Bitcoin Price Prediction Using Bitcoin Ledger Data

Spenser Anderson (aspenser), John Mern (jmern91), John Poothokaran (johnmp) | Stanford University CS 229

Motivation

The Bitcoin (BTC) cryptocurrency is rapidly emerging as an important currency. All BTC transactions are logged in a publicly available ledger. This gives the opportunity to utilize this ledger to extract features to help predict the price of BTC. We train classifiers to inform daily buy-sell decisions for BTC using this network data.

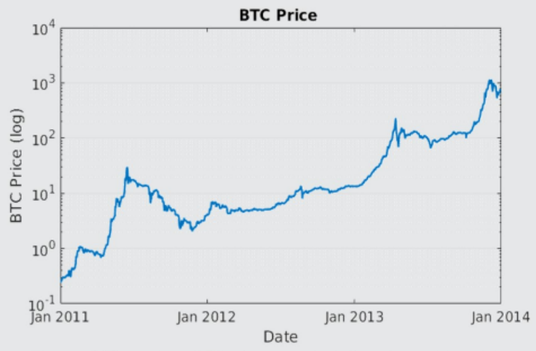

Figure 1: The price of Bitcoin over the interval considered here (3 years starting Jan 2011)

Data

We used the following data on each day:

- Daily closing price for major market indices (SP500, DJIA, VIX, Nikkei 225, Stoxx 600, currency indices)
- Classic financial technical indicators (moving average convergence divergence, etc.)
- Google trends for search terms "Bitcoin," "Bitcoin News," and "Bitcoin Price"
- Data extracted from the Bitcoin ledger network graph as PCA vectors of the transaction adjacency graph
- To predict price change on day *n*, all of these were provided for days *n-1* and *n-2*

Approach

Classification models were trained to predict whether the price would rise or drop the next day. The models surviving an initial weeding process were a **logistic classifier**, a **multilayer perceptron classifier**, and a s**upport-vector classifier.**

Models were evaluated based on ROI, assuming that earned funds so far were used to buy BTC for 1 day at a time each day the model predicted a price rise (and selling at end of day).

Model and Feature Selection

The impact of the following feature categories was studied:

- Data extracted from the bitcoin ledger network
- Google trends information
- Market features (indices, financial indicators, etc)

The logistic classifier (polynomial in all features) performed best.

Network data and Google trends brought substantial improvements in ROI, as seen below.

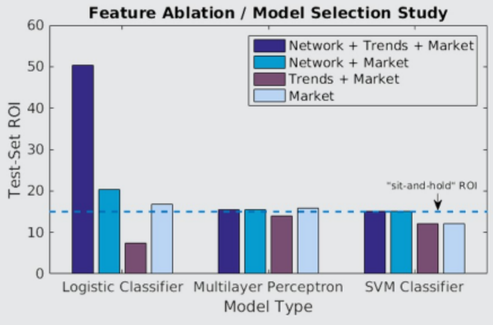

Figure 2: Comparing various feature combinations and model types to a "sit-and-hold" BTC trading strategy

Results

The test-set classification accuracy and ROI are below:

Model	Test Accuracy
Logistic	0.6539
MLP	0.68367
SVM	0.6734

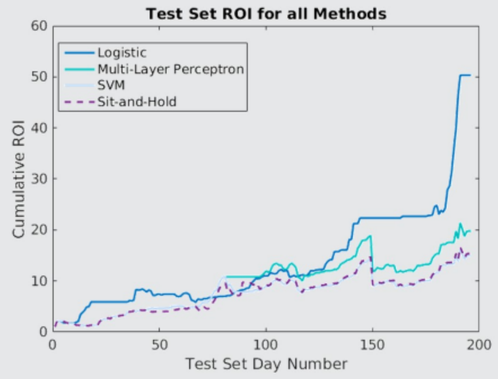

Figure 3: Comparing cumulative test ROI alongside growth of BTC.

Discussion and Future Work

All models shown were able to "beat the market" in the test set, a successful outcome. The network data appears to be a major advantage for predicting BTC price.

Future work should explore more advanced neural network models. A recurrent neural net would be well suited to a time series such as BTC price.

More sophisticated network data would also be worth pursuing, such as auto-encoder based dimensionality reduction.

References

- Kondor, D., Posfai, M., Csabai, I. & Vattay, G. Do the rich get richer? An empirical analysis of the bitcoin transaction network. *PLOS ONE* **9**, 1-10 (2014).
- Baumann, A., Fabian, B. & Lischke, M. *Exploring the Bitcoin Network*, vol. 1 (2014).
- Kondor, D., Csabai, I., Szule, J., Posfai, M., & Vattay, G. Inferring the interplay between network structure and market effects in Bitcoin. *New J. Phys.* **16**, **125003 (2014).**