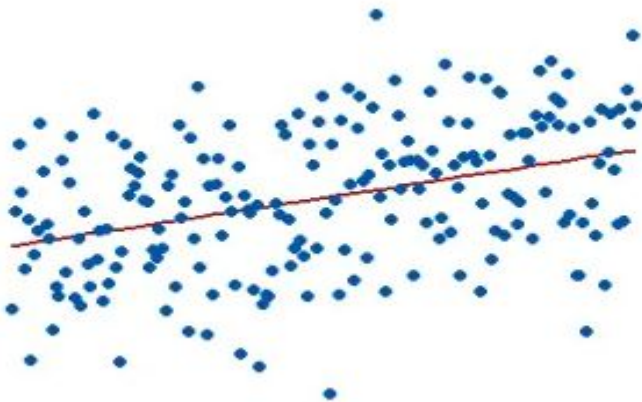


What is R-squared?



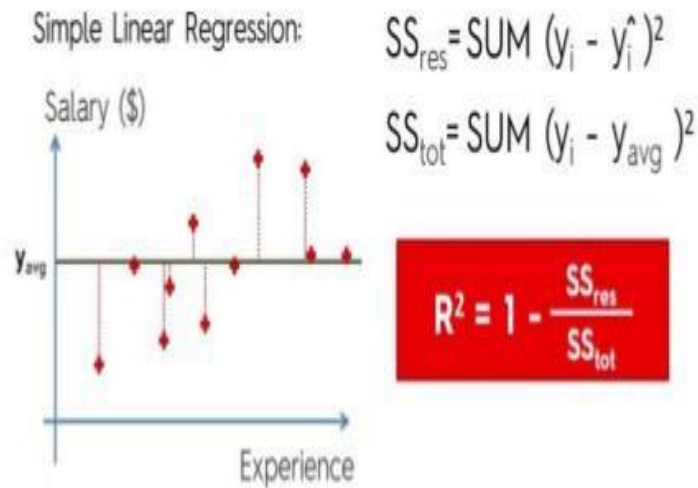
- We evaluate our algorithms (especially the regression algorithms) to see which does the work best for us.
- At times one algorithm might have an edge over the other one but it might not in other cases so to check this we use a method called R-squared intuition which helps to predict how effective our algorithm is
- The R-squared for the regression model on the left is 15%, and for the model on the right it is 85%.



How good is our model?

- Take the sum of differences between the points and our linear model line
- Check if its really less than the differences that the average line

R SQUARED INTUITION



$$R^2 = 1 - \frac{\text{Sum Squared Regression Error}}{\text{Sum Squared Total Error}}$$

$SS_{Regression}$

SS_{Total}

Good R-Squared

R-squared is always between 0 and 100% :-

- 0% indicates that the model explains none of the variability of the response data around its mean.
- 100% indicates that the model explains all the variability of the response data around its mean.
- The value of R-square can also be negative when the models fitted is worse than the average fitted model.

What is a good R-squared?



Problems with R-Squared



INTERNSHIPSTUDIO

- R^2 increases with every predictor added to a model.
- As R^2 always increases, with the more terms you add to the model. This can be completely misleading.
- Similarly, if the model has too many terms and too many high-order polynomials you can run into the problem of over-fitting the data and can lead to misleading projections.
- The idea behind adjusted R-squared is to penalize the score as we add more features to our model.
- Let's look at the formula of adjusted R-squared.

Adjusted R-square

$$adj R^2 = 1 - (1 - R^2) \frac{n - 1}{n - m - 1}$$

n is the number of data points; m is the number of independent features

- Adjusted r-square is a modified form of r-square whose value increases if new predictors tend to improve model's performance and decreases if new predictors does not improve performance as expected.
- It is always lower than the R-squared.
- In the simplified Best Subsets Regression output, where the adjusted R-squared peaks, and then declines. Meanwhile, the R-squared continues to increase.



Q.1 What is of R-squared?

Q.2 What is good R-squared?

Q.3 Pls elaborate the **problems with** R-Squared?

Q.4 How Adjusted R-square overcomes the problems?