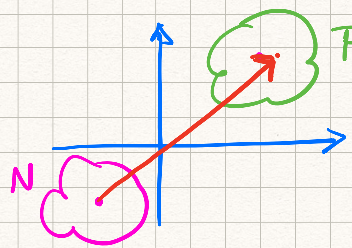
Hierarchy of Classifiers

Data set
Centered data

$$X = \begin{pmatrix} \vec{x}_1^T \\ \vec{x}_2^T \\ \vdots \\ \vec{x}_k^T \end{pmatrix} \qquad \vec{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_k \end{pmatrix}$$

$$X^T = \left(\vec{x}_1 \ \vec{x}_2 \cdots \vec{x}_k \right)$$

First idea



$$\vec{\mu}_P - \vec{\mu}_N = \vec{\beta}$$

If $\vec{x} \cdot \vec{\beta} \geq 0 \Rightarrow P$

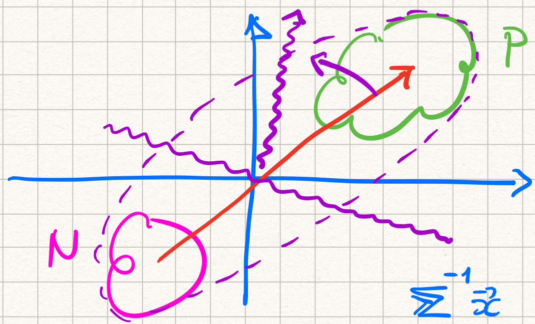
$\vec{x} \cdot \vec{\beta} < 0 \Rightarrow N$

If the classes are balanced

$$X^T \vec{y} \approx \vec{\mu}_P - \vec{\mu}_N$$

Linear regression

$$\vec{\beta} = (X^T X)^{-1} X^T \vec{y} \approx \Sigma^{-1} (\vec{\mu}_P - \vec{\mu}_N)$$



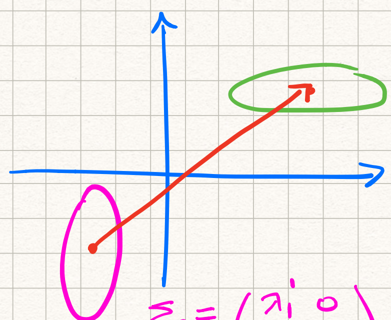
$\Sigma^{-1} \vec{x}$ rotates vector $\vec{x}$ towards the eigenvector with smallest eigenvalue

Ridge Regression

$$\vec{\beta} = (X^T X + \lambda I)^{-1} X^T \vec{y}$$

Fisher discriminant

$$\vec{\beta} = (\underbrace{\Sigma_1 + \Sigma_2}_{\text{average of the two covariance matrices}})^{-1} X^T \vec{y}$$



$$\Sigma_1 = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$$

$$\Sigma_2 = \begin{pmatrix} \lambda_1' & 0 \\ 0 & \lambda_2' \end{pmatrix}$$

$$(\Sigma_1 + \Sigma_2)^{-1} = \begin{pmatrix} \frac{1}{\lambda_1 + \lambda_1'} & 0 \\ 0 & \frac{1}{\lambda_2 + \lambda_2'} \end{pmatrix}$$

Perceptron Learning

$$\vec{\beta} = X^T \begin{pmatrix} \alpha_1 & & & 0 \\ & \alpha_2 & & \\ & & \ddots & \\ 0 & & & \alpha_K \end{pmatrix} \vec{y}$$

$$= X^T \begin{pmatrix} \alpha_1 y_1 \\ \alpha_2 y_2 \\ \vdots \\ \alpha_K y_K \end{pmatrix}$$

The α's are integers (positive)

Logistic regression

$$\vec{\beta} = X^T \begin{pmatrix} \alpha_1 y_1 \\ \vdots \\ \vdots \end{pmatrix}$$

$(\alpha_k y_k)$

the α's are positive and real

Support vector machine

Same as logistic regression, but sparse

(only some α's $\neq 0$