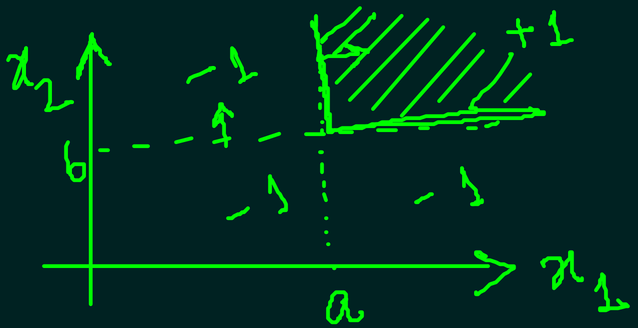
Classification Problems:

$f : X \to Y = \{-1, +1\}$

$E_{in} \rightsquigarrow E_{out}$

$\text{Prob}(y = 1 \mid X)$ or $P(y = 0 \mid x)$



Concept learning

$(x_1 \geq a) \wedge (x_2 \geq b)$

linear Separator

(Hyperplanes)

Linear Models

(Premitive)

$w_1 x_1 + w_2 x_2 + w_0 = 0$

$0.4 + 0.5 x_1 + 0.6 x_2 = 0$

w_1, w_2, w_0

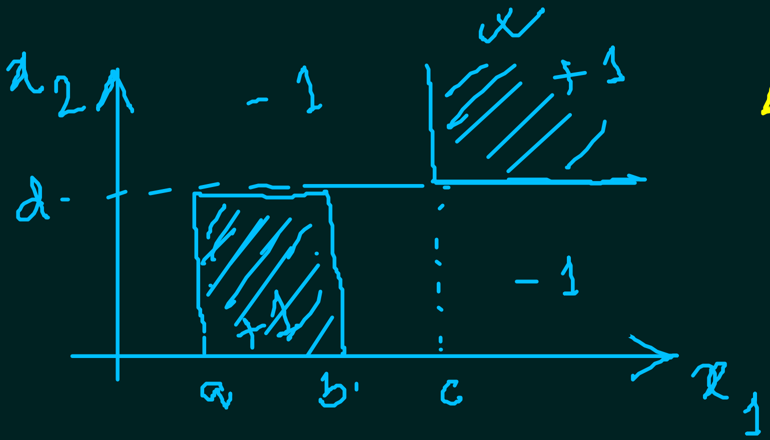
from TE

$w_0 + w_1 x_1 + \dots + w_d x_d = 0$

$X_{new} = \langle x_1^{new}, \dots x_d^{new} \rangle$ for +class ≥ 0

for -class < 0

+/-



Decision Tree Learning

$[(x_1 \geq c) \wedge (x_2 \geq d)] \vee$

