Analysis of executive transaction characteristics based on machine learning cross-validation

Yueqiu Li*, Chunming You
Heihe University, Heilongjiang Heihe, China

*Corresponding author: Yueqiu@163.com

Abstract. Executive trading is a kind of investment strategy that analyses and models the financial market with mathematical methods. Machine learning requires computer programs to improve their performance at specific tasks by learning on data sets. Both require the extraction of information from data, so with the success of machine learning in recent years, there has been a strong interest in the combination of executive trading and machine learning methods in both industry and academia. In this paper, based on machine learning is complete, a more complete model of data processing - build - generated forecast - trading strategy - back-test analysis "research framework, and discusses the structures of branching and merging model structure set up, based on the value relevance of the output value and supervision model selection criteria, independently of the back and trading strategy model assessment method and a series of important issues. This paper focuses on the innovative exploration of the application of the loss function, and puts forward several loss functions based on the portfolio point of view, among which the negative cosine loss function shows significant advantages and strong practicability in stability and gain. Corresponding empirical research also reveals the application value and limitation of other loss functions.

Keywords: Machine learning; Cross validation; Executive transactions; Loss function

1. Introduction
Quantitative investment in China is still in the early development stage, but after many years of development, in the industry has formed four factions of the basic pattern, namely: securities, public funds, private funds and futures. Although the "four factions" are divided by industry, their specific tools and strategy preferences are also quite different due to their different service objects and trading varieties. Futures firms and some private equity firms are also interested in the most aggressive high-frequency trading and deep learning frameworks because they can trade highly leveraged futures assets[1].

Deep learning model has a strong fitting ability. With the iteration of training, the model will almost certainly get a good fitting effect in the training set[2]. However, the real value of the model should be that it can correctly predict the test set data that does not participate in the training. If the loss function is used to describe the fitting ability of the model, and the error on the validation set/test set is taken as the estimation of the generalization error, then theoretically, the underfitting and
overfitting phenomena are shown in Figure 1:

![Figure 1](image1.png)

**Figure 1** Typical cases of underfit and overfit

The identification of under-fitting and over-fitting is an important issue in deep learning practice. Correct identification of fitting is the premise to find the optimal model from the training results of multiple iterations, and it is also the premise to prevent over-fitting and improve the training effect by using early termination of training and dynamic adjustment of learning rate [3].

2. Theory and Algorithm

Cyclic neural network (RNN) is a general term of a class of network structures. RNN is no longer a simple infrastructure network, but a combination of several infrastructure structures (commonly referred to as cells). Taking the common long - short - term memory network LSTM as an example, its structure is shown in Figure 2[4].

![Figure 2](image2.png)

**Figure 2** Structure and stacking of LSTM network cells

The identification of the fitting situation is accomplished by analyzing the statistical record values in the training. Statistical record value refers to the scalar value calculated together with the true value, which is obtained by using the model on the corresponding data set (generally the verification set) after the completion of the phased training of the model. Statistical records value is predicted vector (sequence) \( \hat{y} \) and true value vector (sequence) \( y \) to scalar mappings. Predictive value for the input value \( x \) proceeds through the model, true value for monitoring \( y \), both in accordance with the corresponding order, make up the vector.

\[
\text{statistic} = f(\hat{y}, y) \quad (1)
\]

\( \hat{y} \) computation depends on the model parameters and model parameters in every stage training after the completion of the update, so \( \hat{y} \) relies on the label of training. Therefore, the statistical record value can be regarded as the trajectory equation of the training label, representing the fitting situation of the model[5]. Different mapping \( f \) values corresponding to different statistical records. A special mapping \( f \) is loss function used in the training model; loss function is data description model fitting the best of mapping. Highly non-convex due to deep learning model, its parameters solving method only USES the iterative optimization way - that is, to repeated traversal of data sets, iteratively optimized model parameters, which may involve thousands or even tens of thousands of times the
parameters of the iterative update, statistical records value of every parameter updating calculation is the lack of representativeness, and inefficient. In order to observe the fitting of the model reasonably and efficiently, it is necessary to define "phased training completion" reasonably[6].

There are several batches in each round, and the batch and round counts and their corresponding statistical records can describe the model fitting during the training process. As defined above, any batch is a true subset of the dataset, so the batch cannot reflect the model fitting of a complete dataset iteration. Statistical record values that are representative of the complete dataset can only be obtained at the round level, which is defined as "phased training completion" in this paper. All statistics in this article are performed at the wheel level[7].

The introduction of the models used in this paper focuses on the structural framework and the corresponding considerations. The model is mainly divided into two layers from top to bottom: merge and adjust layer. Different networks can adopt different structures, which need to be adjusted flexibly according to the understanding of the characteristics of the data, which reflects the network designer’s prior knowledge of the understanding of the financial significance of the data. Due to the need to process time series input, it is suggested to use the CONV or RNN series structure as the starting point for the subnetwork design. This paper uses the LSTM network as the basic network for the experiment. The overall structure is shown in Figure 3, and the right side is the data dimension of the corresponding position of the network. For ease of understanding, there is no batch mechanism in the data dimension annotation, and there is no narrowing of the dimensions that have been reduced to 1.

