Linear Regression

Problem: Cars 24 - price prediction

(Features: age, odometer, make, model, ... etc)

Task: Predict the $\boxed{\text{resale}}$ price of the car

→ historical

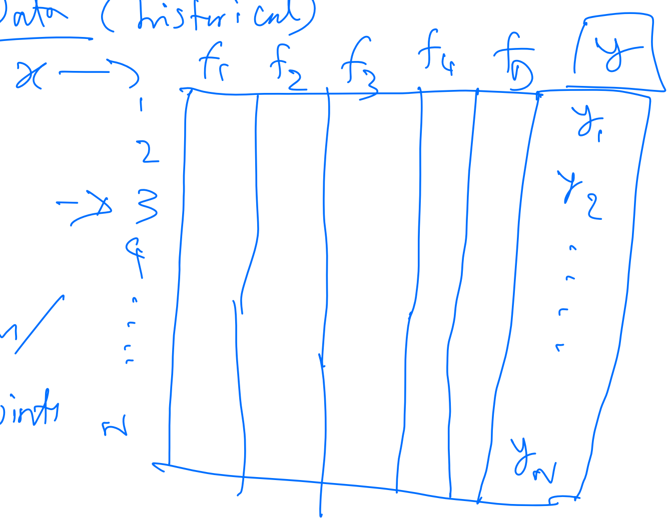
numeric

continuous

$\boxed{\text{regression}}$ or classification

Linear Regression

Data (historical)



d number of features

resale price

target output

N records/ data points

$x_i = [x_{i1}, x_{i2}, x_{i3}, \ldots x_{iD}]$ → d-dimensional input

Goal: $\boxed{f(x) = y}$ → find this function $f(x)$ such that $f(x) = y$

$f(x_i) = y_i$

$f(x_i) = y_i$

$y \rightarrow$ continuous value $\quad y \in \mathbb{R}$ (real-valued)

historical data $\quad f(x) = y$

new data point $x_q \rightarrow$ a very data point

actual value $\longleftarrow$ predict $\boxed{y_q} = \boxed{f(x_q)} \rightarrow$ expected resale price

we will find out later $\quad \boxed{\hat{y}_q} \rightarrow \hat{y}_q = f(x_q)$

Target: $\hat{y}_q$ should be as close as y_q

$$\hat{y}_q \approx y_q$$

$$\rightarrow \hat{y}_q - y_q \approx 0 \quad \text{i.e.} \quad |\hat{y}_q - y_q| \approx 0$$

Train & Test Phase

Train: Use the historical data to find or fit the function $f(x)$

$\downarrow$ train data

$$\rightarrow (\underbrace{x_{i1}, x_{i2}, x_{i3}, \ldots x_{id}}_{d\text{-dim}}, \underbrace{y_i}_{1\text{ dim}})$$

After we find/fit the function $f(x)$, we go to test phase

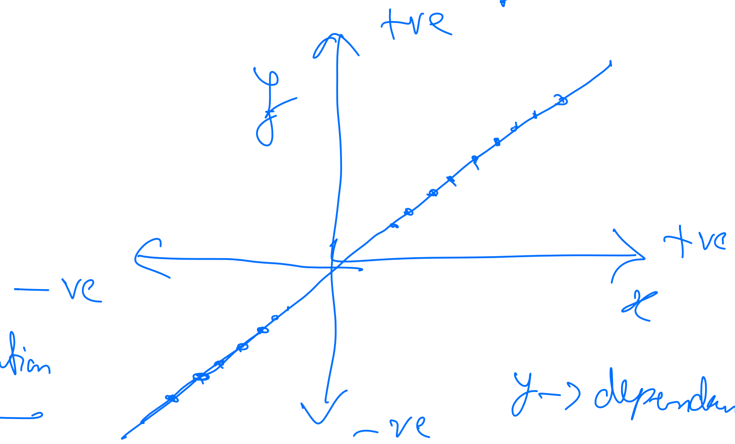
Test Phase : Unseen data $\leftarrow$ test data

i/p: $\underbrace{\left(x_{q1}, x_{q2}, x_{q3}, \ldots, x_{qd} \right)}_{d\text{-dim}}$

