QF 624: Machine Learning for Financial Applications

Introduction & Representation Learning

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

Saurabh Singal

July 2018

SMU
SINGAPORE MANAGEMENT UNIVERSITY

A word of caution

- TED talks are all limited to 18 minutes “because the brain is an energy hog. The average adult human brain only weighs about three pounds, but it consumes an inordinate amount of glucose, oxygen, and blood flow. …you cannot inspire people if you put them to sleep. But scientists are beginning to identify how long most people can pay attention before they tune out. The range seems to be in the area of 10 to 18 minutes…”
  – ***Carmine Gallo**, The Science Behind TED's 18-Minute Rule*
- We will proceed slowly.

SMU
SINGAPORE MANAGEMENT UNIVERSITY

What is Machine Learning?

I have yet to attend a "Introductory Machine Learning" talk which does not ask this question.

- In the beginning, statisticians and computer scientists approached similar problems within their own communities.
- Computer Science community called this discipline *Artificial Intelligence* or *Machine Intelligence* and statisticians called it *Statistical Learning*.
- Current understanding :

Machine Learning = Machine Intelligence+ Statistical Learning; programs that learn from data.

- In this lecture, Machine Learning will be used loosely and broadly to include a wide spectrum of statistical techniques.
- Deep Learning is the latest paradigm but the phrase "shallow learning" sounds very silly, so we will use the word "*Traditional Machine Learnin*g" to denote all "Machine Learning excl-Deep Learning".

SMU
SINGAPORE MANAGEMENT UNIVERSITY

Theme 1: Prediction by Traditional Machine Learning

- Many older Statistical Learning/Machine Learning techniques can be used for pattern recognition (clustering, k-NN, kernel regression) and prediction (SVM, Random Forest, Neural Networks).
- These methods work well on their own, but if the models were more aware of the context (or current market regime or state of the world) their effectiveness could be significantly enhanced.
- Hidden Markov Model can be used to make inference about the current market regime, and we can use trading models specific to current regime.
- News Sentiment Analysis :The plethora of news, and opinions on electronic social media can be mined to form an opinion about market sentiment. This can be automated. (Machine readable news, automated news sentiment analytics).

SMU
SINGAPORE MANAGEMENT UNIVERSITY

Market Sentiment vs Efficient Markets Hypothesis

“ …stock prices often move for irrational reasons. In general, this is because the stock market is driven by human action and naturally reflects the chronic irrationality within human nature. The Efficient Market Hypothesis is a pleasant academic fiction, although successful investing requires being mindful of the inefficient (or say “sentimental”) factors..”

Excerpt from Sentiment Reversals as Buy Signals by John Kittrell

SMU
SINGAPORE MANAGEMENT UNIVERSITY

Studies of News Sentiment using "older" Techniques

- Dow Jones News Analytics(DJNA) has constructed an archive back to 1987 that is not survivor-biased .
- The archive is made-up of tens of millions of financial news stories, where each story is assigned numerous DJNA attribution tags.
- Their basic unit of sentiment is the DJNA MCQ ranking. If a news story *N* mentions a company *C* in a negative light, then the company receives a negative MCQ ranking; if a news story mentions a company in a positive light, then the company receives a positive MCQ ranking. Additionally, a relevance score between 0-100 is also assigned to C depending on how relevant is this story to the company.
- A cumulative MCQ rank weighted by relevance score is useful.

