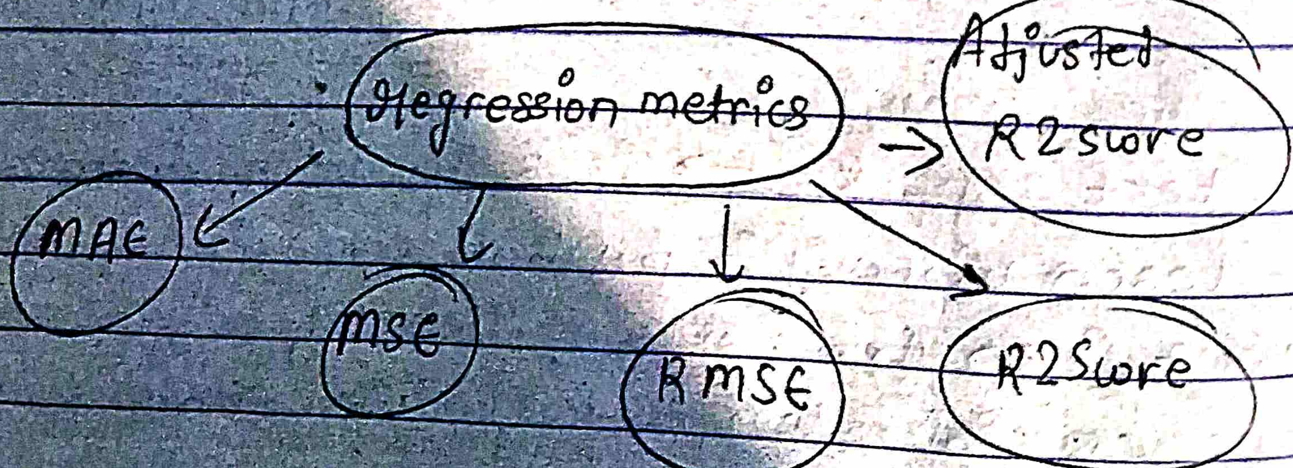
→ Cost Function : $y = a_0 + a_1 x + \epsilon$

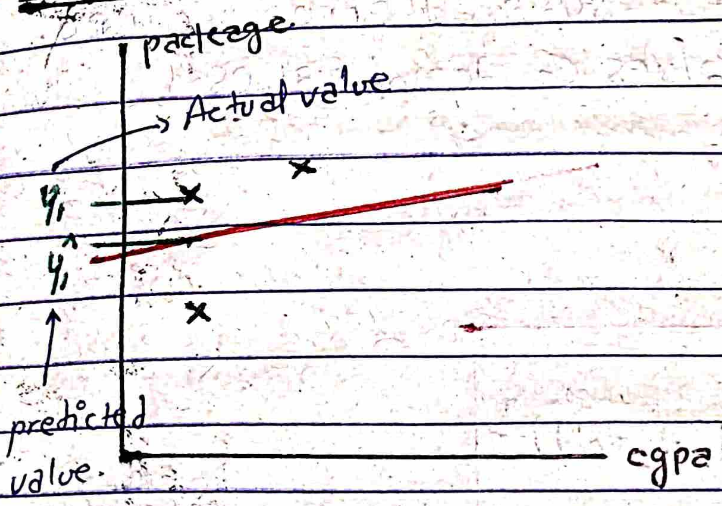
The different values for weights or coefficients of lines (a_0, a_1) gives the different lines of regression and the Cost func. is used to estimate the value of the coefficient for best-line-fit.

Cost func. optimizes the regression coefficient or weights. It measures how a linear regression model is performing.

We can use the cost function to find the accuracy of the Mapping function which maps the input variable to the output variable. The mapping func. is also known as Hypothesis function.



Mean Absolute Error (MAE):



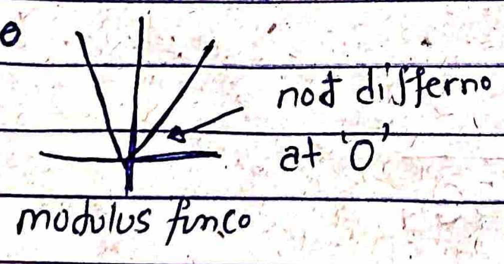
absolute → as it can be anyside up/down

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

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

Cost function
should be less
less C.F → Less Cost

- MAE indicates ~~Const~~ cost func i.e. we have to minimize
- same unit → i.e. dependent var → package
  independent var → CGPA
  → MAE in package unit (in LPA)
- Robust to Outliers
-

