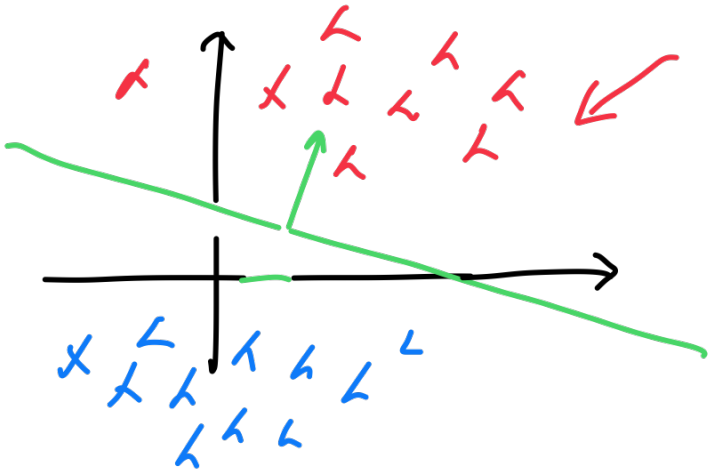
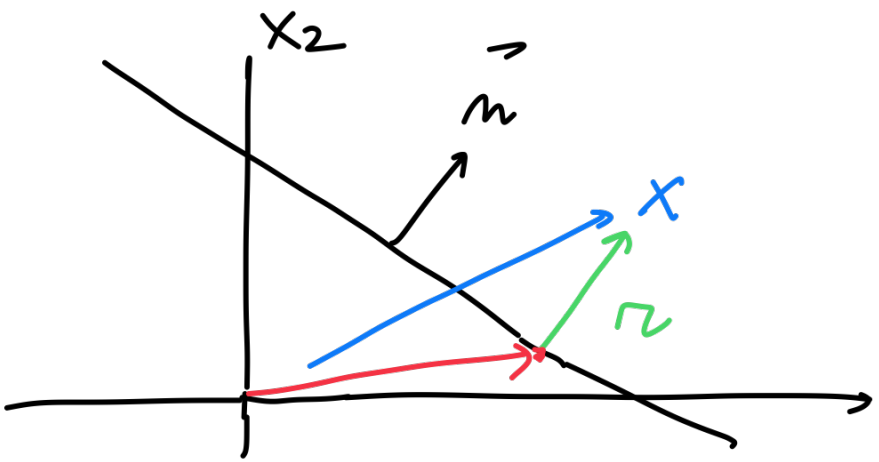
# Classification

$\{t^{(i)}, x^{(i)}\}_{i=1}^{N}$ $\quad t^{(i)} \in \{1, 2, 3, \dots, K\}$






plane equation $\quad$ normal to $\pi = (\beta_1, \beta_2, \dots \beta_D)$

$$\beta_0 + \beta_1 x_1 + \dots + \beta_D x_D = 0$$

if $x$ above plane

$$\beta_0 + \beta_1 x_1 + \dots + \beta_D x_D > 0$$

if $x$ below plane

$$\beta_0 + \beta_1 x_1 + \dots + \beta_D x_D < 0$$

$$x = x_\perp + \frac{\vec{n}}{\|n\|} r$$

$$\beta_0 + \vec{n}^T \vec{x} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_D x_D$$

$$\beta_0 + \vec{n}^T x_\perp + \frac{n^T n}{\|n\|} r = \frac{\|n\|^2}{\|n\|} r > 0$$

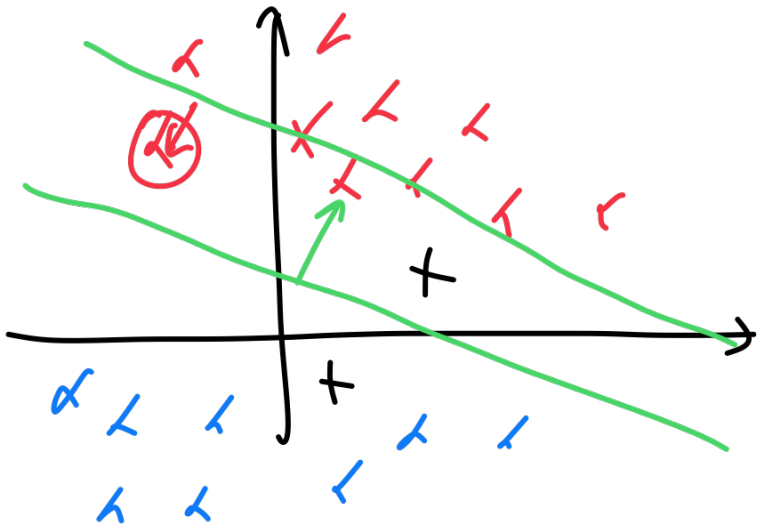
($\vec{n}^T x_\perp = 0$)

$$\beta_0 + \beta_1 x_{\perp,1} + \beta_2 x_{\perp,2} + \cdots + \beta_D x_{\perp,D} = 0$$

