

Home > Course > Improve the Performance of a Machine Learning Model > Evaluate the Performance of a Regression Model


Improve the Performance of a Machine Learning Model

8 hours  Medium

Last updated on 6/29/20



Evaluate the Performance of a Regression Model

 [Log in](#) or [subscribe](#) for free to enjoy all this course has to offer!



In this chapter, we will examine some metrics for evaluating regression algorithms. I will be throwing in some math to help the explanation, but I will also explain things using examples that should be

clear even if the math isn't.

It's important that you understand the concepts behind these metrics, but don't worry if you are not 100% comfortable with the maths. 🧙‍♂️ The Sklearn Python libraries will do all the hard work for us, as we will see later!

Use Mean Absolute Error (MAE) ▼

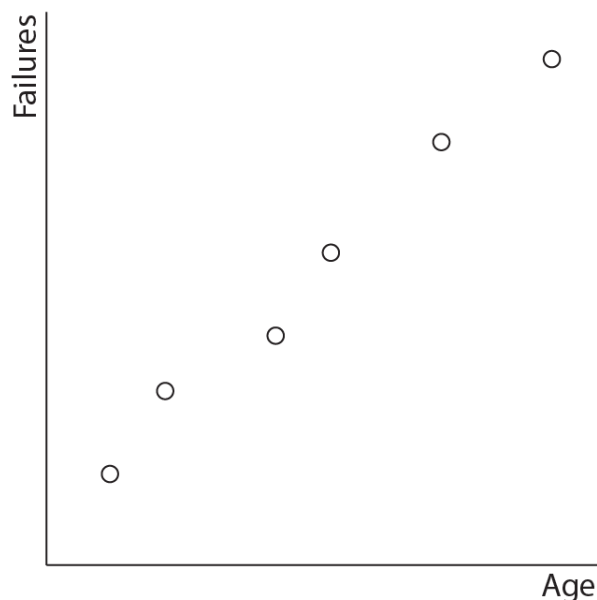
In the previous course, [Train a Supervised Machine Learning Model](#), we evaluated the performance of regression models by computing the mean absolute error (MAE). We defined the MAE as

$$MAE = \frac{\sum |y - \hat{y}|}{N}$$

where y is the actual value \hat{y} is the predicted value and $|y - \hat{y}|$ is the absolute value of the difference between the actual and predicted value. N is the number of sample points.

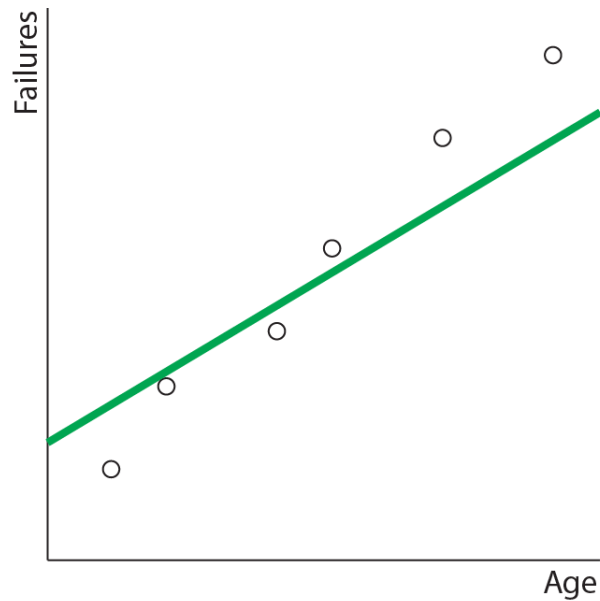
Let's dig into this a bit deeper to understand what this calculation represents.

Take a look at the following plot, which shows the number of failures for a piece of machinery against the age of the machine:



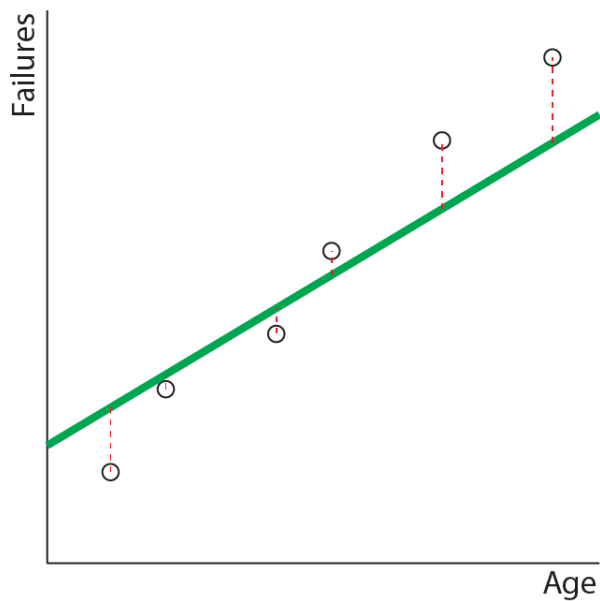
Age	Failures
10	15
20	30
40	40
50	55
70	75
90	90

