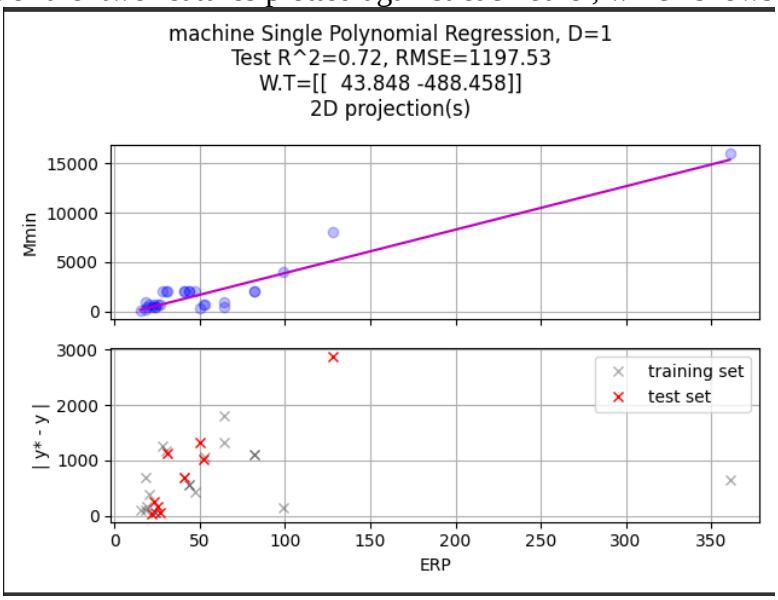
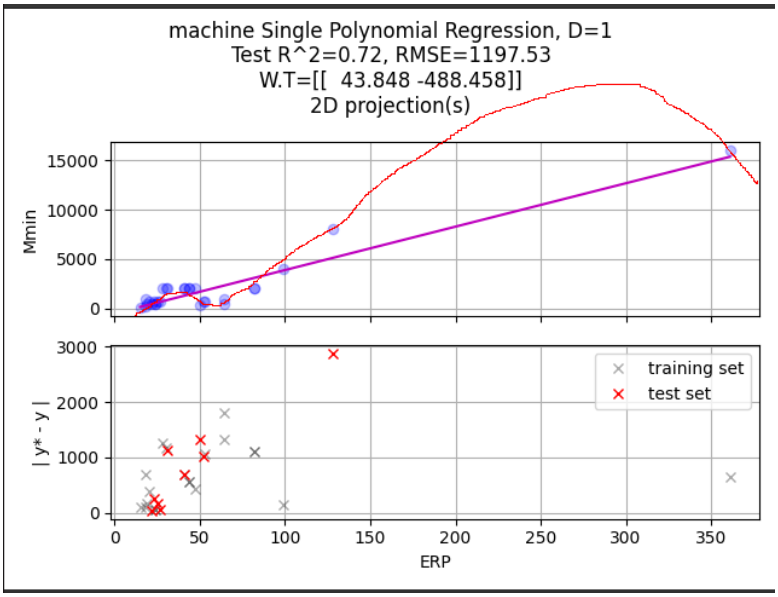
The dataset I am using is the Machines dataset, which looks at the performance of different CPUs by different manufacturers. The features of this dataset are the vendor, the model name, the machine cycle time or MYCT, the minimum main memory or Mmin, the maximum main memory or Mmax, the Cache memory, the minimum channels or ChMin, the maximum channels or ChMax, the published relative performance or PRP, and the estimated relative performance or ERP.
I want to look at which pair of features from the Machines dataset have the closest correspondence, and with what polynomial degree.
Looking first at the pair plots of all the features, ERP and Mmin seem to show an interesting shape. There are other features that I initially looked at that seemed to have a fairly close relationship, however, I picked a fairly small dataset. With other feature pairs, it's difficult to get a consistently strong relationship, because the randomness introduced by the selection of the test and training sets produces large variation in the weights.
I would like to see if I can find a polynomic function that fits to the shape I see on that graph nicely. Below is a screenshot of the two features plotted against each other, which shows the interesting shape.



To me, this looks like it could potentially be a 3rd or 5th order polynomial.
I tried training a model on probably twenty different partitions of the data for polynomial functions with 1st, 3rd, 5th, 7th, and 9th orders. I also tried a couple partitions each on the even numbers for good measure. Despite this seeming like a very clear pattern to my human eyes, the code wasn't able to train a good fit to the data.
I wanted to see something like the below:



I suspect the issue here is that the dataset I chose is not big enough. I picked a different dataset than the one I did lab 1 on, since that dataset was very much a clustering set, and went with this one instead because it was marked as good for regression. However, it is relatively small, with only 200 data points. I thought that would be good enough, but perhaps not.