Linear Regression

Linear predictors

Loss minimization

Stochastic gradient descent

Types of prediction tasks

Binary classification (e.g., email ⇒ spam/not spam):

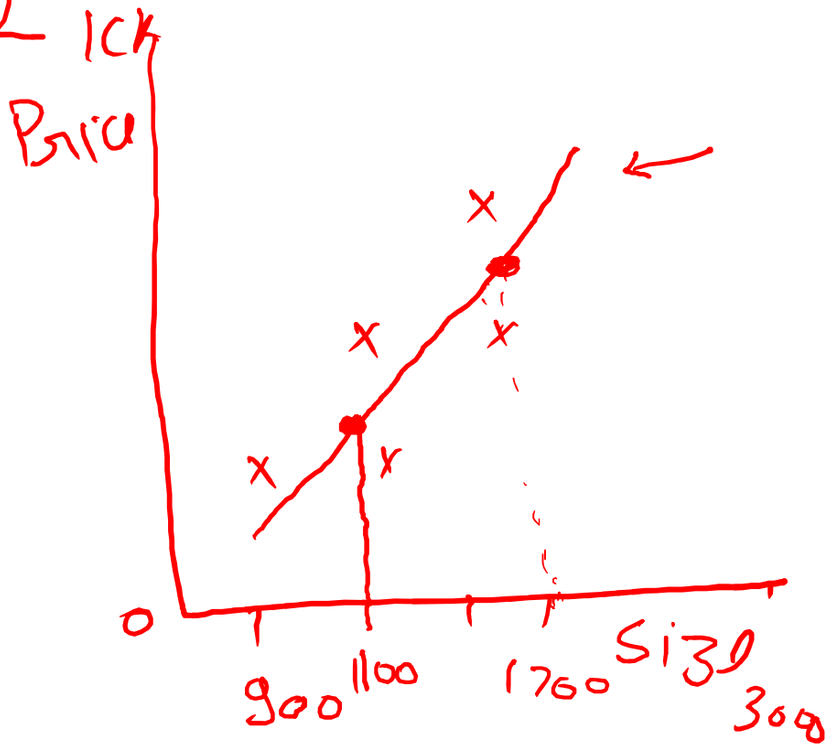
$x \longrightarrow h \longrightarrow y \in \{+1, -1\}$

Regression (e.g., location, year ⇒ housing price):

$x \longrightarrow h \longrightarrow y \in R$

Problem

Input Size(sqft)	#rooms	output Price(Lakh)
900	2	12
1200	2	20
1500	3	30
1800	3	40
2400	4	50



