Till now

→ Python
→ Data toolkit.
→ EDA
→ Statistics
→ introduction ML / feature Eg
→ Core ML.

Agenda

→ Linear Regression

- → intuition
- → Mathematical explanation
- → Implementation.

* Problem Statement ⇄ Data Understanding

⇅

Data Prep / FE

⇅

Model building

↓

model Evaluation

* ML

→ Supervised (y is given)

→ Unsupervised. (y is not given)

Supervised:

→ Regression → y is continuous

→ Classification → y is discrete.

Simple linear Regression (Supervise ML Algorithm)

History * Legendre in 1805 and
Gauss in 1809.

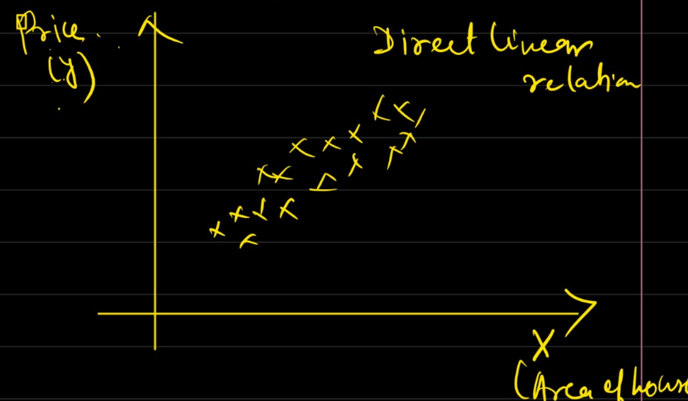
term regression → Francis Galton
* R.A. Fisher.

* Regression → To establish a relationship between the two variables or more than two variables.

* Linear → It establish a linear relationship.

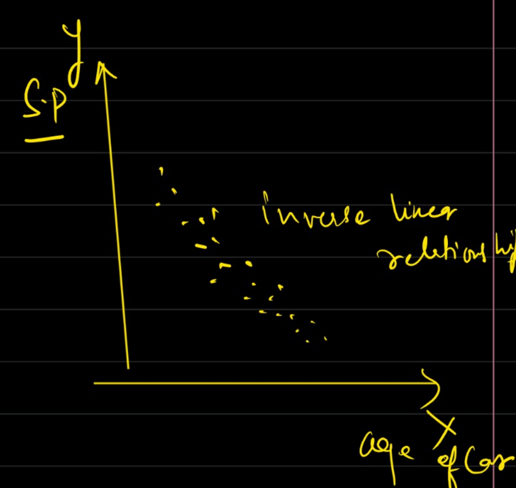
PS-1 → Price of house based on Area of house.

Area of house	y
1160	50
1200	60
1300	70
—	—
—	—
—	—
—	—
→ 1125	?



PS.2 Selling Price of Car wrt age of Car.

X (age of car)	y (SP of car)
10	3.1
9	4.1
7	5.2
—	—

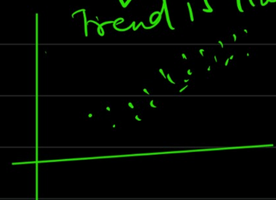


P-S-3 Age left vs Alcohol consumption

PS-4 BMI Vs Sugar level

Simple — X, Y (one X)

Linear — Trend is linear (+ve, -ve) Progression



Regression — relationship b/w X (one or more than one) and Y.

* Only one Independent (X) → Simple Linear Regression
Multiple " → Multiple Linear Regression

features, independent var, predictors → Area of house

Price of house (y) → target, outcomes, dependendent, tags, labels, response.

Area of house	Price of house (y)
1000	50
1100	60
1200	70
1110	75
1000	100
1250	80
1290	82
1111	111
2121	212
1331	33
1122	33

To Understand the mathematical relationship

→ ??

