Weekly Meeting Notes

Table of Contents:

Overview:

- P&L (profit & loss) is a metric used to analyze the effectiveness / profitability of an investment portfolio
- There has been some research in the area that covers P&L using neural networks which is a new approach
- Quantitative researchers find a new analysis for understanding merger arbitrage spreads
- K-means clustering is an unsupervised machine learning approach to clustering
- Within quantitative portfolio management we can use k-means clustering to identify which security to invest in
- Bayesian statistics is one of the widest used probabilistic methods and has tons of application to finance

Quant: Finding Alpha from its Signatures

Key facts

- Quants at JPMorgan working the academic researchers are trying to “signatures” on simulated market data
- They use synthetic data for
    - Machine learning algorithms for option hedging
    - Stress testing, scenario analysis and machine learning training data sets
- Blanka Horvath from King’s College is one of the leading researchers in the area
    - Her team is building a “regime classifier” to adjust options prices

JPMorgan Chase & Co.

Blanka Horvath from King’s College

- Another person working in this area is Maud Lemerceier of the University of Warwick
    - They use a mean reversion parameter in simulated market data to make a distribution regression with signature
    - Mean reversion is used in volatility models: Heston, Dupire, SABR
- The research is being tracked by quants at Citi and Standard Charter, Catley Lakeman Securities
- Some quants think that they can find alpha signals in this by looking at how the coefficients change over time
    - In technical analysis terms this would help them build more rigorous models for head and shoulder patterns
    - The coefficients can possibly be learned by machine learning models which is why they use LSTMs

The concerns

- The computational complexity may to big for most computers
    - The computational time may be to slow or impossible
    - There could be machine learning problems that look at solving this
- The signals are prone to
    - Describing relationships that aren't actually there
- But the general consensus is that buy side and sell side will both benefit from this advancement

What the signature helps solve

- Receiving tick data is hard to work with because you have to view that data as more of a stream and not individual pieces
- Signatures would let you see the data as a path rather than a sequence of point-in-time
- First order signature measure the drift up or down for the sequence
- Second order signatures measure the volatility path
- Higher orders go beyond what can be described
- It also works with data that is missing information

- This could also lead to firms needing less computing power and data storage to maintain all of the information if they can track signatures
- But there is question about going from signatures back to the data
- The main problems that quants have "overfitting"
  - Its when the model can't distinguish between important information and "noise" which is a big problem that plagues unsupervised learning
- The main things that quants want to work out is that signatures may be impossible to compute
  - The way that quants do this is that they usually truncate the signatures into 3 or 4 orders which may lose critical information
  - The loss of critical information may have problems with heavy-tailed data practitioners because they may lose those tail events

Articles:

Risk.net: 'Signatures' promise quants a tool for all jobs here

ArXiv: Distribution Regression for Sequential data here

Risk.net: Synthetic data enters its Cubist phase here

Finance: P&L

Key facts

- Quants are very interested in the future P&L distribution of a portfolio
  - It allows them to manage risk
  - Set aside capital for regulatory reasons
- There are many different types of P&L modelling strategies
  - Parametric modelling
  - Linear and quadratic mapping on risk factors
  - Closed-form analytic approximations

What makes calculating P&L tough

- We may have to use a system of monte carlo simulations that are computationally expensive
- It's much harder to work with when using nonlinear payoffs of path dependent derivatives (American Options)
- The original method was inspired by the least square Monte Carlo (LSM) to estimate via back propagation
- Then came an optimal strategy using polynomial interpolation

Finance: Calculating P&L using Neural Networks

Key facts

- Quants are very interested in the future P&L distribution of a portfolio
  - It allows them to manage risk
  - Set aside capital for regulatory reasons
- There are many different types of P&L modelling strategies
  - Parametric modelling
  - Linear and quadratic mapping on risk factors
  - Closed-form analytic approximations

What makes calculating P&L tough

- We may have to use a system of monte carlo simulations that are computationally expensive
- It's much harder to work with when using nonlinear payoffs of path dependent derivatives (American Options)
- The original method was inspired by the least square Monte Carlo (LSM) to estimate via back propagation
- Then came an optimal strategy using polynomial interpolation

