Final Year Project Plan

Financial Market Prediction by Deep Learning Neural Network

Supervisor: Kwok-Ping Chan

Students: 3035142596 LIU Jiayao

3035183552 ZHANG Yuqing

1. Background

1.1 Objective

We are going to develop a neural network model to predict the trend of DJIA and/ or S&P indices by sectors. We intend to predict whether the value of indices will rise or drop based on the past data with a satisfying accuracy (at least greater than 0.5). Apart from that, an analysis about the performance and the limitation of the model will be conducted.

1.2 Deep learning

Deep learning, as a class of machine learning, has been very prevalent nowadays. The concept of neural networks stems from the idea to imitate the layers of neurons in the neocortex of the brain.

In Figure 1, the virtual neurons of the first layer identify the primitive representations of the data input which will be fed to the next layer to form more sophisticated features until it provides a reliable output in the end.

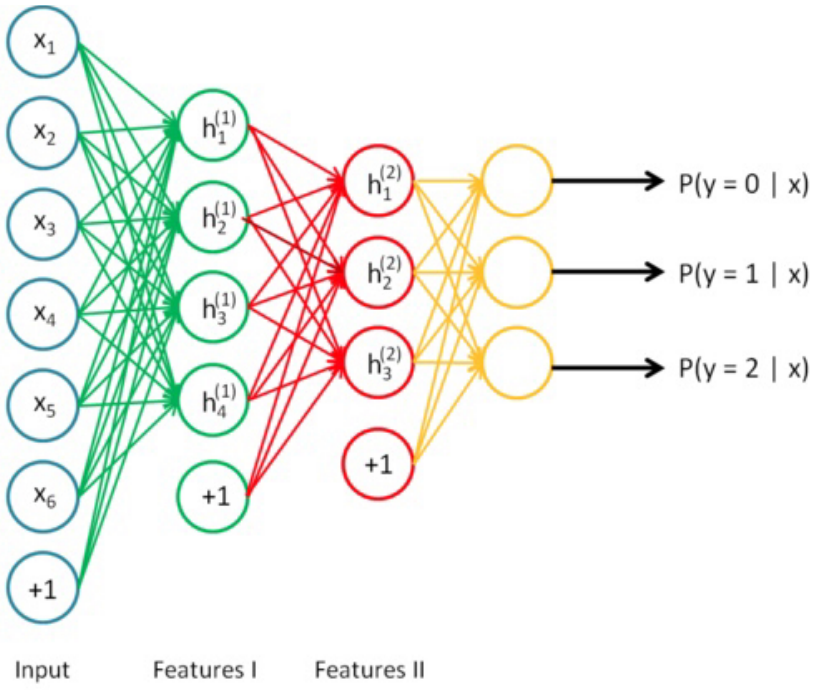

Figure 1. A traditional three-layer neural network

Figure 2 and 3 show two convolutional neural networks which contain three main types of layers in cNN: Convolutional layer, Pooling layer and Fully connected layer. The convolutional and pooling layer first filter and shrink the features and then pass the features to the fully connected layer, where each neuron has an effect on all the neurons of the next layer to determine the output.

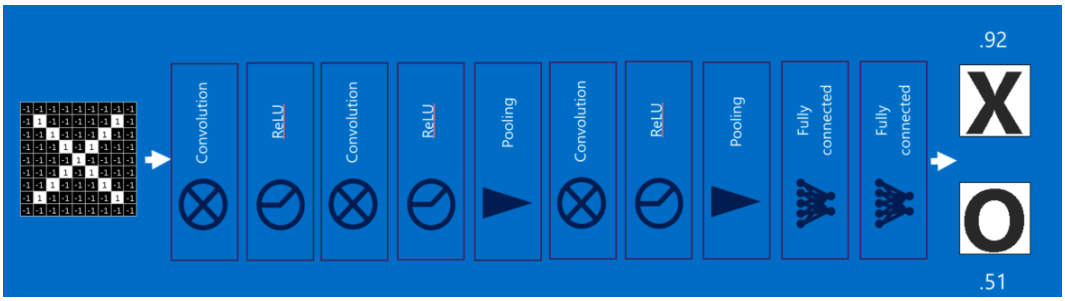

Figure 2. A convolutional neural network to identify X.