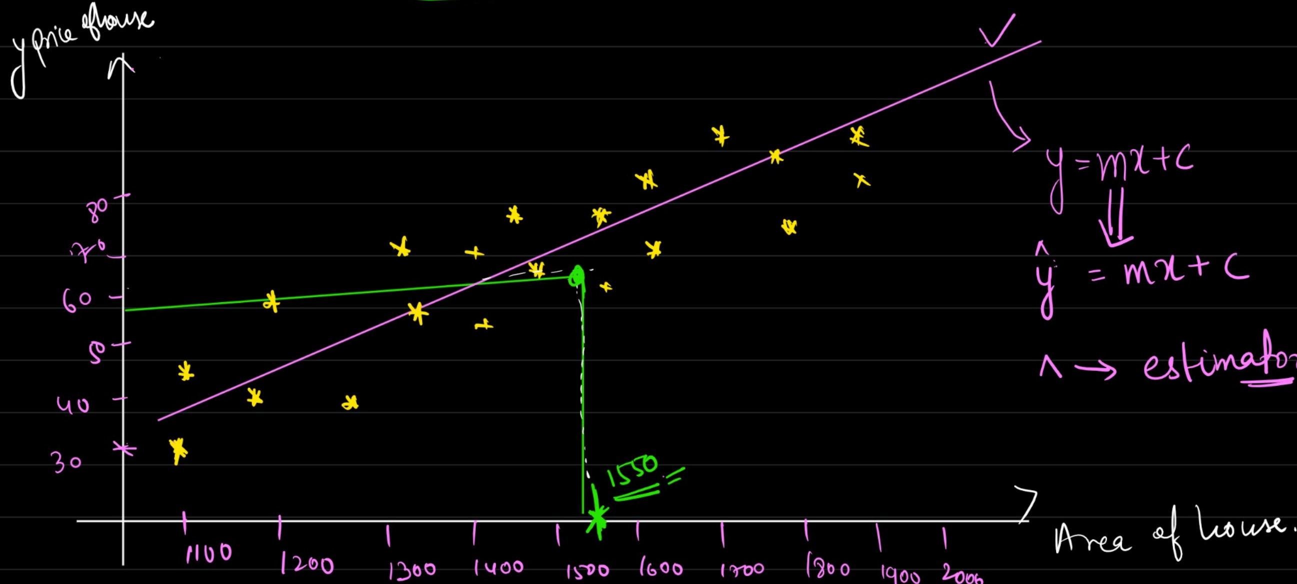
4 dp's → you understood the trend

10L → Understand the relationship !

1210 ??

⇒ Machine learning
⇓
⇒ Mathematical relationship

train (data)
↓
New Area → Model → Price of house.



$y = mx + c$
⇓
$\hat{y} = mx + c$

$\wedge \rightarrow$ estimator.

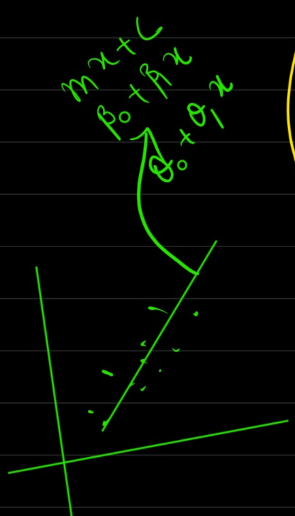
$y = mx + c$

$\hat{y} = mx + c$ (y_{pred})

m slope ⇒ coeff.

