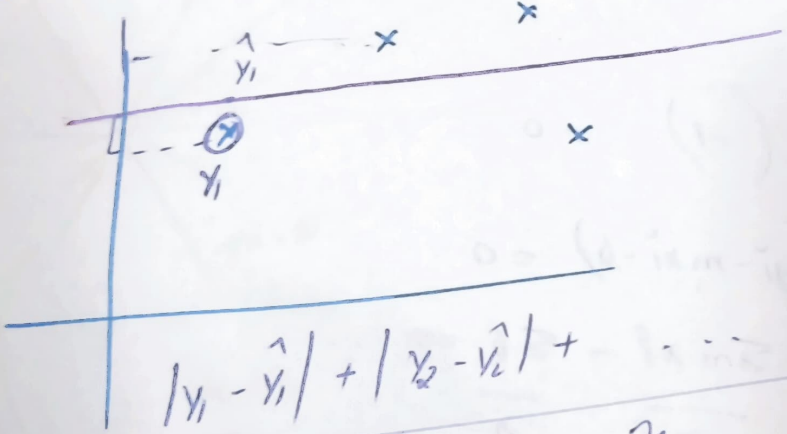
Regression Metrics (MSE, MAE, RMSE) | R2 Score and Adjusted R2

(1) MAE (Mean Absolute Error)



$$\frac{|y_1 - \hat{y}_1| + |y_2 - \hat{y}_2| + \dots + |y_n - \hat{y}_n|}{n}$$

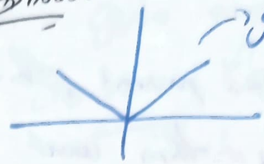
$$mae = \sum_{i=1}^{n} \frac{|y_i - \hat{y}_i|}{n}$$

it gives a number

mae = unit is same as Output Column

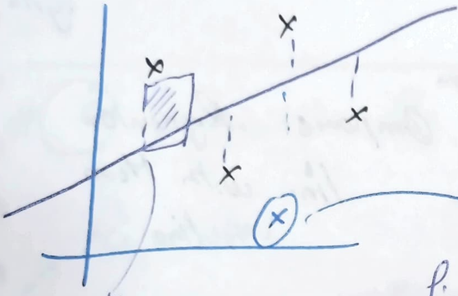
* Advantage
1> Same unit
2> Robust to Outliers

* Disadv



graph not Differentiable

② MSE (Mean Squared Error)



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

$(y_i - \hat{y}_i)^2$ → inko judega fir minimise karna Hai

outlier ko penalise

Advant
- can be used as loss function as Differentiable

Disadvant
- y = unit, mse = (unit)²
- Not Robust to outlier

y = lpa
MSE = (lpa)²

* if kam Outlier → MSE
* jyoda Outl. → MAE.

③ RMSE = $\sqrt{mse}$

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

→ mostly in Deep Learning

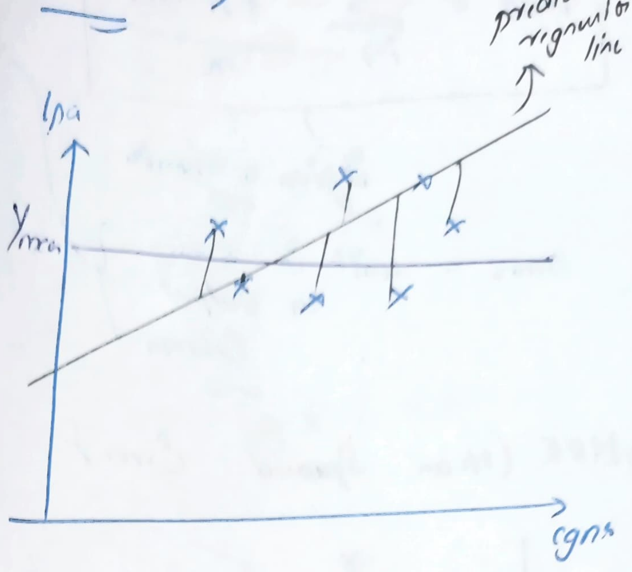
Benefit
- y - lpa
- our RMSE - lpa } same unit

disadv
→ Not Robust to Outliers

lpa | package (lpa)

(independent of Context)

if no cgpa mark, then to predict then worst case scenario is mean package.



