

Predicting the Price of Bitcoin with Machine Learning

CS6350 Machine Learning Spring 2019 Final Project

Caelan Dailey u0881215

April 27th, 2019

GitHub: <https://github.com/caelandailey/CS6350>

Introduction

Bitcoin is a cryptocurrency which utilizes a worldwide peer to peer distributed network protocol that allows two parties to make an instantaneous online transaction without a middle man. The price of bitcoin and other digital currencies fluctuate constantly, especially with the recent increase in popularity. The cryptocurrency market has a market cap that is currently hovering at around 130 billion US dollars and on any given day can easily achieve at least 15 billion US dollars in transactions. Many individuals attempt to trade the currency to make a profit. However, the price is extremely volatile and it's difficult to make an accurate prediction of the true price.



Figure 1: Price of Bitcoin from 2017 to 2019.

Motivation

There are hundreds of different market indicators that traders use to attempt to gain additional insight on certain properties of the market such as Bollinger Bands, which show volatility, or the Ichimoku Cloud, which indicates future price momentum. However, there are too many factors and influences on the market that make it difficult for humans to make an accurate prediction or to utilize all the various market indicators at once. Many of the indicators are fairly simple and not computationally complex. Machine learning models provide a unique way to predict the future price of Bitcoin. It makes it possible to combine all the indicators into one simple market indicator that solely predicts the price using a price movement model it creates based on the historical price data since bitcoin started being traded.

Data

The data for the project is minute to minute updates of the price of Bitcoin dating back to the start of 2012 and goes up to April 2019. This price data is based on the cryptocurrency exchange BitStamp. The categories within the data are as follows, timestamp, open price, high price, low price, close price, volume in bitcoin, volume in USD, and weighted price. Timestamps are in Unix time. Some of the data fields are filled in with NaNs, which means there was not trade activity at that time, or the exchange api was not functioning for one reason or another. These values were simply ignored.

<https://www.kaggle.com/mczielinski/bitcoin-historical-data/data>

Solution

The project created a prediction model of the price of bitcoin by utilizing a neural network model. It used a 2 layer long short term memory. It used gated recurrent unit architecture from a recurrent neural network. It's designed to discover patterns within series of data such as stock markets price data. Since our network is recurrent, every time new data is introduced to the system, there is a feedback loop which propagates through all of the previous predictions. This is perfect for the type of data present in markets and is extremely effective for finding chain-like structures within a list of bitcoin price data.

The model used a 2 layer long short term memory with an adam optimizer and MSE for loss. The data is not shuffled since we need our data in a series and each data position is important. Figure 2 above shows a mapping of the amount of loss compared to the epoch number. It shows that when epoch is around 5, the loss for both training and test is minimized.



Figure 2: Training and test loss.

Experimental Evaluation

In order to explore certain parameters of the price data, some initial data model designing was needed. It's important to understand whether our data is affected by trend or seasonality. This is also helpful because it can be used as a baseline, since this is the trend without any learning or prediction and is simply an average of the price movements. In figure 2 below, the blue line is the price of bitcoin, the orange is the average of bitcoin, and the yellow and green lines show the volatility and difference between the trend and observed price.