Notations

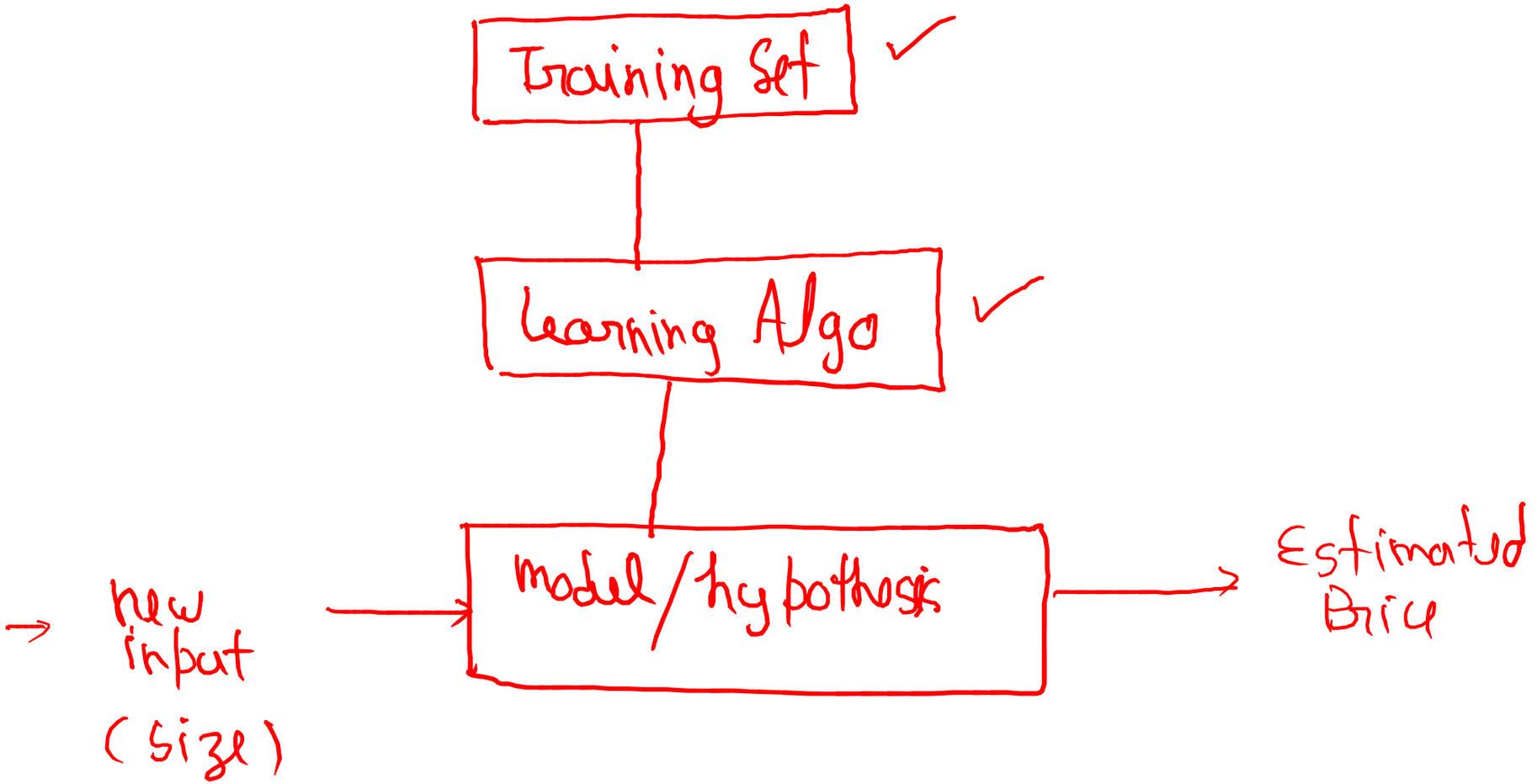
Training set (x, y)

input / feature (x)

output / target variable (y)

m = # training examples.

Hypothesis/Model



Representation of Hypothesis

$y = mx + c$

$$h(x) = w_0 + w_1 x_1$$
$$= w_0 x_0 + w_1 x_1$$

w = Parameters / weight

$$h(x_1, x_2) = w_0 + w_1 x_1 + w_2 x_2$$

$$h(x) = \sum_{j=0}^{n} w_j x_j$$
$$x_0 = 1$$

Job of LA is to find best value of Parameters / weight

$$w = \begin{bmatrix} w_0 \\ w_1 \\ w_2 \end{bmatrix} \qquad x = \begin{bmatrix} x_0 \\ x_1 \\ x_2 \end{bmatrix}$$

