Methodology of the ANN

The training and testing datasets were defined and combined to create the complete dataset.

As a preprocessing step, all predictor variables were standardized and subsequently the Volume and Market Cap were divided by 107 and 108 respectively.

The complete dataset was then lagged by the specified number of lag steps. The training and testing dataset were then selected from the lagged dataset in accordance with the afore defined datasets.

The predictors for the model were selected according to the required model type and the response was selected as the standardized value of the Close price of Bitcoin.

The model was defined with the following architecture. The first layer was the input layer with the input size being the number of predictors. The next layer included 20 nodes and the following layer included 40 nodes. The final output layer used 1 node while all nodes used the hyperbolic tan activation function. The model was trained for 100 epochs with the Mean Absolute Error as the loss function.

After building the main model, the MAPE and the plots of forecasted values were evaluated. If there were no significant improvement a grid search of hyper parameters was utilized to find the best hyper parameters. Finally, the process was repeated for the lag steps ranging from 1 to 8

Point form of ANN

* Defined training set, testing set, and compete dataset
* Performed preprocessing on datasets
* Obtained dataset of lagged values for specific number of lag steps
* Defined predictors and response according to the model type
* Built ANN model with required architecture
* Evaluated model performance
* Continued step 3 to 6 for the number of lag steps ranging from 1 to 8

Methodology of SVM

As the first step, the training set and testing set of the complete datasets were defined and collected. The required predictors were then selected from the datasets according to the model type. The resulting dataset was then lagged by a specific number of lag steps to obtain the lagged dataset. From the lagged dataset training, validation and testing datasets were extracted. A Support Vector Machine based on epsilon regression was fitted using the training set and then later on tuned on the validation dataset using MAPE as the error function. The optimal values for the parameters gamma and cost were found from the range of 2-9 to 210. The best parameters were then used to build the final Support Vector Machine. The model was used to forecast values for the testing set and using the actual values of the forecast the MAPEs and plots of forecasts were generated. In order to find the number of lag steps that provided the best fit, the process was repeated for lag steps ranging from 1 to 8