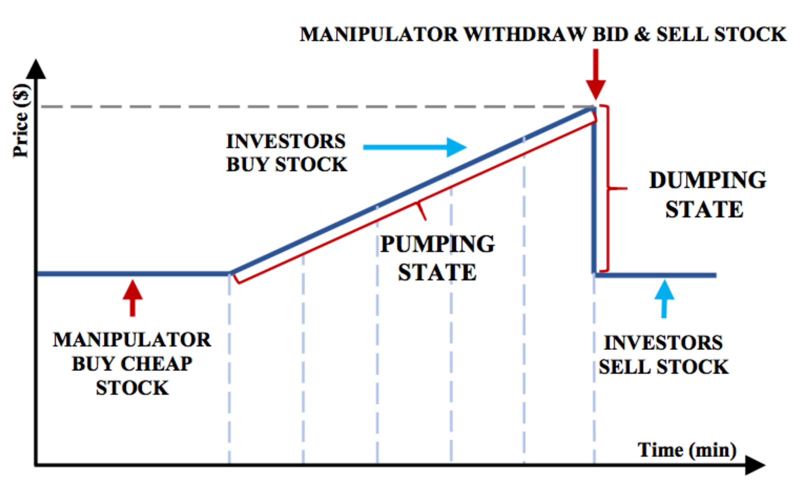
Predicting Market Movements with Public Order Data

CS229 Fall 2017 — Cameron Ramos [calramos@stanford.edu], Noah Golub [ngolub@stanford.edu]

Motivation

- We want to explore the possibility of predicting cryptocurrency market movements, a market with massive volatility and potential upside
- Detecting abnormal price increases makes for a good "pump and dump" detection heuristic, in which actors attempt to artificially inflate the price
- Artificial price increase is hard to detect because order cancellation data is not provided by exchanges



Typical valuation manipulation pattern (T. Leangarun, P. Tangamchit and S. Thajchayapong)

Data

- Order books were scraped from a cryptocurrency exchange called Poloniex, with each row containing a set of $\mathcal{N}$ evenly spaced orders in a 24 hour period
- Rows contain order type (buy or sell), unix timestamp, order volume, and rate
- Each data point is labeled with a ground truth indicating whether market data was associated with a non-trivial price spike (above 15%)
- Currencies considered include Monero, Ripple, Dash, Bitcoin, Litecoin, and Ethereum

Features

- Each input has 3 * $\mathcal{N}$ features representing $\mathcal{N}$ orders (order type, timestamp, volume)
- Timestamps were normalized to index within a 24 hour period
- The price feature was removed in order to train a model that was price agonistic (we wanted to see if it was possible to predict price spikes based on only volume and sell type)

Models

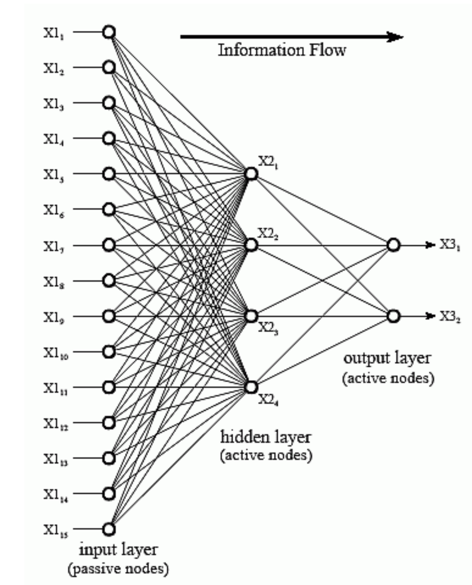
- **SVM** as a fast naive model to gather baseline results.
  - Implemented with sklearn's SVM module and C-Support Vector Classification with the penalty parameter, *C*, left at 1
  - perfect for our large feature dimensionality and relatively limited dataset

Given training vectors $x_i \in \mathbb{R}^p$, i=1,..., n, in two classes, and a vector $y \in \{1, -1\}^n$, SVC solves the following primal problem:

$$\min_{w,b,\zeta} \frac{1}{2} w^T w + C \sum_{i=1}^{n} \zeta_i$$
$$\text{subject to } y_i(w^T \phi(x_i) + b) \geq 1 - \zeta_i,$$
$$\zeta_i \geq 0, i = 1, ..., n$$

