Classification

EE698V - Machine Learning for Signal Processing

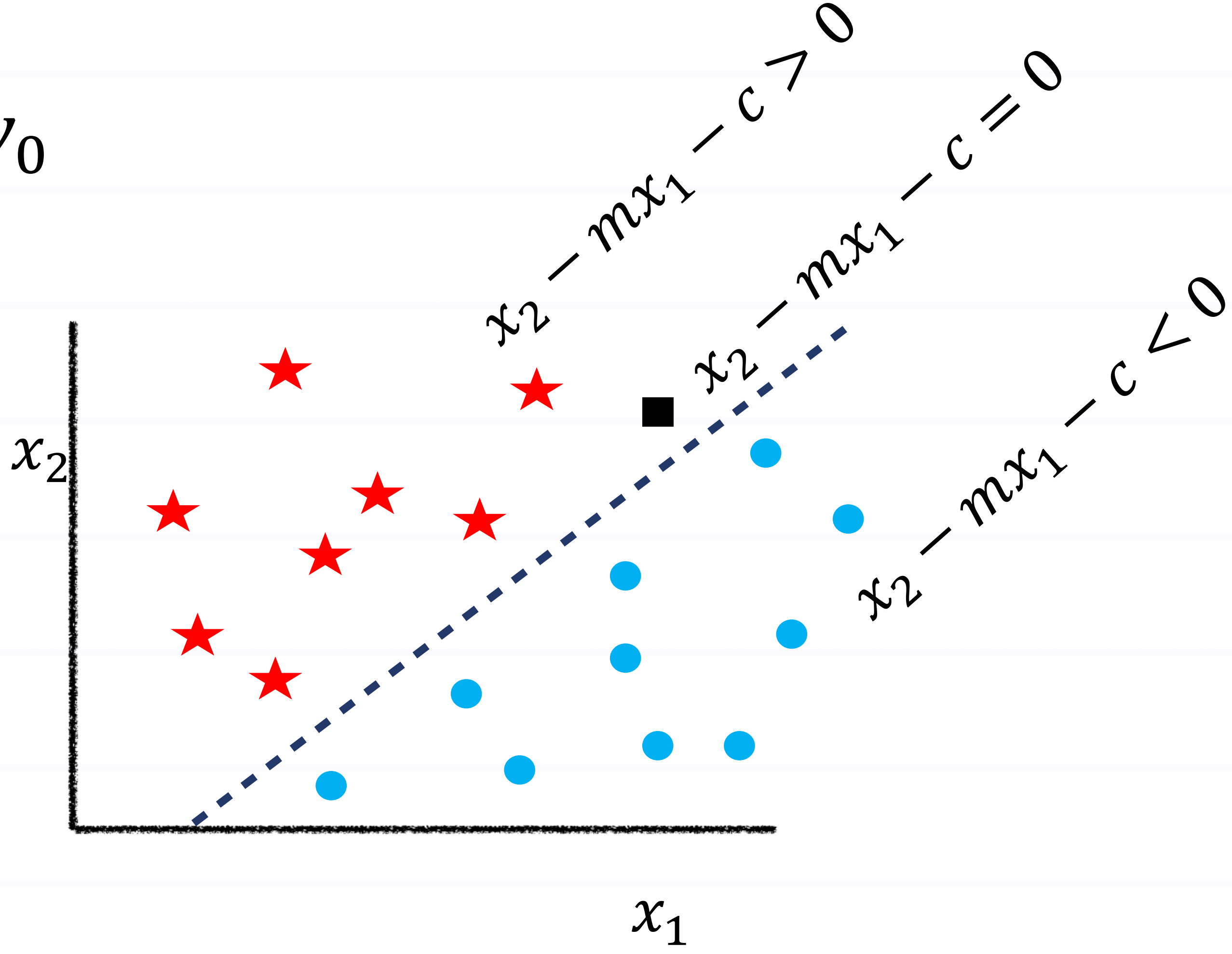
Vipul Arora

Linear Classification

Reference: PRML Sec 4.1

Two-class Classification

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

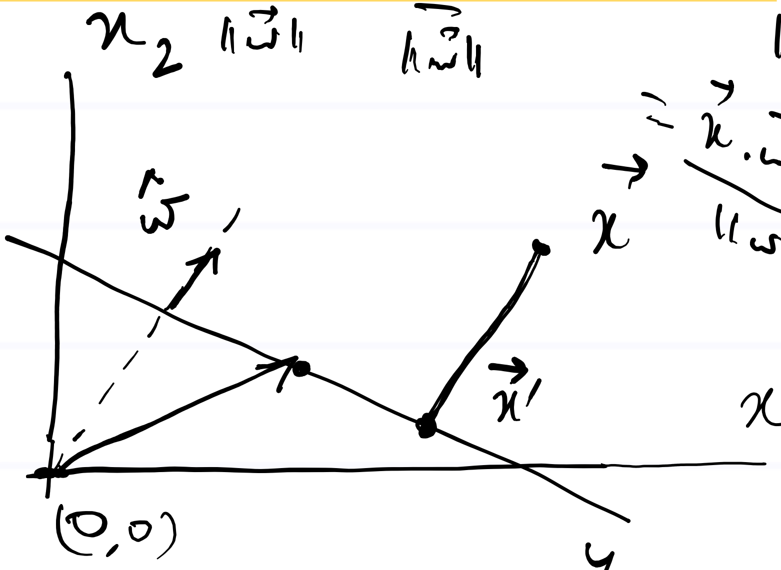


Geometrical Interpretation

- Distance from origin = $\hat{w} \cdot \vec{x}$
- Distance from the line = $\dfrac{y(\vec{x})}{\|\vec{w}\|}$

$$d = \underbrace{(\vec{x} - \vec{x}\,') \cdot \hat{\omega}} = \vec{x} \cdot \hat{\omega} + \frac{w_0}{\|w\|}$$

