Executive Summary:

Predicting Laptop Prices

By Nataniel Moreau

Overview

Due to volatile input markets over the last couple years, I examine if laptop prices can be accurately predicted from its specifications. I train 4 different models with varying levels of success that partially exceeded my expectations.

What is the point?

Ever since the beginning of the pandemic, computer prices have become increasing volatile, more than can be explained by inflation alone. There are multiple possible explanations for why this is. Firstly, adverse supply shocks induced by the pandemic could have created supply bottlenecks that have persisted over time. Secondly, manufacturers could be taking advantage of confusing and wild times to raise prices and subsequently profits. Lastly, scalpers could be insuring manufacturers by removing the risk of offering high prices. Scalpers create a kind of artificial demand where actual consumers don't show interest in purchasing at exorbitant prices, but scalpers take on the risk of not selling stock from the manufacturers.

Data

I sourced my data for this prediction experiment from Kaggle, it came with a total of 10 estimators. Minimal cleaning was necessary. The main challenge of this data set is the large amounts of variation in the predictors. Each pc part has a market of its own with tens if not hundreds of options, this makes predicting onto unique configurations not seen before difficult.

Methods

The first of four model specifications I chose is a simple linear regression, tuned for the number of predictors using sequential stepwise selection, the chosen hyper parameter was 8. The second is a Lasso model which resulted in a penalty of ~3. My third model is an ElasticNet model, tuned for mixture and penalty which were .8 and ~4 respectively. My last model is an ensemble of boosted trees with # trees, tree depth, and learning rate tuned to 100, 9, and .1 respectively.

Results

I am using rmse as my chosen metric for evaluation due to its penalty from larger deviations from the desired price. I was expecting to get mediocre results due to the reasons stated in the second section of this summary. Perhaps because the majority of the variation in computer prices exists in the GPU market combined with the fact that laptops usually aren't GPU intensive, my models performed slightly better than expected. The ElasticNet and Lasso were the best performers and had very similar results at test rmse's of $141 each and oddly training rmse's of $161 and $162 respectively (prices mean: $718 & inter-quart range: $383-$952). There are two main take aways I have; firstly, it is strange that I had better out-of-sample performance. One possible explanation is that due to the smaller size of the hold out set there was less variation in the estimators which allowed for more accurate results. My second take away is that due to the customizability of computers mentioned earlier, it is harder to get excellent performance is this regression setting, perhaps using a different outcome instead of prices in a classification setting would produce better results.