$$MSE = \frac{1}{m}\sum_{i=1}^{m}\left[\hat{y}^{(i)} - y^{(i)}\right]^2$$

Loss or error function — Predicted — Actual

$$h_\theta^{(i)} = \theta_1 x^{(i)} + \theta_0$$

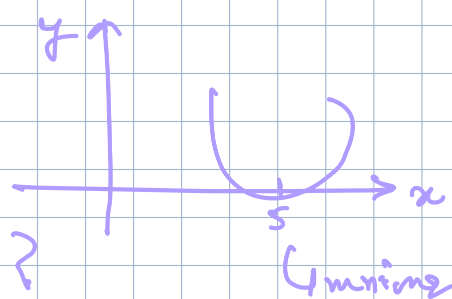
$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m}\left[\theta_1 x^{(i)} + \theta_0 - y^{(i)}\right]^2$$

Loss or Error fxn

- make updation to our θ, so that it becomes a better θ

Gradient Descent (in General)

$y = (x-5)^2$



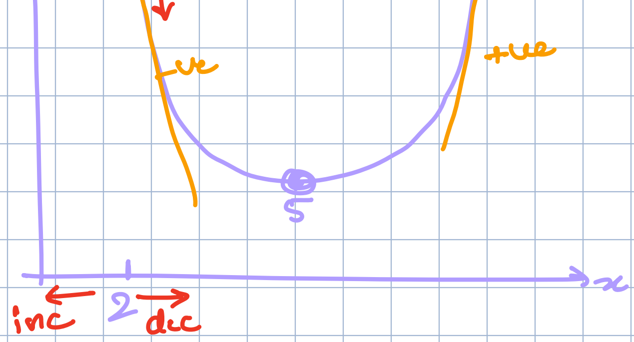
for what value of x this y will be minimum?

$$\frac{dy}{dx} = 2(x-5) = 0$$

$$\underline{\underline{x = 5}}$$

Gradient Descent

adaptive $\eta \to$ lr.



⊙ magnitude (steps) → learning rate

⊙ d^{n} which is minimizing y.