1st approach: learn classifier through least squares

$$\min_{\beta} \frac{1}{N} \sum_{i=1}^{N} \left( t^{(i)} - (\beta_0 + \beta_1 x_1^{(i)} + \cdots + \beta_D x_D^{(i)}) \right)^2$$

$t^{(i)} \in \{+1, -1\}$



$(1-(\ ^{>0}\ ))^2$

---

Possible extensions to Multiclass framework:

→ One vs rest → K-1 linear classifiers

→ One vs One → $\frac{K(K-1)}{2}$ classifiers

Multiple class discriminant  Assume K classes

if $x^{(i)} \in C_k \Rightarrow$ store $t^{(i)}$ as $t^{(i)} = [0 \dots 0, 1, 0 \dots 0]$

($1$ in the $k$th entry)

$$\underline{\underline{T}} = \begin{bmatrix} | & | & & | \\ t^{(1)} & t^{(2)} & \dots & t^{(N)} \\ | & | & & | \end{bmatrix}^T \qquad \underline{\underline{\tilde{X}}} = \begin{bmatrix} — & (\tilde{x}^{(1)})^T & — \\ — & (\tilde{x}^{(2)})^T & — \\ & \vdots & \\ — & (\tilde{x}^{N})^T & — \end{bmatrix}$$

$$B = \begin{bmatrix} | & | & & | \\ \beta^{(1)} & \beta^{(2)} & \dots & \beta^{(K)} \\ | & | & & | \end{bmatrix}$$

$$\underset{B}{\operatorname{argmin}} \; \frac{1}{NK} \sum_{i=1}^{N} \sum_{k=1}^{K} \left( t_k^{(i)} - (\tilde{x}^{(i)})^T \beta_k \right)^2 \leftarrow$$

$$\underset{B}{\operatorname{argmin}} \frac{1}{NK} \operatorname{Tr}\left( (T - XB)^T (T - XB) \right) \quad (*)$$

$$T - XB \overset{N}{=} \begin{bmatrix} t^{(1)} \\ \vdots \\ t^{(N)} \end{bmatrix} - \begin{bmatrix} \tilde{x}^{(1)} \\ \vdots \\ \tilde{x}^{(N)} \end{bmatrix} \begin{bmatrix} 1 & & 1 \\ \beta^{(1)} & \cdots & \beta^{(K)} \end{bmatrix} \} D$$

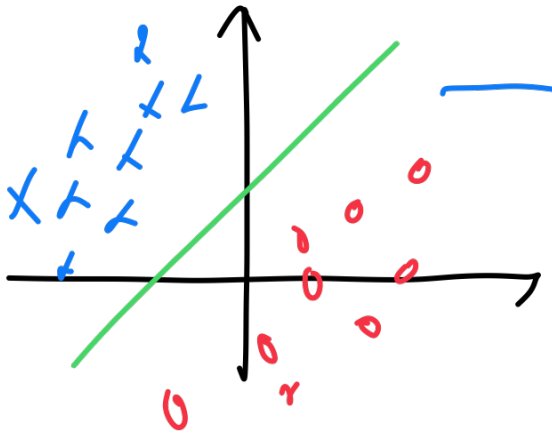
$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}^T \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{21} & a_{31} \\ a_{12} & a_{22} & a_{32} \\ a_{13} & a_{23} & a_{33} \end{bmatrix} \begin{bmatrix} a_{11} & a_{21} & a_{31} \\ a_{12} & a_{22} & a_{32} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

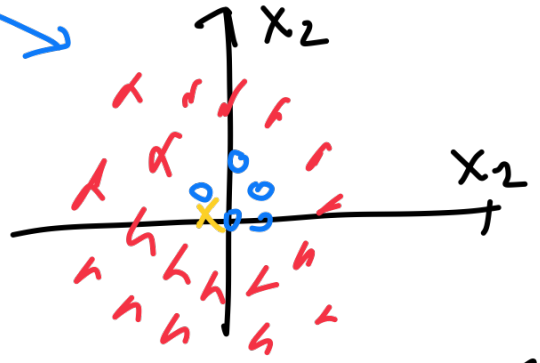
$$= \begin{bmatrix} a_{11}^2 + a_{21}^2 + a_{31}^2 & & \\ & a_{12}^2 + a_{22}^2 + a_{32}^2 & \\ & & a_{13}^2 + a_{23}^2 + a_{33}^2 \end{bmatrix}$$

To find B, compute derivative of $J(*)$ w.r.t B and set it to 0.