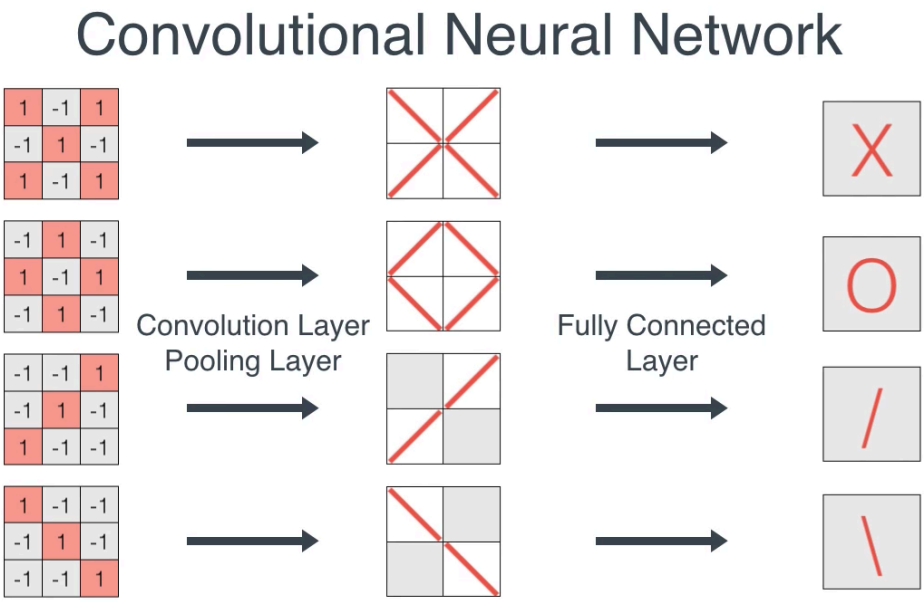

Figure 3. How the data from image pixels are recognized as patterns by the convolutional neural network.

Compared to the traditional machine learning models, deep learning has the capability to identify the features itself, which decreases the need for feature engineering to a large extent, whereas ML models mainly study how to combine the features that are found and formed by humans.

Deep learning has shown its computational power in image and speech recognition. There is a growing trend to apply deep learning to financial market while financial data differs a lot from the image and speech data. Thus, we intend to explore its performance in predicting financial market to see if it can present a more profitable arbitrage opportunity than human beings.

1.3 Finance

Dow Jones Industrial Average (DJIA) serves as a good barometer of the stock market, which reflects the performance of 30 large companies that are representative of American industries. Besides, Standard & Poor's Index of 500 stocks (S&P) is also a good indicator of the overall market activities. The indices differ by the type and number of stocks they cover as well as how they are calculated. Apart from DJIA, there are different Dow Jones indices of other industries and sectors which can be further studied if time allows.

1.4 Scope

The financial market of our project refers to the American stock market. In addition to the stock trading data of the indices, macro information such as GDP, interest rate, unemployment rate and so on can also be embodied. In terms of neural network model, only feed forward and convolutional neural networks are utilized.

2. Methodology

Firstly, financial data of the aforementioned indices with the length of 20 years will be collected from the exchanges or other open source data companies and websites. Data will be cleaned to an appropriate degree and labelled as well as split into training and test datasets as we are going to conduct supervised learning to build our neural network model. After fitting the model, we will optimize the model by tuning the parameters such as the number of hidden layers and the number of units of neurons in a layer. Additionally, the problem of overfitting is expected and we will deal with it by regularization including l2 regularization and

dropout.

The model will be established in the programming language Python, considering that they have frameworks like TensorFlow and PyTorch. A GPU may be needed owing to the large amount of data and the complexity of the model.

Computation time and accuracy of the model will be adopted to evaluate the performance of the model.

3. Risks & Mitigations

Risks	Impact	Mitigation
Failure in data collection	High	Bloomberg, WSJ, Yahoo Finance etc. Might need to manually export the data
Failure in machine setting	High	Python packages and GPU
Poor performance of the model	High	Increase the data size Optimize the model Trial of other models: Time series Logistic regression Tree SVM

4. Schedule and Milestones

1st, Oct	**Deliverables of Phase 1: Inception** - Complete the project plan - Build a project website

1st, Sep to 15th, Oct	**Study the theory & related knowledge** - Deep learning - Python programming - Financial products
16th, Oct to 10th, Jan	**Collect & preprocess the data** - Decide a type of financial product as the predicted object - Collect data from the exchanges or data companies - Clean and label the data
8-12th, Jan	**First presentation**
21st, Jan	**Deliverables of Phase 2: Elaboration** - Conduct preliminary implementation - Complete detailed interim report
10th, Jan to 25th, Feb	**Build the neural network model based on the data** - Fit the model - Tune the corresponding parameters
March to April	**Finalize & Improve the results:** - Optimize the model - Prepare for final presentation - Start to write final report
15th, Apr	**Deliverables of Phase 3: Construction** - Conduct finalized tested implementation - Complete the final report
16-20th, Apr	**Final presentation**
2nd, May	**Project exhibition**