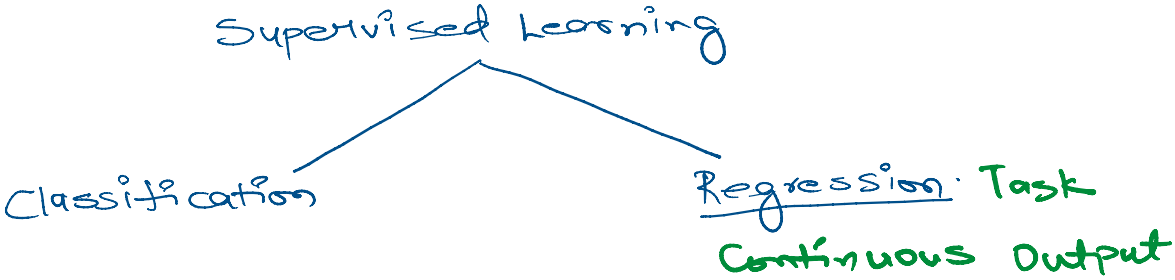
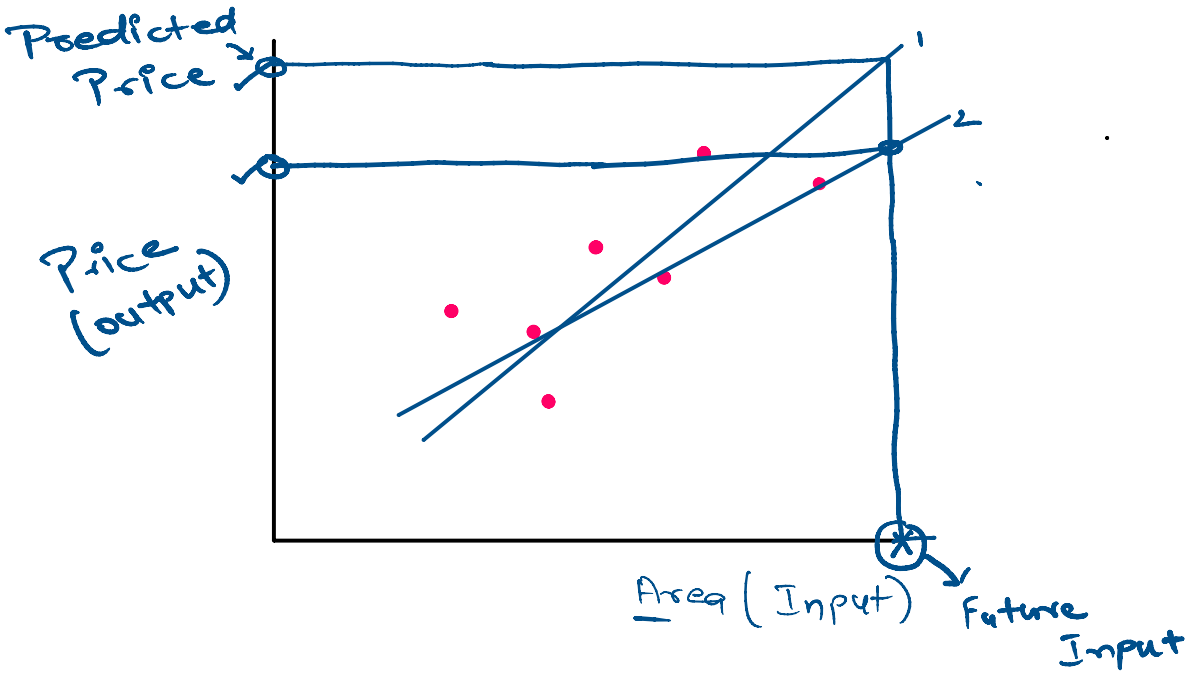
Linear Regression Algorithm

