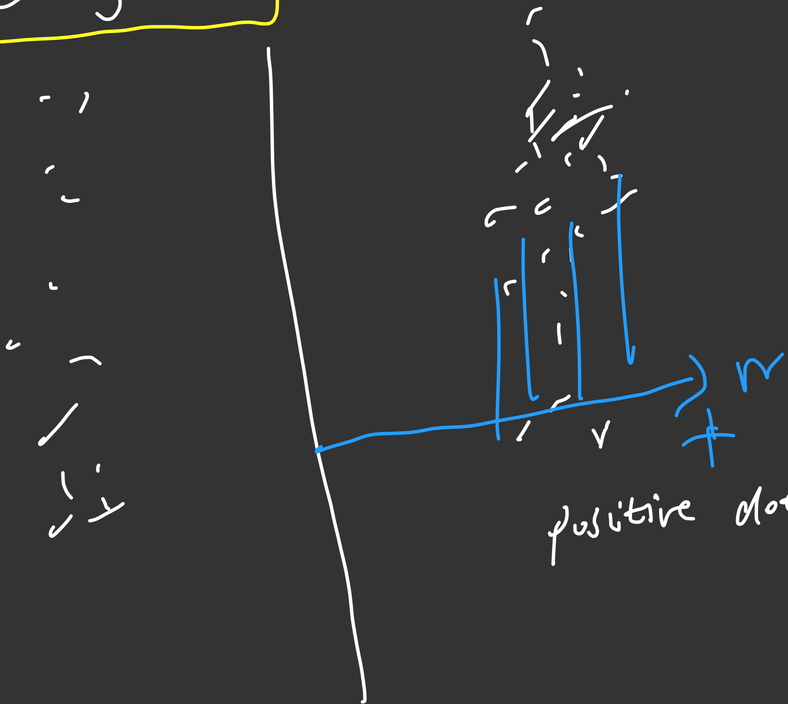
# Logistic Regression

$\boxed{\text{Classification}}$:




positive dot product

$\boxed{\text{binary classification}}$.

$$y = \{-1, +1\}$$

$$\text{sign}(w^T x) \in \{-1, +1\}$$

$$x \rightarrow \text{sgn}(w^T x) = \begin{cases} +1 & w^T x \geq 0 \\ -1 & w^T x < 0 \end{cases}$$

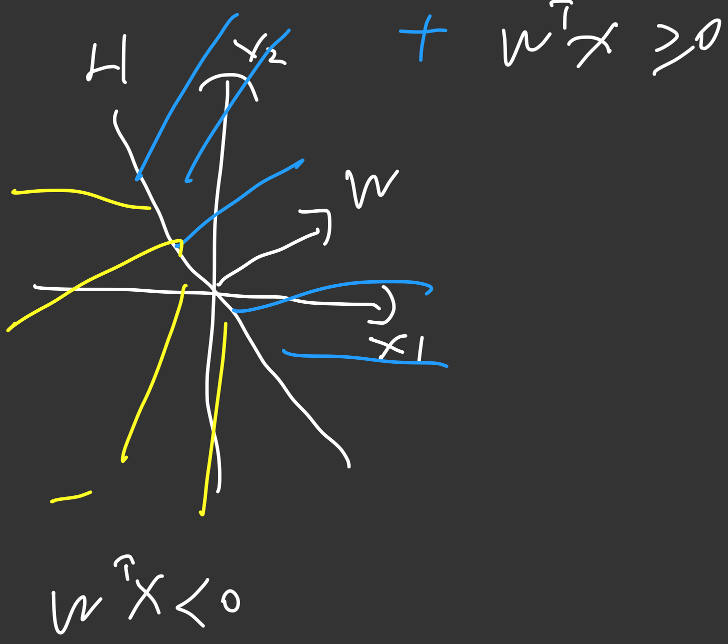
## $\boxed{\text{Linear Classifiers}}$:

Given $w \in \mathbb{R}^d$, predict with

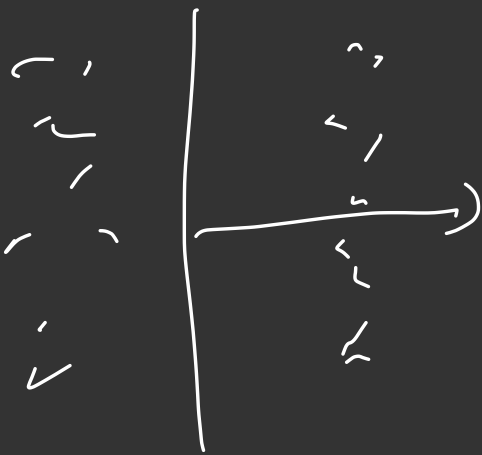
$$x \to \text{sign}(w^T x) \in \{\pm 1\}$$

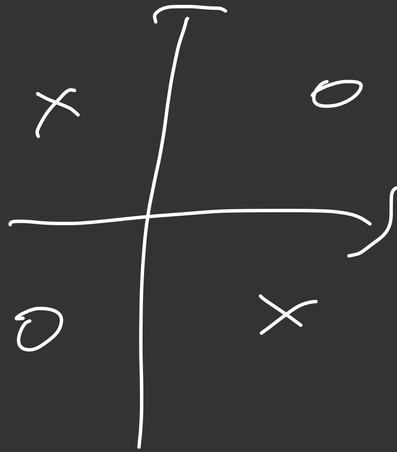
$\boxed{H \text{ (hyperplane)}}$ be orthogonal to $w$:

$$H := \{x \in \mathbb{R}^d : x^T w = 0\}$$



# Linear separability





Linear seperable

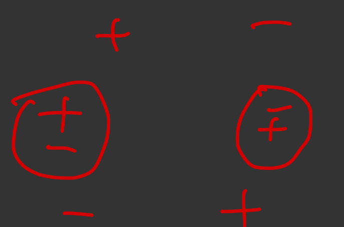
Linear unseperable

# Finding linear classifiers with ERM

$\boxed{\text{Not}}$ with $\boxed{\text{meta-algorithm}}$

$$\underset{w \in \mathbb{R}^d}{\arg\min} \frac{1}{n} \sum_{i=1}^{n} \underbrace{\mathbb{I}[\operatorname{sign}(w^T x_i) \neq y_i]}_{\text{most agreement}}$$