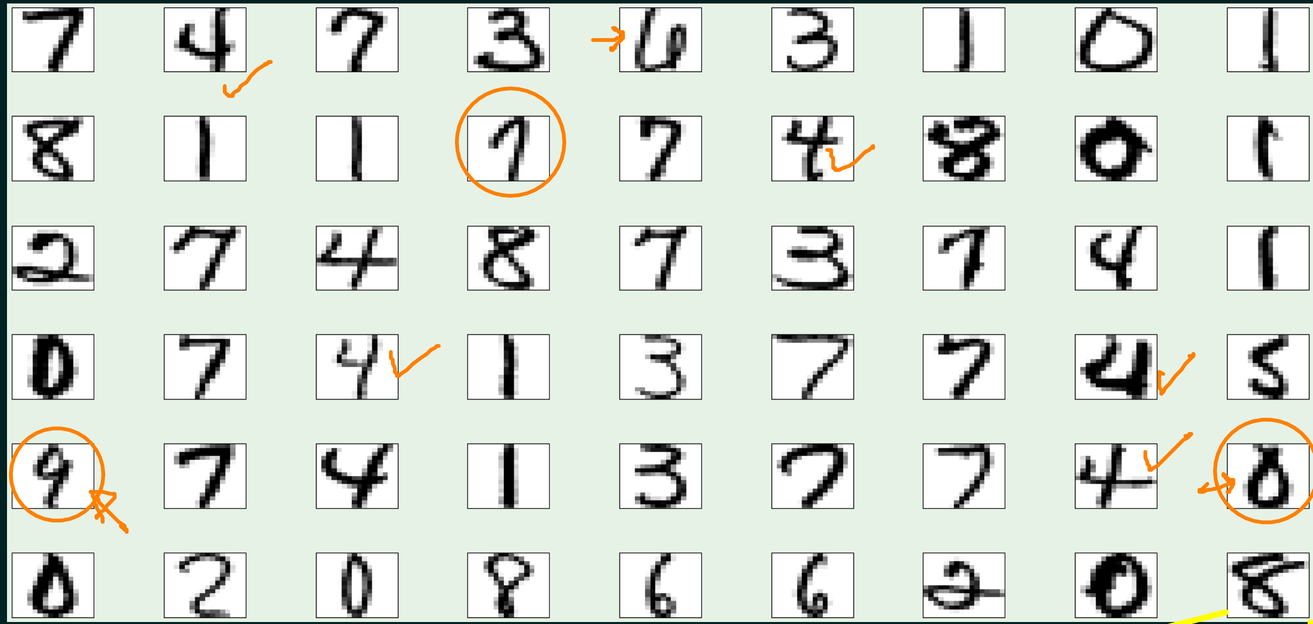
$[a \leq x_1 \leq b \wedge x_2 \leq d]$

$\boxed{721302}$ ← Pincode

$(x_1 \; ----- \; x_{256})$

$$\sum_{i=0}^{256} w_i x_i = 0$$

5 1

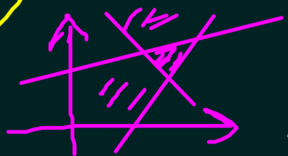


$x_0 = 1$ $\begin{cases} x_1 \rightarrow \text{intensity} \\ x_2 \rightarrow \text{Symmetry} \end{cases}$

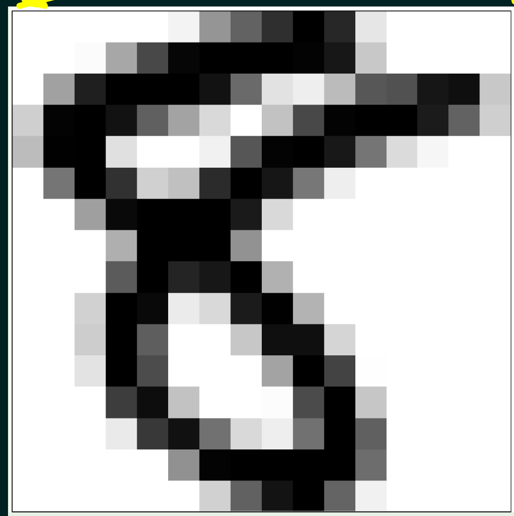
w_0, w_1, w_2

$$W_0 + W_1 x_1 + W_2 x_2 = 0$$

Separator



16 x 16

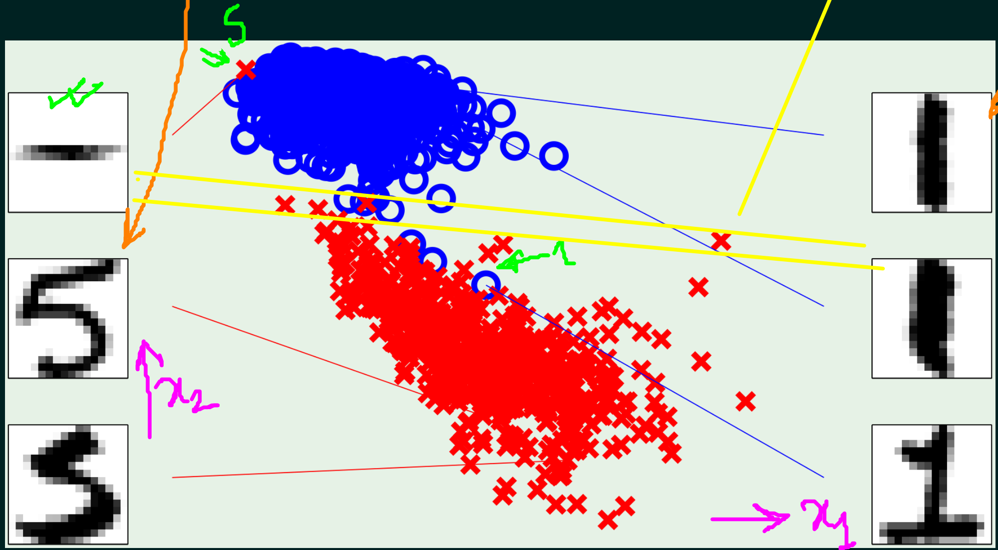


Intensity

0 — 9

1

Symmetry



Perceptrons

$$\delta = \boxed{\sum_{i=0}^{d} w_i d_i = 0}$$

$\left.\begin{matrix} TE_1 \\ TE_2 \\ \vdots \\ TE_N \end{matrix}\right\}$ batch

itr.