$$x = x - \eta \frac{\partial f(x)}{\partial x}$$ → Gradient/slope

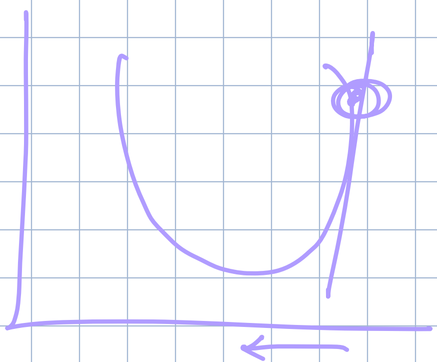
-ve

+ve

learning rate

$\eta \propto \frac{1}{t}$

$\eta \propto \frac{1}{\sqrt{t}}$

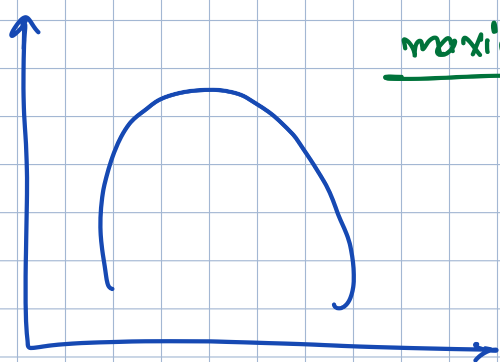


$$x = x - \eta \frac{\partial f(x)}{\partial x}$$

+ve

-ve

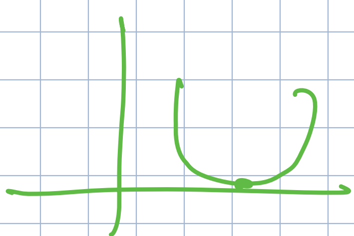
x dec



maxima

Gradient Ascent $x = x + \eta \frac{\partial f(x)}{\partial (x)}$

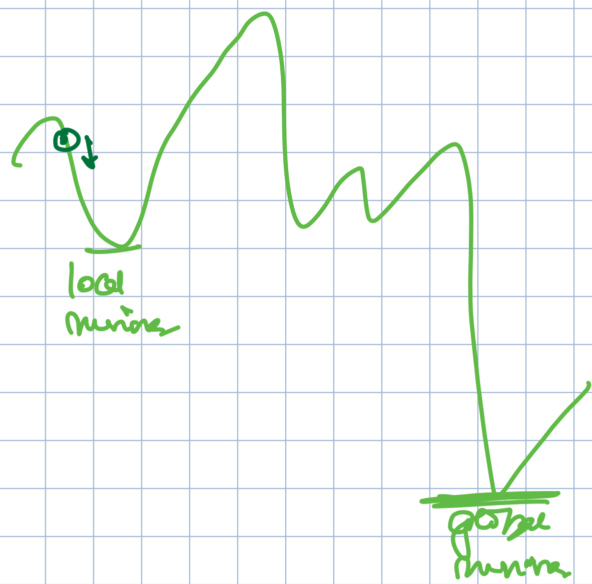
Convex $f x^{n}s$:



local minima = global minima

non convex fn's

G.D. will help to reach to local minima not global minima.

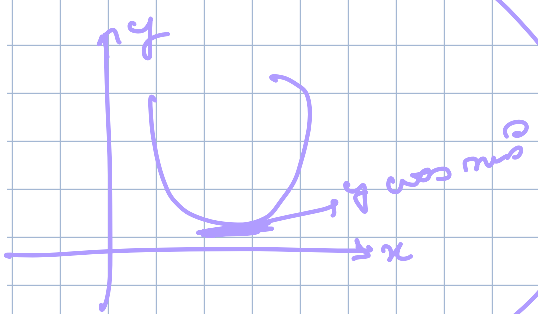


$$J(\theta) = \frac{1}{m} \sum_{i=1}^{m} \left[\left(\theta_0 + \theta_1 x^{(i)}\right) - y^{(i)} \right]^2$$