$\boxed{\text{o/p: } \hat{y}_q}$

1 degree equation

$\boxed{y} = m\boxed{x} + c$



$y \rightarrow$ dependent

$x \rightarrow$ independent

Linear Regression Equation

dependent variables

independent variable

$y = \boxed{mx} + c$

$\underbrace{x_{i1}, x_{i2}, x_{i3}, \ldots, x_{id}}_{\rightarrow \text{independent}}, y_i \rightarrow$ dependent

$$\hat{y}_i = w_1 \times x_{i1} + w_2 \times x_{i2} + w_3 \times x_{i3} + \ldots + w_d \times x_{id} + \boxed{w_0} \rightarrow c$$

$$\boxed{\hat{y}_q = w_0 + w_1 \cdot x_{q1} + w_2 \cdot x_{q2} + \ldots + w_d \cdot x_{qd}}$$

$y = m_1 x + m_2 x^{\boxed{2}} + c \rightarrow$ degree $= 2$

$\rightarrow$ 1 independent $\rightarrow x$

$y = m_1 x + c \rightarrow$ degree $= 1$

$y = m_1 \underline{x} + m_2 \underline{z^2} + k \rightarrow$ degree $= 2$

2 independent variables

d independent variables

$\rightarrow$ $\boxed{\text{variables}}$ we need to find out ? $d+1$

$y = mx + c$, $\boxed{m, c} \rightarrow 2$

$$\boxed{\begin{aligned} 3x_1 + 4x_2 &= 5 \\ 7x_1 + 2x_2 &= 6 \\ 8x_1 + 3x_2 &= 11 \end{aligned}}$$

x_1, x_2

features $= d$

weights $= d+1$

System of linear equations

$$\boxed{\begin{aligned} w_1 \cdot x_{i1} + w_2 \cdot x_{i2} + \cdots + w_d \cdot x_{id} &= y_i \\ w_1 \cdot x_{j1} + w_2 \cdot x_{j2} + \cdots + w_d \cdot x_{jd} &= y_j \\ w_1 \cdot x_{k1} + w_2 \cdot x_{k2} + \cdots + w_d \cdot x_{kd} &= y_k \end{aligned}}$$

N data points, d+1 features

N × (d+1) matrix

$$
\begin{array}{c|c}
 & 1 \quad 2 \quad 3 \quad \dots \quad d+1 \\
\begin{array}{c}1\\2\\3\\ \vdots \\ n\end{array} & \left[\begin{array}{c} \\ \\ \\ \\ \end{array}\right]
\end{array}
= \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}
$$

