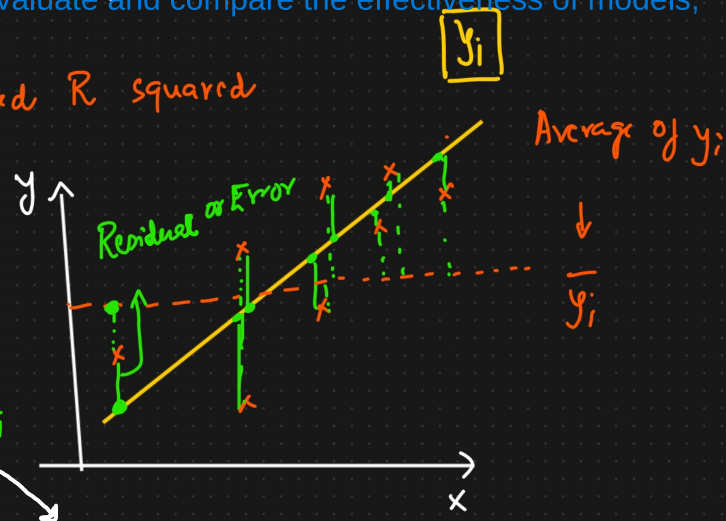
Performance Metrics Used In Linear Regression

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

① R squared ② Adjusted R squared



$$R\,squared = 1 - \frac{SS_{Res}}{SS_{TOTAL}}$$

$\hat{y}_i$

true

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

predicted

SS_{Res} = Residual Sum of Squares

SS_{Total} = Total Sum of Squares

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

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

⇓

Accurate My Model Is?

≈ 1

0.70 ⇒ 70%

0.85 ⇒ 85%

{Overfitting, Underfitting} 0.90 ⇒ 90%

R squared ↑↑↑

Size of the house ↑ Price ↑

$\boxed{\text{+ve Correlation}}$

② Adjusted R squared

Dataset → (Price)

No. of bedroom ↑ Price ↑

↓

+ve Correlation

Gender	Size of the house	No. of bedrooms	Location	Price

R squared = 75% ⇒ 0.75

This is the problem of R squared

R squared $\Rightarrow$ 80% $\Rightarrow$ 0.80

R squared $\Rightarrow$ 85% $\Rightarrow$ 0.85

R squared $\Rightarrow$ 87% $\Rightarrow$ 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.

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

$p=2$ $R^2 = 90\%$ R^2 adjusted $= 86\%$

$p=3$ $R^2 = 92\%$ R^2 adjusted $= 82\%$

- R-square will be always greater than adjusted r-square.

N = No. of data points

P = No. of Independent features