QF 624: Machine Learning for Financial Applications

Examples and Use Cases

Master of Science in Quantitative Finance
Lee Kong Chian School of Business

Saurabh Singal

July 2018

SMU
SINGAPORE MANAGEMENT UNIVERSITY

Example 1 : Uncovering Accounting Fraud by Traditional Statistical Methods

- Fraud can be
  - Overstatement of Sales or revenues, Assets,
  - Understatement of Expense, Liabilities

- Benford's Law: the leading significant digit is likely to be small
  - 1 appears as the most significant digit about 30% of the time, while 9 appears as the most significant digit less than 5% of the time.
  - If the digits were distributed uniformly, they would each occur about 11.1% of the time.
  - The detection of anomalies is derived from the difference between the actual and the theoretical frequency.

✓ Limitation: based on behavioral psychology, almost everything is priced just below a round number e.g., $99.99 and not $100. This can throw off a naïve application using Benford's law into flagging this as fraud.

SMU
SINGAPORE MANAGEMENT UNIVERSITY

Example 1 : Uncovering Accounting Fraud by Traditional Statistical Methods-2

- "Triage in Forensic Accounting using Zipf's Law" by Adeola Odueke & George R. S. Weir
- Zipf Law: If f_1 is the most common term in a collection f_2 is the next most common etc., then the frequency cf_i of the *i*-th most common term is proportional to *1/i*. The *2nd* most frequent word occurs ½ the times as frequently as the 1st, the 3-rd most frequent then occurs 1/3rd as frequently as the first.
- Picalo is an open source software Python library primarily created for auditors/investigators to search for anomalies in datasets.

SMU
SINGAPORE MANAGEMENT UNIVERSITY

Auditing embraces the ***new*** Machine Learning

Deep Learning and the Future of Auditing

By Miklos A. Vasarhelyi and Ting Sun

- KPMG applies IBM Watson's deep learning–powered systems to analyze banks' credit files for commercial mortgage loan portfolios
- Deloitte has allied with Kira Systems to review contracts, leases, invoices, and tweets.

The adoption of deep learning within the accounting profession is still, admittedly, at an early stage.

SMU
SINGAPORE MANAGEMENT UNIVERSITY

Reinforcement Learning

- Reinforcement Learning is a newer area of Machine Learning.
- Reinforcement Learning is inspired by behavioral psychology.
- Reinforcement Learning models how an agent ought to interact with its environment in order to maximize a cumulative reward.
- The agent learns from the environment by interacting with it and receiving rewards for performing actions; rewards can be negative or positive.
- Very successful in games (AlphaGo which beat the human Go champion in 2016, the Deep Q-network in 2015 which mastered a number of Atari 2600 games to superhuman label with only raw pixels and scores as training inputs).

Market Microstructure and Reinforcement Learning

- "Machine Learning for Market Microstructure and High Frequency Trading" by Michael Kearns & Yuriy Nevmyvaka
- This paper discusses machine learning uses for high frequency trading and market microstructure.
- Two applications of Reinforcement Learning are illustrated
  - Optimized Trade Execution
  - Predicting Price Movement from Order Book State

SMU
SINGAPORE MANAGEMENT UNIVERSITY

Example 2: Reinforcement Learning for Optimized Trade Execution

The problem is defined by

- a particular stock, say *S*;
- a share volume *V* ;
- and a time horizon or number of trading steps *T*.
- Goal : buy exactly *V* shares of the stock within *T* steps, while minimizing buying price
- The state variables vector is (*v, t*) where v is the volume left to buy (*v<=V*) and remaining time *t(t<=T)*

SMU
SINGAPORE MANAGEMENT UNIVERSITY

Example 2: Reinforcement Learning for Optimized Trade Execution-2

- if *v* is small and *t* is large (we bought most of our target volume, but have most of our time remaining), we might choose to place limit orders deep in the buy book .
- if *v* is large and *t* is small, we are running out of time and have most of our target volume still to buy, we should perhaps start crossing the spread to meet our target, at less attractive prices.

SMU
SINGAPORE MANAGEMENT UNIVERSITY

Example 2: Reinforcement Learning for Optimized Trade Execution-3

The authors did a full historical order-book reconstruction and also added the following features to *(v, t):*

1. Bid-Ask Spread(difference between the bid and ask)
2. Bid-Ask Volume Imbalance(A signed quantity indicating the number of shares at the bid minus the number of shares at the ask in the current order book)
3. Signed Transaction Volume (A signed quantity indicating the number of shares bought in the last 15 seconds minus the number of shares sold in the last 15 seconds)
4. Immediate Market Order Cost(The cost one would pay for purchasing the remaining shares in the order immediately with a market order.)

