

Machine Learning Engineer Nanodegree

Capstone Proposal

Takayoshi Nishida
22 October, 2017

Proposal

Domain Background

Predicting the stock price has been researched for long. Now many people try to predict with the machine learning algorithms, but there is not a single answer for this and it is still challenging problem.

It is also known that every country's stock market influences each other. So putting the another country's stock market price and predicting the stock price index worth challenging.

Problem Statement

The goal of this project is to predict whether the Nikkei 225 index increases or decreases compared to the previous day. Nikkei 225 is a stock market index for the Tokyo Stock Exchange in Japan.

My hypothesis is that the Nikkei 225 has a strong correlation with the US stock prices (NASDAQ index) and foreign exchange rate (JPY/USD) of the previous day.

Datasets and Inputs

1. Nikkei 225

The data starts from January 1950 to current date. This is obtained by Quandl.

- <https://www.quandl.com/data/NIKKEI/INDEX-Nikkei-Index>

2. NASDAQ Index

The data starts from January 2003 to current date. This is obtained by Quandl.

- <https://www.quandl.com/data/NASDAQOMX/COMP-NASDAQ-Composite-COMP>

3. Currency Exchange – JPY/USD

The data starts from March 1991 to current date. This is obtained by Quandl.

- <https://www.quandl.com/data/CURRFX/USDJPY-Currency-Exchange-Rates-USD-vs-JPY>

Solution Statement

The solution to this problem is to apply deep learning to predict the Nikkei 225 index of the next day.

Data is Nikkei 225, NASDAQ and the currency exchange. Inputting those data from the specific previous days, and predict the next day's Nikkei 225 index increases or decreases.

Benchmark Model

The prediction result is increase or decrease. So that it is only true or false compared to actual Nikkei 225's movement. The random prediction will be 50% true and 50% false. Calculating the accuracy, I'll try to get higher score than it.

Evaluation Metrics

The prediction results are evaluated on accuracy score. It is expected to be higher than random 50%.

Project Design

Data preprocessing

The first step is collecting the data. As discussed above, I will obtain the data from Quandl. The file is CSV and the data is numeric value. But the value among them, it is different value so I need to normalize them.

Model

The second step is to build the model with deep learning. Evaluating with the cross validation set by calculating the accuracy score, I will going to adjust the parameters and the number of hidden layers.