**Lesson 08 Code Talk**

Extrapolation beyond available points is difficult not only in case of polynomial regression. What insight is derived from extrapolation in your workplace, what methods are used and are the approaches reliable?

Unfortunately I do not have a lot of experience with extrapolation of models in the work setting environment, mainly because we were not allowed to extrapolate results from previous validated models to unknown regions. In my particular experience predicting a variable outside the known range, which usually is the range that the model was trained for could lead to significant issues. On the one hand based on the kind of data that we were using to predict engine behavior (thermodynamic data) extreme cases (very low and very high points) were already rare, so the confidence intervals around these ranges were lower than other parts of the data (where most of the data is concentrated). But then leaving the known area all together was a recipe for disaster (at least on the commercial engines field). One of the main reasons for this is given by the very dynamic and not linear behavior of thermodynamics under extreme circumstances (engine running very hot/very cool/very fast) which makes uncertainty about unknown ranges very difficult to predict. One example of this (as mentioned in previous discussion) is the so called “EGT corner point”, which is a temperature value in which the engine (to protect itself) would change its thermodynamic behavior completely, then predicting its behavior using a model trained with pre-corner point data and extrapolating to post-corner point data would lead to huge differences in performance.

On the other hand there are other fields where extrapolation is actually needed, one of those (I believe) would be “trading”. I do not have a lot of experience with that, but I worked on a couple of projects in python to predict long/short positions for a particular stock or a bundle of stocks. In order to have a better idea of this unknown range (the future) multiple mathematical methods are incorporated thru the process of designing a trading strategy (usually based on curing edge research offering a glimpse of the future or a current market behavior that not a lot of people has exploit yet). In this strategy one includes risk factors (description of volatility of assets to minimize/neutralize risk exposures) and alpha factors (deviations from efficient market hypothesis). And even with all the precautions in place, each model has to be tested using the well-known method called “backtesting”, where you take a set of data that simulates the current markets and use that as proxy of how good your investment strategy would perform on unseen (future) data.