SMU
SINGAPORE MANAGEMENT UNIVERSITY

Example 2: Reinforcement Learning for Optimized Trade Execution - 4

Feature(s) Added	Reduction in Trading Cost
Bid-Ask Spread	7.97%
Bid-Ask Volume Imbalance	0.13%
Signed Transaction Volume	2.81%
Immediate Market Order Revenue	4.26%
Spread + signed Volume + Immediate Cost	12.85%

TABLE 1: Reduction in implementation shortfall obtained by adding features to (v, t).

Example: a small spread combined with a strongly negative signed transaction volume would indicate selling pressure (sellers crossing the spread, and filling in the resulting gaps with fresh orders). … we might wish to be more passive in our order placement, sitting deeper in the buy book in the hopes of price improvements.

The results are summarized in Table 1, which shows, for each of the features described above, the percentage reduction in trading cost (implementation shortfall) obtained by adding that feature to the original (v, t) state space

SMU
SINGAPORE MANAGEMENT UNIVERSITY

Example 3: Predicting Price Movement from Order Book State

- Take a problem which is related to, but different from, the Optimal Trade Execution – that of generating profitable state-based models for trading using microstructure features

- Generate alpha by learning
  - when to trade and
  - which direction to trade and
  - how (using what type of order)

SMU
SINGAPORE MANAGEMENT UNIVERSITY

Example 3: Predicting Price Movement from Order Book State-2

The following features were used:

1. Bid-Ask Spread
2. Price
3. Smart Price: A variation on mid-price where the average of the bid and ask prices is weighted according to their inverse volume.
4. Trade Sign: A feature measuring whether buyers or sellers crossed the spread more frequently in recent executions.
5. Bid-Ask Volume Imbalance (as before)
6. Signed Transaction Volume (as before)

SMU
SINGAPORE MANAGEMENT UNIVERSITY

Example 3: Predicting Price Movement from Order Book State-3

The experiment was done for 19 liquid stocks.

- Learning was performed for each name using all of 2008 as the training data.
- Testing of the learned policy for each each name was performed using all 2009 data
- Simple "Learned policies" : buy (or short) at time *T* and close out *t* seconds later.
- The two most important findings are
  - Reinforcement Learning consistently produces policies that are profitable on the test set
  - These policies are broadly similar across stocks

SMU
SINGAPORE MANAGEMENT UNIVERSITY

Example 3: Predicting Price Movement from Order Book State-4

- Profitability is usually better using all six features than any single feature.
- Smart Price appears to be the best single feature

SMU
SINGAPORE MANAGEMENT UNIVERSITY

Example 4: Application of Clustering: Fund of Funds

- Performed ICA (Independent Component Analysis) instead of PCA as hedge fund returns were non-Gaussian
  - Blind Signal Separation (separation of a set of source signals from a set of mixed signals, without the aid of information, or with very little information about the source signals or the mixing process)
  - Cocktail party problem(brain's ability to selectively focus on a particular stimulus while filtering out others, as when a partygoer can focus on a single conversation in a noisy room)
- Then performed Clustering in the factor loadings using PAM (Partioning Around Medoids) and Fuzzy Clustering on hedge funds

See Appendix 1: Clustering

SMU
SINGAPORE MANAGEMENT UNIVERSITY

Finding Groups in Data

This is the title of an old book by Leonard Kaufman and Peter J. Rousseeuw.

- It was popular because the k-medoids algorithm was discussed here.
- It has nice acronyms for several clustering algorithms all based on girls' names
  - PAM (Partitioning Around Medoids),
  - CLARA (Clustering Large Applications),
  - FANNY (Fuzzy Analysis),
  - AGNES (Agglomerative Nesting),
  - DIANA (Divisive Analysis),
  - MONA (Monothetic Analysis)

Example 5: Application of Gaussian Mixture Model (GMM)

- Hong Kong Dollar exchange rate pegged to US dollar
- Several currencies were pegged, mostly to the dollar or German Mark.
- Creeping band, inflation adjusted band etc.
- So either the peg will break by a large amount, or not break.
- So a mixture model seemed appropriate

See Appendix 2: Gaussian Mixture Model

SMU
SINGAPORE MANAGEMENT UNIVERSITY

N-grams as Patterns

- Examples of bigrams
  - Take signs only or **U,D** or **+, -**
- Pentanomial Model: *Parameter Estimation with k-Means Clustering* by Kiseop Lee and Mingxin Xu
- There are five symbols. (***R,+,0,—,C*** : Large Rally; Up; little changed; Down; Large Down or Crash) with frequencies approximately 4%, 24%, 49%, 18% and 5%
- The SDE for capturing the price dynamics has jump processes

SMU
SINGAPORE MANAGEMENT UNIVERSITY

Why is classification important?

