Machine Learning in Finance

Overview

Overview

1. Classification Problem with Big Data: Churn , Part I
2. Classification Problem with Big Data: Churn, Part II
3. Random Forests: Theory and Design, Part I
4. Random Forests: Theory and Design, Part II
5. Random Forests: Predicting Churn, Part I
6. Random Forests: Predicting Churn, Part II
7. Churn Prediction with Neural Nets, Part I
8. Churn Prediction with Neural Nets, Part II

FORDHAM

THE JESUIT UNIVERSITY OF NEW YORK

Gabelli School of Business

Classification Problem with Big Data: Churn, Part I

Classification Problem with Big Data: Churn, Part I

- We will put neural network classification against a **Random Forest** for predicting Churn (percentage of customers who drop their cell-phone service)..
- The Random Forest is an effective classifier.
- It is a development of the .631 Bootstrap.
- When we look at big data, data sets can be big in two ways: a large number of data observations and a large number of classifiers.
- The number of classifiers, of course, can be further expanded if we take interaction terms or polynomial expansions in the regression model

Classification Problem with Big Data: Churn, Part I

- The Random Forest bootstraps on the data: it takes randomly chosen subsets of the data (with replacement) for estimation, multiple times.
- It then examines the predictive or out-of-sample performance with the "left-out" data sets for each bootstrap.
- But the Random Forest also selects subsets of the classifiers randomly out of the full set of classifiers or covariates and builds large numbers of "trees"
- From the various trees, with different combinations of classifiers, it selects the best based on the median out-of-sample performance of its prediction errors.
- Of course Random Forests can be used for simple forecasting as well as classification. But they work best for classification.

FORDHAM

THE JESUIT UNIVERSITY OF NEW YORK

Gabelli School of Business

Classification Problem with Big Data: Churn, Part II

Classification Problem with Big Data: Churn, Part II

- Acme Telephonica (AT) is a mobile phone operator that has customers across every state of the U.S.A.
- AT struggles with customer churn prediction—customers leaving AT for other mobile phone operators.
- In 2010 AT hired a predictive data analytics specialist, to take a new approach to reducing customer churn.
- This case study describes the work carried out by the specialist after a took AT s to develop a predictive data analytics solution to this business problem.
- We have a data set of 32 covariates, with Chun as the Target, 10,000 observations. Mix of categorical and continuous data.
- Statistics on the next two tables.

Classification Problem with Big Data: Churn, Part II

	Median	Std Dev		Median	Std Dev
age	34.00	22.16	**avgOverBundleMins**	3.00	106.37
occupation	0.00	1.80	**avgRoamCalls**	0.00	6.05
regionType	2.00	2.38	**callMinutesChangePct**	-0.10	5.23
marriageStatus	1.00	0.77	**billAmountChangePct**	-0.01	0.76
children	0.00	0.43	**avgReceivedMins**	52.54	169.98
income	5.00	3.14	**avgOutCalls**	13.33	35.67
numHandsets	1.00	1.35	**avgInCalls**	2.00	17.68
handsetAge	339.00	257.08	**peakOffPeakRatio**	1.40	3.88

Classification Problem with Big Data: Churn, Part II

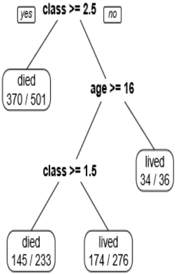
	Median	Std Dev		Median	Std Dev
smartPhone	1.00	0.30	**peakOffPeakRatioChangePct**	0.01	9.97
HandPrice	0.00	57.07	**avgDroppedCalls**	5.33	14.86
creditRating	1.00	1.57	**lifeTime**	17.00	9.61
homeOwner	0.00	0.47	**lastMonthCustomerCareCalls**	0.00	5.75
creditCard	4.00	1.41	**numRetentionCalls**	0.00	0.23
avgBill	49.21	43.89	**numRetentionOffersAccepted**	0.00	0.16
avgMins	359.63	540.44	**newFrequentNumbers**	0.00	0.64
Charge	44.99	23.96	**churn**	0.50	0.50

FORDHAM
THE JESUIT UNIVERSITY OF NEW YORK
Gabelli School of Business

Random Forests: Theory and Design, Part I

- Random Forests is based on **CART** (Classification and Regression Trees)
- To better understand how Random Forests work, we can look at a classifier system from the Titanic survival data
- This is a simple example based on two covariates: age and class of service, for the [0,1] survival probability.
- Trees of course, can built over many more characteristics.

Who survived the Titantic?

