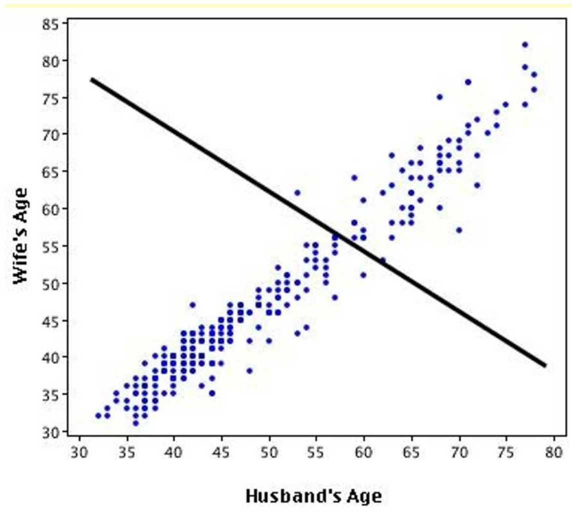

Figure 1: The image shown is an exaggerated example of a model for which R^2 is negative

1 Introduction

I have explained some points of R-Squared, the Coefficient of Determination. Reason to use R-Squared : Earlier, we worked with Mean Square Error, for example :

$$min_\theta = \frac{1}{2}\sum_{n=1}^{N} ||\hat{\xi}^t - \xi^t||^2$$

But as pointed out, it does not consider scale and hence we moved to R-Squared.

2 Background

R^2 is defined as how much of the variance is explained by your model. We use 3 variables: RSS, TSS and ESS.

Calculating RSS (Residual Sum of Squares):

- We predict the value of y for each value of x. Let's call the values of y the line predicts $\hat{y}$.
- The error between what your line predicts and what the actual y value is can be calculated by subtraction and are squared and added up, which gives the Residual Sum of Squares RSS.

$$RSS = \sum(y - \hat{y})^2$$

Calculating TSS:

- We can calculate the average value of y, which is called $\bar{y}$. If we plot $\bar{y}$, it is just a horizontal line through the data because it is constant.
- We subtract $\bar{y}$ (the average value of y) from every actual value of y. The result is squared and added together, which gives the total sum of squares TSS.

$$TSS = \sum (y - \bar{y})^2$$

Calculating ESS (Explained sum of squares):

- The differences between $\hat{y}$ (the values of y predicted by the line) and the average value $\bar{y}$ are squared and added.

$$ESS = \sum (\hat{y} - \bar{y})^2$$

Now,

$$TSS = \sum (y - \bar{y})^2$$
$$\sum (y - \hat{y} + \hat{y} - \bar{y})^2$$
$$\sum (y - \hat{y})^2 + 2 * \sum (y - \hat{y})(\hat{y} - \bar{y}) + \sum (\hat{y} - \bar{y})^2$$

When, and only when the line is plotted with an intercept :

$$2 * \sum (y - \hat{y})(\hat{y} - \bar{y}) = 0$$

$$TSS = \sum (y - \hat{y})^2 + \sum (\hat{y} - \bar{y})^2$$
$$TSS = RSS + ESS$$
$$1 - \frac{RSS}{TSS} = \frac{ESS}{TSS}$$

3 R-Squared

$$R^2 = 1 - \frac{RSS}{TSS}$$

When the line is plotted with an intercept, we can substitute this as $R^2 = \frac{ESS}{TSS}$. Since both the numerator and demoninator are sums of squares, R^2 must be positive.

- When we don't specify an intercept, $2 * \sum (y - \hat{y})(\hat{y} - \bar{y})$ does not necessarily equal 0.
- This means that $TSS = RSS + ESS + 2 * \sum (y - \hat{y})(\hat{y} - \bar{y})$.
- Dividing all terms by TSS, we get $1 - \frac{RSS}{TSS} = \frac{ESS + 2 * \sum (y - \hat{y})(\hat{y} - \bar{y})}{TSS}$.

- Substituting, we get $R^2 = \frac{ESS+2*\sum(y-\hat{y})(\hat{y}-\bar{y})}{TSS}$.
- This time, the numerator has a term in it which is not a sum of squares, so it can be negative.
- This would make R^2 negative. $2 * \sum(y - \hat{y})(\hat{y} - \bar{y})$ would be negative when $y - \hat{y}$ is negative and $\hat{y} - \bar{y}$ is positive, or vice versa.
- This occurs when the horizontal line of $\bar{y}$ actually explains the data better than the line of best fit.
- R^2 can be negative. While a negative value for something with the word 'squared' in it might sound against the rules of maths, it can happen in an R^2 model without an intercept.

3.1 Summary

- - When $R^2 < 0$, a horizontal line explains the data *better* than your model.

3.2 Sources

- http://stats.stackexchange.com/questions/183265/what-does-negative-r-squared-mean
- http://web.maths.unsw.edu.au/ adelle/Garvan/Assays/GoodnessOfFit.html
- http://stats.stackexchange.com/questions/12900/when-is-r-squared-negative
- https://www.researchgate.net/post/Can_we_have_a_negative_R_squared_in_fitting_a_simple_linear_regression _for_exemple_and_how_we_can_explain_this_negative_value

4 R-Squared Greater than 1

On finding out that the value is coming negative, you pointed me to this equation :

$$R^2 = \frac{ESS}{TSS}$$

- Can only we used when $TSS = RSS + ESS$ (explained above).
- If $ESS > TSS$, which is possible when the fit is bad (refer image example above), $R^2 > 1$.
- Math Example on next page.

Values	Predictions
1.01	3.01
0.59	2.59

$$RSS = \sum (y - \hat{y})^2$$
$$RSS = 8$$

$$TSS = \sum (y - \bar{y})^2$$
$$TSS = 0.0002$$

$$ESS = \sum (\hat{y} - \bar{y})^2$$
$$ESS = 8.0802$$

Definition 1:

$$R^2 = 1 - \frac{RSS}{TSS}$$
$$R^2 = -39999$$

Definition 2:

$$R^2 = \frac{ESS}{TSS}$$
$$R^2 = 40401$$