SMU
SINGAPORE MANAGEMENT UNIVERSITY

Early work: Counting words in a Wall Street Journal column

- Paul Tetlock used a quantitative content analysis program, known as General Inquirer, to analyze correlations between sentimental values of words (as determined by the Harvard psychosocial dictionary) in The Wall Street Journal "Abreast of the Market" column and the performance of the Dow Jones Industrial Average.
- Result: media pessimism is shown to predict lower returns on the Dow, at least on the day the pessimistic news is published
- Limitations: inputs cannot be combination of words but only single or individual words; sentiment cannot be estimated by single words (not good, not bad…)

SMU
SINGAPORE MANAGEMENT UNIVERSITY

Dow Jones News Sentiment Analytics

- Dow Jones tied up with RavenPack to form products that rely on algorithmic text analysis and machine-learning to come up with products that can quantify news sentiment.
- These products will use financial news stories as inputs and use various linguistic classification techniques to output sentiment scores.
- Within milliseconds, each news story can be read by a computer program and assigned a quantitative sentiment score.
- These can be then aggregated per company, sector, topic or macro-economic category; we can form time averages of sentiment and create strategies that trade on reversals of sentiments, spike in sentiments etc.

SMU
SINGAPORE MANAGEMENT UNIVERSITY

Google Trends and Predicting Returns-1

- "Quantifying Trading Behavior in Financial Markets Using Google Trends" by Tobias Preis, Helen Moat and H.E. Stanley
- ***Abstract***: *".. By analyzing changes in Google query volumes for search terms related to finance, we find patterns that may be interpreted as ''early warning signs'' of stock market moves. Our results illustrate the potential that combining extensive behavioral data sets offers for a better understanding of collective human behavior."*

SMU
SINGAPORE MANAGEMENT UNIVERSITY

Google Trends and Predicting Returns-2

- "Do Google Trend Data Contain More Predictability than Price Returns?" By Damien Challet and Ahmed Ayed
- ***Abstract* :"***Using non-linear machine learning methods and a proper backtest procedure, we critically examine the claim that Google Trends can predict future price returns. We first review the many potential biases that may influence backtests with this kind of data positively, the choice of keywords being by far the greatest culprit. We then argue that the real question is whether such data contain more predictability than price returns themselves: our backtest yields a performance of about 17bps per week which only weakly depends on the kind of data on which predictors are based, i.e. either past price returns or Google Trends data, or both."*

SMU
SINGAPORE MANAGEMENT UNIVERSITY

Theme 2: Prediction by Deep Learning

- Deep Learning models have started to replace the traditional machine learning techniques that we just talked about.
- The class of networks called Recurrent Neural Networks can perform the same task as an HMM, and more flexibly and powerfully.
- A special type of RNN called LSTM (Long Short Term Memory) is used to model temporal dependence (time series prediction).
- Word Embedding models can be used for Sentiment Analysis.
- CNN's being good at feature-extraction, can aid word embedding models for natural language processing.
- Reinforcement Learning is inspired by psychology and can be used to devise dynamic strategies and actions suited to the environment.

SMU
SINGAPORE MANAGEMENT UNIVERSITY

Examples of Machine Learning Techniques and their uses

- Pattern Recognition: K-NN, kernel regression
- Factors that drive prices: PCA, ICA
- Modes of movement of yield curve: PCA
- Predict Asset direction: SVM, Logistic Regression/Neural Network, Lasso Regression
- Find Assets that move together: Manifold Embedding
- Detect market regime: Hidden Markov Model
- Sentiment of News Story: Bag of Words, Word Embedding +CNN
- What is the topic of an article: Term/Inverse Document Frequency

SMU
SINGAPORE MANAGEMENT UNIVERSITY

Some Examples of Traditional Machine Learning in Finance that we will cover today

- Clustering for Fund of Hedge Funds
- Forensic Accounting
- GMM for options pricing
- HMM for market regime detection
- News Sentiment Analytics
- FX prediction (Soulas and Shasha)
- Trading on Bigrams and Automated pattern recognition using Kernel Regression

SMU
SINGAPORE MANAGEMENT UNIVERSITY

Philosophies of Applying Machine Learning to Problems in Finance

Philosophy 1: Use Machine Learning inside a Massive black box

Example :

- ❑ Black box for predicting with millions of parallel, independent series as inputs;"ParCorr: Efficient Parallel Methods to Identify Similar Time Series Pairs across Sliding Windows"
  Djamel Edine Yagoubi, Reza Akbarinia, Boyan Kolev, Oleksandra Levchenko, Florent Masseglia, Patrick Valduriez and Dennis Shasha
  Data Mining and Knowledge Discovery, 2018

- ❑ A Deep Learning system based on Word2Vec + CNN+ Fully connected layers that detects sentiment using machine readable news.