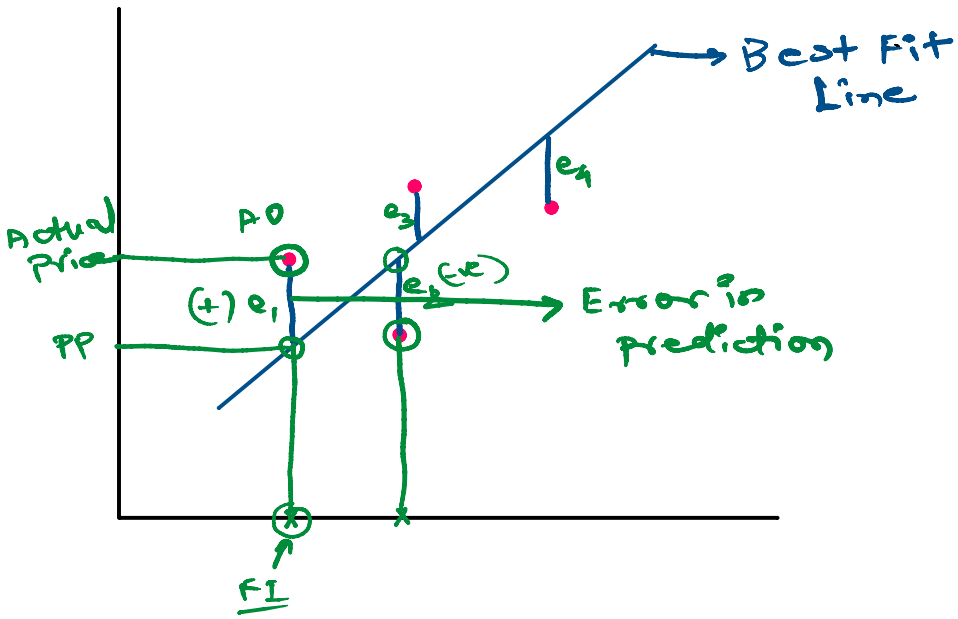
16 August 2024 07:03



Linear Regression



Deciding which is our Best Line going to be used for final prediction →



FI

Error (e) = Actual Price − Predicted Price

$= y - \hat{y}$

Total Error made by the line $= e_1 + e_2 + e_3 + e_4$ modulus

$|-10| = 10$

SSE (sum of squared errors) $= |e_1| + |e_2| + |e_3| + |e_4|$ ✗

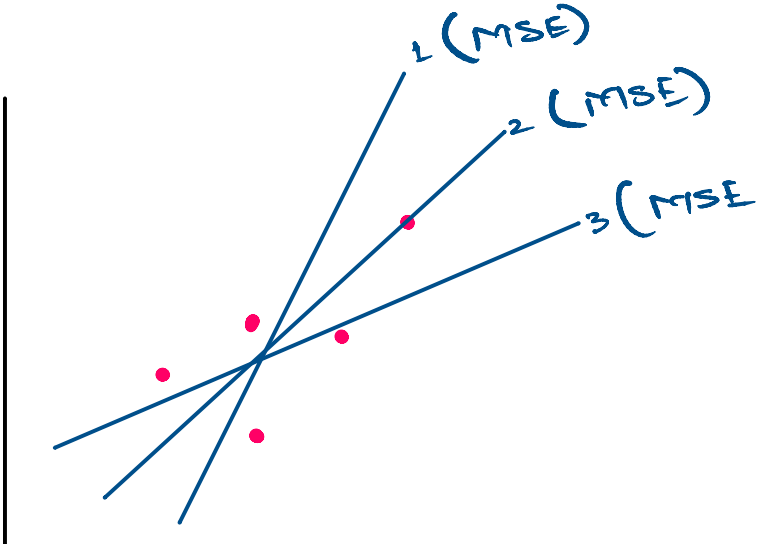
$= e_1^2 + e_2^2 + e_3^2 + e_4^2$

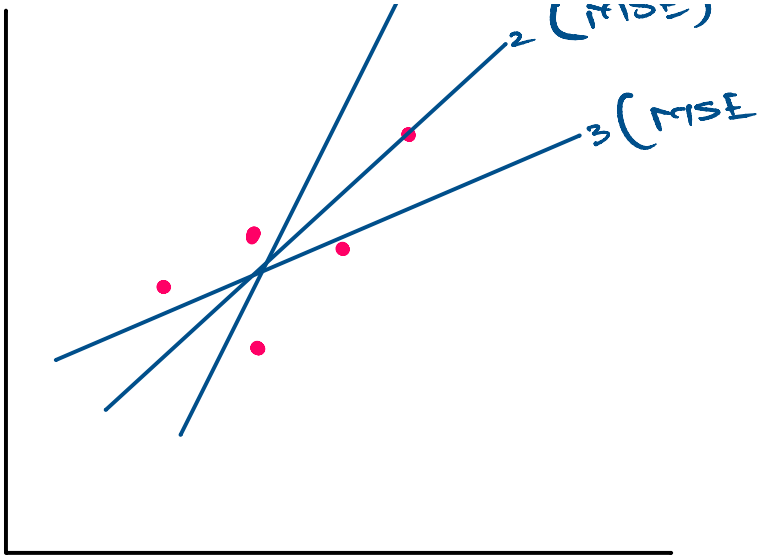
There will be explosion of value if we take square of all error terms specially when we have large amount of values) data.