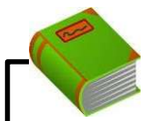
$h(x) \simeq y$

Job of Learning algorithm

find / choose W s.t $h(x) \simeq y$ for the given training set

minimize $J(W) = \frac{1}{2} \sum_{i=1}^{m} (h(x^{(i)}) - y^{(i)})^2$

Loss functions



Definition: loss function

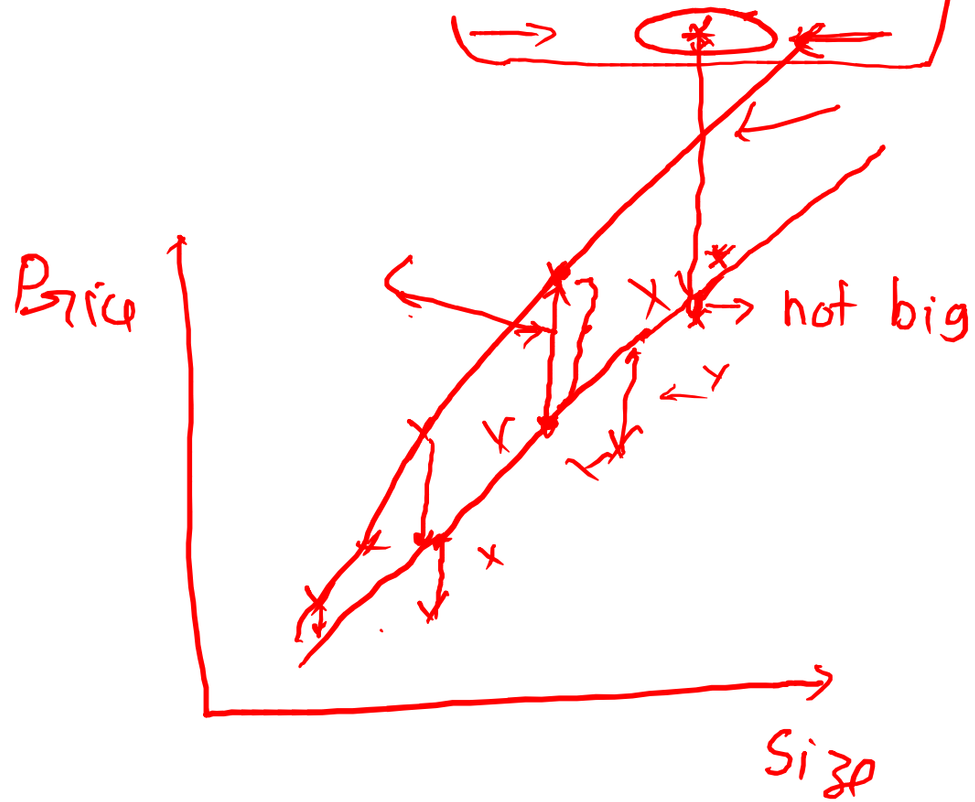
A loss function Loss(x, y, $\mathbf{w}$) quantifies how unhappy you would be if you used $\mathbf{w}$ to make a prediction on x when the correct output is y. It is the objective we want to minimize.

