Performance Metrics used in Linear Regression:

We will be looking at the performance metrics used to determine if the model is good or not for a specific problem statement.

1. R squared

2. Adjusted R squared

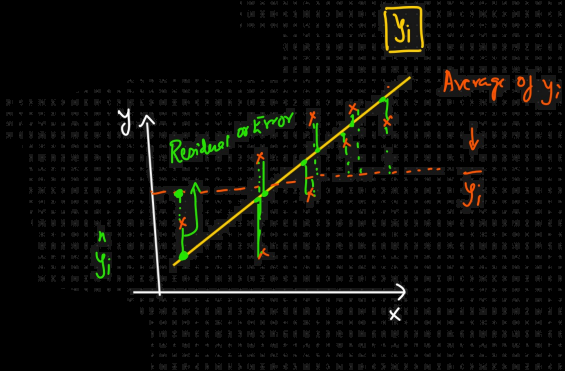
1. R Squared

Where,

-> Sum of squared residual

-> Sum of squared total

Distance between true points and predicted point is called residual



= Average of (true points)

By plotting , we get another line

We are trying to find the distance between true point and average point

= 1 -

= 1 – small number

Approx = 1

Suppose if we get 0.70, it means we are getting 70% accuracy

The closer the value is to 1, the more accurate the linear regression model is

2. Adjusted R Squared

Suppose we have a dataset that has the feature: Size of the house and we have to predict the price of the house

As size increases, price increases

As size decreases, price also decreases

This is an example of good positive correlation

Here R Squared value increases

Suppose we get R Squared = 0.75 or 75% accuracy

Suppose we include another feature, no. of bedrooms

As the number of bedrooms increases, price also increases.

This is also positive correlation

Now R Squared will see an increase. It can be suppose 80% now.

When we add another feature, say location.

There is a chance that the R Squared value will increase again.

Now if we add a random feature such as gender of the occupants which is not at all correlated to price, there is still a chance that R Squared value will increase but it should increase by a large value.

The issue is with R Squared, as we keep on increasing number of features, it’s value also increases.

Even if we add a feature that is not directly correlated with the output feature, the value of R Squared still increases

To prevent this, we use adjusted R Squared. It penalizes with respect to every feature that is not correlated with the output feature.

Formula:

Adjusted R Squared:

Where,

N is the number of data points in our dataset

P is the number of independent features that we are using

For example,

When P = 2, R = 90%, R Squared Adjusted = 86%

When P = 3, R = 92% (Feature that is not correlated is added), R Squared Adjusted = 82%

R Squared adjusted value decreases in such cases

If we apply the R Squared formula, we will get value less than R

If directly correlated, R Squared value increases