n × (d+1)   N × 1

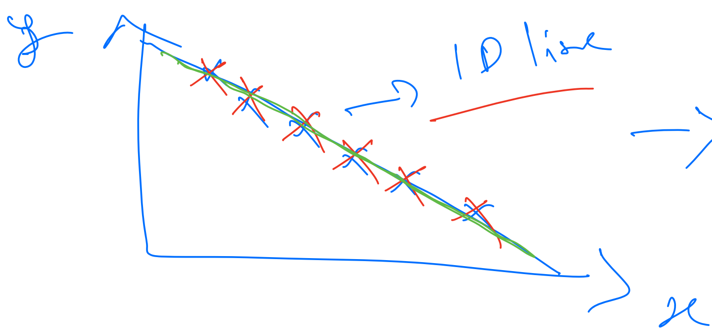
$$\boxed{\text{Gradient Descent}}$$

Geometric Intuition

Resale value   age of the car

$\downarrow$ $\downarrow$

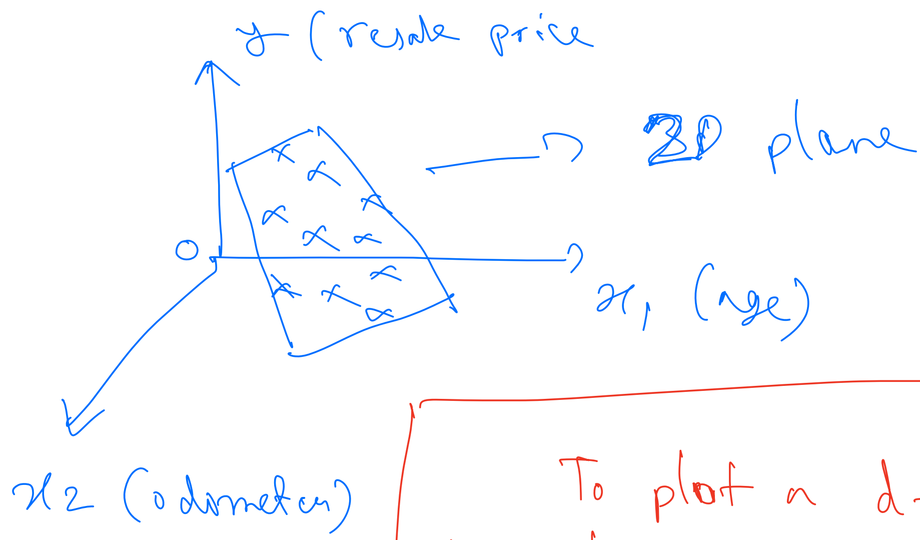
y $\boxed{x} \to 1$



→ 1D line → 2D

Resale value   age, odometer

y $\boxed{x_1, \quad x_2,} \to 2$



To plot a d-dim hyperplane, we require d+1 dimension

Optimization

find w_j ($j = 0, 1, 2, \ldots, d$), such that for every data point i, $\hat{y}_i - y_i$ should be as low as possible.

$f(x_i)$

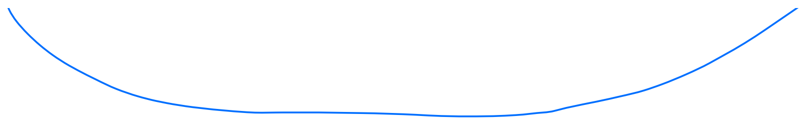
$$\boxed{f(x_i) = w_0 + w_1 x_{i1} + w_2 x_{i2} + \ldots + w_d \cdot x_{id}}$$

$\hat{y}_i - y_i = e_i \rightarrow$ diff for datapoint x_i

$$\boxed{\min \sum_{i=1}^{N} e_i} \rightarrow \text{target}$$

$[x_{i1}, x_{i2}, x_{i3}, \ldots x_{id}]$

Find $w_0, w_1, w_2, \ldots w_d$



find $w_0, w_1, w_2, \ldots w_d$ which will minimize

Sum of errors i.e $\min \sum_{i=1}^{N} e_i$

$e_i = \hat{y}_i - y_i$ → +ve, → −ve

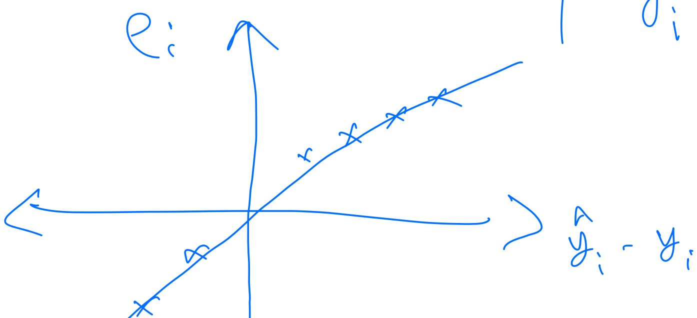
x_i	e_i
1	100
2	−100
3	10,000 ✓
4	−60,000 ✓
5	50,000 ✓
	0