$$\underline{\vec{x}\,' \cdot \vec{\omega}} + \omega_0 = 0$$
$$\|\vec{\omega}\| \qquad \|\vec{\omega}\|$$



$$= \frac{\vec{x} \cdot \vec{\omega} + \omega_0}{\|\omega\|}$$

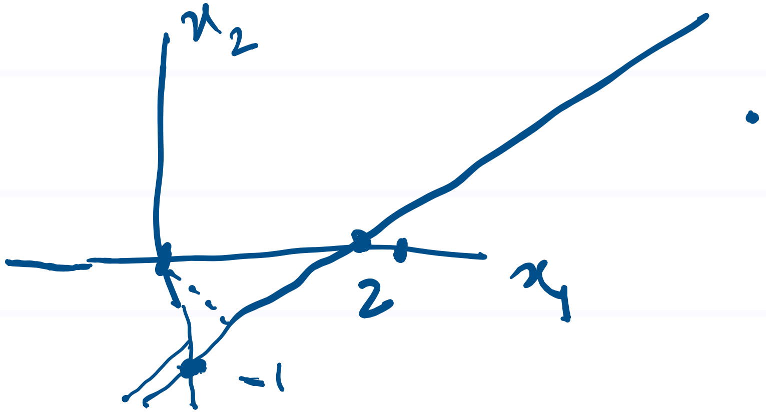
$$y = \vec{\omega} \cdot \vec{x} + \omega_0$$

$$y(\vec{x}) = \vec{\omega} \cdot \vec{x} + \omega_0 \qquad (= \omega^T x + \omega_0)$$

$$\vec{\omega} \to \begin{bmatrix} -1 \\ 2 \end{bmatrix} \qquad \vec{x} \to \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \qquad \omega_0 \to 2$$