Quantitative Finance News: Merger Arbitrage Spread

- Versors Investments looked into the merger arbitrage spreads to analyze them.
- During a merger an acquirer purchases a target at a premium usually expressed in share price.
    - That premium is usually higher than the current share price
    - There is an arbitrage possibility by going long on the target and short on the acquirer. When the acquirer pays the premium you swap the shares at a higher price making an arbitrage
    - A key to merger arbitrage is assuming that the deal will be completed
- Market practitioners and other researchers look at the spread (premium - share price) to give them an idea on what will happen

VERSOR

Analyzing the spread

- Historically if the spread is wider, then it is usually a sign that the merger may fail and vice versa
- Another way of thinking about it
    - If investors think that the merger will be completed they'll buy the share hoping to swap it at the premium price
    - As more investors buy the share they push the price up
    - If they didn't think the deal would go through they wouldn't buy the shares

Versor's approach

- They look at around 4,000 between companies based in US, Canada, UK, and Europe mergers and kept track of
    - Probability that the merger will close
    - Determine downside risk
    - Perform competing bid analysis
- Harford (1999) shows that cash rich firms are more likely to attempt acquisitions
- Another thing to consider is that private equity funds have to spend their cash therefore their dry powder has to be used
- Typical failure rates are around 10%

Versor's findings

- They found that 80% of the spread has no usefulness as a predictor of merger success
- They say that 80% the spread encapsulates a series of information
  - region
  - Nature of deal
- They also found that deals done by private equity backers during times of stress tend to fail more often than other deals
- They also found that it is harder to predict success involving a company based in emerging market

Versor's machine learning approach

- Thye use 2 different undisclosed machine learning algorithms to create a forecast
- They also use Natural Language Processing to analyze news and update their database

<u>Articles:</u>

Risk.net: Machines say: 'Ignore the spread in merger arb' here

Versor Investments: The environment for merger arbitrage: 2021 here

Computer Science: K-means Clustering

Overview:

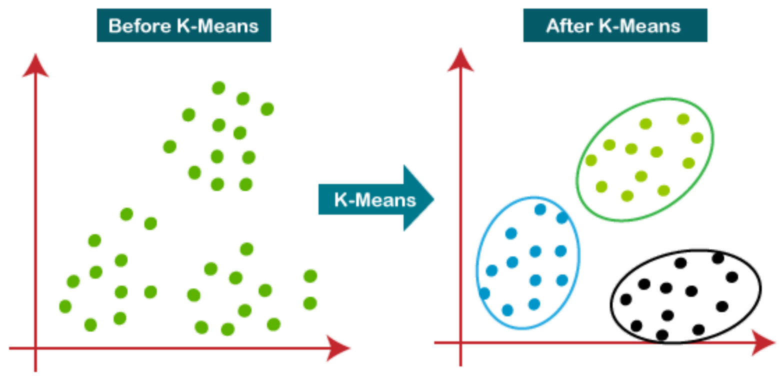
- Clustering algorithm
- Unsupervised learning
- Dimensionality reduction tool

The K-means clustering algorithm is used to find groups which have not been explicitly labeled in the data.

- Easy to apply to even large data sets

- Common data analysis technique to get an intuition about data structure

- Given a set of observations (x1, x2, ..., xn), where each observation is a d-dimensional real vector, k-means clustering aims to partition the n observations into k (≤ n) sets S = {S1, S2, ..., Sk}



Sklearn is the most common python package for kmeans

Algorithms:

- Naive k-means: most commonly use it is defining each cluster via the least squared Euclidean Distance

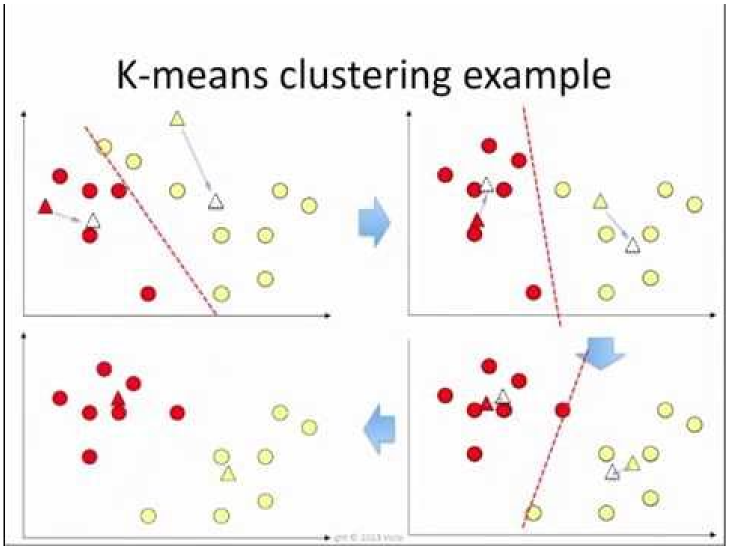
- Hartigan-Wong Method - this locals at finding the local minimum of the the minimum sum-of-squares problems

