

3. Theory

The following chapter is intended to provide the theoretical foundations necessary for our work. It is divided into a part that provides an overview of artificial neural networks. Followed by section 3.2. that shows the background and the ecosystem of Bitcoin. This knowledge should be kept in mind, which should help in understanding the price formation of Bitcoin.

3.1. Neural network

In the context of this work, artificial neural networks are used to answer supervised learning questions that focus on classification of data. This means that a neural network finds a correlation between the data and their labels and optimizes its parameters in order to minimize the error for the next try. This process is called supervised training and is performed with a test data sample. An application example of classification is that a neural network is used for face recognition after it has learned the classification of different faces in the process of supervised training. Predictive analysis works similarly to the classification of labeled data. It estimates future values based on past events and can be trained with historical data. On the other hand, unsupervised learning (clustering) is applied to detect patterns from unlabeled data. Based on these patterns, for example, anomalies can be detected that are relevant in the fight against fraud (fraud detection). Unsupervised learning is not discussed further in this paper.

Section 3.1.1. will demonstrate the functioning of a neural network using a simple perceptron.

3.1.1. Perceptron

The construction of an artificial neural network is demonstrated using a perceptron. It is a simple algorithm for supervised learning of binary classification problems. This algorithm classifies patterns by performing a linear separation. Although this discovery was anticipated with great expectations in 1958, it became increasingly apparent that these binary classifiers are only applicable to linearly separable data inputs. This was only later addressed by the discovery of multiple layer perceptrons (MLP). (Rosenblatt 1958)

Basically, a perceptron is a single-layer neural network and consists of the following five components and can also be observed in figure @ref(fig:perceptron_schema).

1. Inputs
2. Weights
3. Bias
4. Weighted sum
5. Activation function

Inputs are the information that is fed into the model. In the case of econometric time series, it is mostly the current and historical log returns (lags). These are multiplied by the weights and added together with the bias term to form the weighted sum. This weighted sum is finally passed on to the non-linear activation function, which determines the output of the perceptron.

The perceptron can also be represented as a function, which can be seen in equation 1. Analogous to the representation above, the inputs x_i are multiplied by the weights w_i in a linear combination. Then an error term is added so that the whole can be packed into the non-linear activation function $g(S)$. \hat{y} is the binary output of this perceptron. With the aid of an activation function, a binary output is obtained. The Heaviside step function shown in figure 1 is usually only used in single layer perceptrons, which recognize linear separable patterns. For the multi layer neural networks presented later, step functions are not an option, because in the course of the backpropagation algorithm the gradient descent has to be minimized.

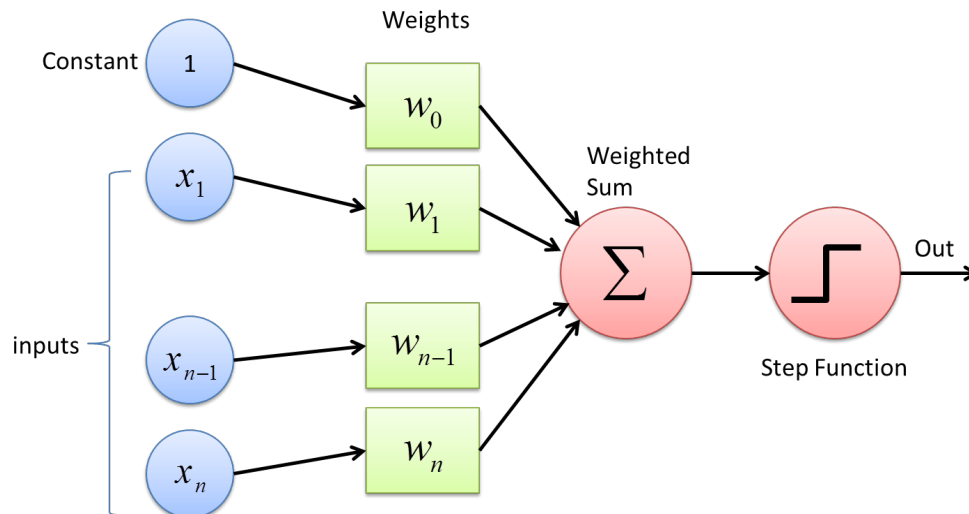


Figure 1: Schematic diagram of a perceptron.

