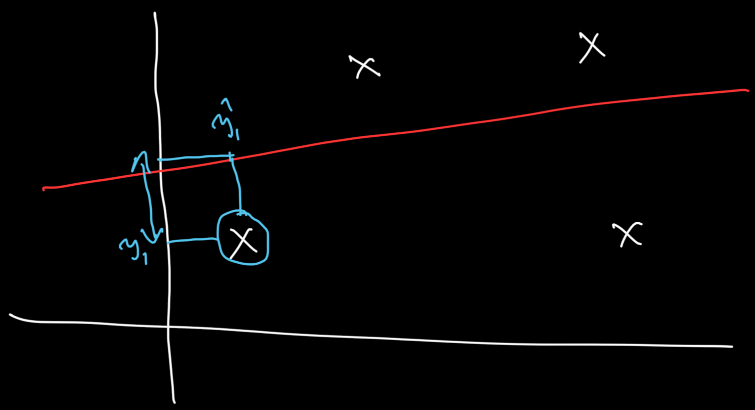
Regression Metrics | MSE, MAE & RMSE | R2 Score & Adjusted R2 Score

Regression Metrics :

1) MAE
2) MSE
3) RMSE
4) R2 Score
5) Adjusted R2 Score

1) MAE (Mean Absolute Error)



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

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

Advantage:

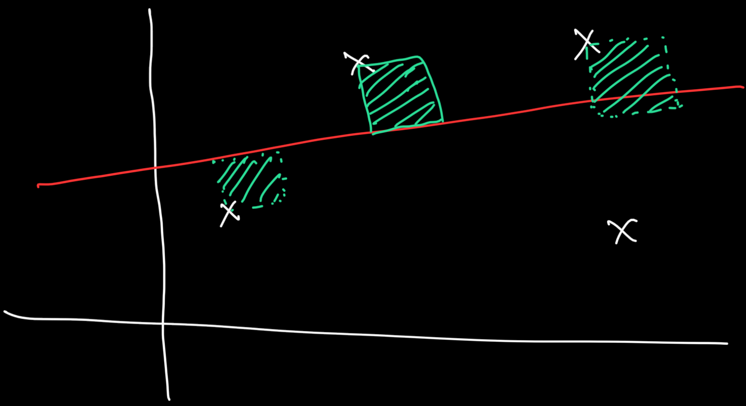
1) Same unit as y (target) [loss in terms of y]
2) Robust to outliers (Handle outliers)

Disadvantage:

1) Modulas Graph is not differentiable at origin
It will not able to converge while using optimization technique like Gradient descent.

2) Mean Squared Error (MSE)

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



Advantage

(i) Square f^n is differentiable at each point. So can be used as a loss function.

Disadvantage

(i) different unit than y (target)
$y \rightarrow dPa$ then $MSE \rightarrow (dPa)^2$

(ii) Penalize outliers (Error increases after square)
Subject to a penalty or punishment
↳ Not Robust to outliers

3) Root Mean Squared Error (RMSE)

$$RMSE = \sqrt{MSE}$$

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

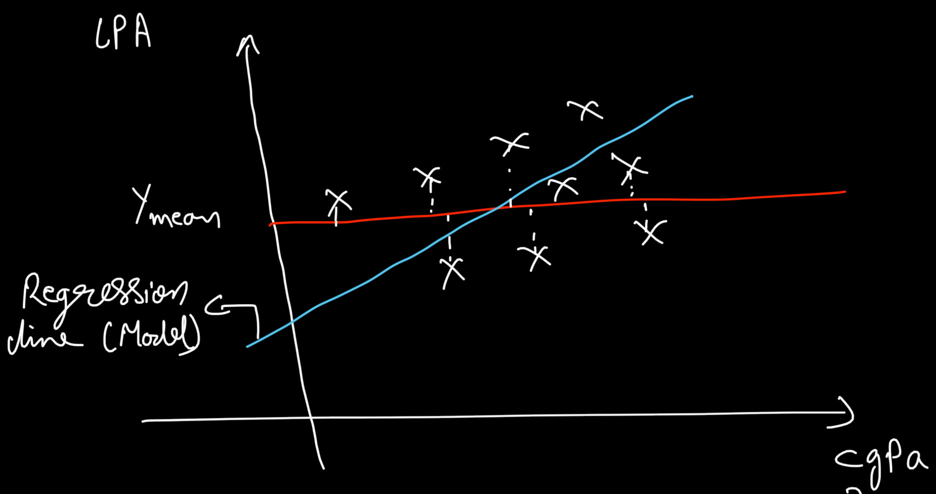
Advantage : Same unit

Disadvantage : not robust to outliers

R2 Score (Coefficient of determination / Goodness of fit)

cgpa	Package (LPA)

R2 Score tells us how good our model is performing (how good is regression line than average line (worst condition))



$$R2 = 1 - \frac{SS_R}{SS_m}$$

SS_R : Sum of Squared error in regression line.

SS_m : Sum of Squared error in mean line.

$$R2 = 1 - \frac{\left[\sum_{i=1}^{n}(y_i - \hat{y}_i)^2\right]_{Reg}}{\left[\sum_{i=1}^{n}(y_i - \hat{y}_i)^2\right]_{m}}$$

$\boxed{\text{When R2 Score} = 0 \text{ then } SS_R = SS_m}$ → worst

When R2 Score = 1 then $SS_R = 0$ (Regression is not making any error)

→ Best (not Possible in real life example)

0 ← R2 Score → 1

Bad ... Good

(The close R2 Score to 1, the best model is)

When R2 Score become -ve ?

If R2 Score is -ve then $\boxed{SS_R > SS_m}$

It means regression model is performing bad than Mean line.

One more interpretation of R2 Score

cgpa | lpa — Suppose R2 Score is 0.80 then we can say that cgpa column able to explain 80% of variance in lpa column.

cgpa | iq | lpa

Explain 80% of variance in lpa column

(20% we don't know, it can be any reason or factor)

80% explaination is coming out from cgpa and iq

R2 Score : This amount of variance in the output column being explained by the input columns.

Problem with R2 Score :

As number of input features increases, R2 Score also increases or remain same, even with irrelevant features.

Adjusted R2 Score

$$R^2_{adj} = 1 - \left[\frac{(1-R2)(n-1)}{(n-1-K)} \right]$$

$R2 \rightarrow$ R2 Score

$n \rightarrow$ no of rows

$K \rightarrow$ total no of independent columns

Scenario

→ Suppose one irrelevant feature added and N^r remain constant but D^r decreases then R^2_{adj} decreases.

$\underbrace{(1-R^2)}_{\text{constant}} \underbrace{(n-1)}_{\text{constant}}$

$\underbrace{(n-1-\underbrace{K}_{\text{increases}})}_{\text{decreases}}$

→ Suppose one relevant column added And

$$\frac{\overset{\text{Increase}}{\underbrace{(1-R^2)}_{\text{decrease}}}\ (n-1) - \text{constant}}{\underbrace{(n-1-K)}_{\text{decrease}}}$$

then [1 − ↓] R2 adj increases