x})\,p(\overline{w}|D)}\,d\overline{w}$$



$$= \mathbb{E}_{\underline{p(\overline{w}|D)}}\left[p(y|\overline{w},\overline{x})\right] \approx \frac{1}{R}\sum_{r=1}^{R} p(y|\overline{w}^{(r)},\overline{x})$$

$$\mathcal{N}(y|\overline{w}^T\overline{x},\sigma^2) \qquad \mathcal{N}(\overline{w}|\overline{\mu}_N,\Sigma_N)$$

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

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

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

(!)

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

$$= \text{Const} - \frac{y^2}{2\sigma^2} - \underbrace{\frac{1}{2}\overline{w}^T\left(\Sigma_N^{-1} + \frac{1}{\sigma^2}\overline{x}\,\overline{x}^T\right)\overline{w}}_{=\Sigma'^{-1}} + \underbrace{\overline{w}^T\left(\Sigma_N^{-1}\overline{\mu}_N + \frac{y}{\sigma^2}\overline{x}\right)}_{=\Sigma'^{-1}\overline{\mu}'} = \boxed{*}$$

$$\overline{\mu}' = \Sigma' \cdot$$

$$= \text{Const} - \ldots(y) - \frac{1}{2}\overline{w}^T\Sigma'^{-1}\overline{w} + \overline{w}^T\Sigma'^{-1}\overline{\mu}' \boxed{-\frac{1}{2}\overline{\mu}'\cdot\Sigma'^{-1}\overline{\mu}'}$$

$$+\frac{1}{2}\overline{\mu}'^T\Sigma'^{-1}\overline{\mu}' = \left(\Sigma_N^{-1}\overline{\mu}_N + \frac{y}{\sigma^2}\overline{x}\right)^T\cdot\cancel{\Sigma'^T}\cancel{\Sigma'^{-1}}\Sigma'(\ldots) =$$

$$= \underbrace{\frac{1}{2}\left(\frac{y}{\sigma^2}\overline{x} + \Sigma_N^{-1}\overline{\mu}_N\right)^T\cdot\Sigma'\cdot\left(\frac{y}{\sigma^2}\overline{x} + \Sigma_N^{-1}\overline{\mu}_N\right)}$$

$$\boxed{*} = \text{Const} - \frac{y^2}{2\sigma^2} - \underbrace{\frac{1}{2}(\overline{w}-\overline{\mu}')^T\Sigma'^{-1}(\overline{w}-\overline{\mu}')} + \downarrow$$

$$\cancel{\log\det\Sigma'}$$

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

$$= \text{Const} - \frac{1}{\boxed{2\sigma^2_{pred}}}\left(y - \boxed{\mu_{pred}}\right)^2$$

$$y^2: \; -\frac{1}{2\sigma^2} + \frac{1}{2\sigma^4}\bar{x}^T\Sigma'\bar{x} = \circledast$$

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

Σ' $\qquad\qquad$ Σ'

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

$$\Sigma_N\bar{x} = (\,-\,-\,)\cdot\Sigma'\bar{x}$$

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

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

$$y^2: \; -\frac{1}{2\left(\sigma^2 + \bar{x}^T\Sigma_N\bar{x}\right)}$$

"

$$-\frac{1}{2\sigma^2_{pred}}$$

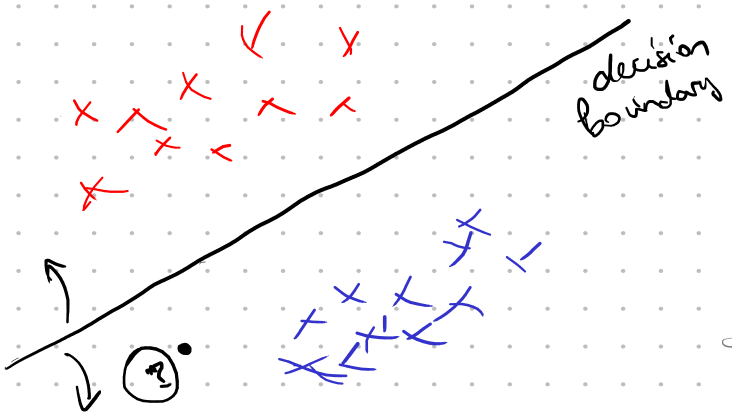
$$\boxed{\sigma^2_{pred} = \sigma^2 + \bar{x}^T\Sigma_N\bar{x}}$$

