*Regression Metrics*

1. *MAE (Mean Absolute Error)*
2. *MSE(Mean Squared Error)*
3. *RMSE(Root Mean Squared Error)*
4. *R2 Score (coefficient of determination bhi bolte hai)*
5. *Adjusted R2 Score*

***MAE***

*Har data point ka galti nikalna through distance*

*Yeah hamesha positive hie rhega*

*Y1 – ^ y1 + Y2 – ^ y2 + Y3 – ^ y3 / n*

**

*Advantage:*

1. *MAE jo hai vo basically Loss hai , jishe humme kamm karna rehta hai, Result jo aata hai vo exactly y ke unit mai aata hai , MAE & Output ka Unit same rehta hai*

*E.g LPA nikalrhe tou jo MAE aayega vo bhi LPA mai hie aayega*

1. *Data Mai Outliers hai tou vo time mai Robust hai yeah Outliers keliye, isko kamm farak padta hai Outliers se*

*Disadvantages:*

1. *Iska Graph is not Differentiable at Zero that’s why MSE ko aana pada , fir hum Optimization Technique use karte hai*

*MSE, Mean Squared Error*

*Square Distance find karte hum isme*

*Isko bhi humme kamm karna rehta hai*

**

*Advantages: We can use it as loss function coz differentiable at 0 hai*

*Disadvantage:*

1. *Suppose Y ka Unit LPA hai tou MSE output aayega LPA ka Square (LPA^2 )*
2. *Bad to deal with Outliers*

*RMSE, Root of MSE*

*Advantage: Same Unit mai aata hai output, Deep Learning mai zyaada use hota hua dikhega.*

*Disadvantage: Not robust to Outlier*

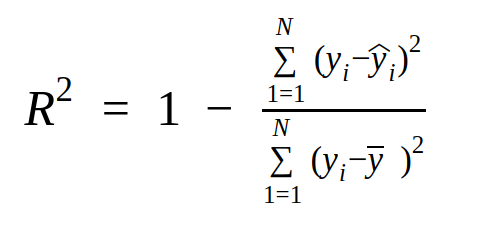
*R2 Score (coefficient of Determination or Goodness of fit)*

*Yeah batata hai ki model kitna acha Perform karrha hai*

*Let’s take a dataset to understand it*

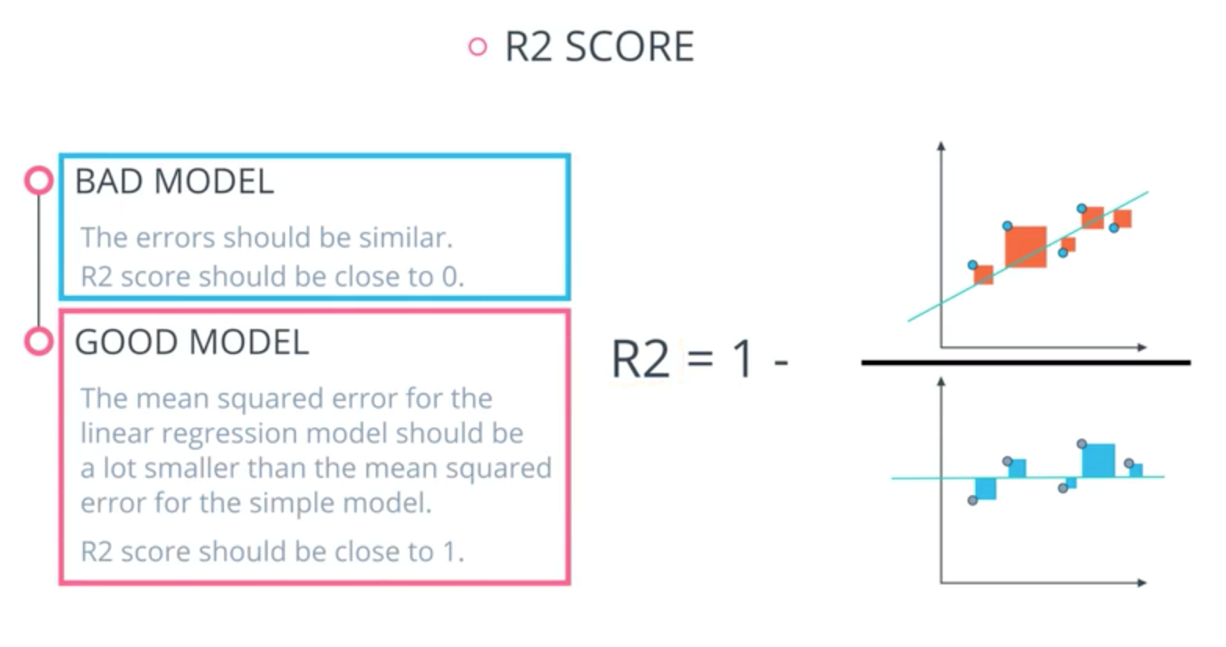
|  |  |
| --- | --- |
| *CGPA* | *Package (LPA)* |
|  |  |
|  |  |
|  |  |

*R2 = 1- SSR/SSM*

**

*SSR- sum of squared Error*

*SSM – Sum of squared of mean*



***If R2 Score negative hai tou Matlab jo Regression Line hai vo Mean Vale Line se bhi zyaada galti karrha hai***

*R2 negative hai mtlb Data bhai Non Linear hai fatte*

*E.g R2🡪 0.80 hai tou mtlb 80% explanation deraha hai cgpa*

*Lpa ka, baki 20% ka idea nahi hai*

***Cgpa| iq| lpa***

***If r2 is 0.80 tou mtlb lpa mai jo data hai uska 80% explanation cgpa & iq derahe hai***

*R2 Score Problem:*

*The Big Flaw*

*Let’s assume we have data CGPA|iq|gender|LPA*

*Jaise Input data badega vaise r2 bhi apne calculation badega*

*That’s okay*

*But the main problem is*

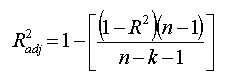
*Suppose 1 column aur hai lyk Temperature of weather*

*Jo ki bhaut irrelevant column hai*

*So ese Irrelevent column ke vajese yatou R2 Badhjata hai yatou utna hie rehta hai*

*So that’s the problem & it is solved by using* ***Adjusted R2***

***Adjusted R2 Score***



*n-> no of rows*

*k-> total no. of independent columns*

*Reliable Metric hai for Multiple Linear Regression.*

*Zyaada Input columns hoo tou Adjusted R2 Score use karo*