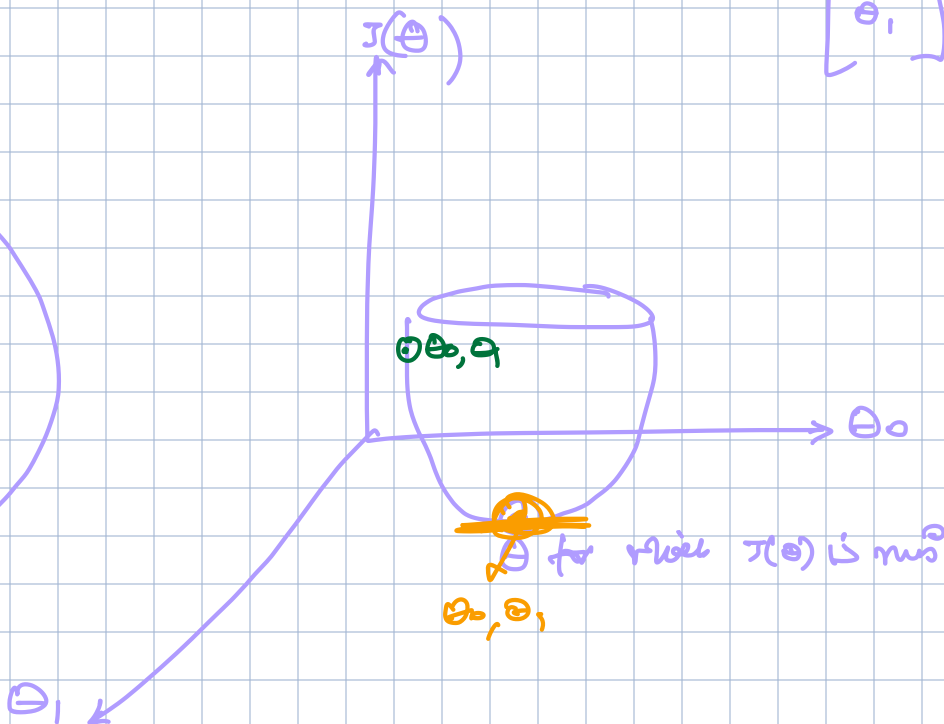
Loss or Error fn

$$\theta = \begin{bmatrix} \theta_0 \\ \theta_1 \end{bmatrix}$$

$y = (x-5)^2 + 2$



$$x = x - \eta \frac{\partial f(x)}{\partial (x)}$$



$$\theta = \theta - \eta \frac{\partial (J(\theta))}{\partial \theta}$$

$\nabla_\theta J(\theta)$

learning rate

gradient

$$\theta = \begin{bmatrix} \theta_0 \\ \theta_1 \end{bmatrix}$$

$$\nabla_\theta J(\theta)$$

$$\frac{\partial J(\theta)}{\partial \theta_0} \qquad \frac{\partial J(\theta)}{\partial \theta_1}$$

1. $$\frac{\partial J(\theta)}{\partial \theta_0} = \frac{\partial}{\partial \theta_0} \frac{1}{m} \sum_{i=1}^{m} \left[\left(\theta_0 + \theta_1 x^{(i)} \right) - y^{(i)} \right]^2$$

$$= \frac{1}{m} \sum_{i=1}^{m} \frac{\partial}{\partial \theta_0} \left[\left(\theta_0 + \theta_1 x^{(i)} \right) - y^{(i)} \right]^2$$

$$= \frac{1}{m} \sum_{i=1}^{m} 2 \left[\left(\theta_0 + \theta_1 x^{(i)} \right) - y^{(i)} \right]$$

$$= \frac{1}{m} \sum_{i=1}^{m} 2 \left[\hat{y}^{(i)} - y^{(i)} \right] \rightarrow \text{grad } 0$$

2. $$\frac{\partial J(\theta)}{\partial \theta_1} = \frac{\partial}{\partial \theta_1} \frac{1}{m} \sum_{i=1}^{m} \left[\theta_0 + \theta x^{(i)} - y^{(i)} \right]^2$$

$$= \frac{1}{m} \sum_{i=1}^{m} 2 \left[\theta_0 + \theta_1 x^{(i)} - y^{(i)} \right] x^{(i)} \rightarrow \text{grad } 1$$

$$= \frac{1}{m} \sum_{i=1}^{m} 2 \left[\hat{y}^{(i)} - y^{(i)} \right] x^{(i)}$$

$$\theta_0 = \theta_0 - \frac{\overset{lr}{\eta}}{m} \sum_{i=1}^{m} 2 \left[\hat{y}^{(i)} - y^{(i)} \right]$$

$$\theta_1 = \theta_1 - \frac{\eta}{m} \sum_{i=1}^{m} 2 \left[\hat{y}^{(i)} - y^{(i)} \right] x^{(i)}$$

$y = mx + c$

$m \rightarrow \theta_1$, $c \rightarrow \theta_0$

θ_0, θ_1 random value

do

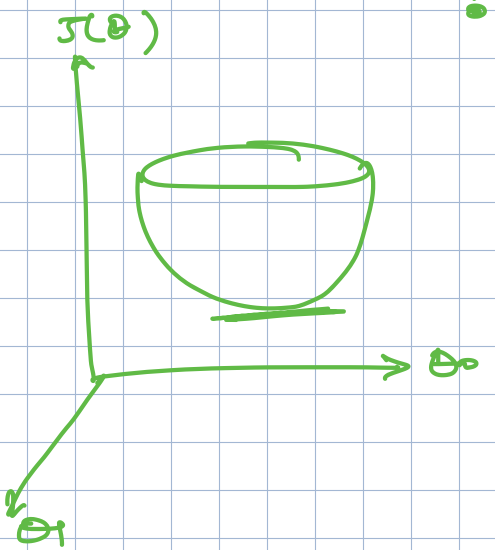
{

loss fxn / error $\rightarrow \frac{1}{m}\sum_{i=1}^{m}\left(\theta_0+\theta_1 x^{(i)} - y^{(i)}\right)^2$

update θ_0, θ_1

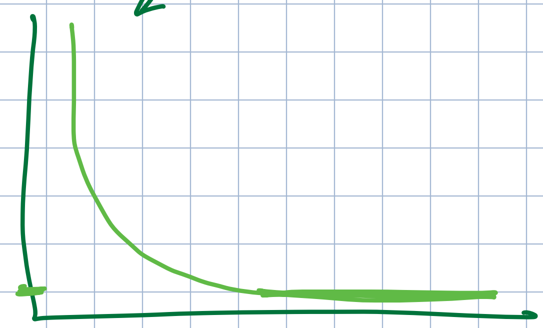
} while (convergence)

?

