Simple Linear Regression

Supervised ML → Regression
Supervised ML → Classification

Dataset

↓ Independent features

Play hours	Study hours	Pass/Fail
7	5	Pass
7	2	Fail
3	8	Pass

Dependent feature ⇓ Pass/Fail ⇐ o/p

Pass, Fail, Pass } ⇒ Classification.

TRAIN ⇓

Play hours, Study hours ⟶ Model ⟶ P/F

⇓

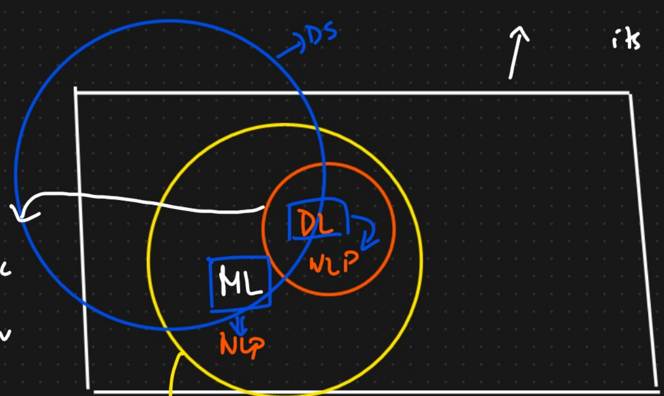
Acc↑↑

House price prediction

No. of Rooms	House size	Price
–	–	150K
–	–	155K
–	–	140K

⇓ Price: 150K, 155K, 140K } Continuous value ⇒ Regression problem Statement

AI Vs ML Vs DL Vs DS



Mimic Human Brain

Multi Layered Neural N/w

Statistical to perform explore, analyze, visualize and perform prediction.

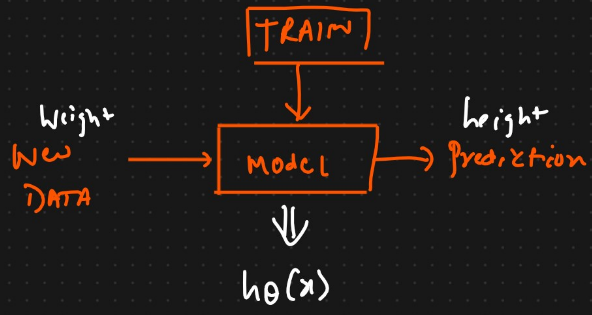
AI ⟶ Smart application that can perform its own task without any human intervention

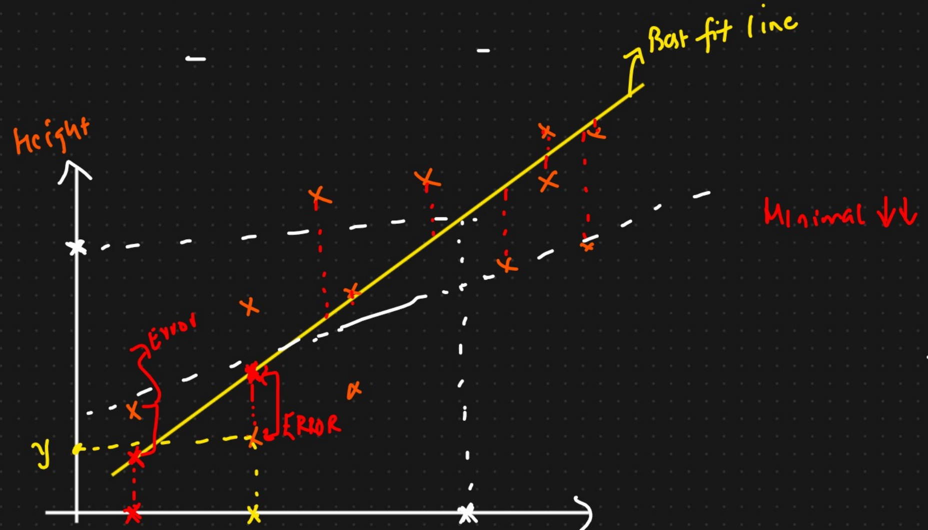
Eg: Self Driving Car, Chatgpt, Alexa, SIRI, Google Home

Eg:- Recommendation system
Weather prediction
Spam detection
Disease prediction

Simple Linear Regression

↓ Independent feature	Dependent feature ↓
Weight	Height (y) Actual value
74	170
80	180cm
75	175.5cm
–	–
–	–