$E_{in} \approx E_{out}$

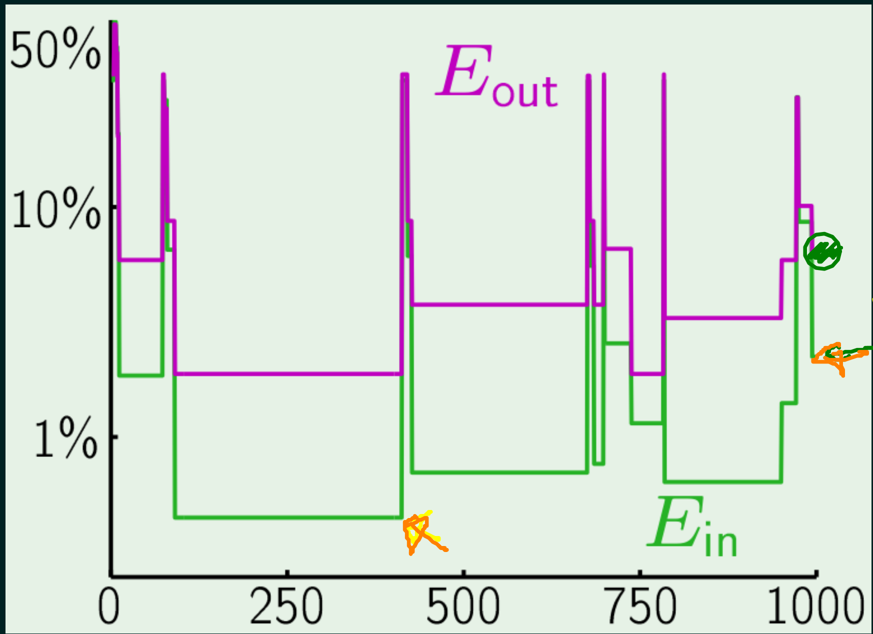
(assume)

