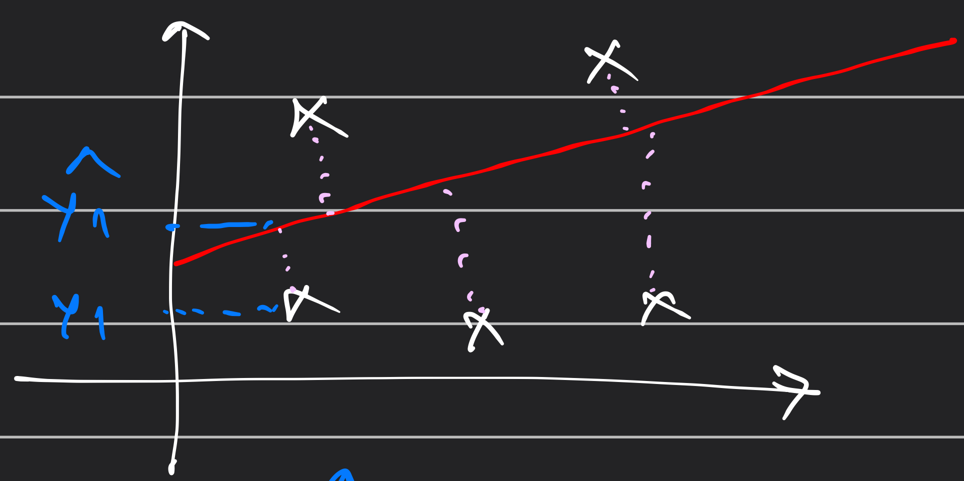
Regression Metrics

⇒ MAE, MSE, RMSE, R2 Score, Adjusted R2 Score. ↳ Loss / Error Functions

- Mean Absolute Error (MAE)



$\hat{y}_1$ → predicted value

y_1 → actual value

Error = $\hat{y}_1 - y_1$

As we are looking for absolute we wont be considering signs here. error could be +ve or -ve but we will take absolute value

Let n be number of observation

$$\text{Mean Absolute Error} = \frac{\sum_{i=1}^{n} |y_i - \hat{y}_i|}{n}$$

Actual — predicted — for calculatⁿ mean

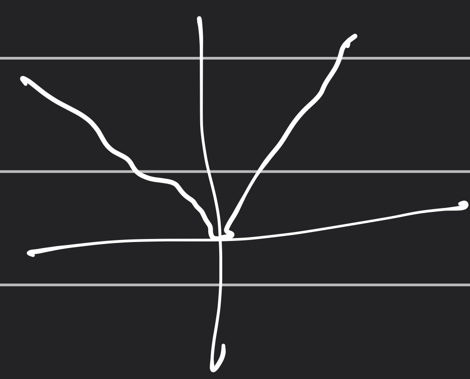
Advantage

i) MAE is basically loss → this should be minimized as possible. If this loss is "zero" that means our LR Line is not Best Fit but perfect line, going through each point.

The perfect scenario.

The loss we gets is always in terms of our O/P. This becomes easy to communicate.

