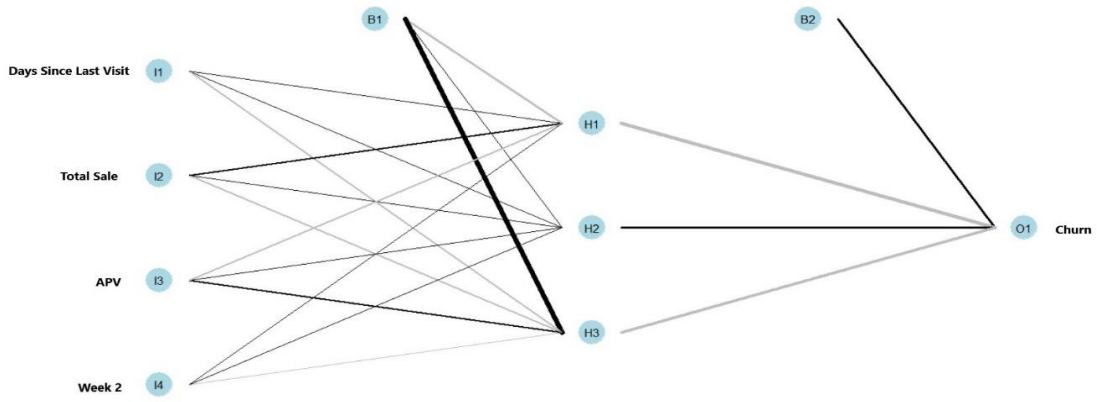


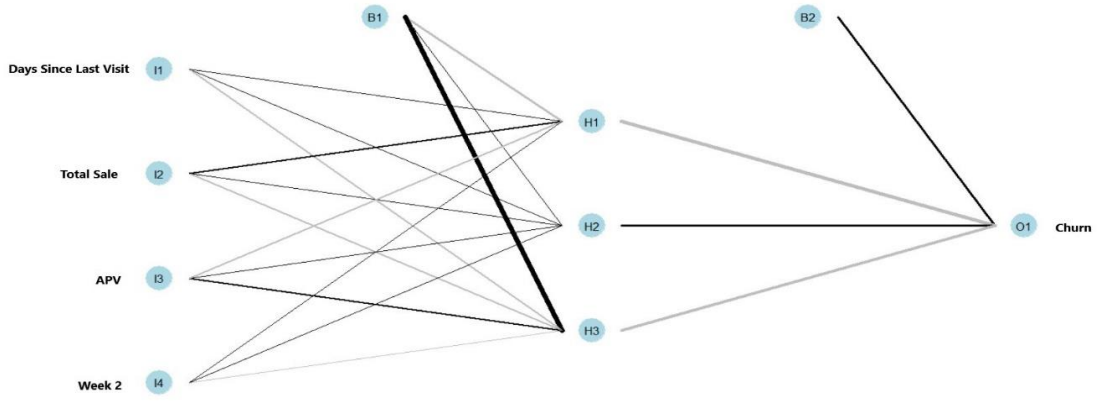
Final Project: Executive Summary

Bradford Simkins

University of Wisconsin, DS740: Data Mining

Goal: Predicting Churn from Customer Data

12/10/2018



Executive Summary

Churn

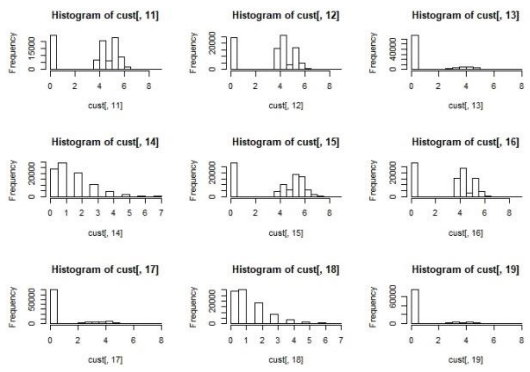
Churn is customer attrition, or the loss of a customer. This is a concern for all businesses, and the ability to predict it could be a gamechanger in customer retention.

Why Predict It?

A model with a low predictive error rate would be very useful for any retail businesses to predict which customers will stay and which will be lost in the churn. They can then focus advertising and customer satisfaction surveys in a more targeted way. As an example, let's say we have a chain of retail stores that sell housewares and kitchen appliances. They can use this model to predict, with a reasonable degree of accuracy, which customers are not likely to come back. Then they can take some steps to either encourage those customers to come back. They can also start asking questions about the customers' experiences, to find out what common factors are causing customers not to return, and maybe make some changes based on the answers. I outline an example in the last section of this document.