$TE_1 \rightarrow +$

$(-)$

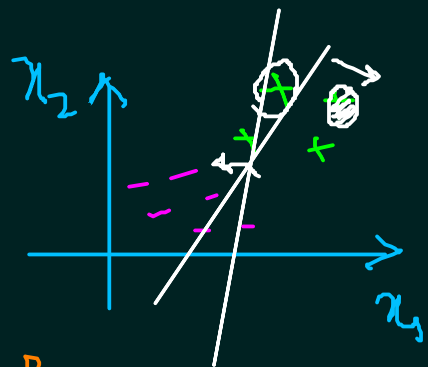
LA → batch

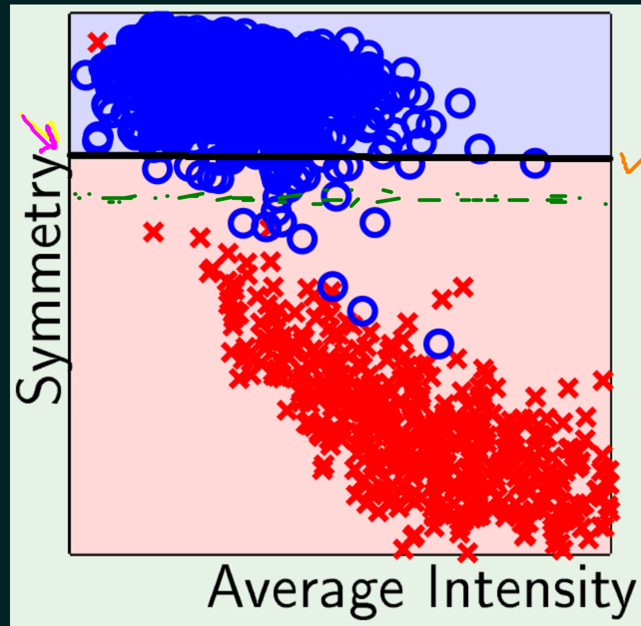
→ Itr ✓

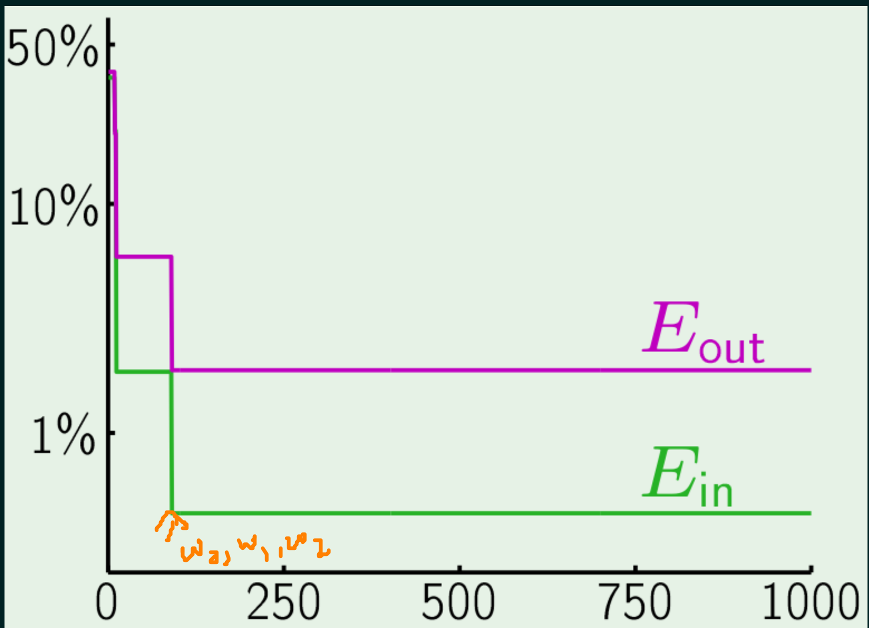


$\boxed{E'_{in}}$

Perceptron Learning Algo



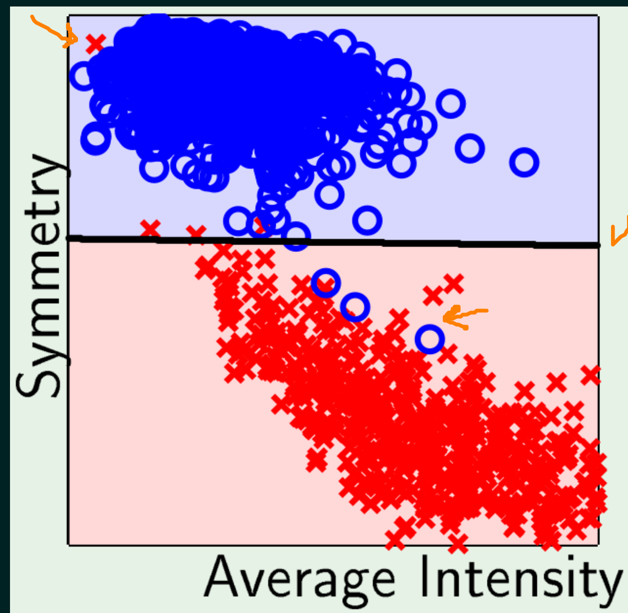




'Pocket' Learning Algo

5 and 6

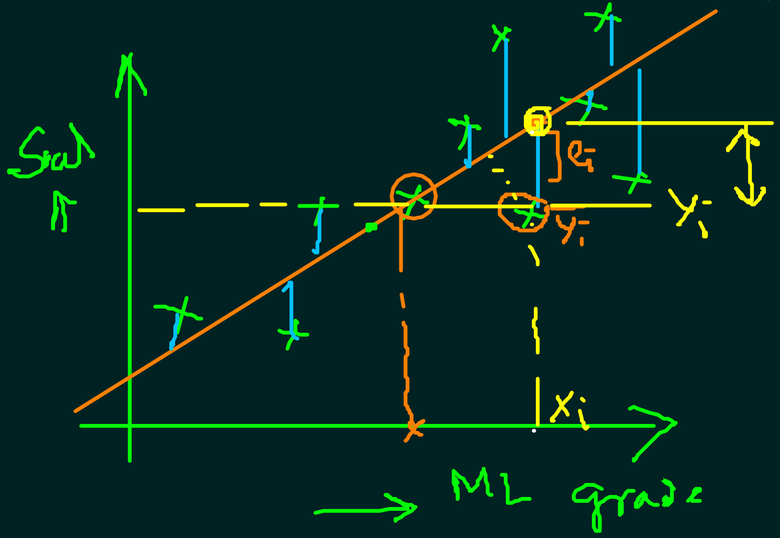
[2.5% error]