intercept

$y_{pred} \leftarrow mX + c$

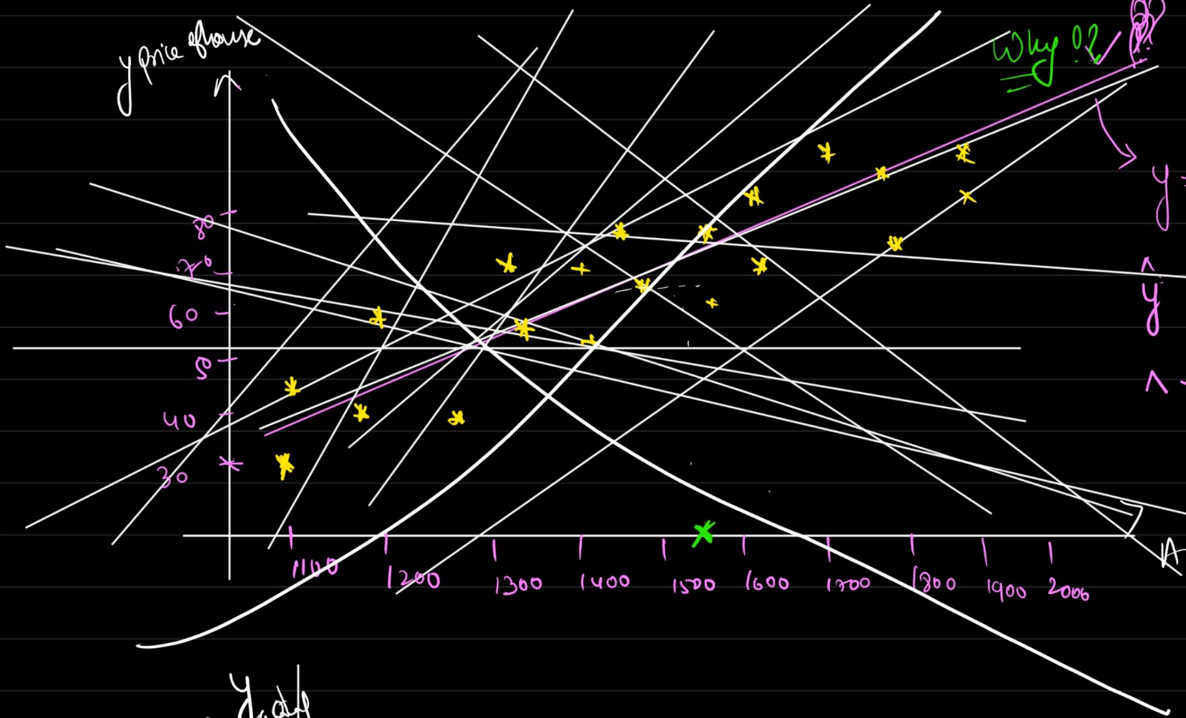


$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 X$ $\quad (\hat{\beta}_1 = m, \ \hat{\beta}_0 - c)$

