Twitter Thread by Jose Jorge



Jose Jorge
@josejorgexl

How many times you have heard that your residuals should be independent, normally distributed with mean zero, and with constant variance?

No matter if your answer is zero or a million times. This thread can help you

It doesn't even matter that you don't know about residuals■

First of all: What is a residual?

It is a synonym for error. It is the difference between the expected output and the output of our model:

Y - Ym

Where Ym is the output of our model. So residuals can be positive or negative and we need them to stay close to zero

1: Independence

We don't want the error in for some input to be dependent on the error for another input

That would mean there is information about the relationship between inputs and outputs that our model is missing and that is present in the residuals2

2: Mean equals zero

Well, this is an intuitive one, isn't it?

We have said already that we need residuals close to zero. There should be both negative and positive residuals.

But we don't want a single huge positive error and then a lot of small negative ones

Let's continue

3: Normally distributed

The mean of the residuals is zero, but that's not enough

We need about 50% of the errors to be negative and the other half to be positive

Also, we need most of the errors to be close to zero

In other words, we need the errors to be normally distributed

4: Constant variance

Last but not least. We need the errors to show a constant variance. This is a property called homoscedasticity ■

That means our model is equally good (or bad) for all inputs

Non-constant variance denotes our model is missing relevant information

Generally, if the residuals of your model don't meet all the above requirements we should dump it

But it might depend on the problem

Note that and include a residual analysis in your machine learning pipeline starting right now!!!

And that's it!

I hope you have enjoyed this thread and you have learned something new

I have been kind of inactive these days but I publish content like this one from time to time

So, if you like ML, AI, algorithms, and Comp-Sci in general, consider following

Stay tuned!