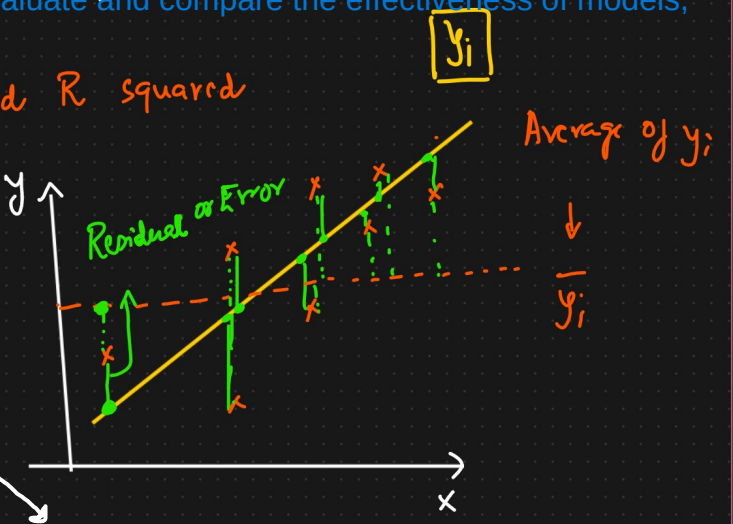
# Performance Metrics Used In Linear Regression

A Performance Matrix is a structured way to evaluate and compare the effectiveness of models, systems.

① R squared ② Adjusted R squared




$$R^2 = 1 - \frac{SS_{Res}}{SS_{TOTAL}} = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y}_i)^2}$$

($\hat{y}_i$: predicted, $y_i$: true)

$$= 1 - \frac{\text{Small number}}{\text{Bigger number}}$$

$$= 1 - \text{Smaller number}$$

$$\approx 1$$

$SS_{Res}$ = Residual Sum of Squares

$SS_{Total}$ = Total Sum of Squares

0.70 ⇒ 70%

0.85 ⇒ 85%

0.90 ⇒ 90%

{Overfitting, Underfitting}

1 ⇓ Accurate My Model Is?

R squared ↑↑↑

Size of the house ↑ Price ↑ → +ve correlation

## ② Adjusted R Squared

Dataset → (Price)

Gender | Size of the house | No. of bedrooms | Location | Price

No. of bedrooms ↑ Price ↑ → +ve correlation

R Squared = 75% ⇒ 0.75

This is the problem of R Squared

R squared => 80% => 0.80

R squared => 85% => 0.85

R squared => 87% => 0.87

Even though this(Location) feature are not at correlated with price. Still R^2 increases.

To prevent this, use adjusted R^2.

- Adjusted R square penalizes(deal with) with respect to every feature that are not correlated with the output feature.

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

- $P=2$ $R^2 = 90\%$ $R^2_{adjusted} = 86\%$
- $P=3$ $R^2 = 92\%$ $R^2_{adjusted} = 82\%$

N = No. of data points

P = No. of Independent features