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

$X \rightarrow dp$

$(m, c), (\beta_0, \beta_1), (\theta_1, \theta_0)$



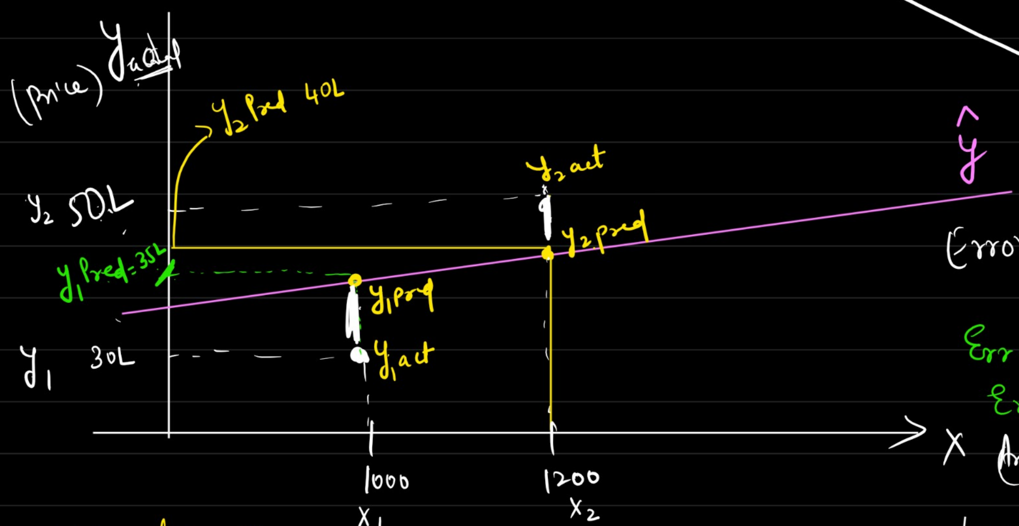
infinit

$y = mx + c$

$\Downarrow$

$\hat{y} = mx + c$

$\wedge \rightarrow$ estimator



Error | residual = $y_{act} - y_{pred}$

Err-1 1st dp – $y_{1act} - y_{pred}$

Err-2 2nd – $y_{2act} - y_{2pred}$

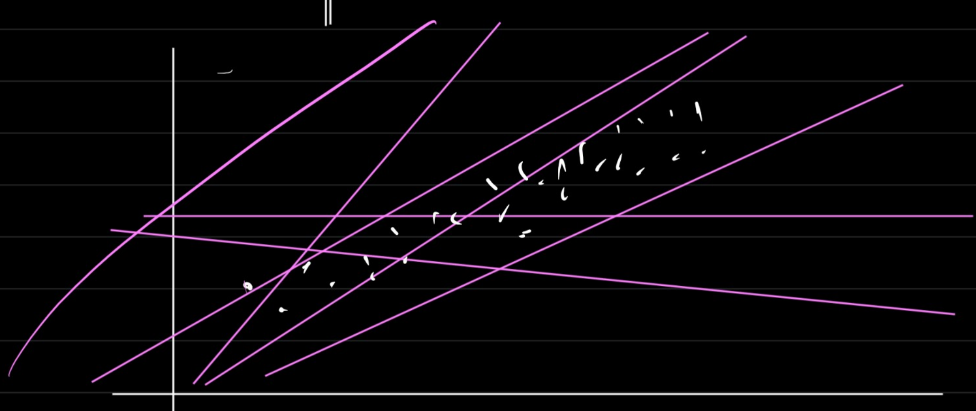
Why this line ??

Residual / Total error = Err1 + Err2

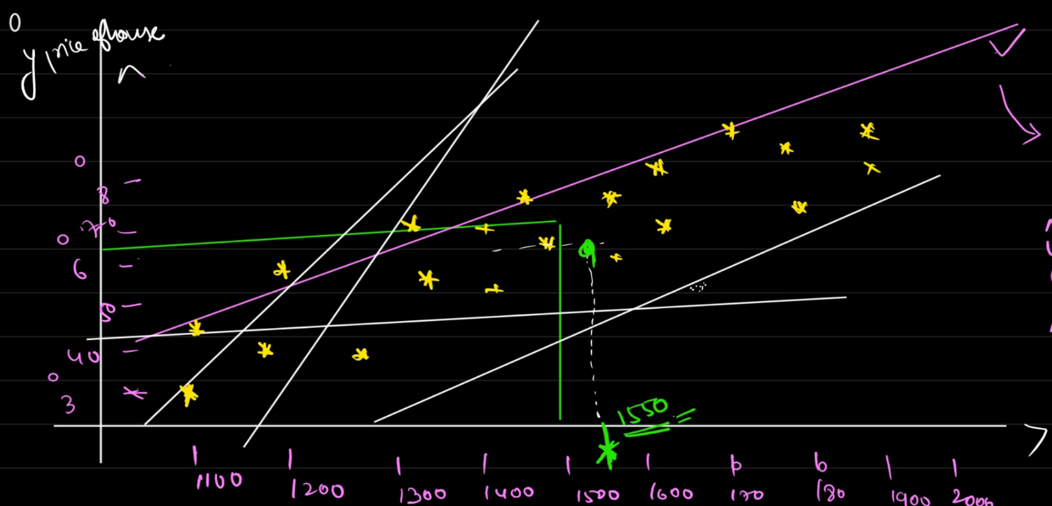
$G_1 \quad G_2 \quad G_3 \quad G_4 \quad G_5$ | $B_1 \quad B_2 \quad B_3 \quad B_4 \quad B_5$.

more red flags

less red flags.



or the line which gives the minimum error.