This requires derivatives of the activation function, which in the case of this Heaviside step function equals 0. Because the foundation for the optimization process is missing, functions like the sigmoid function or the hyperbolic tangent function are used. More about this topic is discussed in chapter 3.1.2.. (Martin Anthony 1999)

$$\hat{y} = g(w_0 + \sum_{i=1}^n x_i w_i) \quad (1)$$

As just mentioned, the aim is to feed the perceptron with the training set and change the weights w_i with each cycle so that the prediction becomes accurate. The output value is compared to the desired value. Finally, the sign of the difference $y - \hat{y}$ determines whether the inputs of that iteration are added to or subtracted from the weights. Ideally, the weights will gradually converge and provide us with an usable model. (Martin Anthony 1999)

3.1.2. Backpropagation algorithm

Finding the optimal weights of the neural network is achieved by finding the minimum of an error function. One of the most common methods for this is the backpropagation algorithm. This algorithm searches for the minimum of the error function by making use of a method called Gradient Descent. The gradient method is used in numerics to solve general optimization problems. In doing so, we progress (using the example of a minimization problem) from a starting point along a descent direction until no further numerical improvement is achieved. Since this method requires the computation of the gradient of the error function after each step, a continuity and differentiability of this function must necessarily be given. The step function mentioned above in section 3.1.1 is therefore out of the question, but a non-linear function such the logistic and the hyperbolic tangent functions (sigmoid). (Rojas 1996)

$$y(v_i) = (1 + e^{-v_i})^{-1} \quad (2)$$

$$y(v_i) = \tanh(v_i) \quad (3)$$

While the target range of the ‘ordinary’ sigmoid function is between 0 and 1, the y of the Tanh function ranges between -1 and 1. Both functions are visible in figure ??.

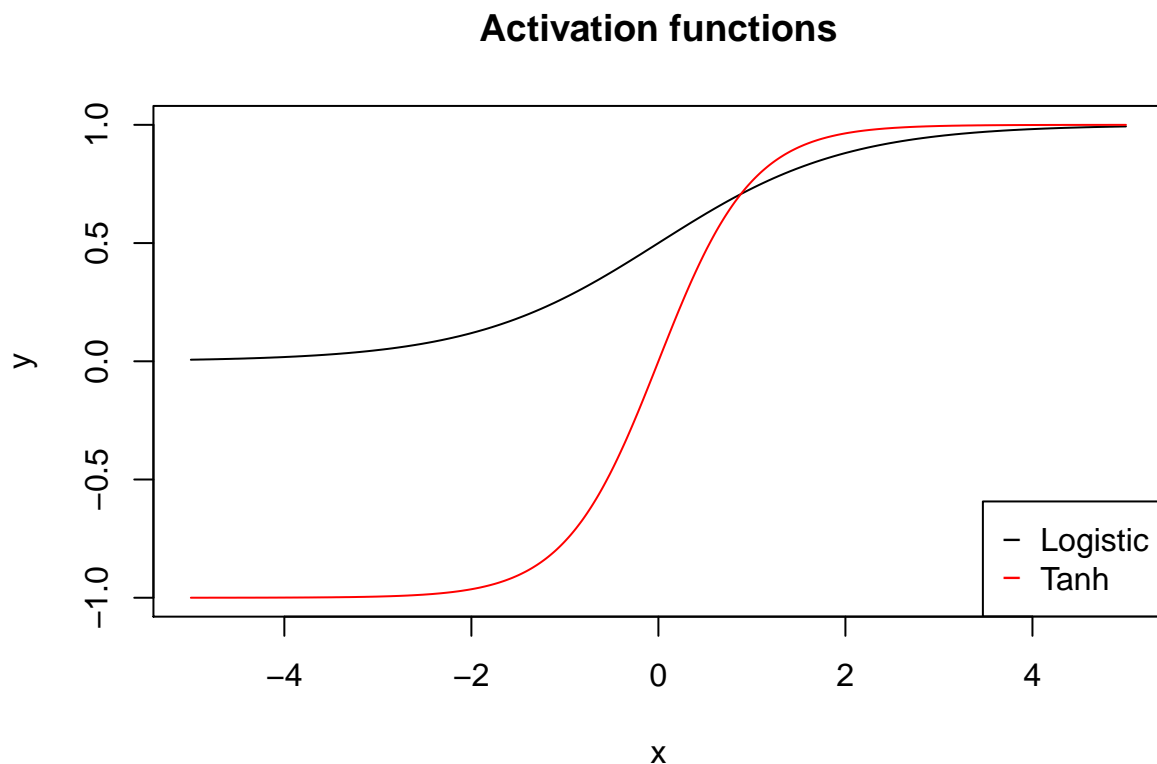


Figure 2: Two common sigmoid activation functions: logistic functions and hyperbolic tangent.

In the course of the error analysis, the output of the neural network respectively the result from the activation function in the output layer is compared with the desired value. The most commonly used error function is the Mean Squared Error (MSE).

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (4)$$

(Yann Lecun 1998)

3.1.3. Types of artificial neural networks

- MLP / DNN (explain nodes and layers)
- Recurrent neural networks (RNN)
- Long-short term memory (LSTM)

3.1.4. Challenges

- Early stopping required to avoid overfitting to in-sample data
- Gradient vanishing problem

Sigmoid suffers from the problem of Vanishing Gradient. The gradients of the NN's output with respect to the parameters become so small, that the NN takes smaller steps towards the minima of the loss function and eventually stop learning.