SMU
SINGAPORE MANAGEMENT UNIVERSITY

Philosophies of Applying Machine Learning to Problems in Finance-2

Philosophy 2: Use Machine Learning in small doses sprinkled throughout the analytical workflow

- For applications to finance, we can use Data Science/Statistical Learning in a universal and broad way.
- Whenever we are faced with an analytical task which lends itself to measurement and data collection, we try to learn the structure embedded in the data, recognize the patterns, and as new data come in, revise our views and hypotheses.
- Example: Online SGD for learning coefficients as "Online Machine Learning Algorithms For Currency Exchange Prediction" *Eleftherios Soulas and Dennis Shasha*

SMU
SINGAPORE MANAGEMENT UNIVERSITY

Philosophies of Applying Machine Learning to Problems in Finance-3

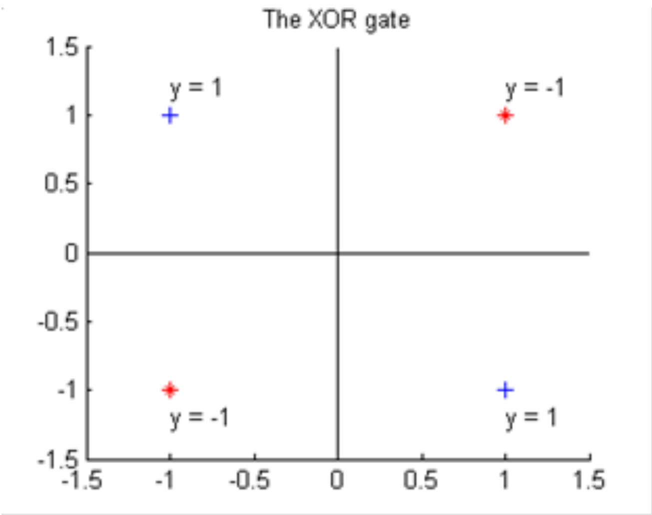
Philosophy 3: Keep it as simple as possible, but not more…

- Often, Occam's Razor is an important guide in all walks of life; and it especially important when fitting models to data.
- An extremely simple method called the *Theta Method* won an interesting competition about forecasting (the M3 competition) , beating many sophisticated machine learning techniques: https://flora.insead.edu/fichiersti_wp/inseadwp1999/99-70.pdf ; the forecasts obtained by the Theta Method are equivalent to simple exponential smoothing with drift. https://robjhyndman.com/papers/Theta.pdf
- That said, we should not hesitate to embrace more complex methods, (e.g., CNN for image recognition) if there is significant gain in performance (but we must be clear what we are doing and why).

SMU
SINGAPORE MANAGEMENT UNIVERSITY

Points which are not Linearly Separable

- A canonical example of data points that are not separable by a linear decision surface is the *XOR* function (*exclusive OR*).
- The *XOR* gate takes a value of 0 when both its inputs are the same and takes a value of 1 otherwise.



The XOR Gate.
There is no straight line that can separate the two classes.