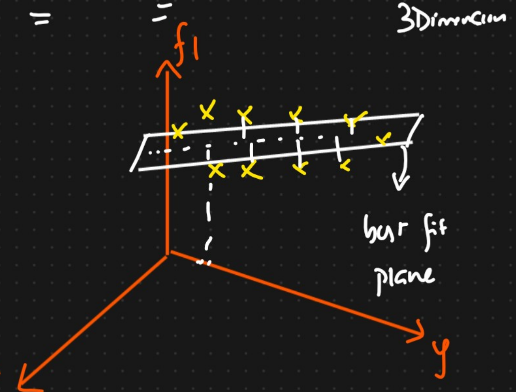


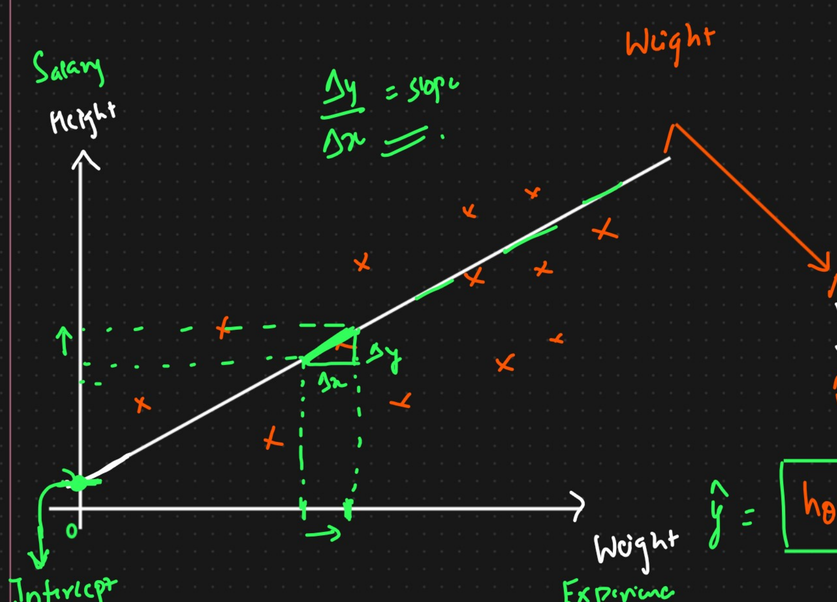


Multiple Linear Regression

f_1 f_2 y

3Dimension





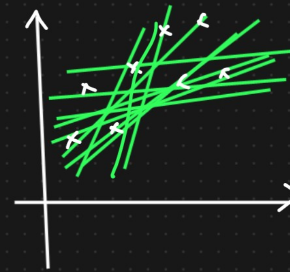
$\hat{y} = mx + c$

$\hat{y} = \beta_0 + \beta_1 x$

$\hat{y} = \boxed{h_\theta(x) = \theta_0 + \theta_1 x_1}$

intercept = 0

$$\left\{ \begin{array}{l} \theta_0 = \text{Intercept} \\ \theta_1 = \text{Slope or Coefficient} \end{array} \right\}$$



$\boxed{\theta_0 \ \& \ \theta_1}$

Cost function

$$J(\theta_0,\theta_1) = \frac{1}{n} \sum_{i=1}^{n} \Big(y_i - \underbrace{h_\theta(x)}_{\hat{y}}\Big)^2 \quad \text{[Mean Squared Error]}.$$

$$J(\theta_0,\theta_1) = \frac{1}{n} \sum_{i=1}^{n} \Big(\underset{\Downarrow}{y_i} - \underset{\Downarrow}{\hat{y}_i}\Big)^2$$

The Data you have Actual predicted.

n = no. of datapoints

$y_i \Rightarrow$ Actual value

$h_\theta(x)$ = predicted Value

Loss function $= (y_i - \hat{y}_i)^2$ {1 data point}.

Final Aim

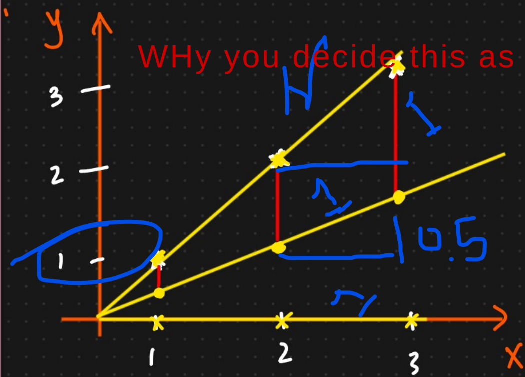
$$\underset{\theta_0,\theta_1}{\text{Minimize}}\ J(\theta_0,\theta_1) = \frac{1}{n} \sum_{i=1}^{n} (y_i - h_\theta(x))^2 \ \downarrow\downarrow\downarrow$$