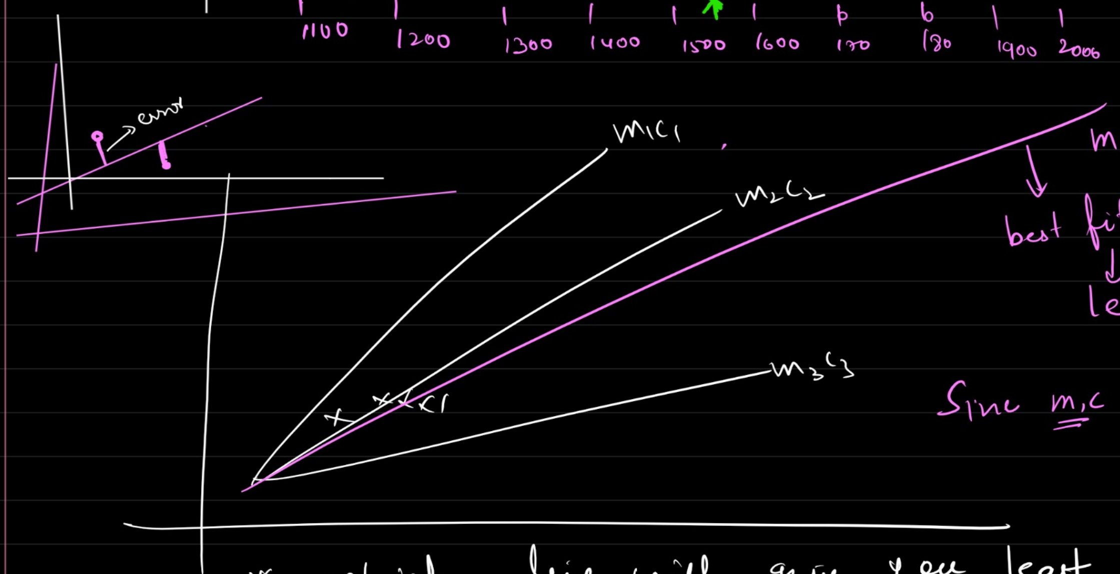


$y = mx + c$

$\Downarrow$

$\hat{y} = mx + c$

$\wedge \rightarrow$ estimator



$mc \rightarrow \hat{y} = mx + c$

↓

best fit line

↓

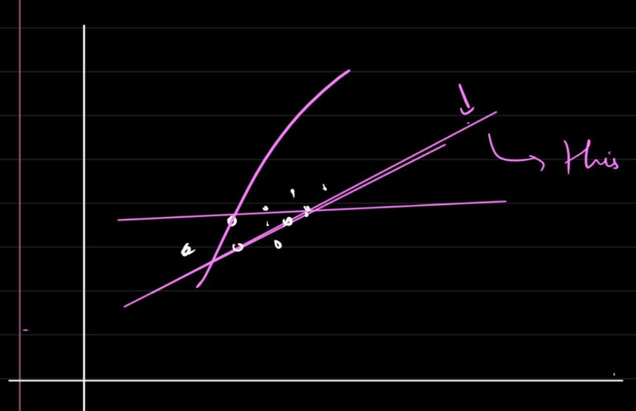
least error

Since $m, c \rightarrow \hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$

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

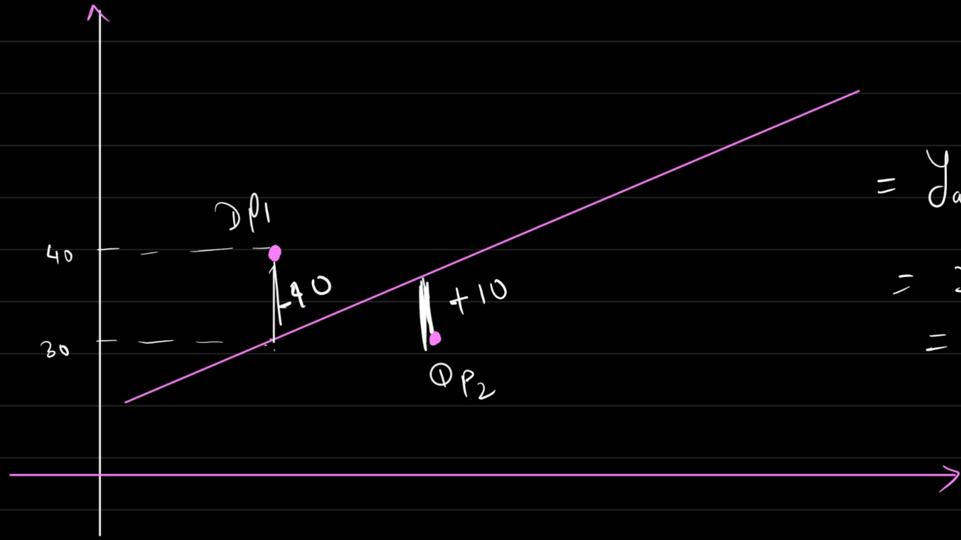
* which line will give you least error ??

→ The line which passes through most of the data point



↳ This line is the best fit line

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



$= y_{act} - y_{pred}$

$= 30 - 40$

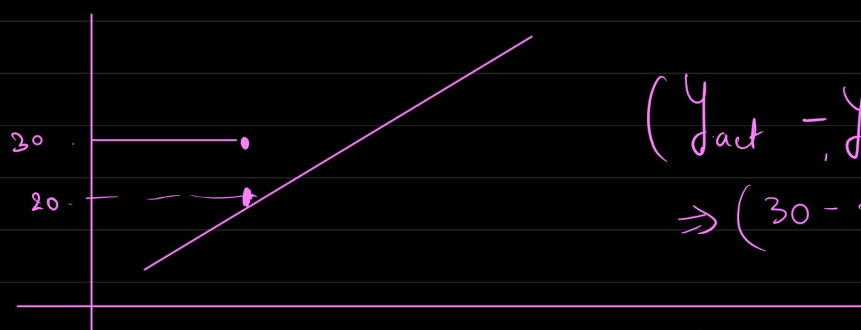