- *Question*: Why is classification useful in finance? Why not focus only on prediction using regression?
- *Answer*: It is difficult to predict the exact magnitude of a market move. Often, knowing the direction of an asset or knowing whether the next three days will be high volatility or low volatility is enough to trade profitably.

Simplest Supervised Learning Model

- The simplest machine learning model is the simple linear regression.
- When there are too many variables, then we extend simple linear regression by introducing penalized regression by using regularization terms to avoid overfitting/spurious coefficients.
  - Lasso – selects only a few variables & uses L_1 regularization (sum of absolute values of the coefficients).
  - Ridge regression cannot zero any coefficients and uses L_2 regularization (sum of squares of the coefficients).
  - Elastic net is in between Lasso and Ridge.

SMU
SINGAPORE MANAGEMENT UNIVERSITY

K-Nearest Neighbours and LOESS

- *K*-NN is a non-parametric technique that can be used to identify historical similar instances. And then we can make predictions by averaging the historical outcomes of the these k-nearest neighbours.
- Almost every one generalizes based on past experience and uses k-NN (perhaps wrongly!) without realizing it.
- Linear regression under fits, k-NN over fits so perhaps a combination would be better.
- LOESS is a localized linear regression in which we first select *k* nearest neighbours that are similar to our given instance. And we then fit a linear regression on the subset of data thus selected. We can even use a Lasso regression or Elastic net or Ridge regression on this subset.

Example 6: Applications of *k*-NN

- Short Term trading:
  - Take 5 days' price bars (30 minutes closes)
  - K-NN can be used for predictions of market direction
  - Moderately effective

- Macro-economic regime: Choose a few macro variables (e.g., Jobless claims, ISM PMI, Long Term Bond yield, short term interest rates, inflation or CPI, Yield curve 10y-2y, Retail Sales, Baltic Dry index), that characterize the state of the world and compare it to the past, drawing *k=5* or *10* nearest neighbours, and use that to model the efficacy of various strategies.
  - Somewhat effective

Example 7: PCA and Yield Curve Movement

- Principal Components Analysis of the changes in the swap rates was performed.
- Just under 85% of the variation was explained by the first principal component, 7% by the second and 3% by the third component.
- These were identified as representing a parallel shift in the yield curve, curve steepening, and twisting of the yield curve.

Random Forest and XGBoost

- A random forest is an ensemble of decision trees. The trees can be classification trees (each leaf node has a class label) or regression tree (each leaf node has a continuous score).
  - RF uses bagging
  - The trees are fully grown
  - The trees are grown in parallel

- XGBoost is extreme gradient boosting.
  - The trees are shallow
  - The trees are weak learners, and
  - They are boosted to become strong learners
  - The trees grown sequentially

Example 8: JP Morgan ETF Strategy (XGBoost)

- JP Morgan introduced a long-short strategy trading nine US sector ETFs: financials, energy, utilities, healthcare, industrials, technology, consumer staples, consumer discretionary and materials.
- The open-source implementation of XGBoost available in R was utilized to predict next day returns based on 8 macro factors: Oil, Gold, Dollar, Bonds; economic surprise index (CESIUSD), 10Y-2Y spread, IG credit (CDX HG) and HY credit spreads (CDX HY).
- After obtaining prediction returns for US sectors (based on previous day's macro factors), sectors were ranked by expected return, and the strategy goes long top 3 and short bottom 3 sectors.
- The strategy used a rolling window of 252 days and rebalanced daily at the market close. This yielded an annualized return of 10.97% and an annualized volatility of 12.29%, for a Sharpe ratio of 0.89. Correlation of the strategy to the S&P 500 was ~7%. *(Source: JP Morgan Quantitative and Derivatives Strategy)*

SMU
SINGAPORE MANAGEMENT UNIVERSITY

Example 9: SVM for Selling Options

- The Support Vector Machine is one of the most powerful classification classifiers.
- SVM's easy to calibrate and powerful at the same time.
- We know that, often, the market movements are small. This is based on the fitting the pentanomial model price dynamics (discussed earlier).
- The SVM can make binary classification over next 3-day window whether the alphabet will be "0" (little changed) or "not-0".
- The SVM along with Random Forest was the most popular Machine Learning classifier until Deep Learning (and XGBoost!) came along.