Minimal

Optimization {Minimizing the Cost function}

DATASET

x	y
1	1
2	2
3	3



WHy you decide this as a Actual Value

$h_\theta(x) = \theta_0 + \theta_1 x_1$

$\boxed{\text{Consider } \theta_0 = 0}$

$\boxed{h_\theta(x) = \theta_1 x_1}$

Let $\theta_1 = 1$ | How we can take this Slope as 1 | Let $\theta_1 = 0.5$ | Let $\theta_1 = 0$

Just Try to Take 1 or 0.5 as A slope and Make me understand in this graph

$x_1 = 1$	$h_\theta(x) = 1(1) = 1$	$h_\theta(x) = 0.5$	$h_\theta(x) = 0$
$x_1 = 2$	$h_\theta(x) = 1(2) = 2$	$h_\theta(x) = 1.0$	$h_\theta(x) = 0$
$x_1 = 3$	$h_\theta(x) = 1(3) = 3$	$h_\theta(x) = 1.5$	$h_\theta(x) = 0$

This is Prediction Value

Cost function

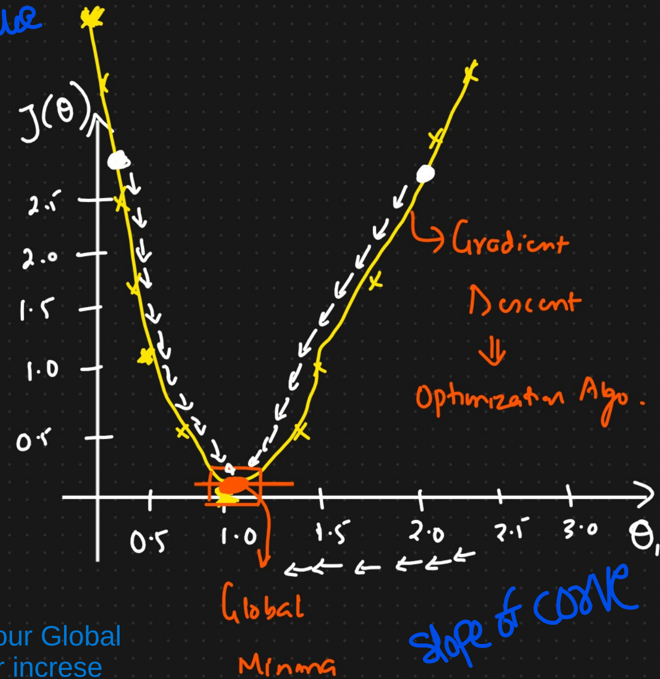
$$J(\theta_1) = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - h_\theta(x)_i\right)^2$$

$n = 3$

how krish sir is finding Actual value and predicted value

$$= \frac{1}{3}\left[(1-1)^2 + (2-2)^2 + (3-3)^2\right]$$

$$= 0$$



How can we decide To reach our Global Minima We need to decrese or increse Our Slope of the corve

Cost fn $\theta_1 = 0.5$

$$J(\theta_1) = \frac{1}{3}\left[(1-0.5)^2 + (2-1)^2 + (3-1.5)^2\right]$$

$$J(\theta_1) = 1.16$$

This is toe to reach our Global minima

Convergence Algorithm

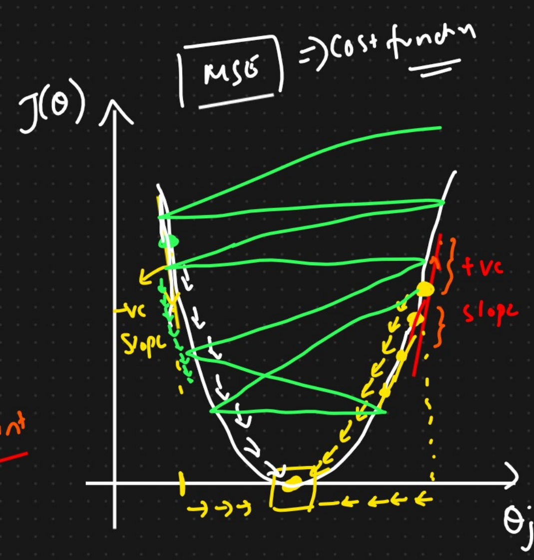
Learning Rate

Repeat until Convergence Global Minima

$\{$

$$\theta_j := \theta_j - \boxed{\alpha}\ \boxed{\frac{\partial}{\partial \theta_j} J(\theta_j)} \Rightarrow \text{Slope at a point}$$