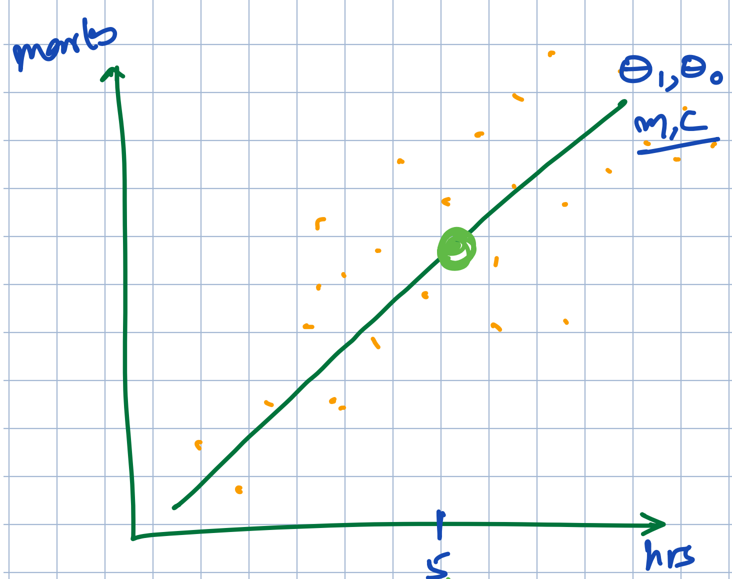


① steps fix

② error / loss fxn



Test Time:



Test Time: Student studied for 5 hrs, marks?

$h_\theta^{(i)} = \theta_0 + \theta_1 x^{(i)}$

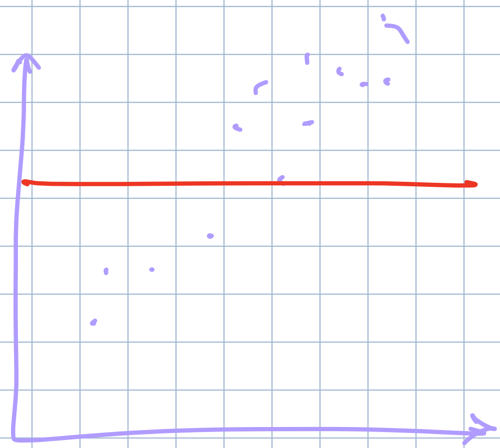
$\underbrace{= \theta_0 + \theta_1 . 5}_{\text{Real no.}}$

metric :

R2 Score:

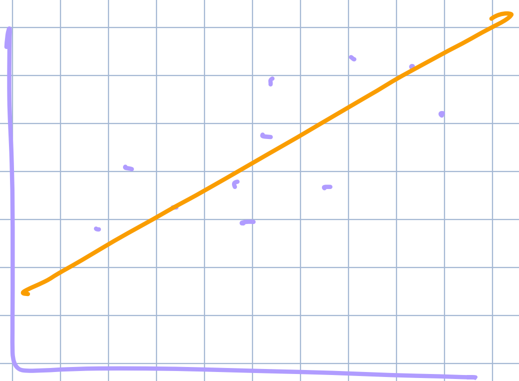
(R Squared or Coefficient of Determination)

$$R2\ Score = 1 - \frac{\sum_{i=1}^{m} \left(y^{(i)} - \hat{y}^{(i)}\right)^2}{\sum_{i=1}^{m} \left(y^{(i)} - y_{avg}^{(i)}\right)^2}$$