See Appendix 3: Support Vector Machines

SMU
SINGAPORE MANAGEMENT UNIVERSITY

Example 10: Random Forest for Stock Selection

- JP Morgan used 14 factors Price/Book ratio, Gross Profit Margin, ROE, Net Margin, Asset Turnover, Gearing, Forward earnings, Earnings certainty, Cash flow, earnings yield, dividend yield, Vol, 1m momentum, 12M-1M momentum and market cap in a Random Forest model to predict 1 month ahead returns.
- Each stock was put in one of 5 portfolios (ranked highest to lowest)
- The spread of the top-bottom portfolios had information ratio of 1.2 and maximum drawdown of 6%. *(Source: JP Morgan Quantitative and Derivatives Strategy)*

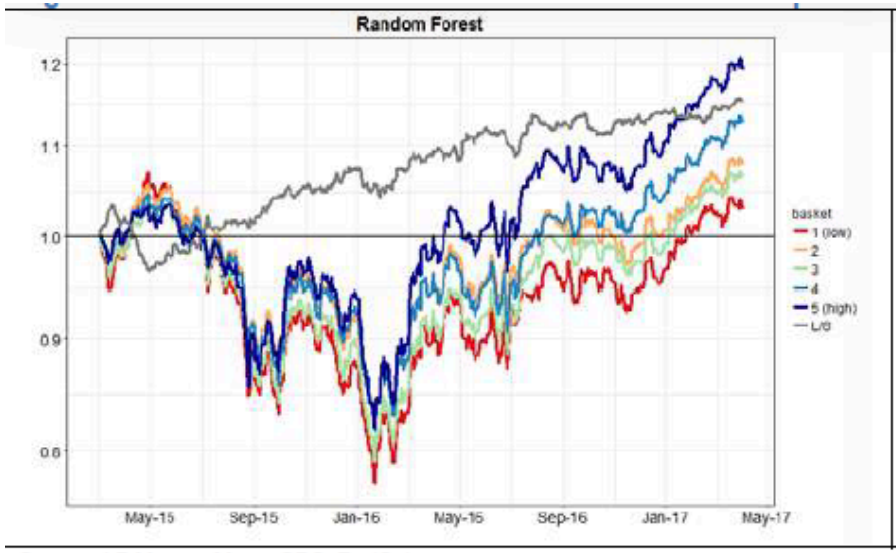
Basket	Cum. Return	CAGR	Volatility	IR	Max Drawdown	Hit Ratio
1 (low)	3.0%	1.0%	11.3%	0.09	27.7%	37.3%
2	7.9%	2.5%	10.7%	0.24	22.7%	37.8%
3	6.4%	2.1%	10.7%	0.19	23.6%	38.3%
4	12.8%	4.1%	10.6%	0.38	21.9%	37.0%
5 (high)	19.2%	6.0%	11.1%	0.54	20.9%	39.5%
L/S	15.4%	4.8%	4.2%	1.16	6.9%	37.7%