To solve the above problem →

$MSE = \frac{SSE}{n}$ → n no of datapoints

(Mean Squared Error)

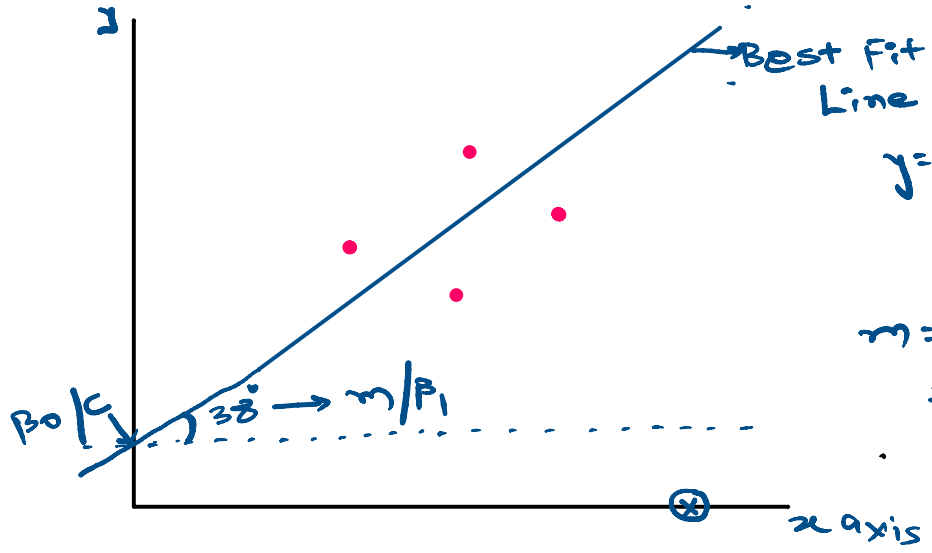




MSE_1 vs MSE_2 vs MSE_3

Take the minimum of these

Suppose MSE_2 is min, so line 2 will be considered as the 'Best Fit Line'.



B: pecta

$y = mx + c$

↓

$\hat{y} = B_1 x + B_0$

Predicted output → $\hat{y} = \beta_1 x + \beta_0$

$= 0.9x + 1.25$

FI

2 acres → $0.9x + 1.25$ → 3.05 crore

FI

$0.9 \times 2 + 1.25$

Predicted Price ($\hat{y}$)

① Simple Linear Regression (1 input & 1 output)