R2 Score ↳ Compare Regression line with Mean line

Coeff of Determination ("goodness of fit")

$$R_2 = 1 - \frac{SS_R}{SS_M}$$

SS_R = Sum of Square error (regression line)

SS_M = ———— (Mean line)

$$R_2 = 1 - \frac{\left(\sum_{i=1}^{n} (y_i - \hat{y}_i)^2\right)_{Regr}}{\left(\sum_{i=1}^{n} (y_i - \hat{y}_i)^2\right)_{mean}}$$

$R_2 = 0$ Mean → $\frac{SS_R}{SS_M} = 1$ } dono same Galti kar rahe hai

$R_2 = 1$ → $\frac{SS_R}{SS_M} = 0$ ⇒ $SS_R = 0$ } mean perfect line

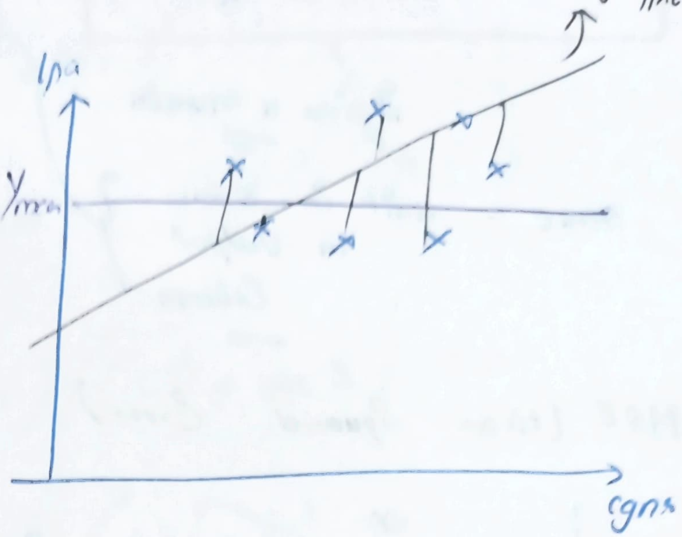
* $R_2 \to 1$

$R_2 < 0$ → $\frac{SS_R}{SS_M} > 1$ → Regr line mean wali line se jyada galti kar raha hai

(4) R_2 Score (independent of Context)

cgpa | package (lpa)

if no cgpa present, then to predict then worst case scenario is mean package.



R_2 Score — Coeff of Determination ("goodness of fit")

↳ Compare Regression line with Mean line

$$R_2 = 1 - \frac{SS_R}{SS_M}$$

SS_R = Sum of Square error (regression line)

SS_M = ———— (Mean line)

$$R_2 = 1 - \frac{\left(\sum_{i=1}^{n} (y_i - \hat{y}_i)^2\right)_{Regn}}{\left(\sum_{i=1}^{n} (y_i - \hat{y}_i)^2\right)_{mean}}$$

($R_2 = 0$) Mean $\frac{SS_R}{SS_M} = 1$ } dono same Galti kar rahe hai

($R_2 = 1$) $\frac{SS_R}{SS_M} = 0$ =1 $SS_R = 0$ } mean perfect line

* $R_2 \to 1$

($R_2 < 0$) ($\frac{SS_R}{SS_M} > 1$) → Regn line mean wali line se jyoda galti kar raha hai

$R_2 \rightarrow 0.80$ ~ [cgpa] explain 80% of Variation in lpa Column

cgpa | lpa

cgpa / iq / lpa — (lpa) → variation 80% cgpa and iq mil ke de rahi hai.

R_2 Score
↳ "This amount of variations in output col explained by input column)."

---

(5) Adjusted R_2 Score.

R_2 Score — cgpa / lpa

if add man input-col in them R_2 Score Badhta hai
Kyuki jyada input-colum increase R_2 Score
} problem with R_2

* if irrelevant Colum added (1%, temp)
↳ Vo badhega ya utna hi rahega

temp / cgpa / lpa

** Adjusted R_2

$$R^2_{adj} = 1 - \left[\frac{(1-R_2)(n-1)}{(n-1-k)} \right]$$

R_2 score
n → no. of rows
k = no. of independent / input Columns

temp colum
k (↑)
R^2_{adj} (↓) $R^2 \simeq$ Same
[Adding irrelevant Column]

iq, Column added
k (↑)
R_2 (↑) Relevant Column
R^2_{adj} (↑)

[Multiple linear Regression]