$$y(\vec{x}) = 0$$

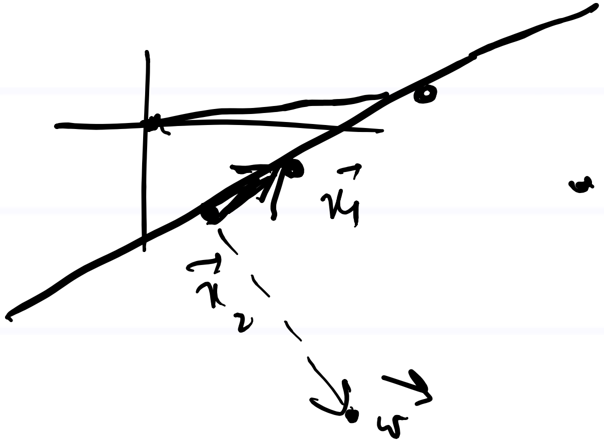
$$-x_1 + 2x_2 + 2 = 0$$



$$\vec{x} = \begin{bmatrix} 4 \\ 1 \end{bmatrix}$$

$$y(\vec{x}_1) = \vec{w}\cdot\vec{x}_1 + w_0 = 0$$

$$y(\vec{x}_2) = \vec{w}\cdot\vec{x}_2 + w_0 = 0$$

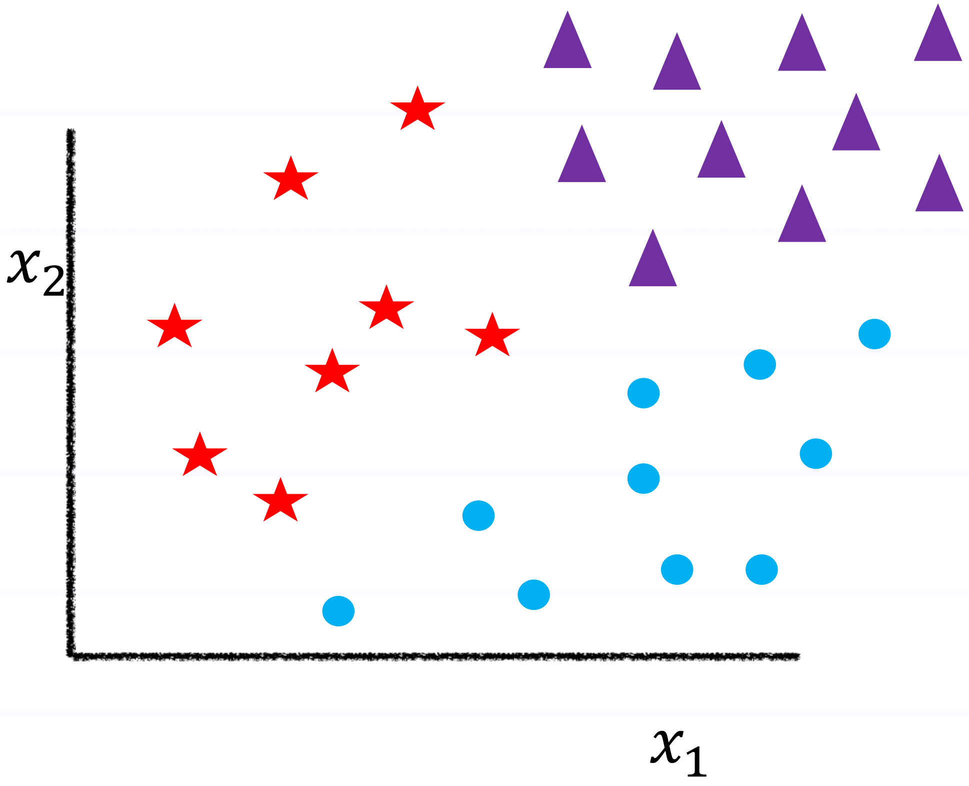