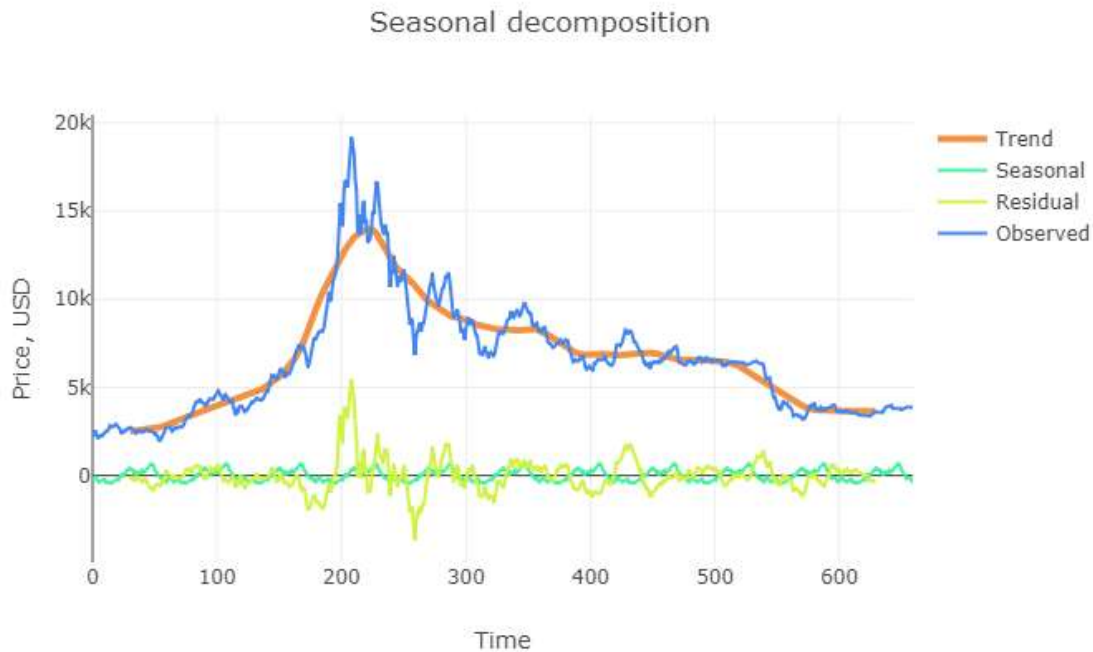
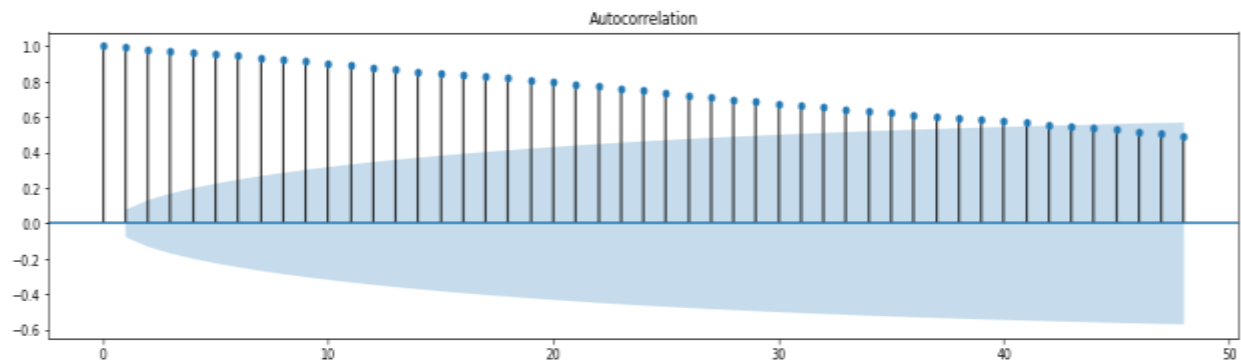


Figure 3: Bitcoin trends in the past 2 years.

Next, figure 3 below represents autocorrelation. It's important to understand whether there are repeating patterns within our price data of bitcoin. These utilize the weighted price data point and has a lag time of 48. The first chart shows a clear trend going slight downwards towards around between 0.6 and 0.8. Meanwhile, the second chart below represents the differences inside of the downward slope of the first chart.



