R^2 – COEFFICIENT OF DETERMINATION

COEFFICIENT → A MULTIPLIER OR FACTOR THAT MEASURES SOME PROPERTY.

COEFFICIENT → COOPERATING TO PRODUCE A RESULT.

R^2 → IS THE PROPORTION OF THE VARIANCE IN THE DEPENDANT VARIABLE THAT IS PREDICTABLE

Jeremy Howard - Intro to Machine Learning: Lesson 2

R-squared intuitively it's how much the model explains the how much it accounts for the variance in the data

So with formulas the idea is not to learn the formula and remember it but to learn what the formula does and understand it right so here's the formula

A data set has n values marked $y_1,...,y_n$ (collectively known as y_i or as a vector $y = [y_1,...,y_n]^T$), each associated with a predicted (or modeled) value $f_1,...,f_n$ (known as f_i, or sometimes $\hat{y}_i$, as a vector f).

Define the residuals as $e_i = y_i - f_i$ (forming a vector e).

If $\bar{y}$ is the mean of the observed data:

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

then the variability of the data set can be measured using three sums of squares formulas:

- The total sum of squares (proportional to the variance of the data):

$$SS_{\text{tot}} = \sum_i (y_i - \bar{y})^2,$$

- The regression sum of squares, also called the explained sum of squares:

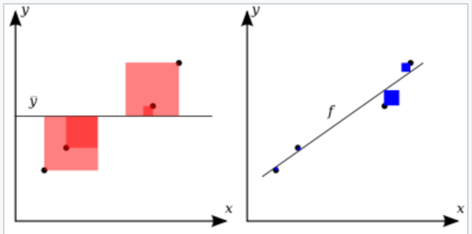
$$SS_{\text{reg}} = \sum_i (f_i - \bar{y})^2,$$

- The sum of squares of residuals, also called the residual sum of squares:

$$SS_{\text{res}} = \sum_i (y_i - f_i)^2 = \sum_i e_i^2$$

The most general definition of the coefficient of determination is

$$R^2 \equiv 1 - \frac{SS_{\text{res}}}{SS_{\text{tot}}}$$



$R^2 = 1 - \frac{SS_{\text{res}}}{SS_{\text{tot}}}$

The better the linear regression (on the right) fits the data in comparison to the simple average (on the left graph), the closer the value of R^2 is to 1. The areas of the blue squares represent the squared residuals with respect to the linear regression. The areas of the red squares represent the squared residuals with respect to the average value.

$$R^2 = 1 - \frac{\text{SOMETHING}}{\text{SOMETHING ELSE}}$$

SS_{tot} →

y_i

3

2

4

1

⋮

SS_{tot} Σ

— minus

AVG → $\bar{y}$ 2.2

IT IS TELLING US HOW MUCH DOES THIS DATA VARY?

SIMPLEST NOT STUPID MODEL → MEAN

RMSE WOULD BE

$$\underbrace{SS_{tot} = \sum (y_i - \bar{y}_i)^2}$$

USES AVG/ MEAN

RMSE OF MOST NAIVE NON STUPID MODEL.

PREDICTION

NOW SS_{res}.

USES PREDICTIONS.

$$SS_{res} = \sum \left(f_i - \bar{y}\right)^2$$

INSTEAD OF NAIVE MODEL

WHAT IS RMSE OF PREDS?

y_i	f_i
3	2.6
2	5.2
1	2.1
⋮	⋮

$\bar{y}$ $\boxed{2.2}$

$$\sum \left(f_i - \bar{y}\right)^2$$

VALUE OF R^2

IF WE WOULD BE EXACTLY AS GOOD AS MEAN IN PREDICTING THE VALUES →

$$R^2 = 1 - \frac{1}{1}$$ EQUAL.

$$\boxed{= 0}$$

IF WE WERE PERFECT.

2 . 1

$$R^2 = 1 - \frac{\cdot}{0}$$

$$= 1 - 0$$

$$\boxed{= 1}$$

POSSIBLE RANGE OF VALUES OF R^2.

IF YOUR MODEL IS WORSE THAN AVG/MEAN IN PREDICTING.

SAY IF YOU PREDICT ∞ FOR EVERY ROW.

RESIDUAL WOULD BE ∞

SO

$$R^2 = 1 - \infty$$

SO POSSIBLE VALUES WOULD BE.

R^2 $\boxed{= \text{NEGATIVE}}$

R^2 IS NEGATIVE WHEN YOUR MODEL

IS WORSE THAN MEAN