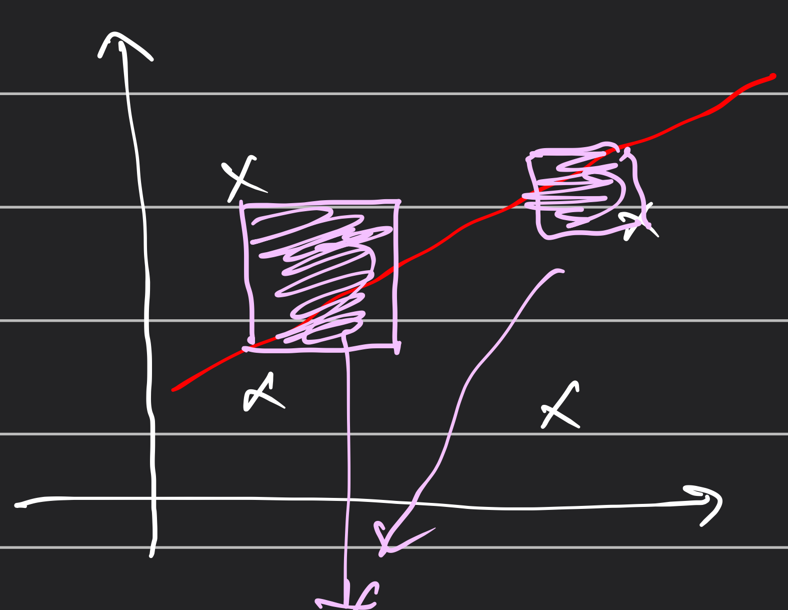
a) Robust to outliers



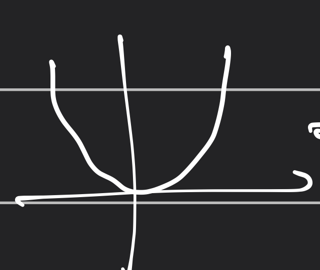
Disadv

1) Graph of modulus is not differential at zero. Gradient decent kind of fails here

Mean Squared Error



Rather than using mod we use Square



⇒ Differential at 0

$(Y_1 - \hat{Y_1})^2$

At the end we sum all the Square and try to minimize it. Divide by N for mean.

Adv

1) used as loss function as it is differentiable

Dis adv

1) MSE is the squared value proportional to o/p variable.
(inserted above "is": unit)

MSE = 11.25 = (square) of Y o/p

2) Penalize the outliers, and will get more impact of outliers. Not robust to outliers.

- more outliers MAE, less outliers we go with MSE

- Root Mean Squared Error

$\sqrt{MSE}$

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

Error/Loss

→ Benefit → o/p is as same unit of o/p of variable

used in Deep learning the most, Not robust to outliers.

- As experienced candiate we stick to one of these and try to minimize this, but as beginner one should try all of this and try to understand what fits best.

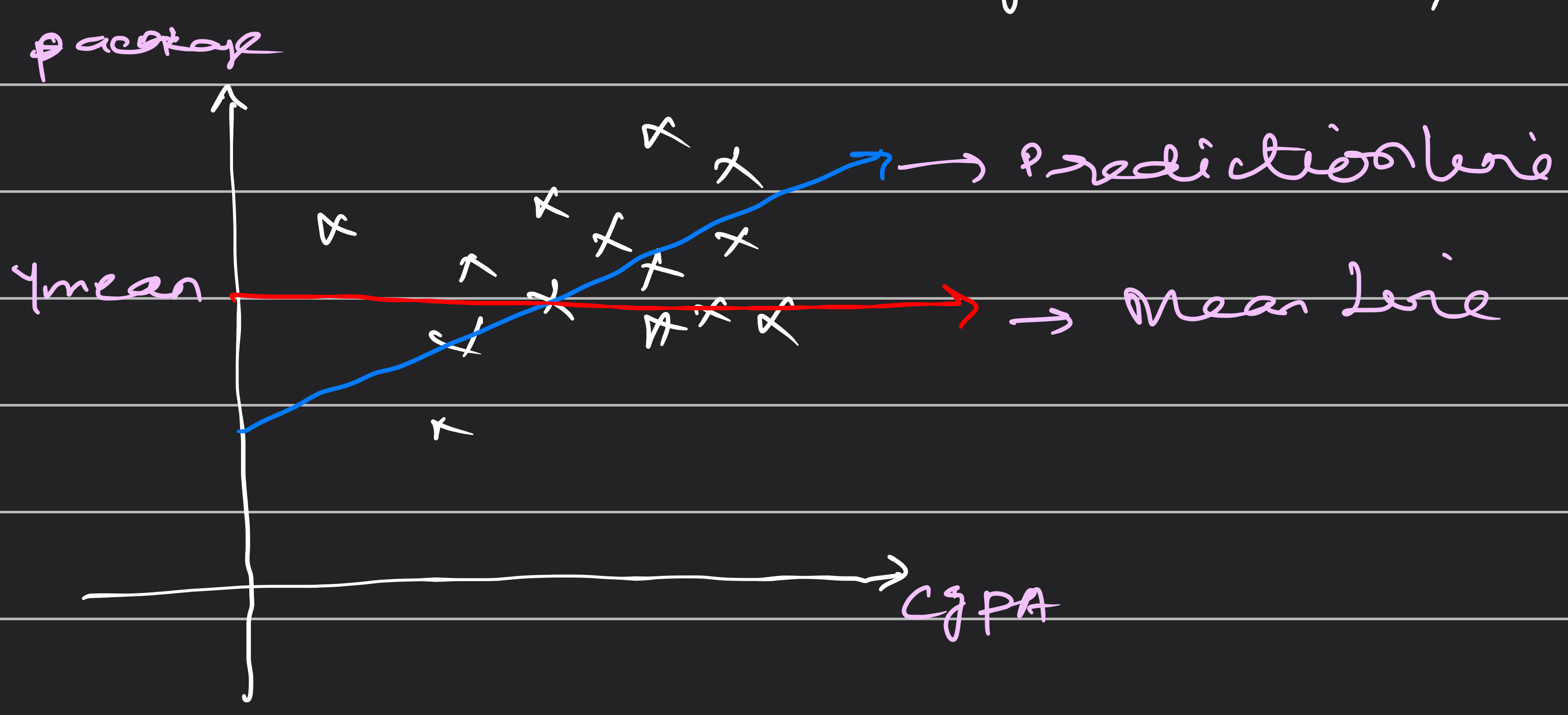
Sometimes minimizing on leads to increasing other.