$\sum e_i = 0$

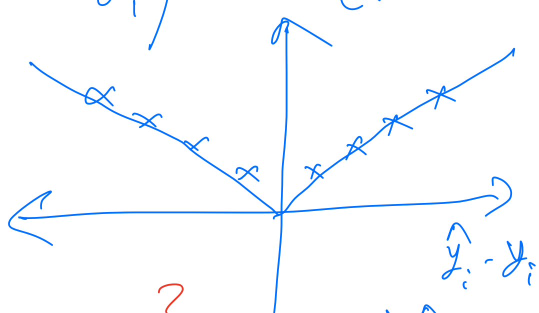
big errors

Better error = abs value

$$e_i = |\hat{y}_i - y_i|$$



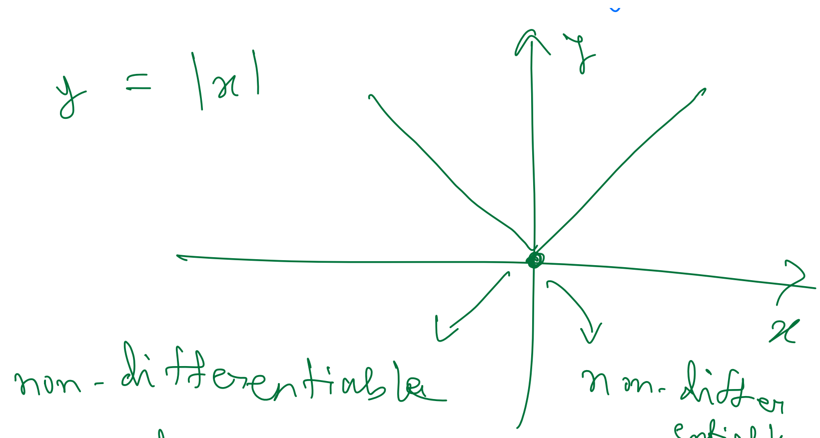
$e_i = \hat{y}_i - y_i$



$e_i = |\hat{y}_i - y_i|$

Abs value $y = |x|$

$x \neq 0$

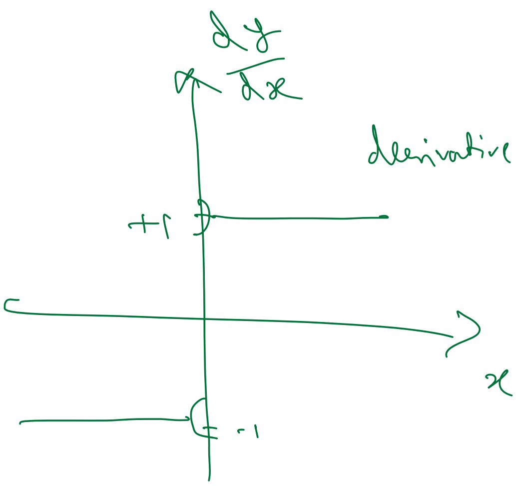


at $x = 0$

but for $x \neq 0$, it is differentiable

$y = |x|$

$$y = 0, \quad x = 0$$
$$= -x, \quad x < 0$$
$$= x, \quad x > 0$$



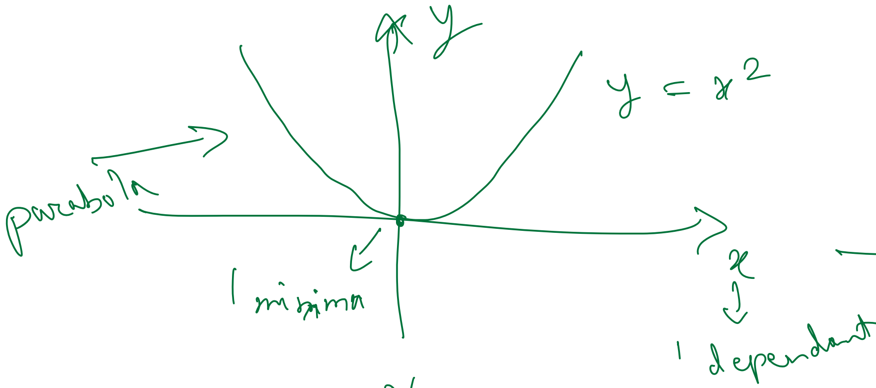
derivative(y) = non-diff

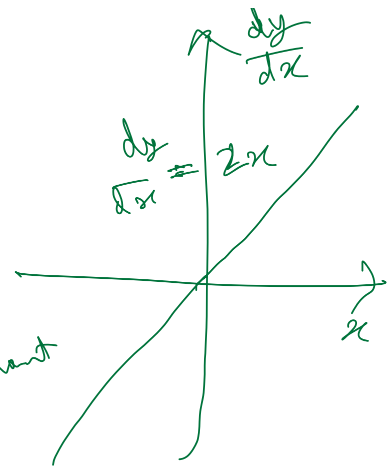
$$= -1, \quad x < 0$$
$$= +1, \quad x > 0$$