Credit Approval → Yes / → No } classify

↑ Yes —— Th —— ↓ No

→ Credit limit → o/p = real value (REGRESSION)



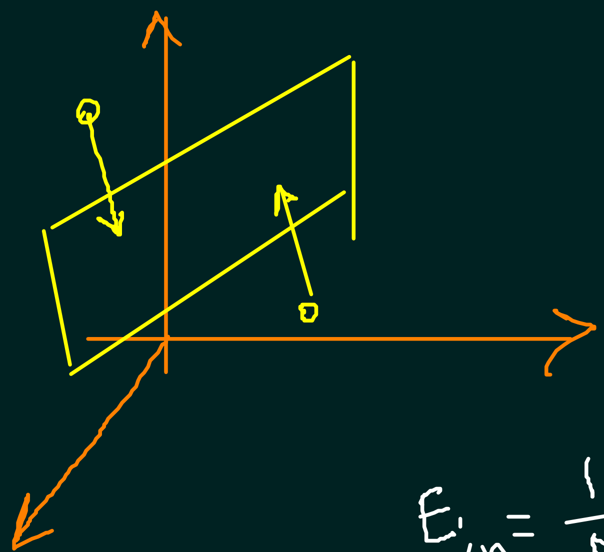
linear Regression

TE: (N)

real

$$X_i = [(x_{i1} \dots x_{id}), y_i]$$

$$y_i = \sum_{i=0}^{d} w_i x_i \Rightarrow W^T . x = y$$



$$\boxed{e_i = \left(y_i - \sum_{i=0}^{d} w_i x_i\right)}$$

minimize for all TEs

$$W = \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_d \end{bmatrix}_{(d+1)\times 1} \qquad X = \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_d \end{bmatrix}_{(d+1)\times 1}$$

$$E_{in} = \frac{1}{N} \sum e_i^2 \longrightarrow \text{minimize}$$

(Sum-sq. error)

$$\boxed{E_{in}(w)} = \frac{1}{N}\sum_{n=1}^{N}\left(w^T x_n - y_n\right)^2 = \frac{1}{N}\|Xw - y\|^2 \qquad w = \begin{bmatrix} w_0 \\ \vdots \\ w_d \end{bmatrix}$$

$$(Xw - y)^T (Xw - y)$$

$$(w^T X^T - y^T)(Xw - y)$$

$$x_i = \begin{bmatrix} x_{i0} \\ x_{i1} \\ \vdots \\ x_{id} \end{bmatrix} \quad (d+1)\times 1$$

$$X = \begin{bmatrix} - x_1^T - \\ - x_2^T - \\ \cdots \\ - x_N^T - \end{bmatrix} \quad N \times (d+1) \qquad y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} \quad N \times 1$$

$$\nabla E_{in}(w) = 0$$

$$\Rightarrow \nabla E_{in}(w) = \frac{2}{N} X^T (Xw - y) = 0$$

$$(AB)^T = B^T A^T$$

$$\Rightarrow (X^T X) w = X^T y$$

$$A^T A: \quad [a_1, a_2, \dots] \begin{bmatrix} a_1 \\ \vdots \end{bmatrix} = a_1^2 + a_2^2 + \cdots$$

$$\Rightarrow \boxed{w} = \underbrace{(X^T X)^{-1} X^T}_{X'} y$$

w: $(d+1)\times 1$

X^T: $(d+1)\times N$, X: $N\times(d+1)$, $X^T X$: $(d+1)\times(d+1)$

X^T: $(d+1)\times N$, y: $(N\times 1)$

$(d+1)\times N$

$$\boxed{w = X' y}$$

$$\boxed{X' = (X^T X)^{-1} X^T}$$

pseudo inverse

$(d+1)\times 1$

$$X' X = (X^T X)^{-1}(X^T X) = I$$

Linear Regression Algo

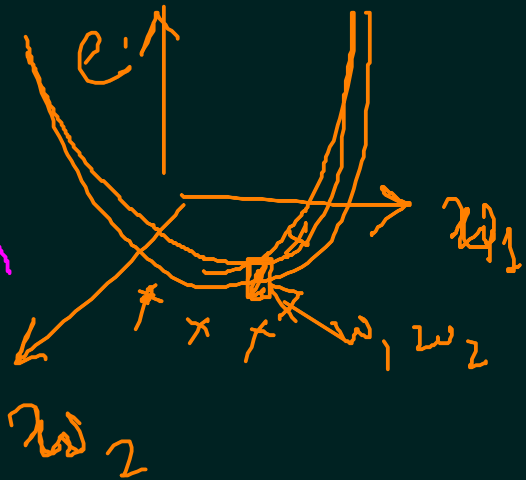
Linear Regression:

