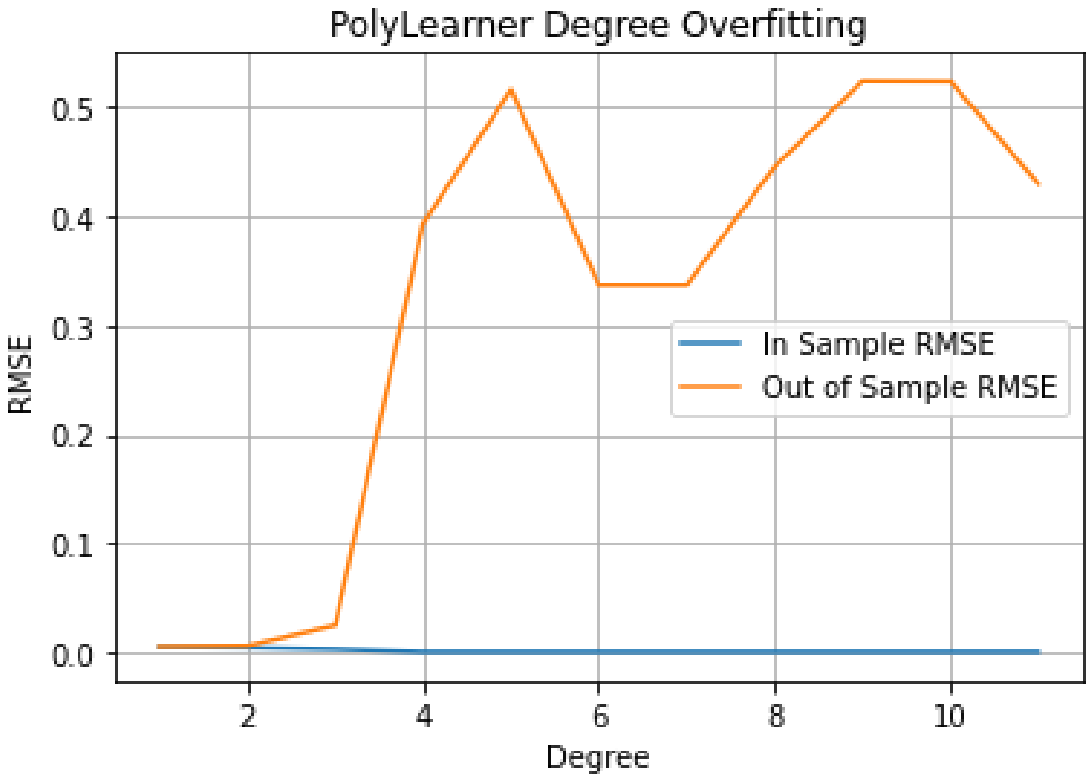
Stephen Crawford
Financial Machine Learning Project 3
2/16/2022

1. We can see that overfitting does occur using the PolyLearner polynomial regression learner as we increase the value of the **degree** hyperparameter. As we can see, after the degree=7 mark, we end up experiencing signs of overfitting where the *in-sample* RMSE remains more or less constant (at approximately 0) but the *out of sample* RMSE begins to increase. Though the *out of sample* RMSE appears to begin to drop again after the degree = 10 mark, this is likely to be an anomaly and not actually indicative of a true trend in the model's fit. Instead, it seems that after we hit a point of best fit at degree=6, the model starts to significantly overfit. The initial low RMSE is not a true fit either since both the *in-sample* and *out of sample* RMSE improve at points after this (there is a slight dowards slope to the *in-sample* and the *out of sample* improves after reaching a realistic value).

   In order to reach this conclusion, I ran a basic experiment where I ran the PolyLearner on the Istanbul dataset with a set random seed of 759941. I then ran 10 separate training models and tested each accordingly. I calculated the *in-sample* RMSE and the *out of sample* RMSE for each of the models run at each separate degree level. The degree of the PolyLearner hyperparameter went from 1 to 11. That is, I recorded both RMSE measurements for 10 models set with degree = 1, then did likewise with degree=2 etc.. I averaged both RMSEs for each set of models trained at a specific degree and then plotted the averaged *in-sample* RMSE and *out of sample* RMSE for each degree specification.



2. Unlike the PolyLearner, it does not appear that the CARTLearner experiences overfitting with regard to the **leaf_size** hyperparameter. As we can see from the *CARTLearner Leaf Size Overfitting* diagram, both the *in-sample* RMSE and the *out of sample* RMSE appear to reach plateaus. The values reach plateaus around the same leaf_size setting though the *out of sample* RMSE may stabilize slightly before the *in sample* RMSE. As we can see, the *in sample* RMSE stabilizes at leaf_size=11, the *out of sample* RMSE stabilizes at leaf_size=5 however. We can tell from the plot that the **leaf_size** hyperparameter does not lead the CARTLearner to overfit since we never see the characteristic behavior of overfitting where the *in sample* RMSE improves while the *out of sample* RMSE becomes worse. Instead, the increase of the leaf size simply stops mattering very much after a certain degree (5 or 11 depending which RMSE you are concerned with; generally 5).

   In order to determine that the CARTLearner did not overfit with the increase of the **leaf_size** hyperparameter, I conducted the exact same experiment that I did with the PolyLearner expect instead of stopping at degree=10 as I did with the PolyLearner, I went all the way up until leaf_size=19 for the CARTLearner.