Error fn $\longrightarrow$ Mean Squared error

1) differentiable everywhere,
2) its value changes with x

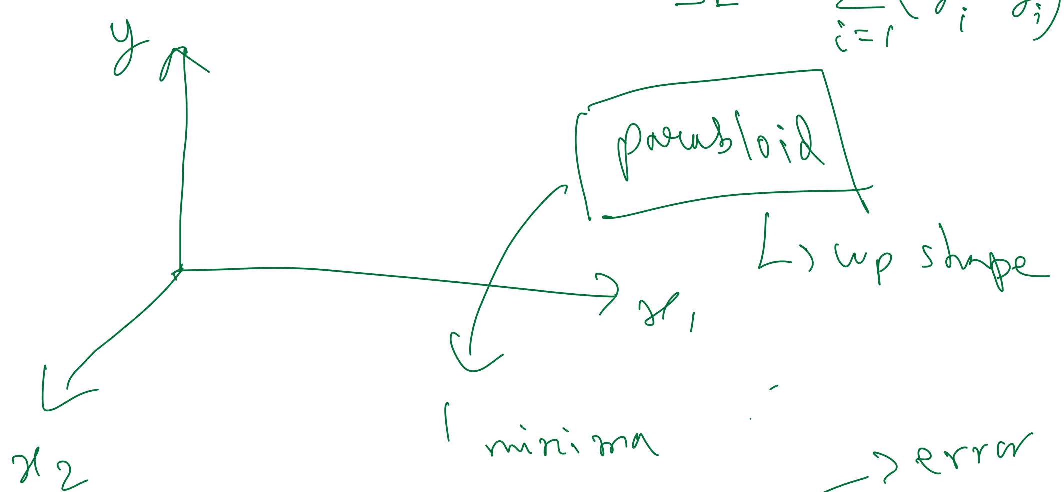
$$y = x^2$$
$$\frac{dy}{dx} = 2x$$





$$MSE = \frac{1}{N} \sum_{i=1}^{N} (\hat{y}_i - y_i)^2$$

$$SE = \sum_{i=1}^{N} (\hat{y}_i - y_i)^2$$



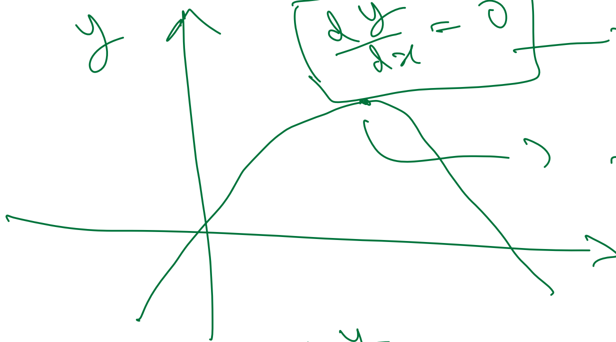
MSE → convex $\boxed{\text{loss}}$ function → error

Solve for w_j

$$MSE = \frac{1}{N} \sum_{i=1}^{N} (\hat{y}_i - y_i)^2$$

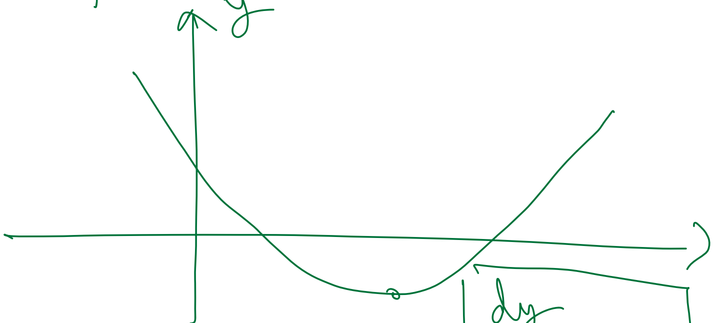
$$L = \frac{1}{N} \sum_{i=1}^{N} \left[(w_0 + w_1 x_{i1} + w_2 x_{i2} + \cdots w_d x_{id}) - y_i \right]^2$$