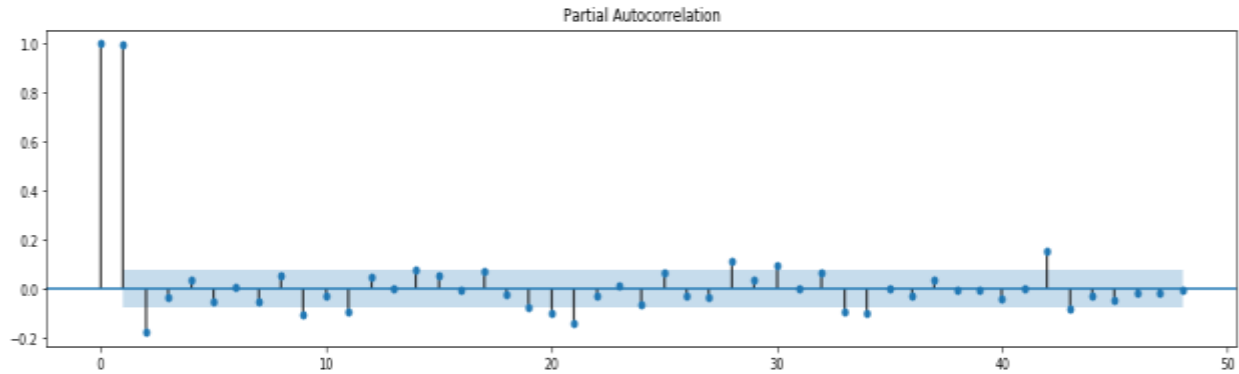


Figure 4: Repeating patterns in data time lag observation similarities.

Using these models, a prediction for the price of bitcoin is created. Below in figure 5, it shows the real actual price of bitcoin compared to the price that is predicted using the model. The dates below show the last 3 months of data within the data set. It shows that there is a slight error, which is less than 50 USD at most and is almost exact in certain situations. The model was training on the previous two years between 2017 and 2018, and then the test for the model below is for 2019.



Figure 5: Bitcoin price prediction.

Future plan

In the future, this project could be extended to data that is more relevant to today, as the data used in the project only goes up to April and is almost 2 months old. Another extension could be, to use second by second data, instead of minute by minute data. Many of the accuracies in the predictions in this project is correlated to how accurate the data is. Some data wasn't able to be capture for technical reasons and this reduced the overall accuracy as well. The project could also be used on other cryptocurrencies. Bitcoin was chosen for the project because it's the most well known cryptocurrency and has the highest market cap. However, there are also other cryptocurrencies with a large market cap, such as Ethereum, which is the second largest with a market cap of around 17 million USD as of April 27th, 2019. Furthermore, it may be smarter to use these models on the overall market cap, which would encompass every currency.

References

A beginner's guide to LSTM and recurrent neural networks.

<https://skymind.ai/wiki/lstm>

Understanding LSTM networks

<http://colah.github.io/posts/2015-08-Understanding-LSTMs/>

Empirical Evaluation of Gated Recurrent Neural Networks on Sequential Modeling

<https://arxiv.org/pdf/1412.3555v1.pdf>

The Unreasonable Effectiveness of Recurrent Neural Networks

<http://karpathy.github.io/2015/05/21/rnn-effectiveness/>

Guide on using machine learning methods to estimate price

<https://www.kaggle.com/smitad/bitcoin-trading-strategy-simulation>

Website with a list of various bitcoin machine learning projects

<https://www.kaggle.com/mczielinski/bitcoin-historical-data/kernels>

Article on how to get a prediction of the bitcoin price by the neural network model.

<https://medium.com/activewizards-machine-learning-company/bitcoin-price-forecasting-with-deep-learning-algorithms-eb578a2387a3>

Compares NN and ARIMA modeling to predict bitcoin price.

<https://www.kaggle.com/zikazika/using-rnn-and-arima-to-predict-bitcoin-price>

Scholarly journal that attempted to apply machine-learning algorithms to predict Bitcoin price.

<https://pdfs.semanticscholar.org/e065/3631b4a476abf5276a264f6bbff40b132061.pdf>

Scholarly journal that investigated the predictive power of blockchain network-based features on the future price of Bitcoin.

http://snap.stanford.edu/class/cs224w-2015/projects_2015/Using_the_Bitcoin_Transaction_Graph_to_Predict_the_Price_of_Bitcoin.pdf

Study of modeling and predicting bitcoin prices with a bayesian neural network.

<https://ieeexplore.ieee.org/abstract/document/8125674>