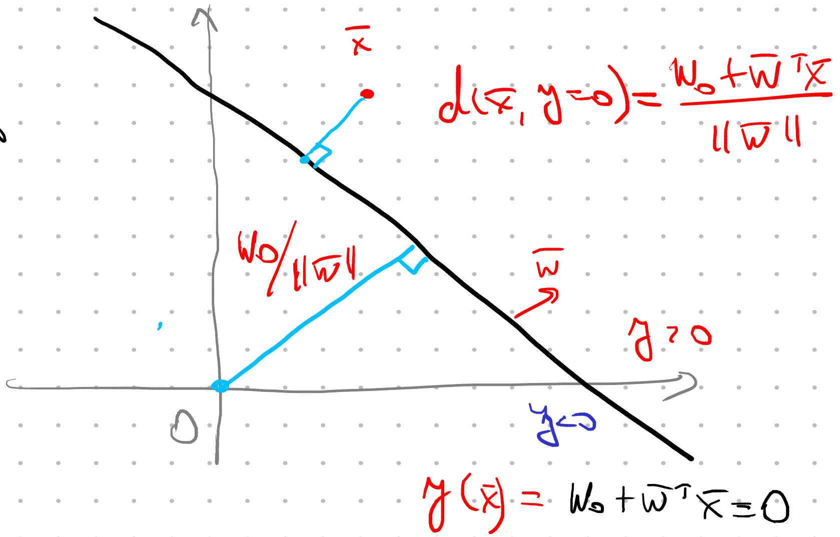
$$\text{"}\;\cancel{\Sigma_N}\bar{x}\cdot\frac{1}{\,-\,-\,}$$