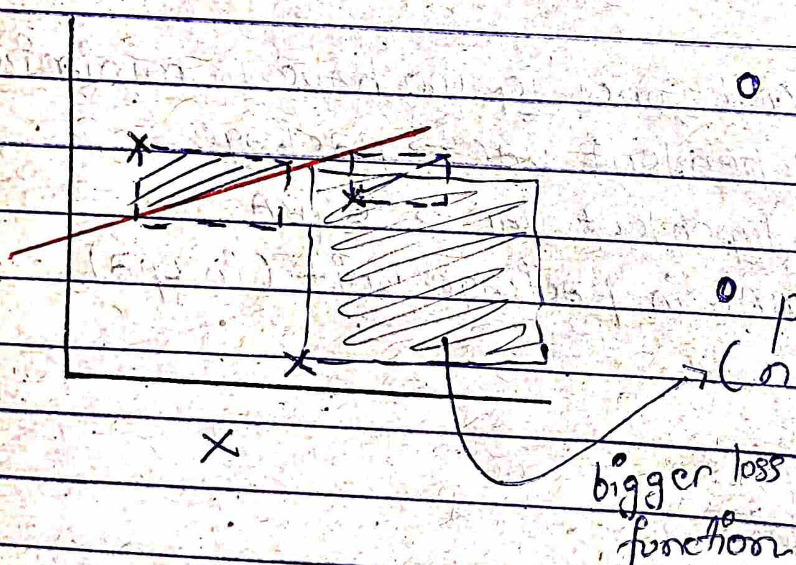


Mean Squared Error (MSE) : Is the Average of squared error occur between the predicted values & Actual values.

$$MSE : \frac{\sum_{i=1}^{n} (y_i - (a_1 x_i + a_0))^2}{n}$$

or

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



- differentiable func
- panelize outliers (not robust to outliers)

Root mean squared MSE (RMSE) :

$$\sqrt{MSE}$$

or

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

gives the same unit

Model Performance: The goodness of fit determines how the line of regression fits the set of observations. The process of finding the best model out of various models is called optimization. It can be achieved by below method:

R-squared Method: R-squared is a statistical method that determines the goodness of fit

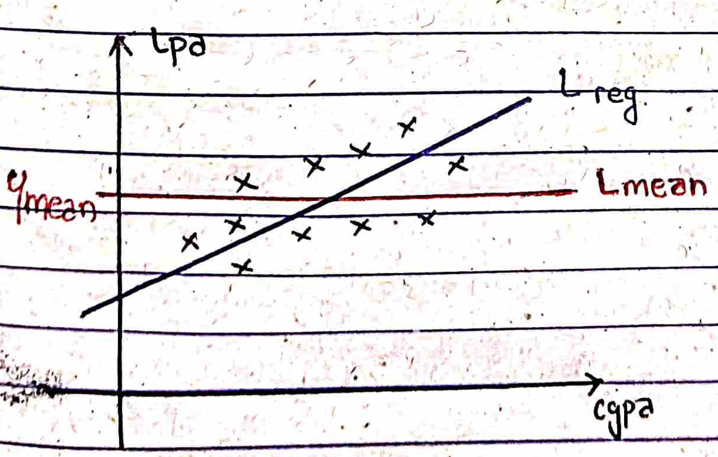
It measure the strength of relationship b/w 0-100%.

The high value of R-squared determines the less difference b/w the predicted values & Actual values & hence represents a good model.

It's also called a coefficient of determination or coefficient of Multiple determination for multiple regression

R-squared $\Rightarrow \dfrac{\text{Explained Variation}}{\text{Total variation}}$

R2 score:



R2-score $\Rightarrow 1 - \dfrac{SS_R}{SS_m}$

comparing Reg line with mean line.

$$\boxed{1 - \dfrac{\text{sum of squared error in Regression Line}}{\text{sum of squared error in mean Line}}}$$

worst-case: calculate mean

$$R2 \rightarrow 1 - \frac{\left(\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}\right)_{reg}}{\left(\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}\right)_{mean}}$$

R2 score = 0 ie 1 − 1 = 0 reg line = mean line
Hence, model is worst of use.

R2 score = 1 perfect regression Line give
exact data to every input.

0 → → → → → 1
(worst) (best)