$$\vec{w}\cdot(\vec{x}_1 - \vec{x}_2) = 0$$

$$\frac{\vec{w}\cdot 0 + w_0}{\sqrt{w_1^2 + w_2^2}}$$

Multi-class Classification

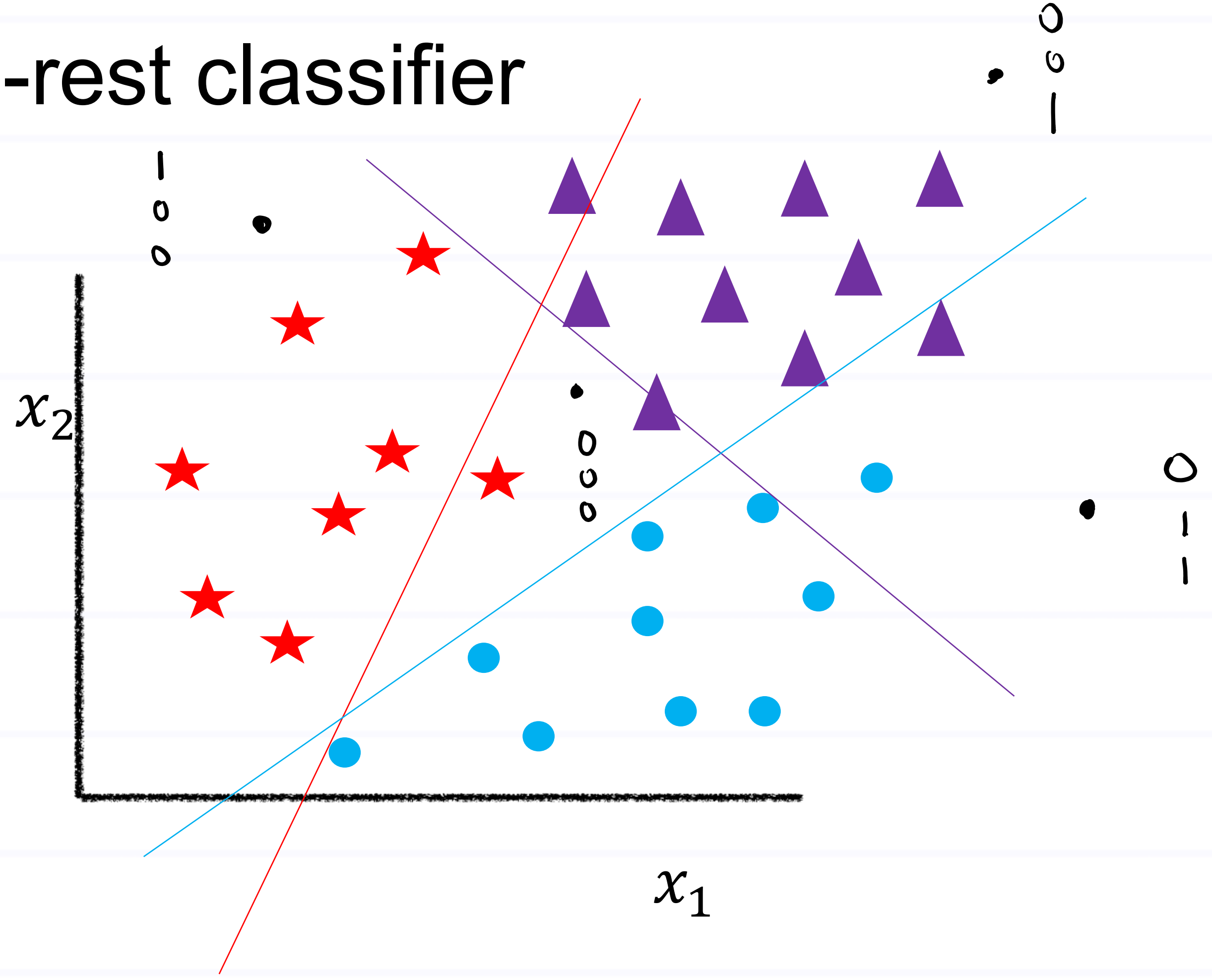
- One-versus-the-rest classifier



Multi-class Classification

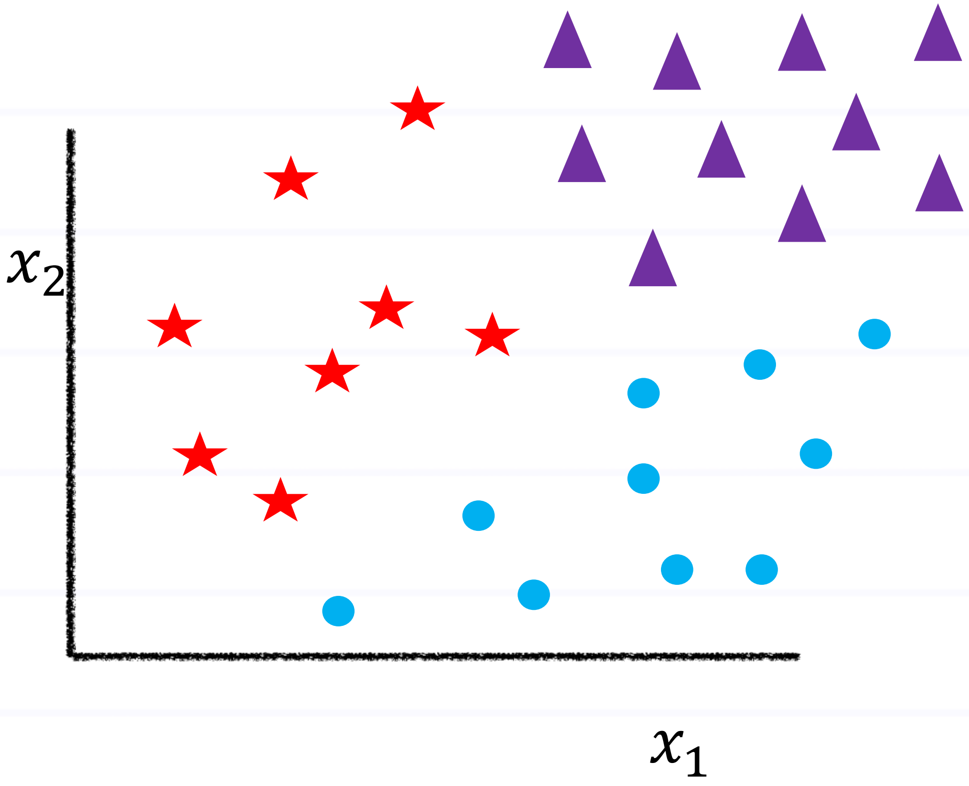
- One-versus-the-rest classifier