R2 Score → to test how good our Model is performing.

independent of context as this is absolute for observation.

what happens when someone ask for prediction based upon previous observation
⇒ worst case scenario model to give Average of your o/p.

R2 Score → comparing our Prediction (LR Line) with the mean of our o/p data.



R2 Score is also called Coefficient of determination ⇒ Goodness of fit.

$$R_2 \text{ Score} = 1 - \frac{SSR}{SSM}$$

SSR ← Sum of Squared Error Regression Line

SSM ⇒ Sum of Squared Error Mean Line

$$R_2 \text{ Score} = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}$$ → Bestfit / regression line → mean line

How R2 Score is interepreted

1) R2 == 0 ⇒ $\frac{SSR}{SSM} = 1$ ⇒ Amount of Error created by mean line is same as regression line

so at the end we giving o/p from mean value.

2) R2 == 1 ⇒ $\frac{SSR}{SSM} = \frac{0}{\text{some value}}$ ⇒ we have perfect line

Our prediction line has no errors.

So the more we move towards 1 the better our model is.

3) R2 == -ve ⇒ $\frac{SSR\uparrow}{SSM\downarrow} = > 1$ ie

our Regression Line is making more mistakes than our mean line.

Ex → Non Linear Data pe Linear Regression use krdiya.

Q. How do interepret R_2 = 0.80 lets say cgpa / LPA terms?

This ⇒ that CGPA can explain 80% of Variance of data in LPA (O/P) rest of

20% we dont know.
∴ CGPA can explain why 3.8 one's salary LPA
2-B has 3.0 LPA package but it cannot
explain why 2-0 has 4.8 LPA package.
that 20% can be effected by different
things [cant explain mathematically]

Adjusted R2 score

There is one flaw with R2 score, the
time you increase the input variable,
now there more explanation for variance
in o/p.

For CGPA, IQ | LPA → previously if it was .80 it can increase to 0.90.

what Happens when you add Irrelevant
columns?

So problem with R2 is that it remains
still or gets increased even when we
add irrelevant input ex Temperature. While
ideal behaviour would have to be to
get decreased.

Solution ⇒ Adjusted R2 Score

$$\text{Adj R2 Score} = 1 - \left[\frac{(1 - R2)(n-1)}{(n-1-K)}\right]$$

R2 = normal R2 score

n = number of Rows

k = Number of Independent (Input variables)