- Variations only converge to a local minima of minimum-sum-of-squares cluster problem

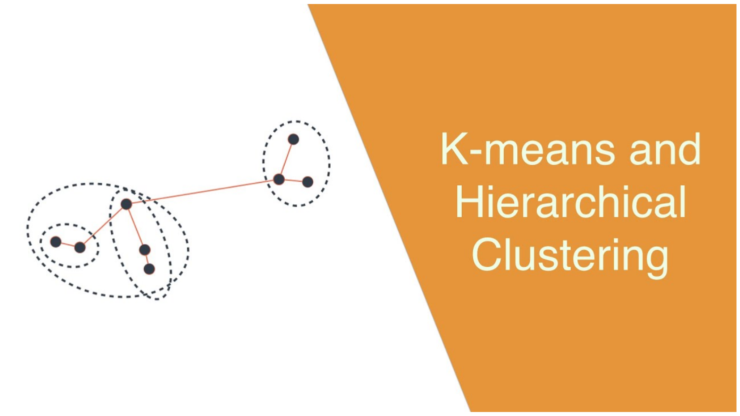
- And there are many other variations

- Standard Naive K-means algorithm alternates between an assignment step and an update step
    - Assignment: assign observations to clusters with nearest mean (euclidean distance)
    - Update: Recalculate means

Victor Lavernko: K-means clustering: how it works



Lusi Serrano: Clustering: K-means and Hierarchical



StatQuest: K-means clustering

Codebasics: Machine Learning Tutorial Python - 13: K Means Clustering Algorithm

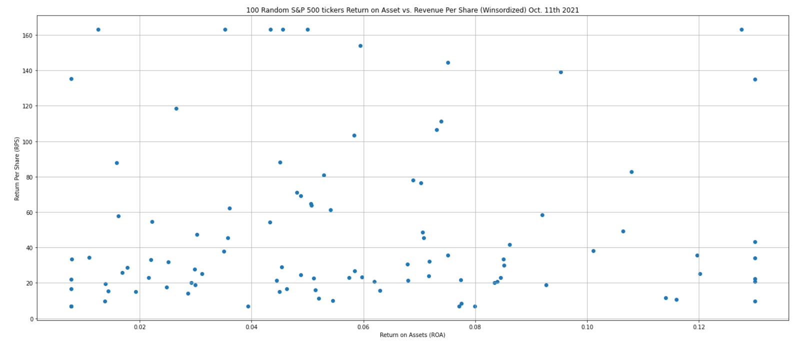
Kmeans clustering file (here)

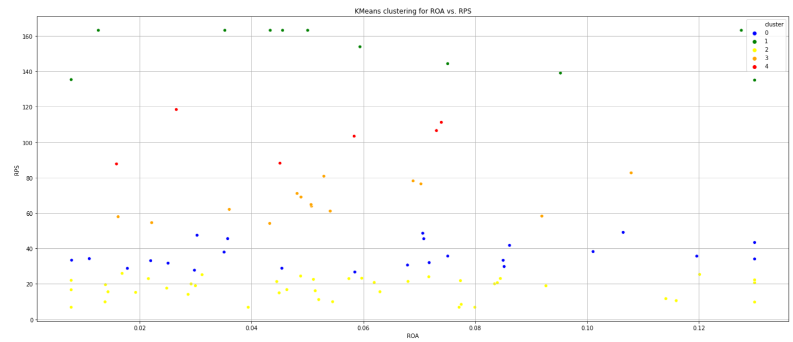
Quant: K-Means clustering allocation

The goal

- Clustering algorithms have ton of application in asset allocation models
- Allocation strategies using clustering algorithms would be
    - Principal component analysis to allocation positions that best “offset” each other or their eigenvectors are orthogonal
    - Kmeans clustering to find best securities that are the best “centroids” for each cluster