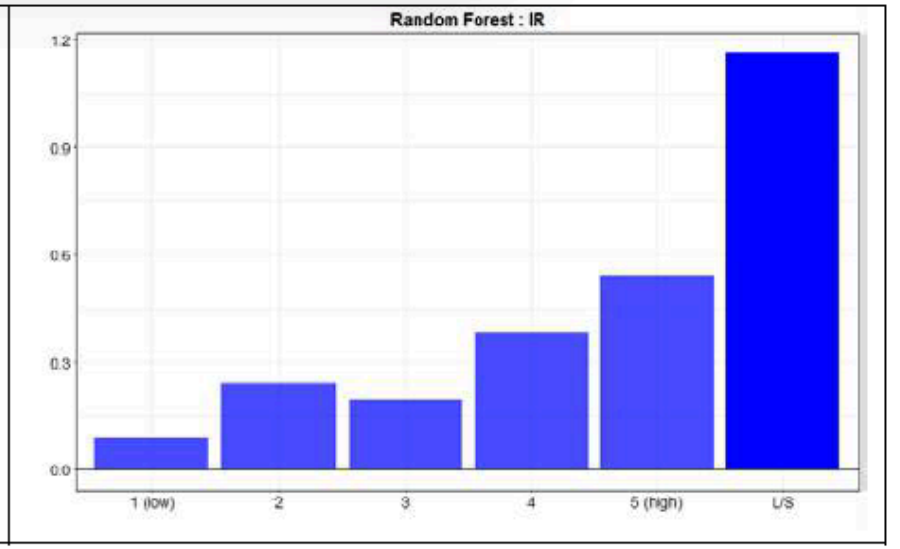
Source: J.P.Morgan Macro QDS, FactSet

SMU
SINGAPORE MANAGEMENT UNIVERSITY

Example 10: Random Forest for Stock Selection-2

- The figure on the left shows the performance (equity curve) of each of the 5 quintiles and the Long-Short portfolio. The right figure shows the corresponding Information Ratios.





Source: J.P.Morgan Macro QDS, FactSet

SMU
SINGAPORE MANAGEMENT UNIVERSITY

Example 11: Random Forest to improve Trading Models Performance

Figure 1. Trading Signals enhanced by Random Forest **Proprietary Random Forest Classifier**

We form a trading algorithm that continuously makes trading decisions. The algorithm has no knowledge of a specific set of variables ("withheld variables"). We then train a random forest classification model using the withheld variables. Based on the output of the trained model, we accept or reject certain trading signals. This random forest filter is able to improve the original model substantially.

SMU
SINGAPORE MANAGEMENT UNIVERSITY

Example 12: Prediction of Bankruptcy

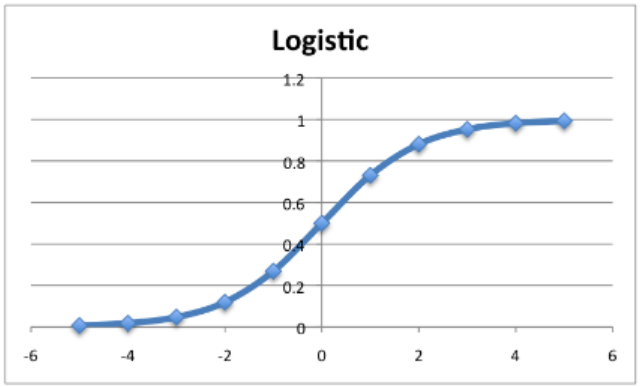
- The Altman Z-score was the first attempt to model risk of corporate bankruptcy.
- Ohlson used a 9 variable O-score which is more accurate.
- As logistic regression (or logit) is useful when the response is binary, it is used to model the probability of default in various settings (corporate default, credit card default, etc.) using different input variables that may be appropriate.
- Other Generalized linear models besides logit, e.g., probit can be used.
- “On a new logistic regression model for bankruptcy prediction” by Elena Belyaeva, Dec 2014

SMU
SINGAPORE MANAGEMENT UNIVERSITY

What is Logistic Regression?

- Logistic Regression is one of the most basic and important statistical techniques used in Machine Learning.
- It is closely related to neural networks, because the **logistic** function, also called the **Sigmoid** function, is heavily used as the ***activation function*** in Neural Networks. The logistic function is S-shaped and is defined as

$$\sigma(t) = \frac{e^t}{e^t + 1} = \frac{1}{1 + e^{-t}}$$



SMU
SINGAPORE MANAGEMENT UNIVERSITY

Logistic Regression-contd.

- Logistic function is very useful in modelling probabilities since its range is [0,1]
- Hence it is very useful in Binary Classification as it predicts *p*, the probability of a data point being in class 1. (And therefore also predicts 1, which is *1-p*)
  - If we define $t = \beta_0 + \beta_1 X$
  - Then the definition of logistics function becomes

$$\frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}}$$

SMU SINGAPORE MANAGEMENT UNIVERSITY

Softmax Function

- The **Softmax** function is a multinomial generalization of the Logistic function. It is useful when dealing with multi-category classification instead of binary classification. For example, classifying handwritten digits into one of ten classes.

$$\sigma(\mathbf{z})_j = \frac{e^{z_j}}{\sum_{k=1}^{K} e^{z_k}} \quad \text{for } j = 1, \ldots, K.$$

- The **Softmax** function "squashes" a *K*-dimensional vector ***z*** of arbitrary real values to a *K*-dimensional vector ***σ(z)*** of real values in the range (0, 1) that add up to 1.

SMU
SINGAPORE MANAGEMENT UNIVERSITY