**Eloy’s ideas:**

**1**

Replicate paper: A SVM APPROACH TO STOCK TRADING <https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=2&cad=rja&uact=8&ved=0ahUKEwjXyIvw8avaAhUHKqwKHbOkAAsQFgg9MAE&url=http%3A%2F%2Fcs229.stanford.edu%2Fproj2009%2FKingVandrotWeng.pdf&usg=AOvVaw2SD-lEstrWTbunS9Vs-JqC>

This paper uses SVM to create a long/short trading strategy investing only in the 7 oil firms of the SP500 index. We could replicate the paper, but based on another industry and see if the strategy holds.

Data needed: Stock prices (Bloomberg) and accounting data (Datastream). We should be able to find all the variables needed, but I would check before we jump to the pool. (Spanish expression literally translated, no idea if Americans understand it).

**2**

Replicate paper: Quantifying Trading Behavior in Financial Markets Using Google Trends

<https://www.nature.com/articles/srep01684>

This is the paper that we replicated in ML1 HW2 Problem 3. We could try to do the same, but instead of predicting DJIA returns, cryptocurrency returns. If it is not possible to predict returns, we could try to predict volatility or volume.

Data needed: Some cryptocurrency prices and Google trends data (should be available online)

**3**

Do as ML1 HW6 Problem 4 Section e). Use SVM to find anomalies in the price of cryptocurrencies.

**4 This one may be too much wishful thinking**

Use SVM to predict when the following trade will occur. We have a set of bids and asks. We cross-validate the C that would better separate the two groups of points (bids and asks). Based on the magnitude of C, estimate when the next trade will occur.

FX ML ideas applied to Crypto:

**5 Replicate (/build on) paper**

**Analysis of a network structure of the foreign currency exchange market**

analyze structure of the world foreign currency exchange (FX) market viewed as a network .

Group together all the exchange rates with a common base currency and study each group separately. Analyze the temporal evolution of the network and detect that its structure is not stable over time. **Apply to crypto: across different currencies and exchanges.**

[**https://arxiv.org/pdf/0906.0480.pdf**](https://arxiv.org/pdf/0906.0480.pdf)

**6 Bootstrap Methods for FX prediction.**

<http://ieeexplore.ieee.org/document/4371141/?reload=true>

**7 - Justin - Spectral Clustering (Can be combined with #5)**

* **Data:** A *time x exchange x asset* dataset for cryptocurrencies, physical commodities, and/or traditional currencies (or whatever asset classes we want). These are time series with a high level of granularity (hopefully hourly) for each *asset,exchange* pair.
* **Method/Idea:** Use spectral clustering, which is an improvement on the clustering we have seen in class so that you are viewing the data as a graph with vertices and edges, to partition the data into clusters. I was thinking we would have a rolling window on which we get a similarity measure for all the the *asset,exchange* pairs. An interesting one I found is called TWED (see: [paper on similarity measures for time series](https://arxiv.org/pdf/1401.3973.pdf)) but there are others we could use like radial basis or a simple [correlation measure](https://en.wikipedia.org/wiki/Canonical_correlation). In the paper, TWED performs the best classification tests. The similarity matrix gets fed into the spectral clustering method and you get a graph and a clustering of the vertices in the graph. If we wanted to extend it to use a neural net we could do something like [this paper](https://arxiv.org/pdf/1801.01587.pdf).
* **Combination with #5:** I was thinking we could use the paper and the idea from #5 to see how the network changes over time between different currencies and/or exchanges and then see if we could come up with some type of actionable idea for what we do when the structure changes. If it’s something like go long x and short x and z we could backtest that strategy to see if it is plausible in the market.
* **Overall Questions:** Can we cluster financial time series better than the methods we learned in class? Do specific markets lend themselves to better classification than others and why? Do the networks exhibited in these markets change over time in a specific way? Can we take action on this information that is profitable?

**Andrew’s ideas:**

1. **Anomaly Detection of Crypto Time Series using Support Vector Machines**
   1. Collect time series data for Ethereum/USD exchange rates across top 6 most liquid exchanges for the past 6 months
   2. Find a way to identify when significant divergences occur by using Support Vector Machines (akin to final problem of Statistical Machine Learning 1, Homework 4, to identify anomolies)
   3. Refine using the windopane method in which a new normalization occurs on a weekly or shorter basis and anomalies are detected from the new normal
   4. This has real world application in helping to develop a statistical arbitrage strategy with trading signals developed by an anomaly detected method
2. **Using K-means, and more advanced clustering methods to identify the relationships and dissimilarities of Ethereum with other currencies (crypto and fiat) as well as commodities and major equity exchanges**
   1. Analyze the time series of various currencies and other securities to see which might be clustered as similar or different for the importance of finding buy signals in the data for future cryptocurrency movement
   2. Find those assets that have distinct time series which might serve as an effective hedge or make ethereum and effective diversifying security
3. **Time Lag, identifying tag lag in a signal and how a change in pricing on one exchange changes across the various exchanges**

Eloy’s thoughts:

Andrew 2-3. I would not go for anything that we do not have a prior paper/previous work where to base our project on.

Andrew 3. I do not know how would ML be applied there.

6. I like it. The only con that we have not studied this neural networks yet.

5. I like it. Just worried that the methods used are ML and wonder if we will be able to implement them.

7. I like it. I just hope it will not be too complex.