② Multiple Linear Regression

x_1 = # of rooms

x_2 = # of floors

x_3 = Area of House

x_4 = Age of the House

x_1 # of Rooms	x_2 # of floors	x_3 Area	x_4 Age	y Price

Never

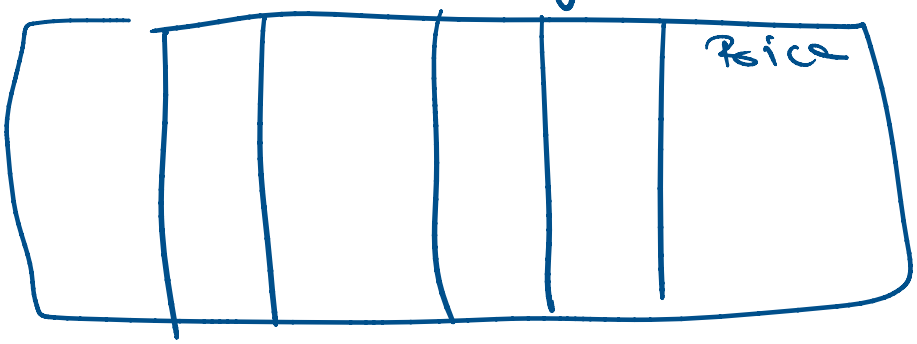
$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4$$