$= -10$

Total error

$-10 + 10 = 0$

$$\sum_{i=1}^{n} \left(y_{i\,act} - y_{i\,pred} \right)^2 \Rightarrow (-10)^2 + 10^2$$

100 + 100 = 200 units.



$$\left(y_{act} - y_{pred} \right)^2$$

$$\Rightarrow (30 - 20)^2 = 10^2 = 100$$

$$\underset{TE}{\sum_{i=1}^{n} \left(y_i - y_{pred_i} \right)^2} = (y_1 - y_{1\,pred})^2 + (y_2 - y_{2\,pred})^2 + \dots + (y_n - y_{pred\,n})^2$$

$$(y_{act} - y_{pred})_1 + (y_{act} - y_{pred})_2$$

$$-10 + 10 = 0$$

(negate each other)

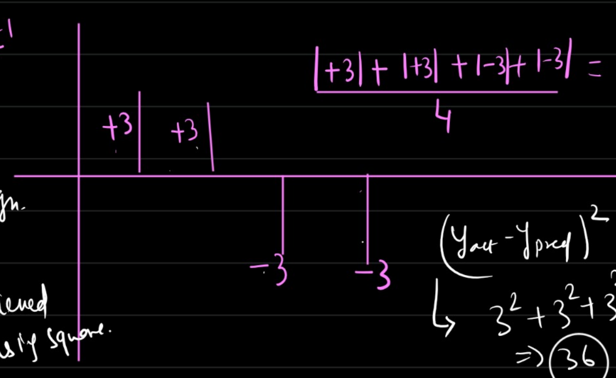
$$TE = \sum_{i=1}^{n} \left(y_i - y_{pred_i} \right)^2$$

$\rightarrow |y_i - y_{pred_i}|$

$\rightarrow |y_i - y_{pred_i}|^3 \rightarrow$ −ve sign.