sklearn C-Support Vector Classification notes

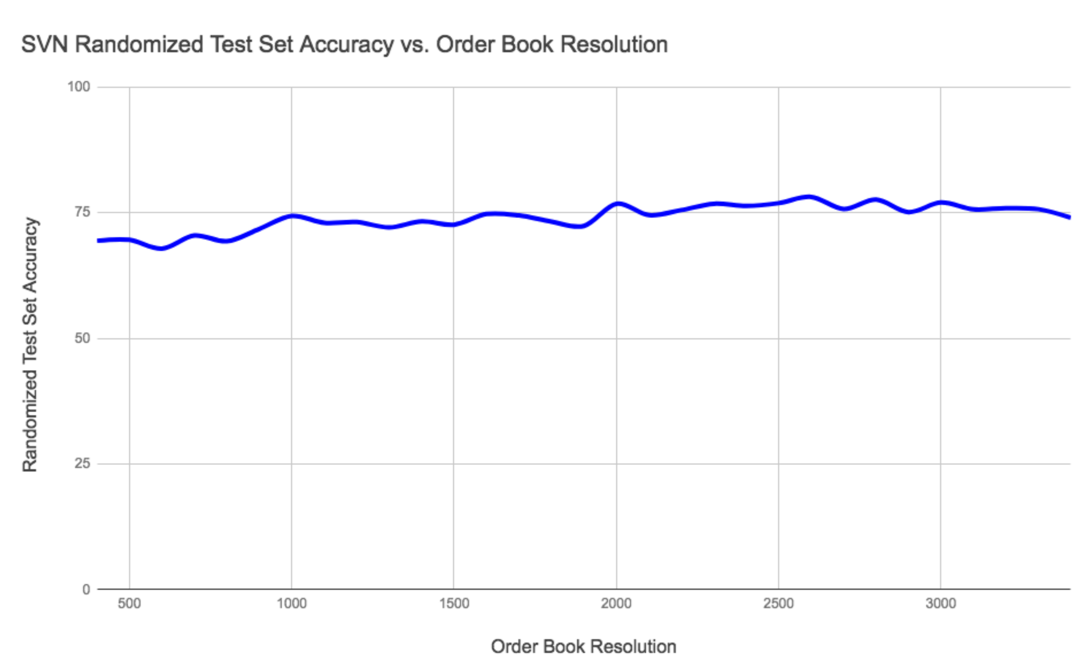
- **Neural Net** to predict 24 hour price movements from 12 hours of order data
  - Implemented with TensorFlow's DNN classifier
  - 3 hidden layers or respective sizes: NUM_ORDERS, 20, 2