$$= 1.8 + 2.1 x_1 + 0.3 x_2 + 1.6 x_3 + (-1.71) x_4$$

Base Price

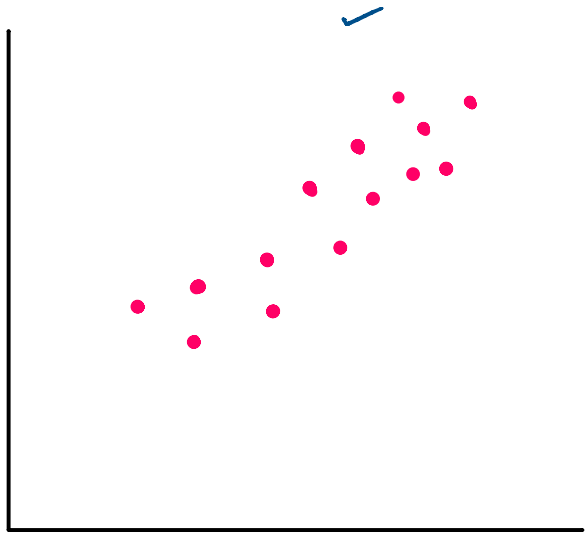
Bangalore

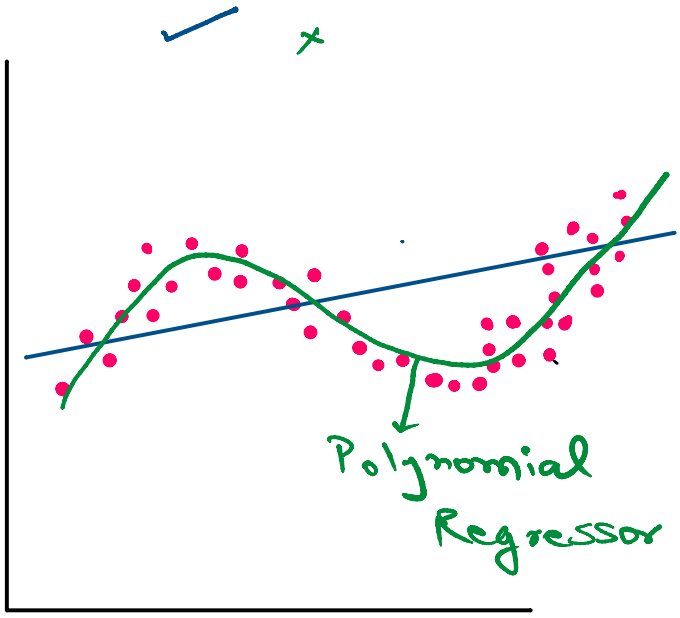
Houses of Bangalore



Assumptions →

1





2

$$e_1 + e_2 + e_3 + e_4 + \dots$$

↓

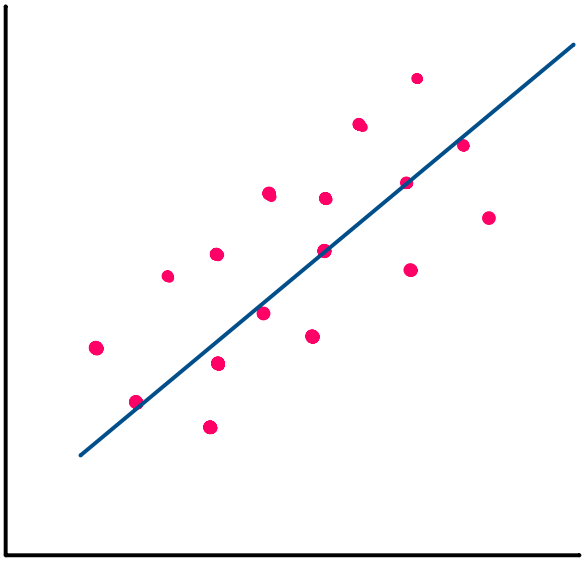
↓

zero or close to zero

↓

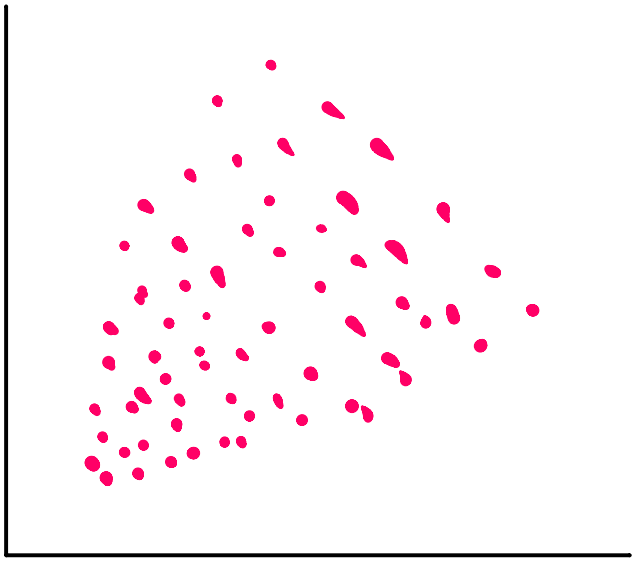
If this condition is satisfied, the data is linear and hence Lin-Reg can be applied.

(3)



Homoscedasticity

↑



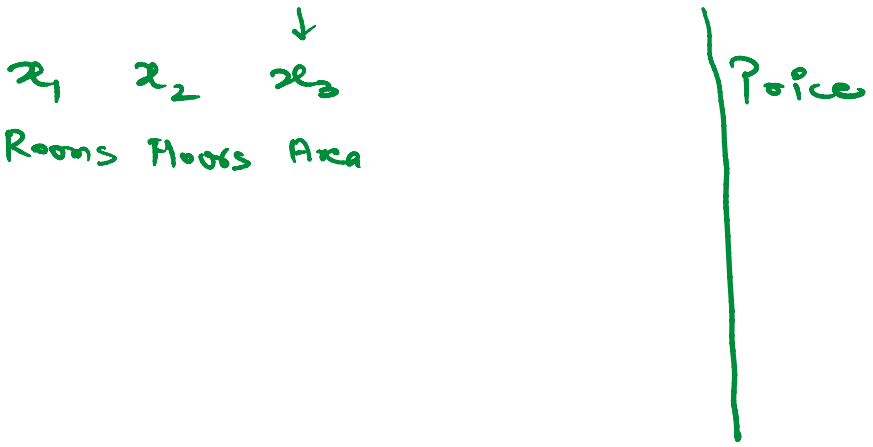
Heteroscedasticity

×

(4) Multicollinearity should not be present in the data.

↓

when input cols are highly correlated with each other, that is called as multicollinearity.



Area = Length × width

VIF (variance Inflation factor) score

VIF score > 5

↓

Delete the column