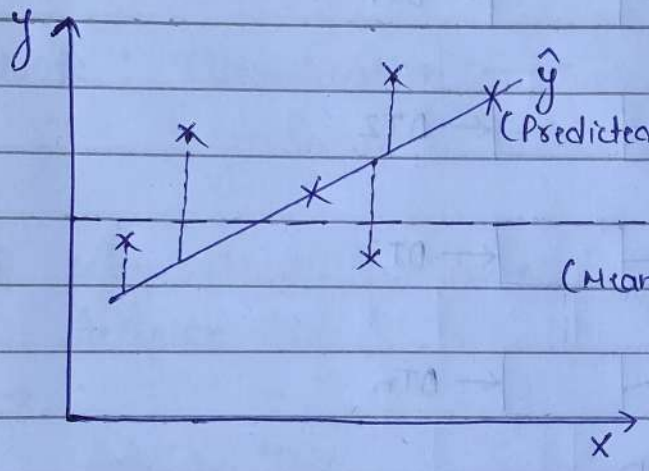
R Squared and Adjusted R Squared

* **R Squared** : ⟶ Ranges between 0 and 1.

$$R^2 = 1 - \frac{RSS}{TSS}$$

⟶ RSS : Residual Sum of Square
TSS : Total Sum of Square.



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

Note : • $R^2 = 0.20$:- 20% of the variance in y is predictable from $x(x_1, x_2, \cdots x_n)$.

• $R^2 = 0.98$:- 98% of the variance in y is predictable from $x(x_1, x_2, \cdots x_n)$.

⟶ Best fit line : {lower error value}

∴ $1 - \frac{\text{Small value}}{\text{Big value}}$ } ⟶ Small value.

↑ 1 − Small value ⟶ value closer to 1 ≈ 0.90 - 0.99

* Adjusted R Squared :

Dataset

X (Size, # of bedrooms) | y (Price) → R^2 = 76% {Let say}

Size, # of bedrooms, location (↑ {Important Feature}), Price → R^2 = 80% ↑↑

Gender (↑ {Not Important Feature}), Size, # of bedrooms, location, Price → R^2 = 81% ↑ (↓ Increased but very less)

→ R^2 value is still increased by adding an irrelevant column, it should not be the case. Therefore, to penalize this, there is a mechanism called adjusted R^2.

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

- N : # of datapoints
- P : # of independent features
- R^2 : R^2 score

- Relevant Features : R^2 → 0.75, Adjusted R^2 → 0.73
  - ↓ Adding more relevant feature
  - R^2 → 80% ↑, Adjusted R^2 → 76% ↑
    - ↓ Adding irrelevant feature
    - R^2 → 81% ↑, Adjusted R^2 → 70% ↓