Kmeans clustering example





The method used in this example

- The goal of this is to find a way to make a “scatter” plot of each stock. In this case we used
    - Return on Asset = Net Income / Net Assets
    - Return per Share = Revenue / Common Shares
- It can really be any calculation, and with python’s yfinance API they offer a whole list of statistics that we can pull

- In this case we were able to separate the securities and then find which one is best
- The model is good at picking securities, but there isn’t framework for how much to allocate
- Of course that can be solved via any allocation optimization model

Python file:

Data collection file (here)

Kmeans allocation jupyter notebook (here)

Computer Science: Bayesian Statistics

Bayesian inference is a method of statistical inference in which Bayes' theorem is used to update the probability for a hypothesis as more information becomes available.

- The posterior (new) probability is a consequence of two antecedents:
  - A prior (old) probability
  - And a likelihood function

- H is the hypothesis
- P(H) is the prior probability
- P(E | H) is the likelihood function
- E is the evidence corresponding to new data
- P(E) is the model evidence

Bayes' Theorem

$$P(H \mid E) = \frac{P(E \mid H) \cdot P(H)}{P(E)}$$

Although it looks easy, it is one of the cornerstones to probabilistic methods.

- In finance, Bayesian Inference has been applied to problems of prediction such as how changes in interest rates affect the value of an index

- Bayesian probability models for forecasting are liked due to logical rigor, general reliability, and intuition

- Some issues with Bayes:
  - Choice of prior takes work
  - Models involving many variables are computationally intensive
  - Posterior distributions are difficult to incorporate
  - Predictions are not always precise (there is room for error)

Bayesian statistics notebook (here)

The Content is for informational purposes only, you should not construe any such information or other material as legal, tax, investment, financial, or other advice. Nothing contained on our Site constitutes a solicitation, recommendation, endorsement, or offer by CU Quants or any third party service provider to buy or sell any securities or other financial instruments in this or in in any other jurisdiction in which such solicitation or offer would be unlawful under the securities laws of such jurisdiction.

All Content on this site is information of a general nature and does not address the circumstances of any particular individual or entity. Nothing in the Site constitutes professional and/or financial advice, nor does any information on the Site constitute a comprehensive or complete statement of the matters discussed or the law relating thereto. CU Quants is not a fiduciary by virtue of any person's use of or access to the Site or Content. You alone assume the sole responsibility of evaluating the merits and risks associated with the use of any information or other Content on the Site before making any decisions based on such information or other Content. In exchange for using the Site, you agree not to hold CU Quants, its affiliates or any third party service provider liable for any possible claim for damages arising from any decision you make based on information or other Content made available to you through the Site.

There are risks associated with investing in securities. Investing in stocks, bonds, exchange traded funds, mutual funds, and money market funds involve risk of loss. Loss of principal is possible. Some high risk investments may use leverage, which will accentuate gains & losses. Foreign investing involves special risks, including a greater volatility and political, economic and currency risks and differences in accounting methods. A security's or a firm's past investment performance is not a guarantee or predictor of future investment performance.

As a convenience to you, CU Quants may provide hyperlinks to web sites operated by third parties. When you select these hyperlinks you will be leaving the CU Quants site. Because CU Quants has no control over such sites or their content, CU Quants is not responsible for the availability of such external sites or their content, and CU Quants does not adopt, endorse or nor is responsible or liable for any such sites or content, including advertising, products or other materials, on or available through such sites or resources. Other web sites may provide links to the Site or Content with or without our authorization. CU Quants does not endorse such sites and shall not be responsible or liable for any links from those sites to the Site or Content, or for any content, advertising, products or other materials available on or through such other sites, or any loss or damages incurred in connection therewith. CU Quants may, in its sole discretion, block links to the Site and Content without prior notice.

YOUR USE OF THIRD PARTY WEB SITES AND CONTENT, INCLUDING WITHOUT LIMITATION, YOUR USE OF ANY INFORMATION, DATA, ADVERTISING, PRODUCTS, OR OTHER MATERIALS ON OR AVAILABLE THROUGH SUCH WEB SITES, IS AT YOUR OWN RISK AND IS SUBJECT TO THEIR TERMS OF USE.