In order to predict the number of failures from the age, we would want to fit a regression line such as this:



Age	Failures	Prediction
10	15	26
20	30	32
40	40	44
50	55	50
70	75	62
90	90	74

In order to understand how well this line represents the actual data, we need to measure *how good a fit it is*. We can do this by measuring the distance from the actual data points to the line:



Age	Failures	Prediction	Error
10	15	26	11
20	30	32	2
40	40	44	4
50	55	50	-5
70	75	62	-13
90	90	74	-16

You may recall that these distances are called residuals or errors. The mean size of these errors is the MAE. We can calculate it as follows:

Age	Failures	Prediction	Error	abs(Error)
10	15	26	11	11
20	30	32	2	2
40	40	44	4	4
50	55	50	-5	5
70	75	62	-13	13
90	90	74	-16	16

Mean abs(Error)	8.5
------------------------	------------

The mean of the absolute errors (MAE) is 8.5.

Why do we take the absolute value?

To **remove** the sign on the error value! If we don't, the positive and negative errors will tend to cancel each other out, giving a misleadingly small value for our evaluation metric.

If mathematical symbols are not your strong point, you may not immediately see how this calculation relates to the formula at the start of this chapter:

$$MAE = \frac{\sum |y - \hat{y}|}{N}$$

So here is how the table and formula relate:

	y	\hat{y}	$y - \hat{y}$	$ y - \hat{y} $
Age	Failures	Prediction	Error	abs(Error)
10	15	26	11	11
20	30	32	2	2
40	40	44	4	4
50	55	50	-5	5
70	75	62	-13	13
90	90	74	-16	16

Mean abs(Error)	$\frac{\sum y - \hat{y} }{N}$	8.5
-----------------	--------------------------------	------------

Mean Absolute Error (MAE) tells us the average error in units of y , the predicted feature. A value of 0 indicates a perfect fit, i.e. all our predictions are spot on.

The MAE has a big advantage in that **the units of the MAE are the same as the units of y** , the feature we want to predict. In the example above, we have an MAE of 8.5, so it means that on average our predictions of the number of machine failures are incorrect by 8.5 machine failures. This makes MAE very intuitive and the results are easily conveyed to a non-machine learning expert!

Use Root Mean Square Error (RMSE)

Another evaluation metric for regression is the **root mean square error (RMSE)**. Its calculation is very similar to MAE, but instead of taking the *absolute* value to get rid of the sign on the individual errors, we *square* the error (because the square of a negative number is positive).

The formula for RMSE is:

$$RMSE = \sqrt{\frac{\sum (y - \hat{y})^2}{N}}$$

Here is the calculation for RMSE on our example scenario:

	y	\hat{y}	$y - \hat{y}$	$(y - \hat{y})^2$
Age	Failures	Prediction	Error	Error ²
10	15	26	11	121
20	30	32	2	4
40	40	44	4	16
50	55	50	-5	25
70	75	62	-13	169
90	90	74	-16	256

Mean of Error²	$\frac{\Sigma(y - \hat{y})^2}{N}$	98.5
Square root of Mean of Error²	$\sqrt{\frac{\Sigma(y - \hat{y})^2}{N}}$	9.9

As with MAE, we can think of **RMSE as being measured in the y units**. So the above error can be read as an error of 9.9 machine failures on average per observation.

MAE vs. RMSE

Compared to MAE, RMSE gives a higher total error and the gap increases as the errors become larger. **It penalizes a few large errors more than a lot of small errors.** If you want your model to avoid large errors, use RMSE over MAE.

Root Mean Square Error (RMSE) indicates the average error in units of y , the predicted feature, but penalizes larger errors more severely than MAE. A value of 0 indicates a perfect fit.

You should also be aware that as the sample size increases, the accumulation of slightly higher RMSEs than MAEs means that the gap between these two measures also increases as the sample size increases.

Use R-Squared

I stated above that an advantage of both MAE and RMSE is that they can be thought of as errors in the units of y , the predicted feature. This is helpful when relaying the results to non-data scientists. We can say things like "our model can predict the reliability of our machinery to within 8.5 machine failures on average" or "our model can predict the selling price of a house to within £15k on average".

But take heed! This advantage can also be considered a disadvantage!

It says nothing about whether an error of 8.5 machine failures or an error of £15k on a house price is *good or bad*. We can't compare *how good* different models are for different scenarios.