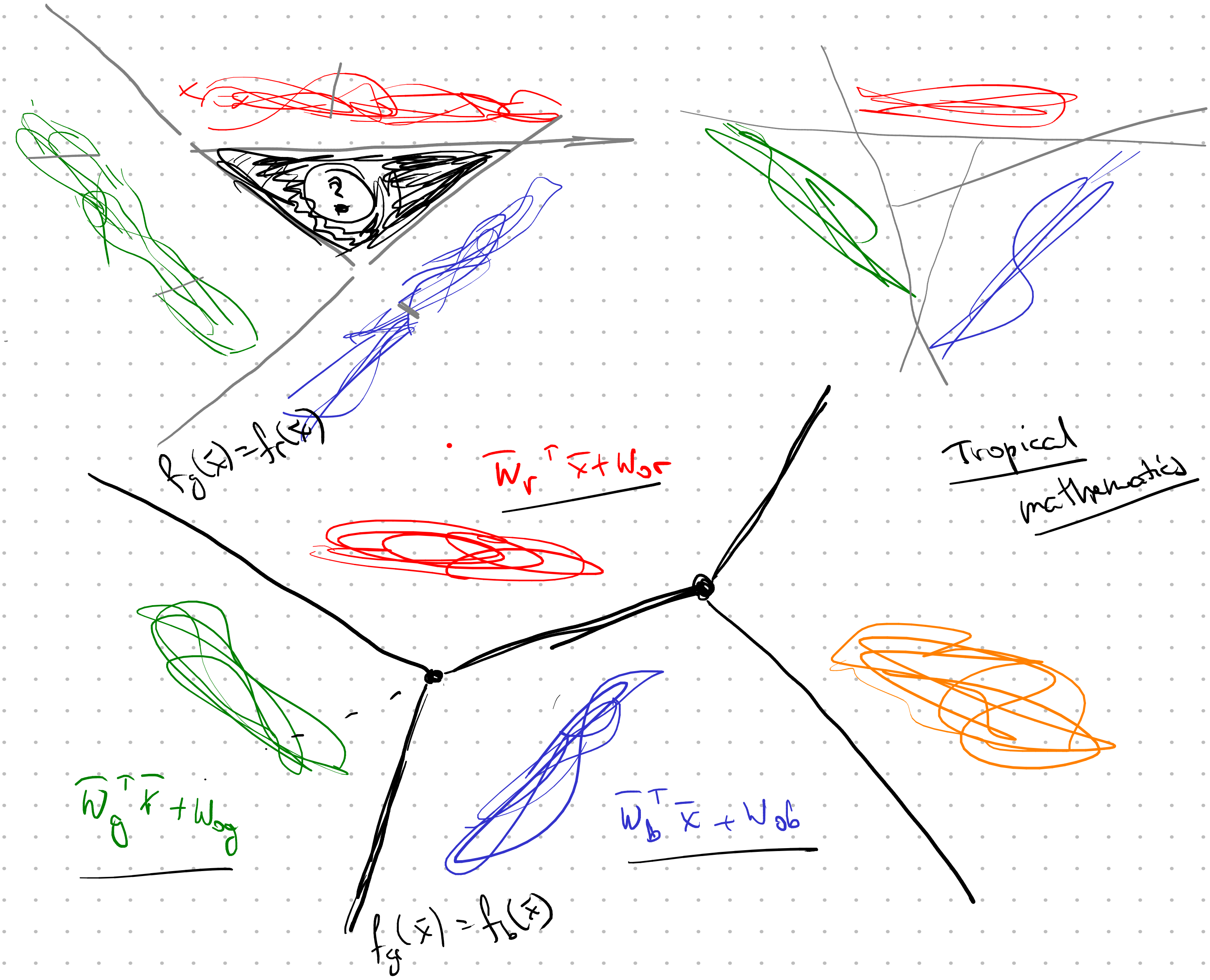
$$y: \; \frac{1}{\sigma^2}\bar{x}^T\Sigma'\Sigma_N^{-1}\bar{\mu}_N = \frac{1}{\sigma^2}\bar{\mu}_N^T\cancel{\Sigma_N^{-1}}\left(\Sigma'\bar{x}\right) = \qquad \boxed{\bar{\mu}_{pred} = \bar{\mu}_N}$$

$$= \frac{1}{\sigma^2}\,\frac{\bar{\mu}_N^T\bar{x}}{1 + \frac{1}{\sigma^2}\bar{x}^T\Sigma_N\bar{x}} = \frac{\bar{\mu}_N^T\bar{x}}{\sigma^2 + \bar{x}^T\Sigma_N\bar{x}} = \frac{\boxed{\bar{\mu}_N^T}\,\bar{x}}{\sigma^2_{pred}}$$