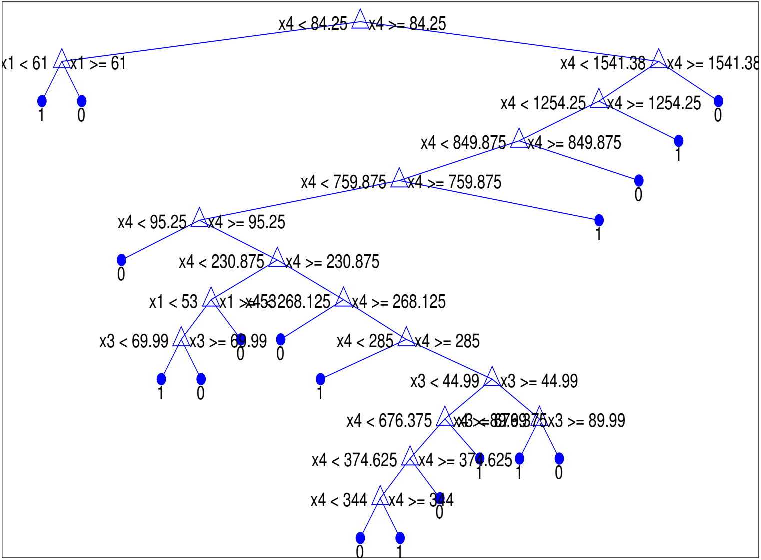


FORDHAM
THE JESUIT UNIVERSITY OF NEW YORK
Gabelli School of Business

Random Forests: Theory and Design, Part II

- For the Churn example, we can draw a tree based on four covariates [children no, roaming calls, average call minute, home owner]
- As becomes clear, the classification trees become wider and deeper
- We can start several classification trees starting with different variables.

Random Forests: Theory and Design, Part II



Random Forests: Theory and Design, Part II

- As noted by Kelleher, MacNamee, and D'Arcy (2020) in Machine Learning for Predictive Data Analytic s, "data understanding" is a key first step.
- In their data set, the researcher sampled data over a period of 5 years.
- The final data set of 10000 randomly sampled customers was split evenly between those who churned and those who did not.
- In the full data set the ratio is about one in ten, but 50-50 in a balanced **ABT (**Analytic Base Table).
- We first estimate the full sample without and with cross-validation. Then we estimate 75% of the data set (randomly sampled) and evaluate the predictive performance with the left-out data. We repeat this 500 times and evaluate the predictive accuracy.

FORDHAM

THE JESUIT UNIVERSITY OF NEW YORK

Gabelli School of Business

Random Forests: Predicting Churn, Part I

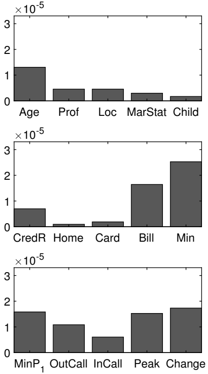
Random Forests: Predicting Churn, Part I

- Let's first look at the in-sample performance of the Churn Estimation.
- The total percentage of wrong decisions, both false positives and false negatives, is less than 10%.

Confusion Matrix		
	State of World	
Decision:	Churn=1	Churn=0
Churn=1	.461	.037
Churn=0	.038	..462

Random Forests: Predicting Churn, Part I

- For the Decision Tree, we can assess the relative importance of each of the 31 covariates: Age of phone is most important.



FORDHAM

THE JESUIT UNIVERSITY OF NEW YORK

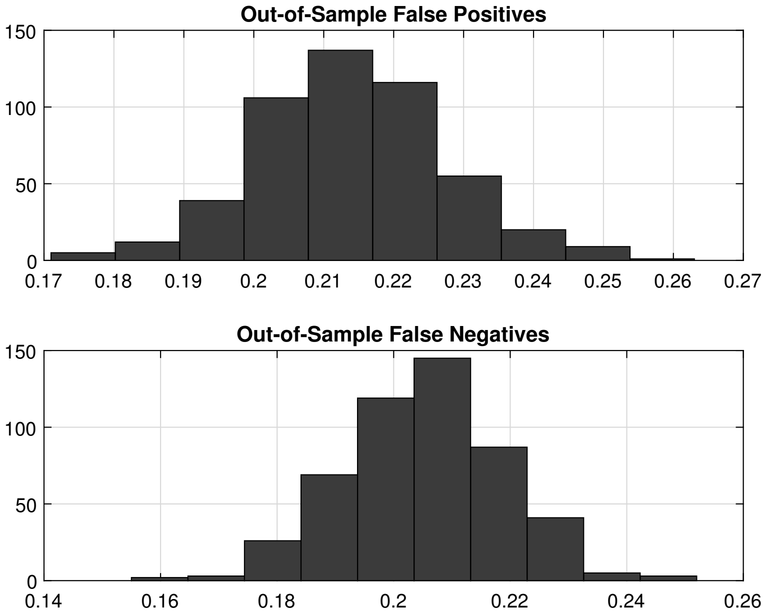
Gabelli School of Business

Random Forests: Predicting Churn, Part II

Random Forests: Predicting Churn, Part II

- We see that the hand phone age, not surprisingly, is the biggest predictor of Churn
- The amount of the bill and total minutes are also important predictors
- While this information is interesting we want to know how the model does with out-oi-sample data.
- The ABT had a 50-50 split of Churn and Non-Churn results. In reality the Churn rate is 10%
- Lets try random selections of 9000 for estimation with the remaining 1000 for prediction, and repeat this 500 times.