Gradient Descent Approach ✓



$$\boxed{\frac{dy}{dx} = 0} \rightarrow x_1 \rightarrow \frac{df(x_1)}{dx_1} = 0$$

$$\rightarrow x_1$$



$$\boxed{\frac{dy}{dx} = 0}$$

$$\rightarrow w_0, \; w_1, \; w_2, \; \dots \; w_d$$

$$\left.\begin{array}{l} w_0 \leftarrow \dfrac{\partial L}{\partial w_0} = 0 \\[2ex] w_1 \leftarrow \dfrac{\partial L}{\partial w_1} = 0 \\[2ex] w_d \leftarrow \dfrac{\partial L}{\partial w_d} = 0 \end{array}\right]$$

$d+1$ equations to solve

E.g

Resale value $\longrightarrow y$

Age of the car $\longrightarrow x$

$$y = w_0 + w_1 x$$

$$\Rightarrow L(w_0, w_1) = (y - \hat{y})^2 = (y - (w_0 + w_1 x))^2$$

$$\frac{\partial L}{\partial w_0} = \frac{\partial (y - (w_0 + w_1 x))^2}{\partial w_0}$$

$$= 2(y - (w_0 + w_1 x)) \cdot \frac{\partial (y - (w_0 + w_1 x))}{\partial w_0}$$

$\downarrow$ -1

$$= 2(y - (w_0 + w_1 x))(-1)$$

$$= -2(y - \hat{y})$$

$$\boxed{\frac{\partial L}{\partial w_1}} = \frac{\partial (y - (w_0 + w_1 x))^2}{\partial w_1}$$

$$= 2(y - (w_0 + w_1 x)) \cdot \frac{\partial (y - (w_0 + w_1 x))}{\partial w_1}$$

$\downarrow$ $(-x)$

$$= 2(y - \hat{y})(-x)$$

$$L = \sum \qquad \frac{\partial L}{\partial w_1} = \sum$$

$$MSE = \frac{1}{N} \sum_{i=1}^{N} L(w_0, w_1)$$

$$\boxed{\begin{aligned} \frac{\partial MSE}{\partial w_0} &= \frac{1}{N} \sum_{i=1}^{N} \frac{\partial L}{\partial w_0} \quad \ldots (1) \\ \frac{\partial MSE}{\partial w_1} &= \frac{1}{N} \sum_{i=1}^{N} \frac{\partial L}{\partial w_1} \quad \ldots (2) \end{aligned}}$$

$$w_0, w_1 = 0, 0$$

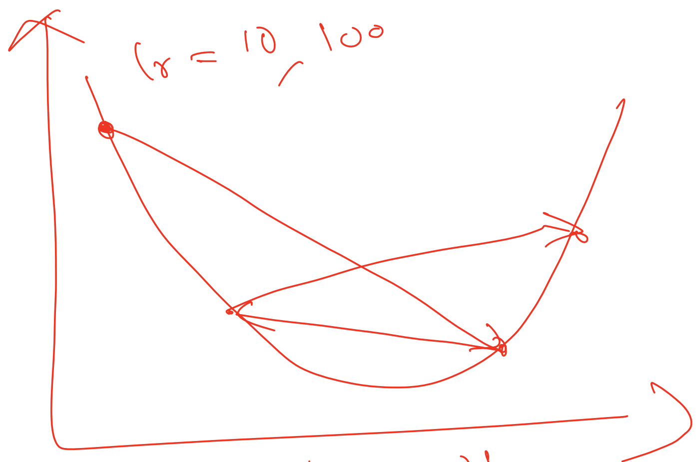
Gradient Descent Approach

$$\boxed{\begin{aligned} w_0 &= w_0 - \boxed{\eta} \times \boxed{\frac{\partial L}{\partial w_0}} \\ w_1 &= w_1 - \eta \times \frac{\partial L}{\partial w_1} \end{aligned}}$$