$$\begin{bmatrix} y_r \\ y_b \\ y_p \end{bmatrix}$$



Multi-class Classification

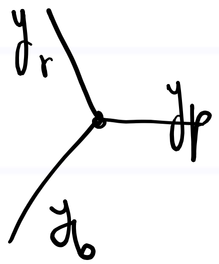
- One-versus-one classifier



Multi-class Classification

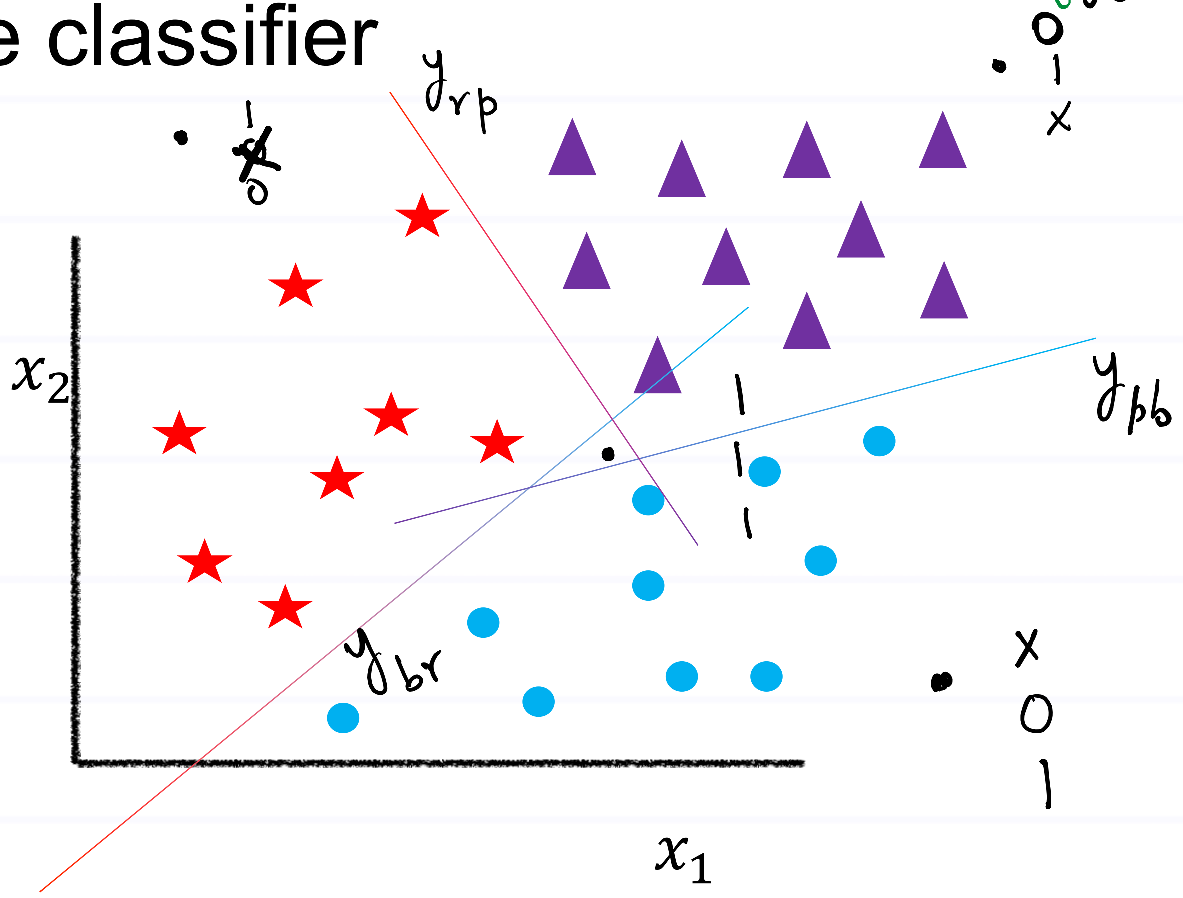
$$\text{predicted class} = \arg\max_{c} y_c \qquad c \in \{r, p, b\}$$

