

Analysis and Prediction of Bitcoin Price Development Using Linear Regression and Neural Networks

Katja Bouman & Tom Lotze

Background:

In 2008 the Bitcoin (BTC) was introduced, a revolutionary cryptocurrency that uses the Blockchain to enable a peer-to-peer, decentralized financial system. The popularity of not only Bitcoin, but all kinds of cryptocurrencies has skyrocketed over the past 5 years. The price of 1 BTC was \$1000 in the beginning of 2017, and recently the price reached \$11K.

However, a drawback for the actual implementation of BTC as a functional currency is the very fluctuating value. At the same time, the volatile price developments enable huge profits, but also losses. Crucial for investing is optimal timing considering buying and selling. In this project, we will analyze the historical price data of Bitcoin and with that, we hope to be able to train different classifiers and predict the future Bitcoin price.

Data:

For this research, we have found a dataset that covers the timespan 2011-now, and it includes per day the following attributes: ['Date', 'Open', 'High', 'Low', 'Close', 'Volume (BTC)', 'Volume (Currency)', 'Weighted Price']. Another possibly valuable attribute to use would be the moving average. The moving average is the average price of one Bitcoin over a specific time period, such as 5 or 10 days. The data contains 2276 data points and the prices are given in US dollars. We plan to split this set into a training set, cross-validation set and a test set in the ratio 60 - 20 - 20.

Classifier:

Considering output, since predicting an exact price (continuous) is very prone to overfitting, we will also predict the BTC price in terms of different classes. If x represents the percentual increase we could distinguish the following classes: such as stable ($-1\% < x < 1\%$), small increase ($1\% < x < 10\%$), small decrease ($-10\% < x < -1\%$), big increase ($x > 10\%$) or big decrease ($x < -10\%$). For this output, we would either use logistic regression, neural networks, or a discretized form of decision trees. Most likely we will train all of the mentioned classifiers and compare their performances on different situations and outcome variables (using confusion matrix and recall and precision scores)