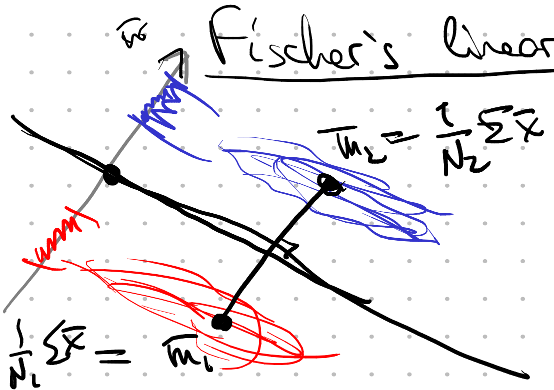
Classification

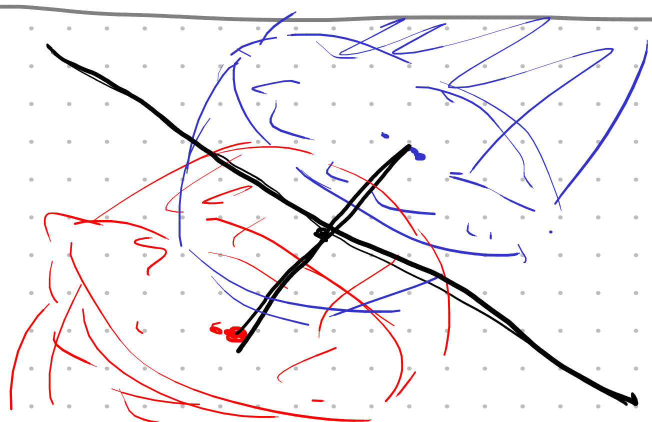


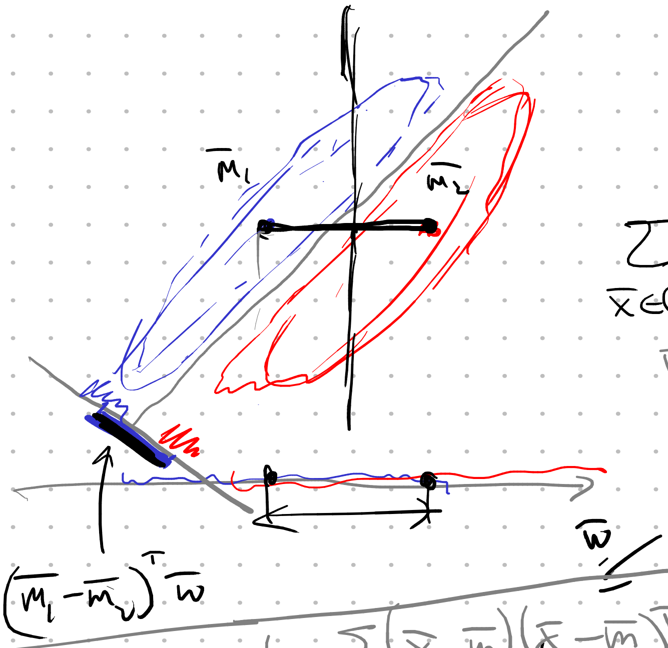




Fischer's linear discriminant







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

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

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

$$\mathcal{S}_B+\mathcal{S}_W=\mathcal{S}=\sum_{\bar{x}}(\bar{x}-\bar{m})(\bar{x}-\bar{m})^T$$

$$J(\bar{w})=\frac{\bar{w}^T\overbrace{(\bar{m}_1-\bar{m}_2)(\bar{m}_1-\bar{m}_2)^T}^{\text{between-class covariance}}\bar{w}}{\bar{w}^T\underbrace{\left(\sum_{\bar{x}\in C_1}(\bar{x}-\bar{m}_1)(\bar{x}-\bar{m}_1)^T+\sum_{\bar{x}\in C_2}(\bar{x}-\bar{m}_2)(\bar{x}-\bar{m}_2)^T\right)}_{\text{within-class covariance}}\bar{w}}$$

$$J(\bar{w})=\frac{\bar{w}^T\mathcal{S}_B\bar{w}}{\bar{w}^T\mathcal{S}_W\bar{w}} \xrightarrow{\bar{w}} \max$$

$$\nabla_{\bar{w}}J=\frac{2\mathcal{S}_B\bar{w}\cdot(\bar{w}^T\mathcal{S}_W\bar{w})-2\mathcal{S}_W\bar{w}\cdot(\bar{w}^T\mathcal{S}_B\bar{w})}{(\bar{w}^T\mathcal{S}_W\bar{w})^2}=0$$

$$\underbrace{(\bar{w}^T\mathcal{S}_W\bar{w})}_{\in\mathbb{R}}\cdot\mathcal{S}_B\bar{w}=\underbrace{(\bar{w}^T\mathcal{S}_B\bar{w})}_{\in\mathbb{R}}\cdot\mathcal{S}_W\bar{w}$$

$$\mathcal{S}_B\bar{w}\propto\mathcal{S}_W\bar{w}$$

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

$$\mathcal{S}_W\cdot\bar{w}\propto\bar{m}_1-\bar{m}_2$$

$$\boxed{\bar{w}\propto\mathcal{S}_W^{-1}(\bar{m}_1-\bar{m}_2)}$$