Mapping to a Higher Dimension

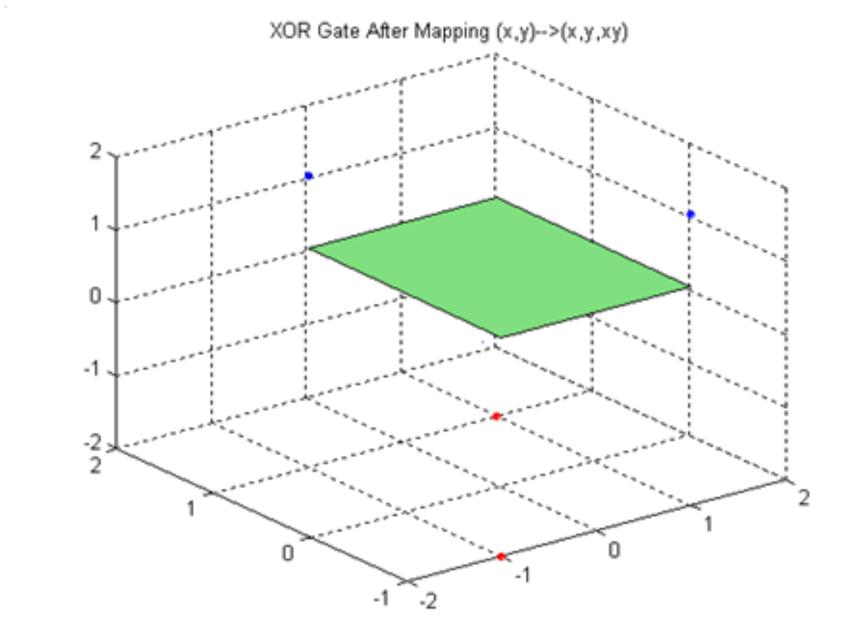
- These points cannot be linearly separated : we cannot draw any straight line that will separate the classes.
- But there is a trick we can use
- Define *x3= x1.x2* and then augment the 2 dimensional input vectors by this third number, *x3*. In the *XOR* example, we get

x=(x1,x2,x1.x2)	*y*
(1,1,1)	-1
(-1,-1,1)	-1
(1,-1,-1)	1
(-1,1,-1)	1

Table 2: The modified XOR Gate

SMU
SINGAPORE MANAGEMENT UNIVERSITY

Mapping to Higher Dimension-2



Separation of the XOR Gate After Mapping to Higher Dimension.

The Importance of Data Representation

One more very useful thing we should always think about is how to represent data.