Data and Method Selection

The Customer Churn Prediction Analysis data set (Huzaif, 2018) is a massive set of customer data that I am using to create and compare two supervised learning models to predict churn. The data set contains 91,698 observations, 1 response variable (CHURN), 34 predictor variables, and a column for customer ID number. We have considered several methods to predict churn using the 34 predictor variables, or a subset of predictor variables.



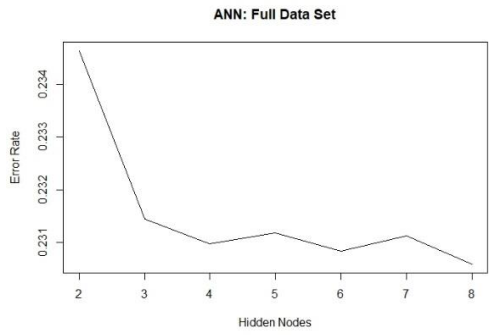
Taking the grand tour of the dataset reveals that most of the predictor variables are not even close to being normally distributed. You can see this for yourself in the lack of a bell-shaped curve in the sample of 9 variables' distributions to the left. This has ruled out parametric methods like linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), and multiple linear regression. Some of these methods would possibly work with major data transformations, but fortunately there are some methods that don't require that step.

This tour also turned up significant correlations between several of the predictor variables, such as Historical Visits with Total Sale, Week 3 Sale with Week 3 Maximum Sale and Week 3 Minimum Sale. So, while decision trees seem like a good way to classify churn, these correlations would cause high variance, even with the use of bagging. For this reason, I have chosen random forests as one of the methods to create a model with, since they are not as strongly affected by correlations.

K-Nearest-Neighbors would likely perform decently on this data set, but because of the size of the set I have chosen Artificial neural networks (ANN) as my second method. ANNs are best when used on large data sets with a lot of nonlinear relationships. A match made in heaven.

The Analysis

Random Forests
Running a random forest with 10-fold cross-validation on the full dataset took 8+ hours. I took a representative, but more manageable random sample of 1000 customers to work with, and get an idea of what I was looking at. I got a general idea of the range of the error rate and the number of trees (60) needed to level it off. Then moved the model onto the full dataset. After running the 10-fold cross-validation I found that the models with a size of 3 minimize (see graph to right) the **error rate at just over 23.1%.**



Artificial Neural Networks (ANN) – While the ANN with 10-fold cross-validation on the full dataset didn't take as long as the random forest, it did still take a couple of hours to tune the number of hidden nodes and the decay rate. So, once again I started with the sample of 1000 customers. After I got an idea of what a good range would be for hidden nodes and decay rate, I did a 10-fold cross-validation on the full dataset and found that while 8 hidden nodes (see left) minimizes the error rate, anything between 3 and 8 is very close, so I went with 3 to keep the model as simple as possible. A decay rate of 1.8 and 3 hidden nodes minimizes the **error rate at just under 23.1%.** Working with 32 predictor variables is unwieldy, so I used a Garson function to get an idea if there were a select few high-importance variables that could do just about as good of a job. I remade the model using the decay rate of 1.8 and 3 hidden nodes as well as the 4 variables with the highest importance: Days since last visit, Total Sale, APV, and Week 2. The **error rate moved up slightly to 23.6%.** This is a pretty insignificant increase of 0.5%.

Where do we go From Here?

Amazingly, or not, the random forest model and the ANN had almost exactly the same error rate on the full dataset, using the parameters I tuned via cross-validation. This means that they are equally good models for predicting customer churn with a 23±% error rate.

For the sake of the simplicity of adding new data to the model, I would recommend the ANN, since it only needs 4 variables. The random forest only uses 3 variables, but we don't get to know which ones are being randomly selected to minimize error in this computational process. So new data would include all 32 variables.

- **Example**: A 23±% error rate is not amazing, but it is not bad either. In this dataset about 34% of the customers are churned out. Returning to my chain of retail stores that sell housewares and kitchen appliances example from the Why Predict it? section; from a group of 1000 random customers, we could expect approximately 340 to not return. Using either of these models we could expect to correctly identify about 262 of these customers. Once we have identified them, we could send coupons and targeted advertising and promotions to say 131 of them and see if a larger portion of them than expected make a return trip. For the other 131 we could send them a customer satisfaction survey, along with some incentive to fill it out, maybe a gift card. Perhaps we find out that 40% of the customers predicted to churn feel that the items we sell are too expensive. In that case we might make the decision to sell a few bargain products, along with our higher quality items.

Reference

Huzaif, Tila (2018). Customer Churn Prediction Analysis [csv data file]. Retrieved from https://www.kaggle.com/huzaiftila/customer-churn-prediction-analysis/