Kofoid, Eric

Assignment 6

Question 1)

The presentation was about his experience discovering how to predict whether the economy will be in recession or not using forecasting techniques. It seems that the most difficult part of the process was figuring out what data to use (just like all data science projects). He started off with Logistic Regression and moved on to Naïve Bayes.

I was most surprised about how little data he used to create his models from. Four economic indicators seem like such a small number to use. Furthermore, the S&P 500 didn’t have a significant impact on the results of the model. It surprised me that he didn’t just throw that variable out and find another one. I also found the error charts fascinating. I think it was the Bayes model that converged to a consistent error faster, but the other model took its sweet time to converge. Charting error like that is very useful to show why the model that doesn’t predict as well at first can still be the better model overall.

Lastly, it really surprises me that he didn’t seem to have a problem that his results may go unused. He spent all that time and effort and it possibly won’t even be used. I guess when we do have the next recession he will be able to pull out his model and see if it predicted it.

Question 2)

At my company, we are currently forecasting sales revenue every day for each day of the rest of the current month. Currently this is being done completely in SQL. I am not working on this problem, but my co-worker is. Every day the model is run using actual sales data and then forecasts sales for the rest of the days of the month. I am still learning what factors make up the model and what our leadership uses it for. It is definitely important because our leadership team is constantly asking questions about the report. Being exact on any single day in the month isn’t as important as being as close as possible to the end of month total.

The first problem I noticed was with the errors. The model was almost always predicting high. So, he added in a way to weight the model better to lower the error and gained a bit of accuracy.

I would like to use this data in some of the models we have been learning to see how it does in comparison to the current manual model. To start with, I think just the actual sales numbers by day may be enough information, since our business is very seasonal and we have had a positive trend (not exponential, yet!). Our B2B sales are to schools with traditional semesters and our B2C customers test at the end of the semesters and renew every 2 years. I suspect that any one of our forecasting models will do a pretty good job. I haven’t worked with daily data yet, so I am wondering if I will run into some new concept I haven’t seen yet.

The current model does use a calculation to include future sales for the month into the model. I am very curious to see how including those sales is different in forecasting than using just past data alone.

Question 3)

We could use binary classification. We have all the independent variables for each point in time and we have the dependent variable of Yes/No in a recession for all the same points in time. We could train the model on the independent variables but use either the same time dependent variable or a future value of the dependent variable. For instance, if I wanted to predict 3 months from now, I would use training variables of time tx, but the dependent variable values for time tx+3.

A quick search of the web shows that there is an area in using neural networks for forecasting (<http://www.neural-forecasting-competition.com/index.htm)>. Our text even has a small section dedicated to it. It explains that differences in lagged values are used in neural network training. Seasonality is also used as inputs. Seems similar to the differencing we just learned.

I would hope that doing this in conjunction with the forecast models we could find a model that does the best job of predicting future recessions.