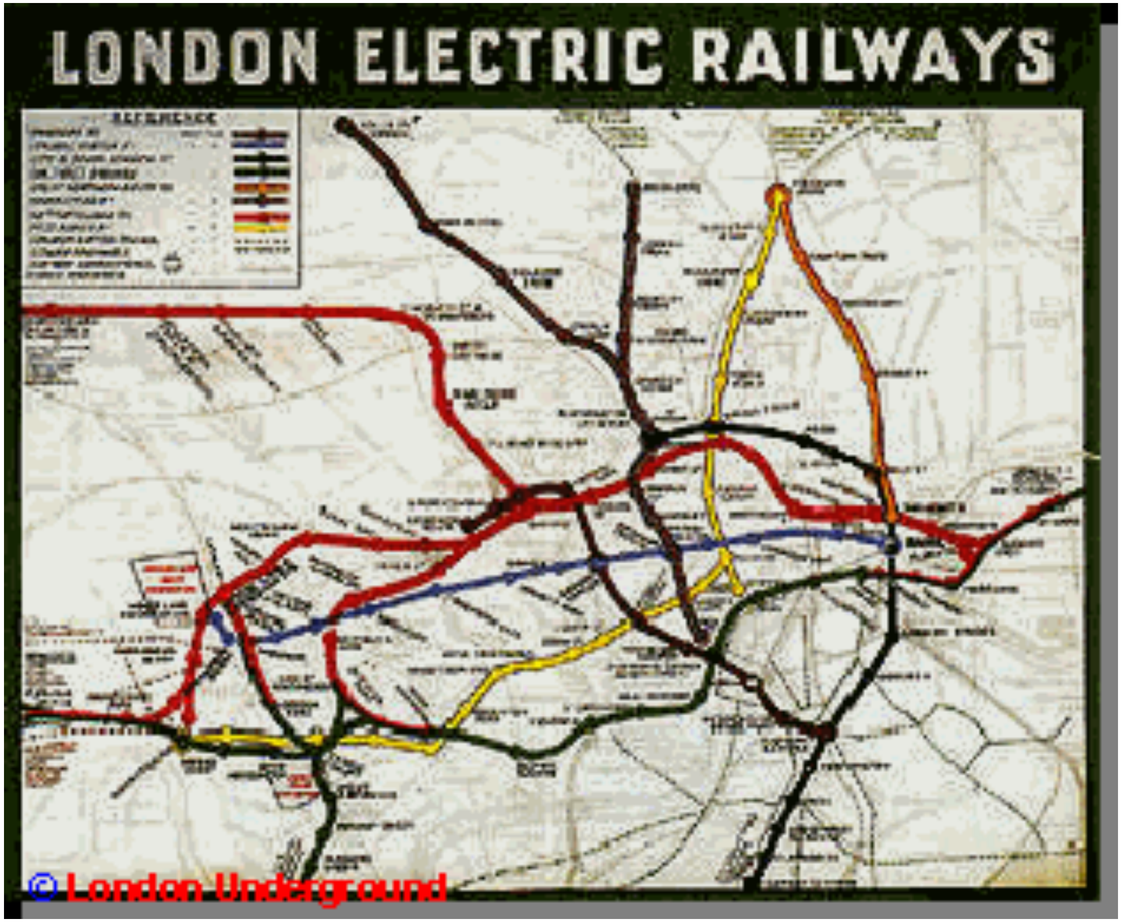
For example

- SVM can solve problems requiring non-linear separation, because of a clever idea called the kernel-trick.
- Google's word2vec represents each word as a vector in a vector space of dimension several hundred but words similar in meaning are close to each other in the vector space representation.
- ***No problem can be solved until it is reduced to some simple form. The changing of a vague difficulty into a specific concrete form is a very essential element in thinking. (attributed to J.P. Morgan)***

Abstract representations can allow more concrete visualization.

SMU
SINGAPORE MANAGEMENT UNIVERSITY

Real Representation, but Less Interpretable



Original Map of the London Underground Train ("tube") System

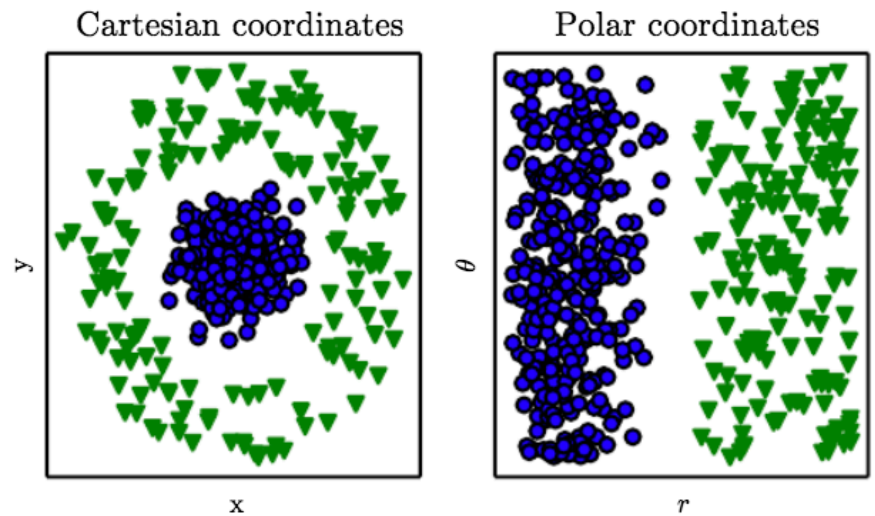
SMU
SINGAPORE MANAGEMENT UNIVERSITY

Less Real, but more Interpretable…

Harry Beck's Map of the London Underground Train ("tube") System

SMU
SINGAPORE MANAGEMENT UNIVERSITY

Representation Learning



Let us try to perform linear separation on the points.

- Left: Representation in Cartesian coordinates *(x,y)*
- Right: Representation in polar coordinates *(r,θ)*

(Source: Deep Learning by Goodman, Bengio & Courville)

SMU
SINGAPORE MANAGEMENT UNIVERSITY