- Dying Relu Problem (?)

3.2. Bitcoin

In this section bitcoin as a crypto-currency is introduced. The historical data is analyzed and commented. Further the technology in and around crypto-currencies is briefly explained. A detailed explanation would require a paper itself, therefore the explanation is done as simple as possible.

In the following work bitcoin as a cryptocurrency is mentioned in its short term BTC, by the meaning of US Dollars per Bitcoin.

3.2.1. Historical analysis The story of bitcoin began with a paper published by the name of Satoshi Nakamoto (Nakamoto 2008). The publisher of the document cannot be assigned to a real person, therefore the technology inventor remains mysteriously unknown until today. In 2009 the first bitcoin transaction was executed. On account of the opensource technology of bitcoin, lots of alternative currencies were created. Until 2013 the cryptocurrencies operated under the radar of most regulatory institutions. Because of the anonymity of the transactions, criminals were attracted by the newborn payment method. Headlines, such as the seizure of 26,000 bitcoins by closing the “Dark-Web” Website Silkroad through the Drug Enforcement Agency, followed more often in the newspapers. Nevertheless in 2014 more companies, such as Zynga, D Las Vegas Casinos, Golden Gate Hotel & Casino, TigerDirect, Overstock.com, Newegg, Dell, and even Microsoft (Chohan 2017), began to accept bitcoin as a payment method. In 2014 the first derivative with bitcoin as an underlying was approved by the U.S. Commodity Futures Trading Commission. 2015 an estimated 160,000 merchants used bitcoin to trade. It is observed that the value of bitcoin is very volatile, we will discuss this in a FURTHER XYXY section. Let us first look at the price in ?? and the log(price) ?? and get a sense of the chart. Note: The data in the charts start in 2014 where it was listed in coinmarket, events between 2009 and 2014 are described without visualization.

Around 2010 bitcoin had the first increase in price as it jumped a 100% from 0.0008 USD to 0.08 Dollar (Edwards n.d.). In 2011 the price rose from 1 USD to 32 USD within 3 months and recessed shortly after to 2 USD this can be referred as a first price bubble in bitcoin, for the next year the price climbed to 13 Dollars and reached a never seen level of 220 USD, only to plunge to 70 USD within a half month in April 2013. By the end of the year a rally brought btc up to a peak of 1156 USD. The following year brought bad news and the price slowly decreased to 315 USD in 2015 after an observed drop of 20% after news from the trial of Ross Ulbricht, founder of Silk road marked in Letter **A**. From this point in time, things began to change, more volume was flushed in the market and the price of BTC began to ascend and the real rally began, the BTC rose up to 20k USD / BTC on 17th September 2017 **B**. After the rise comes the fall and BTC lost value for more than a year until **C** 2018-12-15 the trend reverted and found its peak after 6 months in **D** 2019-06-26, but once more it was not lasting for long as bitcoin lost **D** 2020-03-12 nearly half its value in 4 days. But the story wasn't over by now after the drop, the price of the cryptocurrency regained value, passed previous levels and shortly after exploded, after companies like tesla and signal bought a big chunk of bitcoins, into a maximum of 58,000 USD per bitcoin.



Figure 3: Schematic diagram of a perceptron.



Figure 4: Schematic diagram of a perceptron.

3.3.2. SHA256 Hash

- Block
- Blockchain
- Distributed Blockchain
- Token
- Coinbase Transaction
- Public/Private Key -> Signing
- Signature (sign, verify)
- Transaction

3.7. Reference

3.7.1. Figure Reference

3.7.2. Equation Reference

3.7.3. Table Reference

3.7.4. Section Reference

- Chohan, Usman W. 2017. *A History of Bitcoin*. University of New South Wales, Canberra.
- Edwards, John. n.d. "Bitcoins Price History." Accessed March 1, 2021. <https://www.investopedia.com/articles/forex/121815/bitcoins-price-history.asp>.
- Martin Anthony, Peter L. Bartlett. 1999. *Neural Network Learning: Theoretical Foundations*. Cambridge University Press.
- Nakamoto, Satoshi. 2008. *Bitcoin: A Peer-to-Peer Electronic Cash System*. online: www.bitcoin.org.
- Rojas, Raul. 1996. *The Backpropagation Algorithm*. Springer Berlin Heidelberg.
- Rosenblatt, Frank. 1958. *The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain*. Psychological Review.
- Yann Lecun, Genevieve B. Orr, Leon Bottou. 1998. *Efficient BackProp*. Image Processing Research Department AT& T Labs.

3.7.5. Literature Reference