$\}$



$\Uparrow$ huge

$\theta_j := \theta_j - \alpha(+ve\ value)$

$\theta_j := \theta_j - (+ve\ value)$

$\theta_{new} << \theta_{old}$

$\boxed{\alpha = 0.01}$ $\Leftarrow$ usually we take it

$\alpha \Rightarrow$ Learning Rate $\Rightarrow 1.00$

$\Downarrow$

Speed of Convergence.

$\theta_j := \theta_j - \alpha(-ve\ value)$

$= \theta_j + \alpha(+ve\ value)$

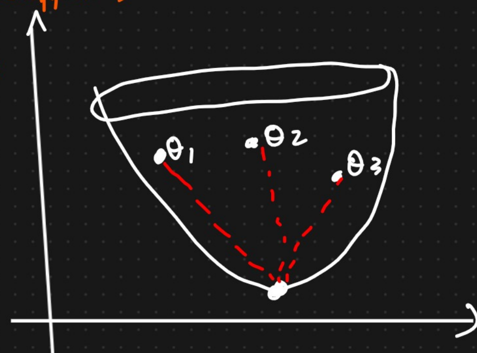
$\theta_{new} >> \theta_{old}$

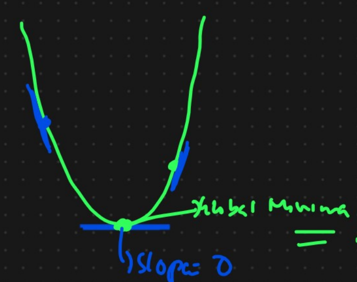
$f_1 \quad f_2 \quad f_3 \quad y$

$h_\theta(x) = \theta_0 + \theta_1 f_1 + \theta_2 f_2 + \theta_3 f_3$ {Multiple Linear Regression}

$\theta_1, \theta_2, \theta_3 \Rightarrow$ Coefficients

$\theta_0 \Rightarrow$ intercepts



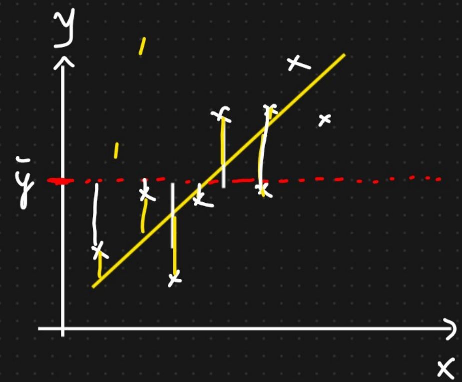


$\boxed{\theta_{new} = \theta_{old}}$

Performance Metrics to check Accuracy of our Model

① R squared ② Adjusted R Squared

① R squared



R squared = $1 - \dfrac{SS_{Res}}{SS_{Total}}$ {Best fit line} / {Average of y}

SS_{Res} = Sum of square Residuals or Errors

SS_{Total} = Sum of Squared Total

$$\boxed{R\ squared = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2}}$$

$\Rightarrow$ Small value / Big value $= 0.7$

$\Downarrow$

70% Accurate

$\boxed{f_1}$ $\boxed{f_2}$ $\boxed{f_3}$ f_4 f_5 $\boxed{O/P}$

R squared = 70%

75%

76%

Adjusted R squared

Then also R^2 is ↑

Not ↑ corelated

Not a think

e.g [illegible]

Size of house | No.of Rooms | Location | Gender → Price

That's why we calculate R square

$R^2 = 70\%$, $R^2 = 75\%$ $R^2 = 78\%$ $R^2 = 79\%$

Adjusted R square < R squared

$$\text{Adjusted R square} = 1 - \frac{(1-R^2)(N-1)}{N-p-1}$$

N = no.of data points

P = No.of independent predictors

$R^2 = 80\%$ $N = 11$ $p = 2$

$$\text{Adjusted R squared} = 1 - \frac{(1-0.8)(10)}{11-2-1} = 0.75 \Rightarrow 75\%$$

$p=2$ $R^2 = 80\%$ Adjusted $R^2 = 75\%$

$p=3$ $R^2 = 85\%$ Adjusted $R^2 = 78\%$

$p=4$ $R^2 = 86\%$ " " $= 76\%$

⇓

Feature is not important