Loss function

$h(x) - y$ – margin

→ $|h(x) - y)|$ → $Loss_{absaluty}$

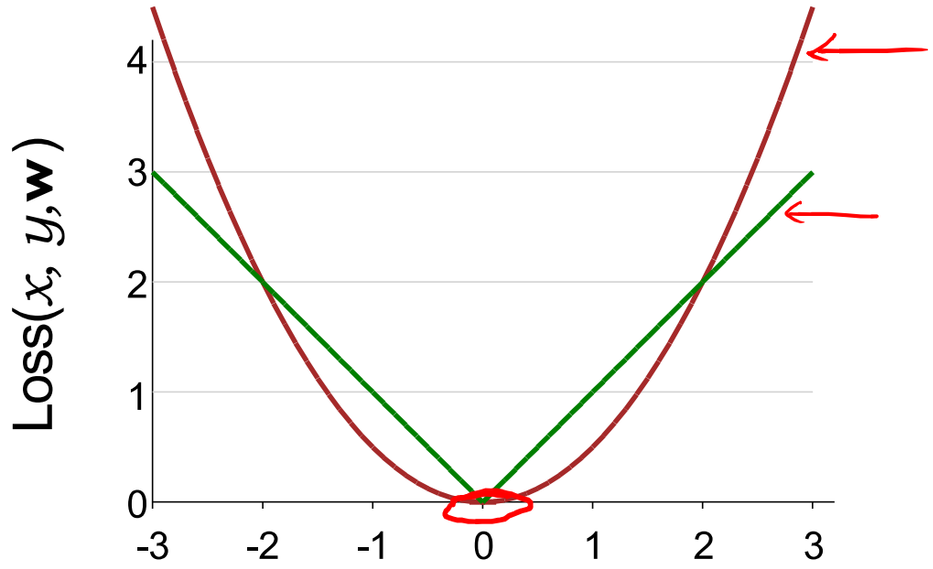
→ $(h(x) - y)^2$ = Loos squared loss



$h(x)$ → predicted value

y → ground truth / original value

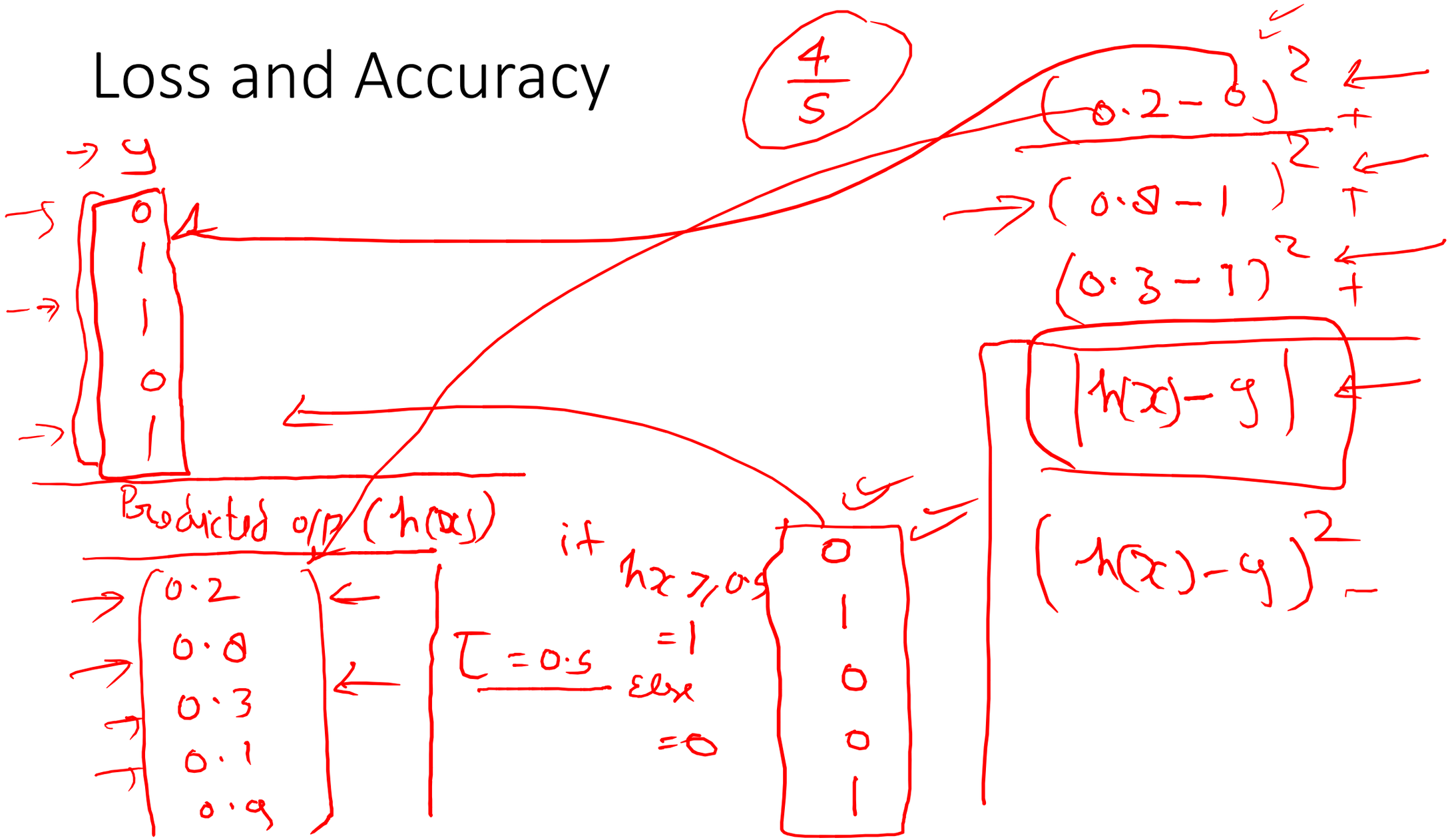
Regression loss functions



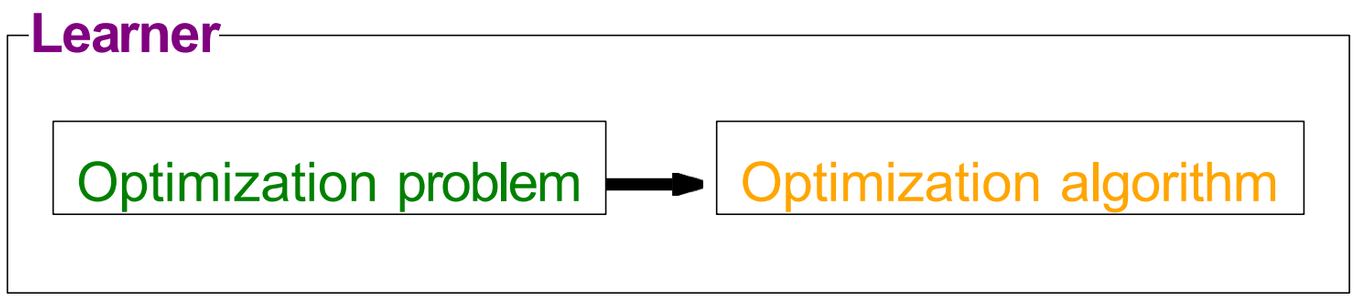
$$\text{Loss}_{\text{squared}}(x, y, \mathbf{w}) = (\mathbf{w} \cdot \varphi(x) - y)^2$$