$\rightarrow |y_i - y_{pred_i}|^4 \rightarrow$ achieved using square.

Case-1



$$\frac{|+3| + |+3| + |-3| + |-3|}{4} = 3 \quad 12$$

$(y_{act} - y_{pred})^2$

$\hookrightarrow 3^2 + 3^2 + 3^2 + 3^2$

$\Rightarrow 36$

Though error spread is different,
Total error is same.

Case

$\frac{|8|+|1|+|-2|+|-1|}{4} = 3$

12

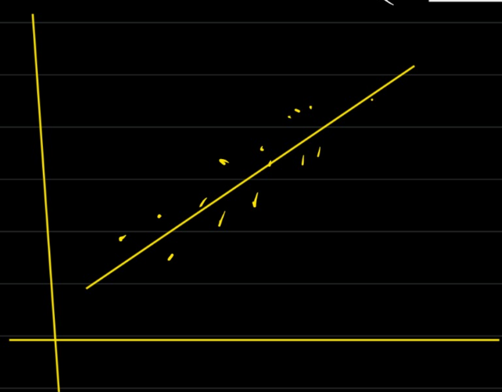
$(y_{act} - y_{pred})^2 = 70$

$(y_{act} - y_{pred})^2$

↳ magnify the error.

↓

It penalises more (more red flag)



* We want to have that combination of (m, c) | (θ_0, θ_1) | $\beta_0 \beta_1$ where the overall error is least and the line will be best line.

OLS $= \min \sum_{i=1}^{n} (y_i - y_{pred})^2$

↓

Ordinary Least Squared Error.

↳ This is the simplest | most initial formulae to find best line.

$$\min \underbrace{\sum_{i=1}^{n} \left(y_i^{act} - y_i^{pred}\right)^2}$$

minimise the sum
square of difference y_{act} and y_{pred} for all the datapoints.

$$= \min \sum_{i=1}^{n} \left(y_i - (\underset{\theta_0}{\beta_0} + \underset{\theta_1}{\beta_1} x_i)\right)^2$$

$y_{pred_i} = \beta_0 + \beta_1 x_i$

$\downarrow$

optimal $\boxed{\beta_0, \beta_1}$ $\rightarrow$ best fit line.

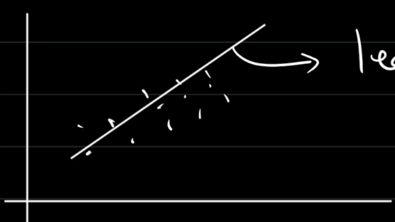
Which technique is to find best fit line in linear regression $\Rightarrow$ OLS.

$$\text{Mean Squared Error} = \min \frac{1}{n} \sum_{i=1}^{n} \left(y_i - (\beta_0 + \beta_1 X_i)\right)^2$$

$\Downarrow$

Cost Function

more red flag ↑ most cost ↑

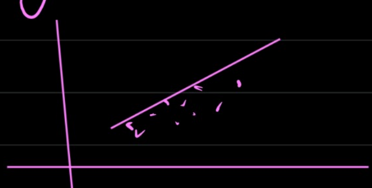


$\rightarrow$ least red flg

↳ loot your money $\Rightarrow$ it will cost you

High MSE (red flg) ↑ High Cost

* Cost Function $\rightarrow$ A function | metric | Expression that ensures to penalize you whenever you do something wrong. (more error)



Cost function

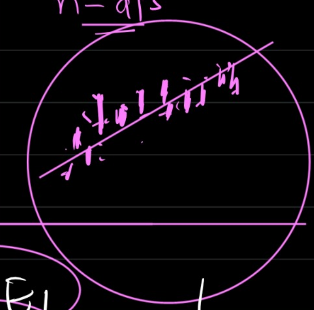
↳ Mean Square Error.