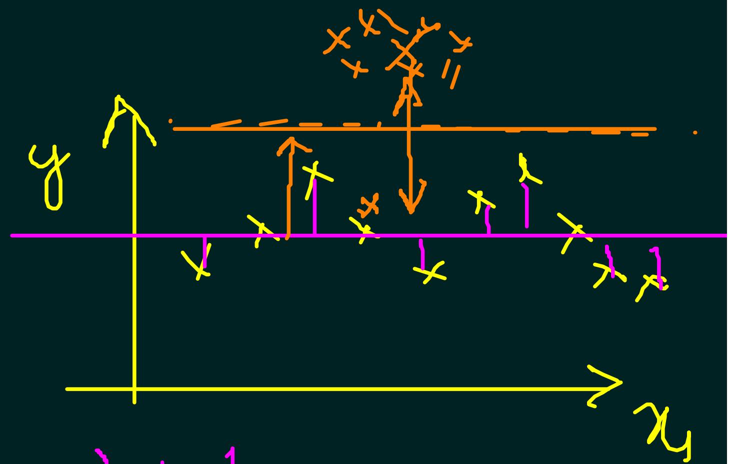
$h(x) = y \rightarrow$ Real value

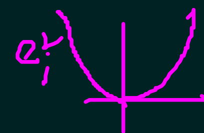
$\rightarrow \{-1, +1\}$ Classification

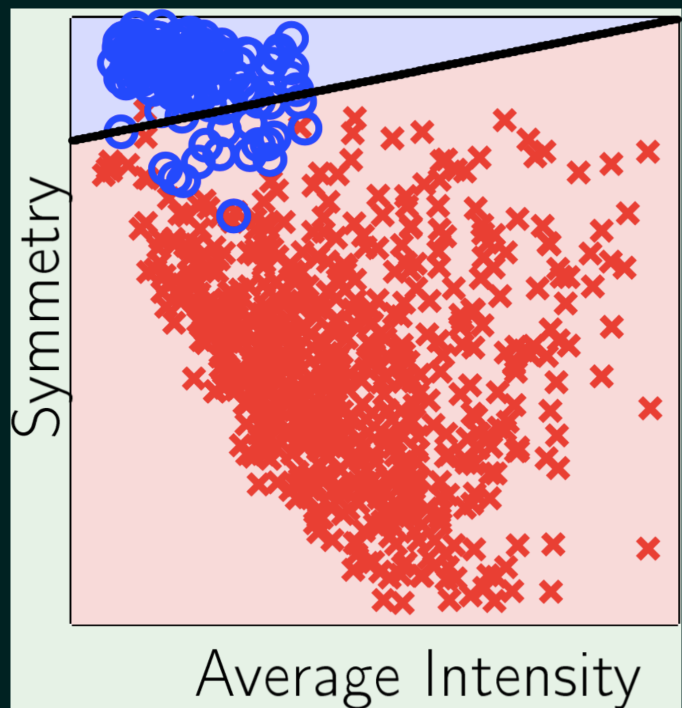
Can I use it for classification?

(Yes)



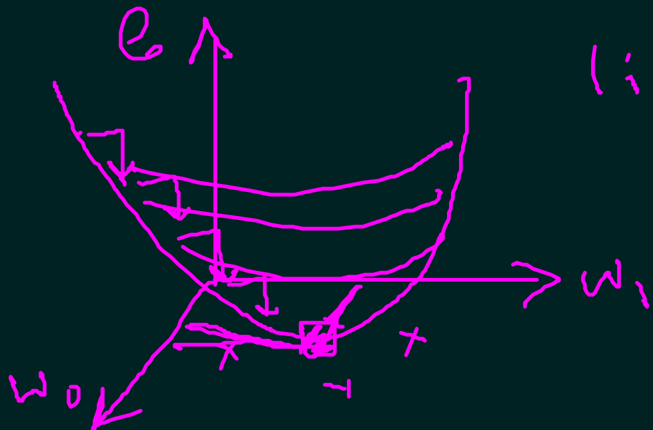






$w_1 x_1 + w_2 x_2 + w_0 = 0$

linear w.r.t. w



$\boxed{\text{gradient descent}}$

batch

stochastic