$$y = \begin{bmatrix} y_r \\ y_p \\ y_b \end{bmatrix}$$



- One-versus-one classifier

$$\begin{bmatrix} y_{rp} \\ y_{pb} \\ y_{br} \end{bmatrix}$$



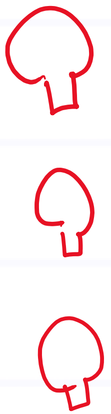
$$r = \begin{matrix} 1 & 1 \\ 0 & 1 \\ 0 & 0 \end{matrix} \qquad p = \begin{matrix} 0 & 0 \\ 1 & 1 \\ 0 & 1 \end{matrix} \qquad b = \begin{matrix} 0 & 1 \\ 0 & 0 \\ 1 & 1 \end{matrix}$$

Multi-class Classification

$$\mathbf{y} = \begin{array}{c} \\ \begin{bmatrix} \cdot \\ \cdot \\ \cdot \end{bmatrix} \end{array} \begin{array}{c} r \\ b \\ b \end{array} \qquad \mathbf{x} = \begin{bmatrix} \cdot \\ \cdot \end{bmatrix}$$

$$W \to 3 \times 2 \qquad \mathbf{w}_0 \to 3 \times 1$$

- K-class discriminant



$$\mathbf{y}(\mathbf{x}) = W\mathbf{x} + \mathbf{w}_0$$

$$k^* = \underset{k}{\operatorname{argmax}} \; \mathbf{y}(\mathbf{x})$$

$$= \underset{k}{\operatorname{argmax}} \; y_k(\mathbf{x})$$

$$= \underset{k}{\operatorname{argmax}} \; \mathbf{w}_k^T \mathbf{x} + w_{0,k}$$

(kth row of W) (kth element of $\mathbf{w}_0$)

$$\mathbf{y} = \begin{bmatrix} y_0 \\ y_1 \\ \vdots \end{bmatrix} = y_k$$

$k = 0, 1, \ldots$

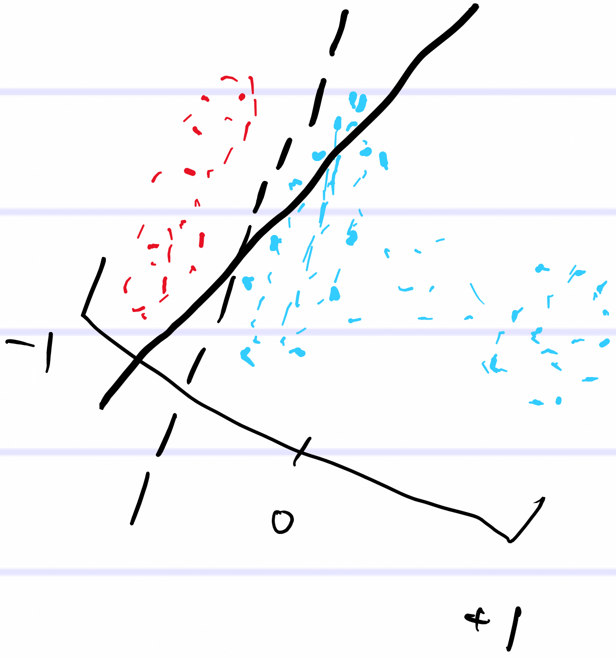
Least Squares

$$E = \sum_{\text{samples}} \| t(x_i) - y(x_i) \|^2$$

$$y(x_i) = \underbrace{\begin{bmatrix} W & w_0 \end{bmatrix}}_{\tilde{W}^T} \underbrace{\begin{bmatrix} x_i \\ 1 \end{bmatrix}}_{\tilde{x}_i}$$

$$\tilde{W}^T = (\tilde{X}^T \tilde{X})^{-1} \tilde{X}^T T$$

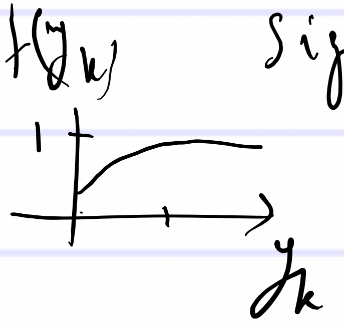
Problems with Least Squares Loss



$$\|t_k - y_k\|^2 \qquad t_k - f(y_k)$$

1	5
1	4
1	6
-0	-.1
-0	-.2
-0	-.1

sigmoid



Why can't I use sign(·)?

non diff