$\hat{y}^{(i)} = y_{avg}^{(i)}$

R2 Score = 1-1=0

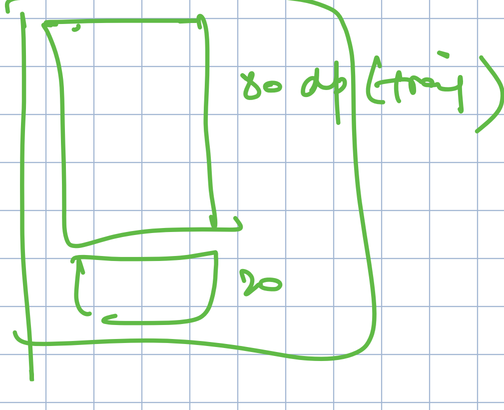


$\hat{y}^{(i)} = y^{(i)}$

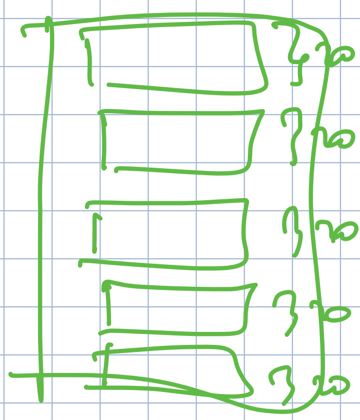
R2 Score = 1-0=1

R2 Score: 0.17 → 17%

100 data points



5 fold cross validation



Linear Regression with Multiple features

Eg: House Price Prediction

↳ Aim: Predict the price of house.

features = 4

$n = 4$

	$\overbrace{\text{\# Bedrooms} \quad \text{\# Floors} \quad \text{\# Area} \quad \text{\# Age}}^{x}$				$\overbrace{\text{Price}}^{y}$
	# Bedrooms	# Floors	# Area	# Age	Price
$x^{(1)}$	2	3	200	5	2 Cr
$x^{(2)}$	5	2	500	10	1.5 Cr
⋮					

$$X = \begin{bmatrix} \text{———} & x^{(1)} & \text{———} \\ \text{———} & x^{(2)} & \text{———} \\ & \vdots & \\ \text{———} & x^{(m)} & \text{———} \end{bmatrix}$$

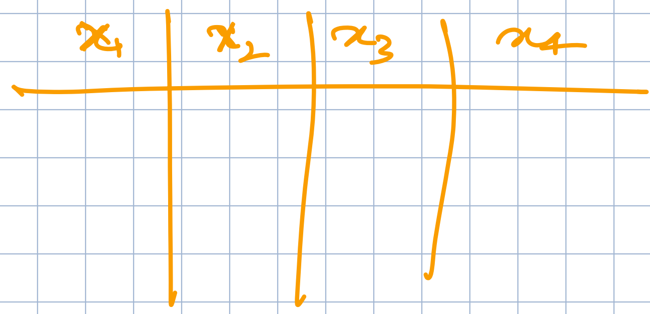
m examples

$$X = \begin{bmatrix} x_1^1 & x_2^1 & x_3^1 & \cdots & x_n^1 \\ x_1^2 & x_2^2 & x_3^2 & \cdots & x_n^2 \\ \vdots & & & & \\ x_1^m & x_2^m & x_3^m & \cdots & x_n^m \end{bmatrix}_{m \times n}$$

$x_j^i \rightarrow$ ith eg jth feature

Hypothesis

$$\hat{y} = h_\theta(x) = \theta_0 + \theta_1 x$$



$$h_\theta(x) = \theta_0 + \underbrace{\theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \theta_4 x_4 \ldots}_{\text{weight to every feature}}$$