$$\text{Loss}_{\text{absdev}}(x, y, \mathbf{w}) = |\mathbf{w} \cdot \varphi(x) - y|$$

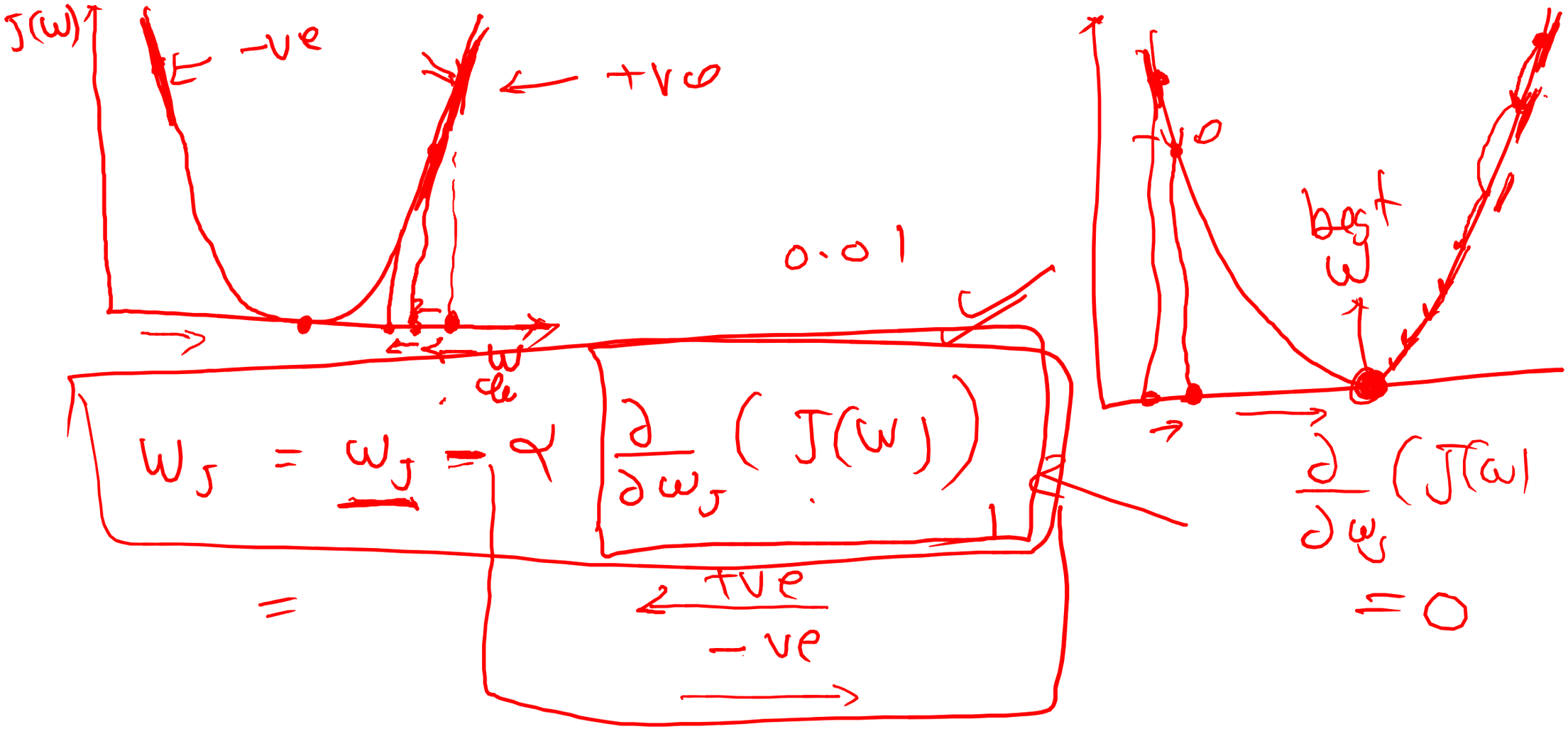
Loss and Accuracy



Learning as optimization