$\boxed{\text{NP-hard !!!}}$



# Relaxing the ERM problem

① remove sign(·)

$$\frac{1}{n}\sum_{i=1}^{n}\underbrace{\mathbb{1}[\overbrace{\text{sign}(\overbrace{w^Tx_i}^{0}) \neq \overbrace{y_i}^{1}}^{1}]}_{\text{sometimes, 0 at 0}} \longrightarrow \underbrace{\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}[y_i(w^Tx_i)\le 0]}_{=1 \text{ whenever } y_i w^T x_i = 0}$$

$w=0$

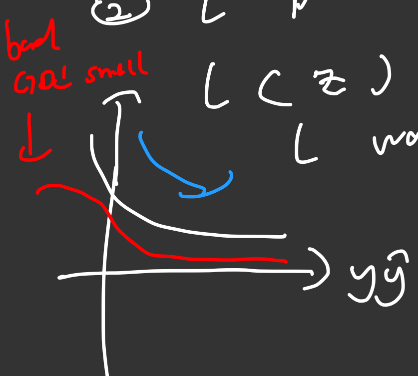
② $\hat{R}_{zo}(w) = \frac{1}{n}\sum_{i=1}^{n} \ell_{zo}(y_i w^T x_i)$

where $\ell_{zo}(z) = \underbrace{\mathbb{1}[z \le 0]}_{\text{not continuous! so no GD!}}$

# Logistic Loss

① loss should be continuous.

② L pushes predictions in correct direction




$L(z) > 0$ when $z \le 0$

L wants $\underbrace{z}_{y_i w^T x_i} > 0$

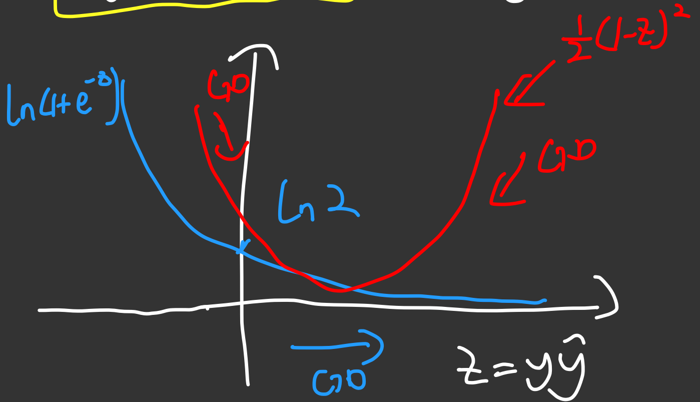
$\underset{\{-1,+1\}}{y} \cdot \hat{y} > 0$

# Losses for classification

**squared loss**: $l_{ls}(z) := \frac{1}{2}(1-z)^2$

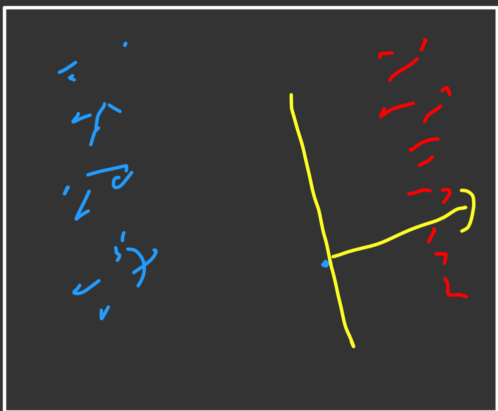
$$2\, l_{ls}(y\hat{y})^2 = \underbrace{(1-y\hat{y})^2}_{\text{it wants } y\hat{y} \text{ to be } 1.} = y^2(1-y\hat{y})^2 = (y-\hat{y})^2$$

**logistic loss**: $l_{log}(z) = \ln(1+\exp(-z))$



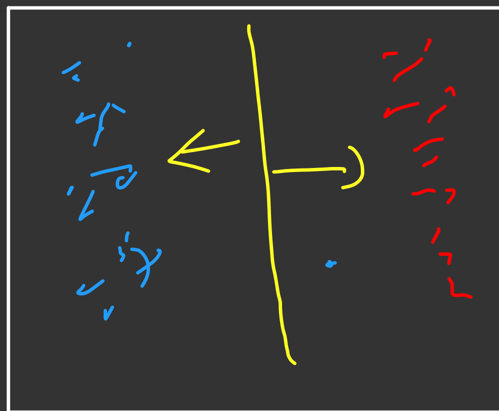

torch.CrossEntropyLoss( )( $\hat{y}$, $y$ )

logistic loss



logistic loss forced to push all line at blue side

squared loss



sqr loss choose to make a mistake because it generally wants $y\hat{y} = 1$

## Least Squares and Logistic ERM

**least squares**: ① normal equations ② OLS ③ GD

**logistic loss**:

$$\nabla_w \hat{R}_{log}(w) = \nabla_w \frac{1}{n}\sum_{i=1}^{n} \ln(1+\exp(-z)) = 0 \ : \ \text{GD}$$