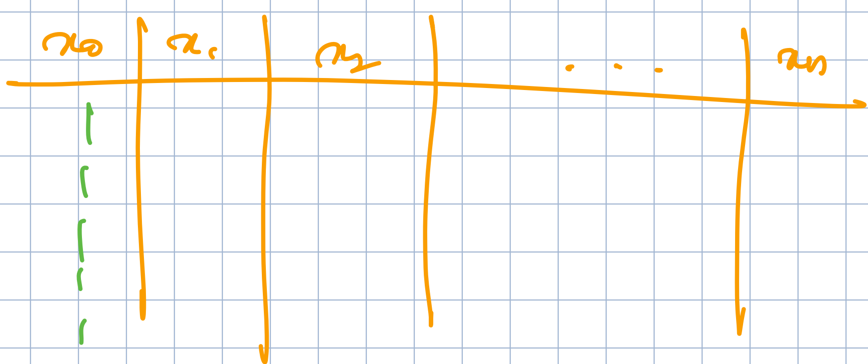
bias

$$h_\theta(x) = \theta_0 + \sum_{i=1}^{n} \theta_i x_i$$

$$h_\theta(x) = \theta_0 x_0 + \sum_{i=1}^{n} \theta_i x_i \qquad x_0 = 1$$

$$h_\theta(x) = \sum_{i=0}^{n} \theta_i x_i$$

n+1 features



$$\theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \vdots \\ \theta_n \end{bmatrix} \qquad x = \begin{bmatrix} 1 \\ x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

$\theta^T x$

$$[\theta_0 \ \ \theta_1 \ \ \theta_2 \ \cdots \ \theta_n] \begin{bmatrix} 1 \\ x_1 \\ x_2 \\ \vdots \\ \vdots \\ x_n \end{bmatrix}$$

$$\theta_0 + \theta_1 x_1 + \theta_2 x_2 \ \cdots \ \theta_n x_n$$

$$h_\theta(x) = \sum_{i=0}^{m} \theta_i x_i = \theta^T x$$

Loss fx^n:

$$J(\theta) = \frac{1}{m} \sum_{i=1}^{m} \left(\hat{y}^{(i)} - y^{(i)}\right)^2$$

$$J(\theta) = \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta(x^i) - y^i\right)^2$$

Update θ → Gradient Descent Rule

$$\theta = \theta - \eta \, \nabla_\theta J(\theta) \longrightarrow \frac{\partial J(\theta)}{\partial \theta}$$

GRADIENT?

$$x = x - \eta \frac{\partial f(x)}{\partial x}$$

$\nabla_\theta J(\theta)$?

$$\frac{\partial J(\theta)}{\partial \theta_j} = \frac{\partial}{\partial \theta_j} \frac{1}{m} \sum_{i=1}^{m} \left(\hat{y}^{(i)} - y^{(i)}\right)^2$$

$$= \frac{\partial}{\partial \theta_j} \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)}\right)^2$$

$$= \frac{1}{m}\sum_{i=1}^{m} \frac{\partial}{\partial \theta_j}\left(\theta_0 x_0^{(i)} + \theta_1 x_1^{(i)} + \cdots \theta_j x_j^{(i)} \cdots + \theta_n x_n^{(i)} - y^{(i)}\right)^2$$

$$= \frac{1}{m}\sum_{i=1}^{m} 2\Big(\underbrace{\theta_0 x_0^{(i)} + \theta_1 x_1^{(i)} \cdots\cdots \theta_n x_n^{(i)}}_{h_\theta(x^i)} - y^{(i)}\Big)\, x_j^{i}$$

$$\frac{\partial J(\theta)}{\partial \theta_j} = \frac{1}{m}\sum_{i=1}^{m} 2\left(h_\theta(x^i) - y^i\right) x_j^i$$

Final Gradient Update Rule

$$\theta_j = \theta_j - \eta \frac{1}{m}\sum_{i=1}^{m}\left(\hat{y}^{(i)} - y^{(i)}\right) x_j^i$$

$j = 0 \longrightarrow n$

$$\begin{array}{lll}
0 \to \frac{\partial J(\theta)}{\partial \theta_0} & & \\
1 \to \frac{\partial J(\theta)}{\partial \theta_1} & \longrightarrow & \begin{bmatrix} __ \\ __ \\ \vdots \\ __ \end{bmatrix} \ (n+1) \\
2 & & \\
3 & & \\
\vdots & & \\
n \to \frac{\partial J(\theta)}{\partial \theta_n} & &
\end{array}$$

$[\theta_0 \ \theta_1 \ \cdots\cdots \ \theta_n]$ Randomly

do

{

loss fxn

update θ

} while (convergence)