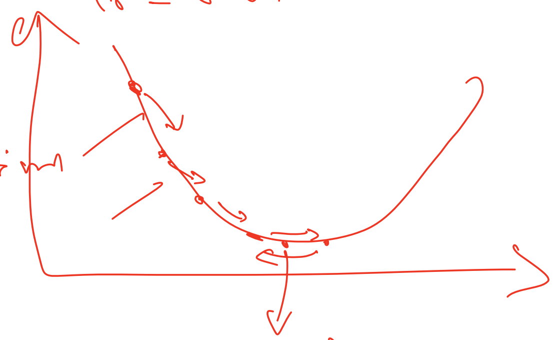
learning rate

$$\eta = 1, \; 100, \; 0.1, \; 0.01$$

ideally smaller values



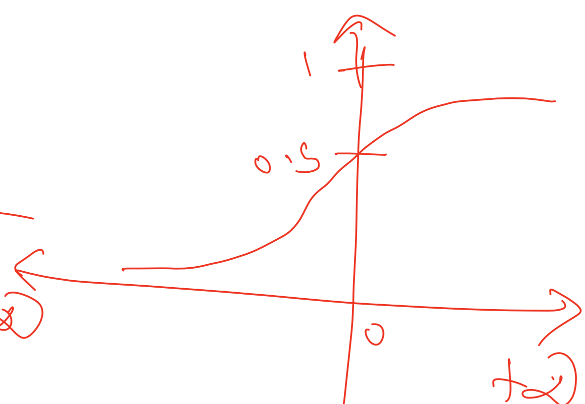
lr = 0.01

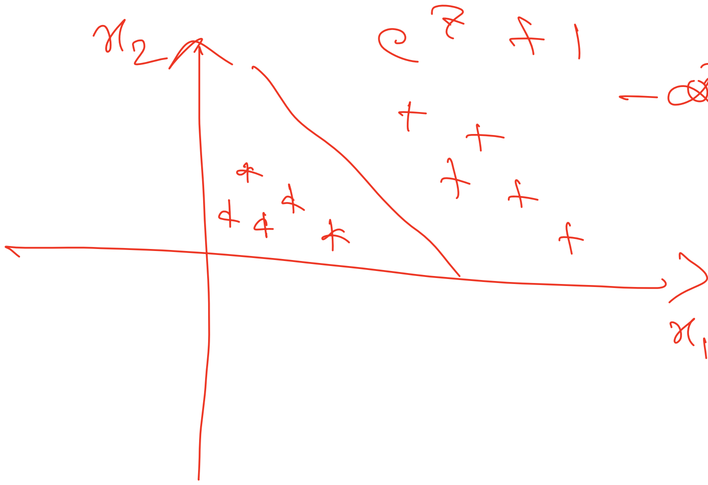


$x \rightarrow$ independent variable

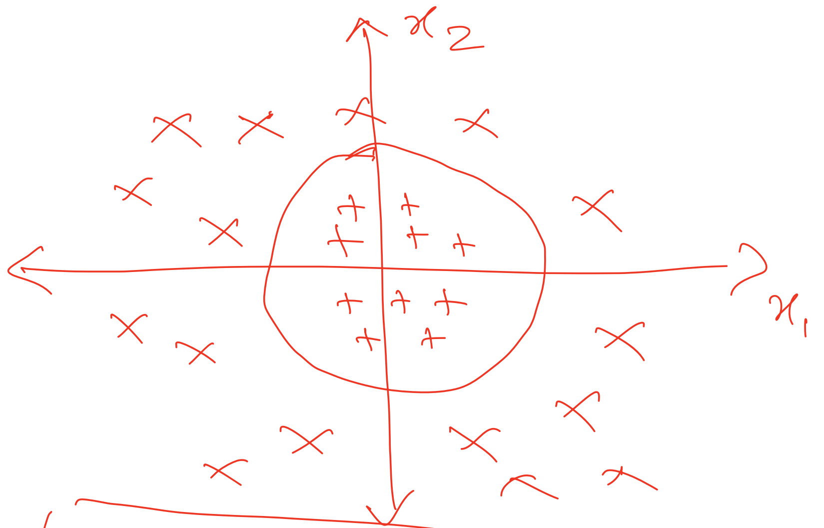
$$LT = \boxed{w_0 + w_1 x} = z$$

$$\text{Sigmoid} = \frac{e^{z}}{e^{z} + 1}$$





$$\boxed{\begin{aligned} y &= c \\ y &= m_1 x + c \end{aligned}}$$



$$\boxed{z} = w_0 + w_1 x_1 + w_2 x_2 + w_3 x_1 x_2 + w_4 x_1^2 + w_5 x_2^2$$

sigmoid (z)