Random Forests: Predicting Churn, Part II



Random Forests: Predicting Churn, Part II

- We see that the error rates are much larger when we move to out-of-sample or test-set evaluation.
- Given the presence of outliers, it is recommended that one use the median values of the percentage errors rather than the mean.
- Total accuracy: .580; False Positive: .214; False Positive: .206
- Results are better than a 50-50 coin flip.

FORDHAM

THE JESUIT UNIVERSITY OF NEW YORK

Gabelli School of Business

Churn Prediction with Neural Nets, Part I

Churn Prediction with Neural Nets, Part I

- Let's see if a neural network can beat a Random Forest, both for in-sample and out-of-sample accuracy.
- We will first try a complex neural network of many hidden layers, with the following structure: [20 15 10 5]
- We will estimate the network for the full sample and then evaluate its performance on the basis of bootstrapping.
- We randomly select 9000 observations for estimation and evaluate the out-of-sample performance for the remaining 1000.

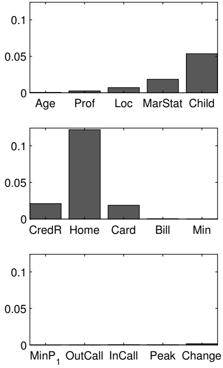
Churn Prediction with Neural Nets, Part I

- To evaluate the network we will look at the accuracy and the importance of the variables.
- However with the nonlinear network, we have to do a numerical perturbation from the median values of each covariate.
- We first obtain the network prediction, $\hat{y}=f(x^{median})$
- For each of the covariates, $i = 1,\ldots.31$, we calculate $x^{i^*} = x^{median}+.00001$,for $i == i^*$, else $x^{i^*}= x^{median}$ for $i \neq i^*$
- We then calculate $y^{i^*} = f(x^{i*})$ and obtain the differences $(y-y^{i^i})/.00001$

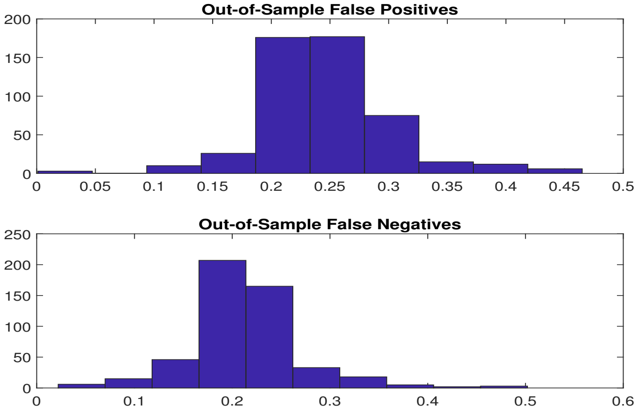
- In-sample performance of complex network: Overall Accuracy Rate: .560. OK better than flipping a fair coin.

Confusion Matrix		
	State of World	
Decision:	Churn=1	Churn=0
Churn=1	.2902	.229
Churn=0	.2098	..2705

- The network gives another picture of the relative importance with Smartphone, home and children playing a bigger role as predictors of likelihood of churn



- We see that the out-of-sample accuracy of the net is better than that of the Random Forest. Well slightly better!



FORDHAM
THE JESUIT UNIVERSITY OF NEW YORK

Gabelli School of Business

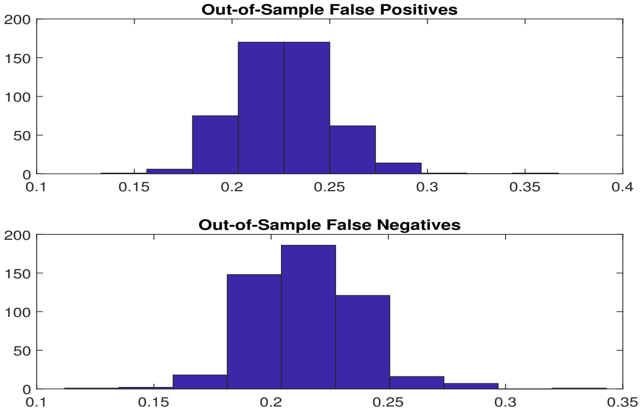
- Let's try a **simple net** with just one layer of 31 neurons.
- Accuracy rate is .58. Performance is better in-sample than that of the deep net.

Confusion Matrix

	State of World	
Decision:	Churn=1	Churn=0
Churn=1	.278	.194
Churn=0	.202	..305

Chun Prediction with Neural Nets, Part II

- Out-of-sample prediction errors are worse than the more complex net.



FORDHAM
THE JESUIT UNIVERSITY OF NEW YORK
Gabelli School of Business