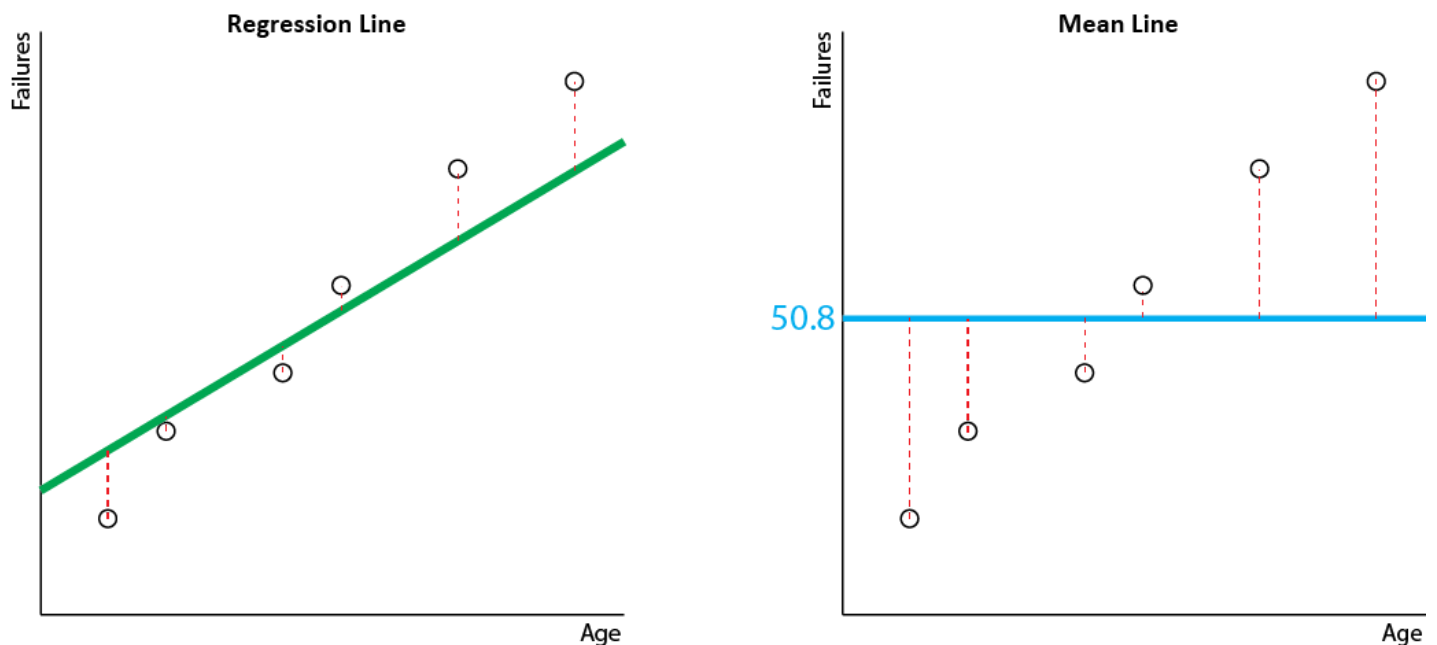
This is where R-squared or R^2 comes in. Here is the formula for R^2 :

$$R^2 = \frac{\Sigma(y - \bar{y})^2 - \Sigma(y - \hat{y})^2}{\Sigma(y - \bar{y})^2}$$

R^2 computes **how much better the regression line fits the data than the mean line**. Another way to look at this formula is to compare the *variance* around the mean line to the variation around the regression line:

$$R^2 = \frac{\text{var}(\text{mean}) - \text{var}(\text{line})}{\text{var}(\text{mean})}$$

Take our example above, predicting the number of machine failures. We can examine the errors for our regression line as we did before. We can also compute a mean line (by taking the mean y value) and examine the errors against this mean line. That is to say, we can see the errors we would get if our model just predicted the mean number of failures (50.8) for *every* age input. Here are the regression and mean lines, and their respective errors:



You can see that the regression line fits the data better than the mean line, which is what we expected (the mean line is a pretty simplistic model, after all). But can you say *how much better* it is? That's exactly what R^2 does! Here is the calculation.

				Regression Line	Mean Line	Regression Line	Mean Line
	y	\hat{y}	$y - \hat{y}$	$y - \bar{y}$	$(y - \hat{y})^2$	$(y - \bar{y})^2$	
Age	Failures	Prediction	Error	Error	Error ²	Error ²	
10	15	26	11	-35.8	121	1281.6	
20	30	32	2	-20.8	4	432.6	
40	40	44	4	-10.8	16	116.6	
50	55	50	-5	4.2	25	17.6	
70	75	62	-13	24.2	169	585.6	
90	90	74	-16	39.2	256	1536.6	
Mean of Error ²					$\frac{\Sigma(y - \hat{y})^2}{N}$ 98.5	$\frac{\Sigma(y - \bar{y})^2}{N}$ 661.8	
R^2					$\frac{\Sigma(y - \bar{y})^2 - \Sigma(y - \hat{y})^2}{\Sigma(y - \bar{y})^2}$	0.85	

Notice something? Most of this is the same as the calculation of RMSE.

The additional parts to the calculation are the column on the far right (in blue) and the final calculation row, computing R^2 .

So we have an R-squared of 0.85. Without even worrying about the units of y we can say this is a decent model. Why? Because the model explains 85% of the variation in the data. That's exactly what an R-squared of 0.85 tells us!

R-squared (R^2) tells us the degree to which the model explains the variance in the data. In other words, how much better it is than just predicting the mean.

Here's another example. What if our data points and regression line looked like this?