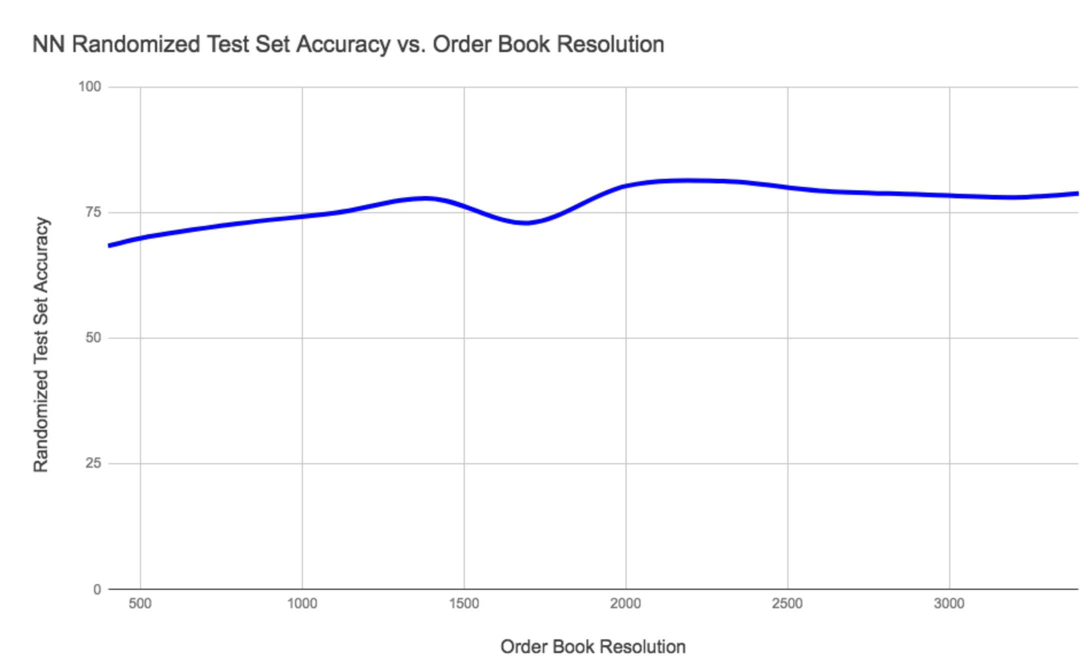


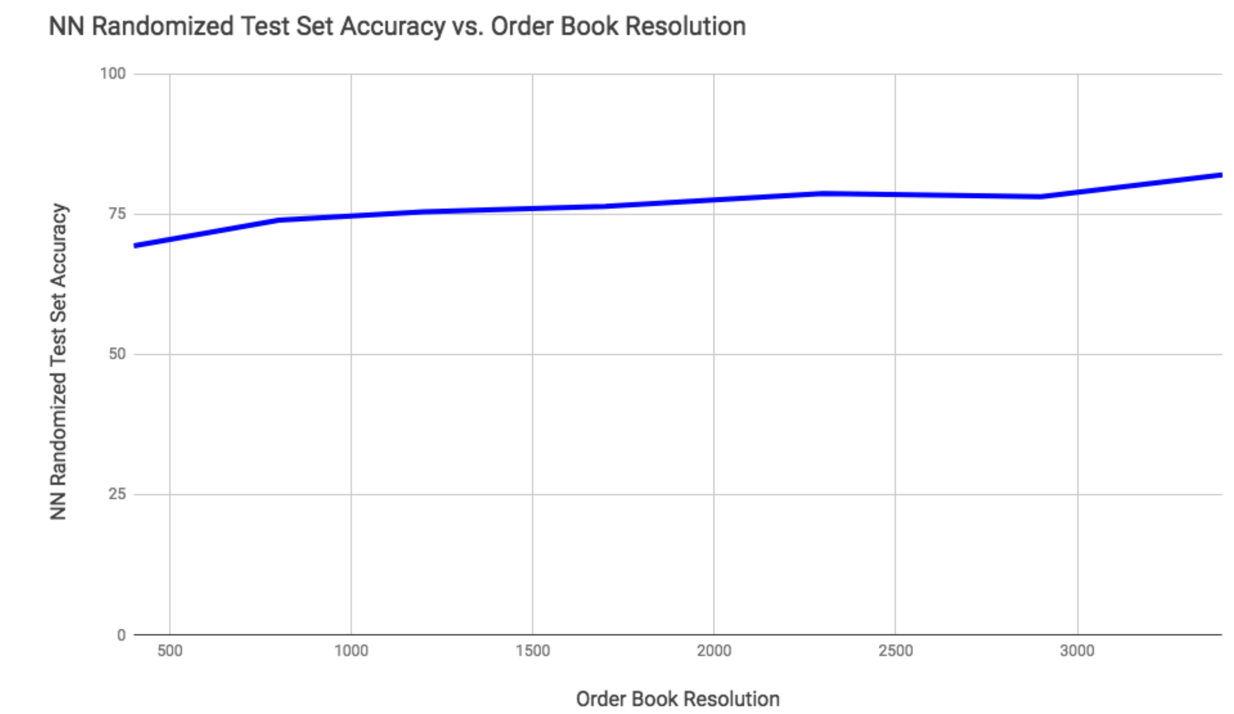
Results

Model	Order Book Size	Avg. Training Accuracy	Avg. Test Accuracy	Order Book Size
SVM (24 hour training)	2600	94.6	78.13	2600
NN (24 hour training)	2300	82.8	81.245	2300
NN (12 hour training)	3400	78.2	82.5	3400

- Results where gathered for a range of order book sizes and averaged over 9 runs using a randomized training and test set of equal size
- Table includes top results for each model. Figures below depict accuracy over the entire range for each.







Discussion

- It is possible to predict 24 hour price movement with a limited view of the order-book
- Results were better than expected, especially for the SVM which cannot capture complex feature relationships
- A higher resolution order book (with some exceptions) tended to improve results. This was true even though higher resolution order books reduced the size of our dataset.
- Our model is rate invariant! This suggests that price movements can be solely predicted based on the volume and type of orders placed

Future Work

- **Tracking Order Cancellations.** Our dataset does not track whether orders are cancelled, a feature that indicated wether a "pump and dump" is occurring
- **Incorporating Model Into Trading Strategy.** Our model can be used in an MDP like trading strategy that chooses actions based on confidence that asset price will rise.
- **Leverage our model** against a larger number of exchanges

References

T. Leangarun, P. Tangamchit and S. Thajchayapong, "Stock price manipulation detection using a computational neural network model," *2016 Eighth International Conference on Advanced Computational Intelligence (ICACI)*, Chiang Mai, 2016, pp. 337-341.