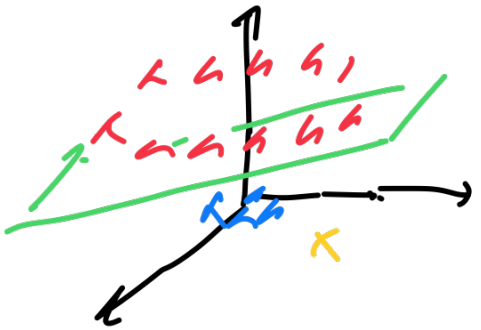
$\Rightarrow$

$$B = (X^T X)^{-1} X^T T$$

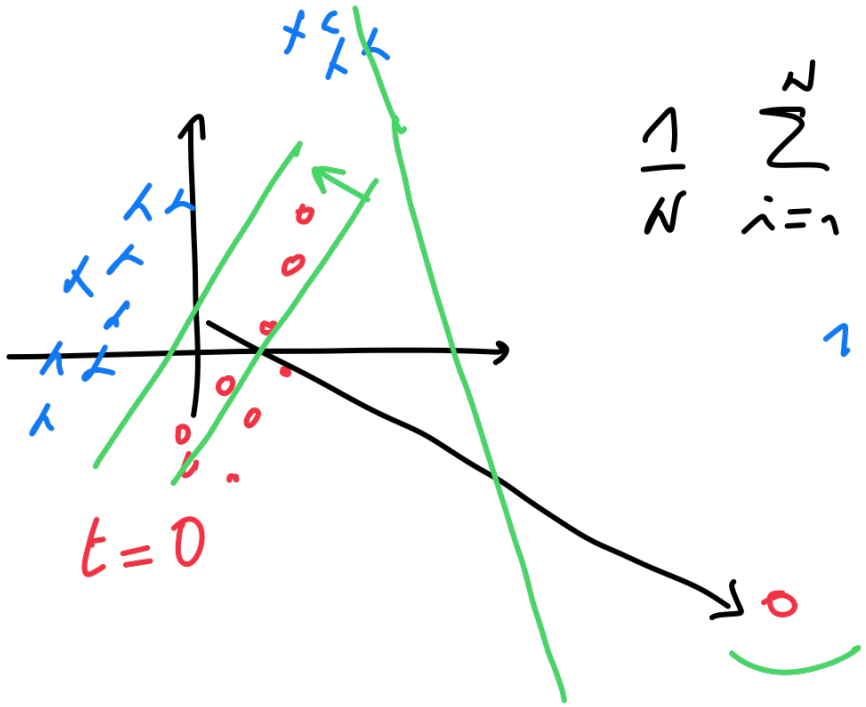





$x_1, x_2, x_1^2 + x_2^2$



$x_1, x_2, x_1^2 + x_2^2$



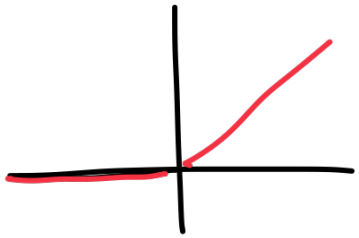

$$\frac{1}{N} \sum_{i=1}^{N} (t^{(i)} - \beta^T \tilde{x}^{(i)})^2$$

2 Motivations : presence of outliers

interpretability of classification

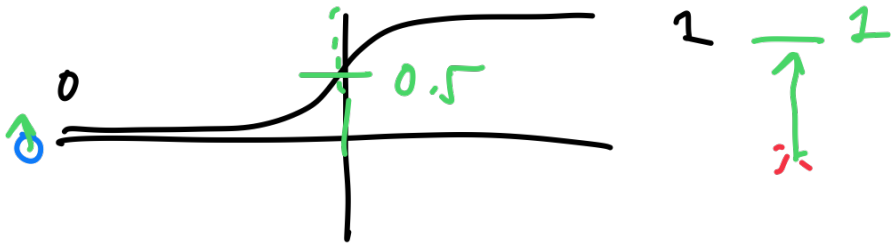
(quantitative criterion)

Solution : take an activation function



Sigmoid function

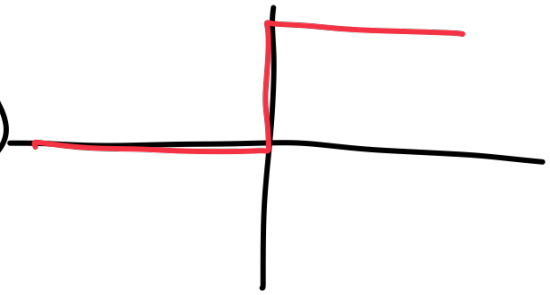
$$\sigma(x) = \frac{1}{1 + e^{-x}}$$




## Logistic regression

$$h_\beta(x) = \sigma(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_\Delta x_\Delta)$$

$$= \sigma(\beta^T \tilde{x}) \quad \in [0, 1]$$



$$p(t^{(i)} = 1 \mid x^{(i)}) = \sigma(\beta^T \tilde{x}^{(i)})$$

$$p(t^{(i)} = 0 \mid x^{(i)}) = 1 - \sigma(\beta^T \tilde{x}^{(i)})$$