![Figure 3](image)

**Figure 3** The network form of shunt + merge corresponds to the data dimension of each level

The above structure and its principles still give the design a great deal of freedom, and in essence it can be a general framework for multi-factor inputs. This paper calls it a shunt + merges structure. Each pathway sub network is used to process a specific batch of factors, and the factor data in the same batch should have a strong correlation, while the correlation between the factors in different batches is relatively weak. The definition of "correlation relation" here is vague, and the work of factors in batches is where the prior knowledge of human beings is put to use. For example, when processing fundamental information, stock information and flow information can be divided into different batches, indicators related to leverage and cash flow can be divided into different batches, factors generated by using four operations and factors generated by using complex methods such as regression can be divided into different batches.

3. Conclusion
From the perspective of research, CSI 300, as one of the most important stock market indexes in
China, has irreplaceable representation, so this paper takes its constituent stocks as the research object. The research period of this paper includes two complete years: 2015 and 2018. In these two years, stock crash level market turbulence occurred respectively. The major transformation of the market situation provides a good behavior for studying the behavior of the deep model in the real market. The purpose of the predictive model is to provide signals for trading, so the back test analysis is necessary. However, as repeatedly emphasized above, trading strategies have a significant impact on returns, so it is one-sided to evaluate the merits of the predictive model solely from the perspective of back test. Therefore, this paper only uses the simplest continuous market trading mode to provide a basic reference.

In order to reflect the variation of correlation more intuitively, this paper smoothest the sequence of correlation coefficients, that is, the average of the calculated historical value of fixed window length at a certain point is taken as the average of this point. In the following picture, the window length is set to 20 days, the line is the moving average, and the scatter is the correlation coefficient at a point in time.

**Figure 4** Correlation analysis under the loss function of neg-Cos

Based on the correlation analysis images of multiple loss functions, the stability of the smoothed correlation coefficient graph line is extremely important. Neg-COS basically kept stable above 0 before 2017, while its return rate and risk performed well before 2017. Both ABS and Combine have sharp fluctuations in the images, and ABS is even better than Combine in terms of the height of positive peaks. However, the stability of ABS is too poor, and Combine can guarantee positive correlation in most of the time in the fluctuations. Considering that there is no mechanism that can make all stocks with high correlation can be traded, and the observation period corresponding to the monitored value is not the same as holding it, it is a reasonable analysis method to maintain long-term and stable correlation and make the holding period of stocks close to the observation period of the monitored value in terms of trading strategy.

An intuitive and effective method is to deal with the correlation of each model in its test period by using a calculation method similar to T statistics, that is, the mean value and standard deviation of a series in the test period are taken as the measure of model training state, so that a small series can be converted into a scalar. Obviously, the larger the value of the "T-like statistic", the higher the correlation and better stability of the model during its operating period.

Firstly, this paper introduces the setting of model training, prediction and back test, and explains the parameter values, calculation methods and trading strategies that have important influence on model validation. Secondly, this paper analyses the model based on the back test results formed by trading strategies. The analysis results show that the performance of the NEG-COS loss model based on the portfolio view is significantly better than the MSE loss model based on the individual stock view, as well as the ABS and Rank loss model based on the portfolio view. Thereafter, independent of trading strategy and back test, this paper formed a correlation analysis method of predictive value-supervised value with financial significance only from the model's fitting to the supervised value, and obtained a means to evaluate the merits of the model. Then, by analysing the statistical record values
recorded in the training process of the model, the differences in the convergence stability of the loss functions of MSE, Neg-COS, ABS and RANK are shown, which proves that Neg-COS is a good measure of the correlation relationship from the perspective of portfolio. Finally, this paper further discusses the application range of ABS and RANK loss based on the order relationship characterization of simulated grade difference. By constructing hypothesis test, it is confirmed that ABS has a statistically significant relationship with the market volatility during model training and testing, which provides an important basis for the use and improvement of this kind of loss function.

Acknowledgements
The Project of Year 2018 Scientific research business fee project of Heilongjiang Provincial Education Department New risk characteristics of border enterprises in Heilongjiang Province under the new normal Project number: 2018-KYYWF-1264.

References
[1] Hao H, Guo Q. “State-Frequency Memory Recurrent Neural Networks”. Sydney: PMLR, pp. 1568-1577, 2017.
[2] Ben M J. “Michael Herrmann, Gbenga Ibikunle. Lagged correlation-based deep learning for directional trend change prediction in financial time series”. Expert, vol. 120, pp. 197-206, 2019.
[3] Thomas F, Christopher K. “Deep learning with long short-term memory networks Research for financial market predictions”. European Journal of Operational, vol. 270, no. 2, pp. 654-669, 2018.
[4] Sotirios P C, Vassilis S, Anastasios P, et al. “Forecasting stock market crisis events using deep and statistical machine learning techniques”. Expert Systems with Applications, vol. 112, pp. 353-371, 2018.
[5] Guizhu S, Qingping T, Zhang H, et al. “Deep Learning with Gated Recurrent Unit Networks for Financial Sequence Predictions”. Procedia Computer Science, vol. 131, pp. 895-903, 2018.
[6] Hiransha M, Gopalakrishnan E A, Vijay K M, et al. “NSE Stock Market Prediction Using Deep-Learning Models”. Procedia Computer Science, vol. 132, pp. 1351-1362, 2018.
[7] Eunsuk C, Chulwoo H, Frank C. “Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies”. Expert Systems with Applications, vol. 83, pp. 187-205, 2015.