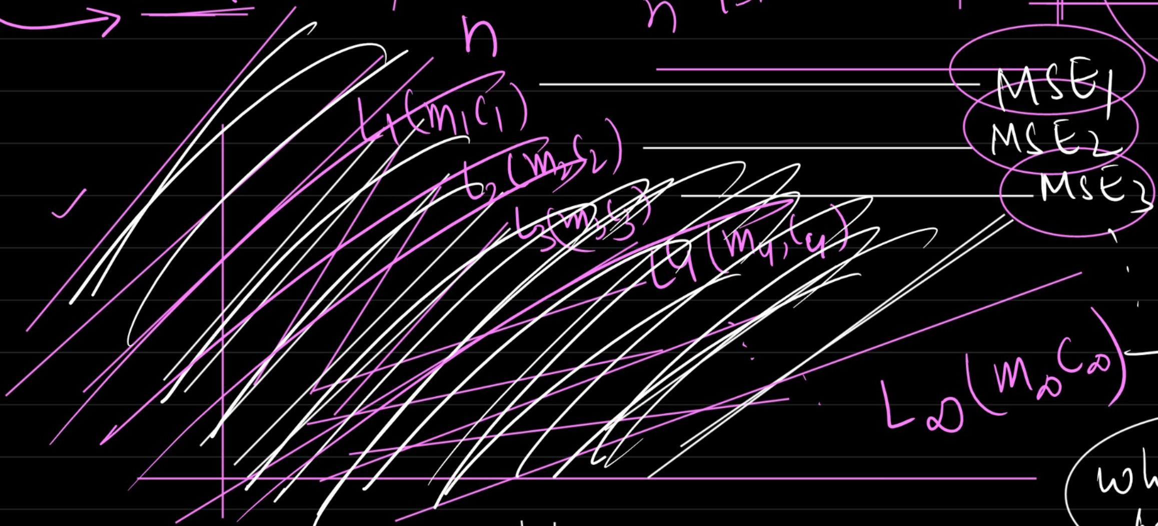
what is avg error at overall model level | for all the datapoints ?

OLS (SSE) $= \sum_{i=1}^{n} \left(y_{act} - (\beta_0 + \beta_1 x_i)\right)^2$
(technique)

MSE $= \dfrac{SSE/OLS}{n} = \dfrac{1}{n} \sum_{i=1}^{n} \left(y_{act} - (\beta_0 + \beta_1 x_i)\right)^2$

n-dps

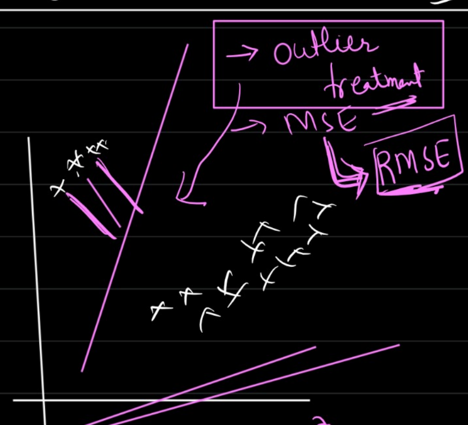




whichever has least MSE

Optimization Process

Gradient Descent



$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - y_{pred})^2}$

$9 \approx \sqrt{9} = 3$

$OLS \rightarrow SSE \sim \frac{SSE}{n} \rightarrow Avg(MSE) \sim \sqrt{MSE} \sim RMSE$

Rsquare