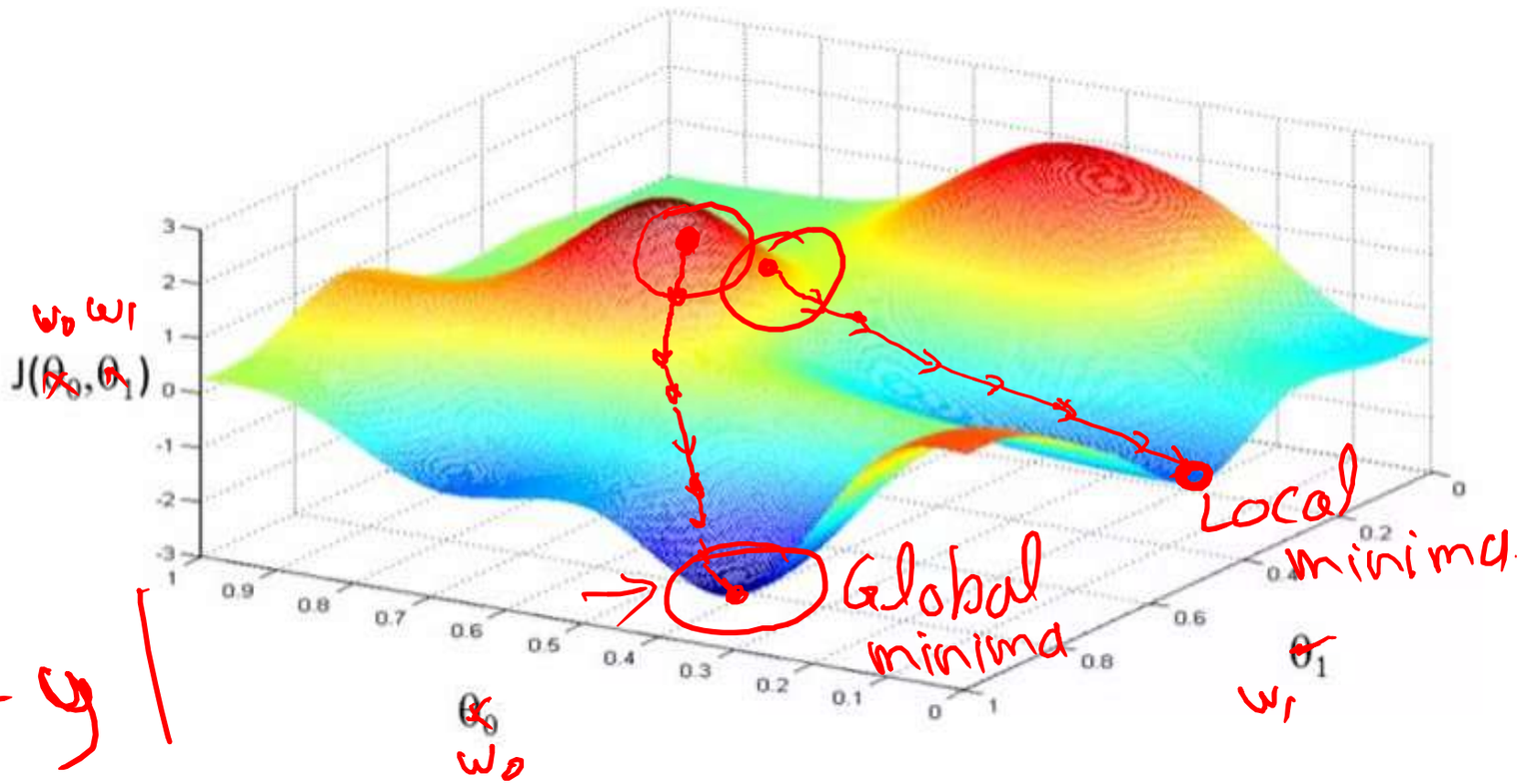


Learning algorithm : Gradient Descent



** Initialization of weight

Visualization



$J(w_1, w_2)$

$|h(x) - y|$