How Regression Models Can Guide our Decisions Tech Insights – Mandy

I chose “How to Choose the Best Regression Model” as I felt that will be one of the biggest challenges in my learning to start with. Though I do notice it may be specifically relating to linear regression models specifically from the beginning once I’ve started it, I’m hoping it will still give me thoughts for what things I’m considering even outside of linear options.

Just to tie my learning into my thoughts the beginning talks about:

**Too few**: An underspecified model tends to produce biased estimates. (overfitting I believe)

 **Too many**: An overspecified model tends to have less precise estimates. (underfitting I believe)

Just right – what we are aiming for to make meaningful decisions from data

I guess I feel like first I’ll also need to get a better feel for which variables (features) we are testing… based I think on correlation to target (label) for example. So making sure you are having relevant features and not too many or too few to start with, choosing those features, we discussed a bit when I was working on the breast cancer dataset.

Which they do bring up under the complications “The best model can be only as good as the variables measured by the study. The results for the variables you include in the analysis can be biased by the significant variables that you don’t include”

I feel like the key to going from where we are now to where we will be able to identify these better will be practicing with existing data that we can do our own testing but then compare our results to what other people chose to get an idea of how we can refine it better? I worry a bit because this program is a lot to learn in a short time that while learning the basics and the module assessments I won’t be able to also do practice in a way to get feedback once it’s all put together beyond that to get to the next step a bit. However perhaps as we progress that will be nearer the end or there’s a way to do that afterwards. I struggle with feeling right now that there will be enough time to do that step until after the program as between work, class and projects I don’t feel I have the mental space to do more on top of that currently.

Back to the article it mentions that after you fit you need to determine whether it aligns with theory and to then make adjustments. I believe that’s by testing different models and comparing their outcomes overall. Interesting to see they say studies show that simpler models typically produce more precise predictions as off hand to begin with one would typically think the opposite. I like the concept of starting simple and adding a bit at a time to see if/where it helps the prediction.

I’m not quite sure however what they mean by residual plots do they mean plots that aren’t used to make sure that they shouldn’t have been? Possibly good to research further there to make sure I understand.

The summary appears to be there there isn’t a single measure, but that testing is integral, and a bit that intuitively as you improve you will get better at identifying. Practice makes improvement? I won’t say perfect that goes too far!