

I first learned about R^2 back in my first course in statistics in University. To be honest I did not have a good grasp of it for the longest time; I knew the definition of R^2 , and simple properties such as how it monotonically increases as you add regressors. My knowledge of R^2 was stuck there and did not advance.

Years after, I started picking up some Machine Learning. One would notice that Machine Learning is very similar to statistics (the scope of similarity is beyond the scope of this blog post), and the similarity made me question why some concepts of statistics did not translate to machine learning. In particular, what about coefficient of determination, R^2 ?

Why is R^2 so important in context of linear regression?

R^2 is a widely understood model performance metric in the context of linear regression. To see this, let us define three important sum of squares:

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

$$SSR = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$$

$$SST = \sum_{i=1}^n (y_i - \bar{y})^2$$

A key property (proof omitted) in **linear regression** is that $SSR + SSE = SST$, meaning SST can be decomposed to two different parts, regression sum of squares SSR and error sum of squares SSE . This property allows us to define R^2 to be $\frac{SSR}{SST}$, which can be interpreted as the percentage of variations in response variable which is explained by variations of the regressors. The closer to 1, the better.

Why don't we use R^2 in machine learning to assess model performance?

The simple answer is that $SSR + SSE = SST$ does not necessarily hold anymore when model is **non-linear**, and that if R^2 continues to be defined by $\frac{SSR}{SST}$, it would not be very useful and we do not get that nice interpretation about variations anymore. The frustrating part is that I read many websites about this and this was ultimately hand-waved without any examples. As a result I came up with a quick example on a Saturday afternoon:

```
{% include figure.liquid loading="eager" path="assets/img/R_Squared/SVR.png" title="example image"
class="img-fluid rounded z-depth-1" %}
```

I used the support vector regression package in Python and observed that $SST = 22 > 15.07 = SSE + SSR$. More details can be found on my [notebook](#).