ISSN 2090-3359 (Print)
ISSN 2090-3367 (Online)

ΑΔΣ

Advances in Decision Sciences

Volume 22(A)
22nd Anniversary Special Issue
December 2018

Michael McAleer

Editor-in-Chief
University Chair Professor
Asia University, Taiwan

Published by Asia University, Taiwan

ADS@ASIAUNIVERSITY

Artificial Intelligence and Economic Growth*

Shigeyuki Hamori

Graduate School of Economics

Kobe University

Kobe, Japan

Takahiro Kume

Graduate School of Economics

Kobe University

Kobe, Japan

Revised: October 2018

* The authors are grateful to the Editor, Michael McAleer, Ms. Minami Kawai, the participants of The Second International Forum on East Asia Macroeconomic Studies (Xiamen, China), and SIBR 2018 Hong Kong Conference on Interdisciplinary Business & Economics Research (Hong Kong, China), for helpful comments and suggestions. This work is supported by JSPS KAKENHI Grant Number 17K18564 and 17H00983.

Corresponding author: hamori@econ.kobe-u.ac.jp

Abstract

This paper describes the use of five machine learning methods for predicting economic growth based on a country's attributes and presents a comparison of their prediction accuracy. The methods used are four neural network (NN) methods with different activation functions, and eXtreme Gradient Boosting (XGBoost). Their performance is compared in terms of their ability to predict the economic growth rate using three measures (prediction accuracy rate, area under the curve (AUC) value, and F-score). The results obtained can be summarized as follows: 1) XGBoost outperforms the NNs in terms of prediction accuracy and F-score for original data; 2) data standardization enhances the reliability of NNs, improving their prediction accuracy, AUC-value, and F-score; 3) XGBoost has smaller standard deviation of prediction accuracy rate than that of NNs; and 4) "Political institution", "Investment and its composition", "Colonial history", and "Trade" are important factors for cross-country economic growth.

Keywords: Economic growth, machine learning, XGBoost, neural network.

EL: E10, C45.

1. Introduction

Artificial intelligence (AI) and machine learning are often featured in the media. This indicates that AI and machine learning permeate every facet of society and business activity. The trigger of this movement may be the victory of Deep Blue against Garry Kasparov (the chess grandmaster) in 1997. In addition, AlphaGo gained a victory over Lee Sedol, the Korean Go champion, in 2016.

The rapid development of AI has resulted in machine learning methods being widely applied to various research activities, for example, Razi and Athappilly (2005), Kurt, Ture, and Kurum (2008), and Austin, Tu, Ho, Levy, and Lee (2013) for medical sciences; and Tso and Yau (2007) and Ahmad, Mourshed, and Rezgui (2017) for energy analysis. Additionally, machine learning methods have been widely applied to various aspects of economic analysis, especially in the field of finance. For example, Oberlechner and Hocking (2004), Hegazy et al. (2013), Ding et al. (2014), Zhuge et al. (2017), Roondiwala et al. (2017), and Santos Pinheiro et al. (2017) analyzed stock price forecasting using machine learning methods. Angelini et al. (2008), Khashman (2009), Yeh and Lien (2009), Khashman (2010), Khemakhem and Boujelbène (2015), and Hamori et al. (2018) analyzed credit risk assessment using machine learning methods.

This study analyzes the problem of economic growth in terms of machine learning. Economic growth is undeniably one of the most important problems in economic analysis. Two typical theories are the neoclassical growth theory and the endogenous growth theory. Neoclassical growth theory attempts to explain long-range economic growth by considering capital accumulation, labor or population growth, and growth in productivity, commonly referred to as technological progress (Solow, 1956; Swan, 1956). Endogenous growth theory suggests that economic growth is primarily the result of endogenous rather than external forces and emphasizes the importance of the concept of human capital (Lucas, 1988; Romer, 1986, 1990). Endogenous growth theory holds that investment in human capital, innovation, and knowledge are significant contributors to economic growth. The theory also focuses on positive externalities and spillover effects of a knowledge-based economy, which can lead to economic development.

Sala-i-Martin (1997) focused on cross-country growth regressions and shed light on the importance of such models for empirical growth research. He identified a number of variables to explain economic growth and divided these variables into nine categories: geography; political institution; religion; market distortions and performance; investment and its composition; dependence on primary products; trade; market orientation; and colonial history.

Fernández et al. (2001) analyzed cross-country growth regressions using a Bayesian model and found that certain variables are important regressors for explaining cross-country growth patterns, which is consistent with the findings of Sala-i-Martin (1997).

In this study, cross-country growth patterns are analyzed using machine learning methods. Six machine learning methods are used to predict economic growth based on a country's attributes, and their prediction accuracy is compared. Specifically, the methods used are four neural network (NN) methods with different activation functions, and the eXtreme Gradient Boosting (XGBoost) method. We compare their performance in terms of their ability to predict the economic growth rate using three measures (prediction accuracy rate, area under the curve (AUC) value, and F-score).

The results obtained can be summarized as follows: 1) XGBoost outperforms NNs in terms of prediction accuracy for original data; 2) data standardization enhances the reliability of NNs, improving the prediction accuracy, AUC-value, and F-score of NNs; 3) XGBoost has a smaller standard deviation of prediction accuracy rate than that of NNs; and 4) "Political institutions", "Investment and its composition", "Colonial history", and "Trade" are important factors for cross-country economic growth.

Section 2 explains the experimental design. Section 3 indicates the experimental results. Section 4 summarizes the study and states our conclusions.

2. Experiment

2.1. Methods for Comparison

An NN is a network structure comprising multiple connected units. It consists of an input layer, a middle layer(s), and an output layer. The NN configuration is determined by the manner in which the units are connected; different configurations enable a network to have different functions and characteristics. The feed-forward NN is the most frequently used NN model; it is configured by the hierarchical connection of multiple units. When the number of middle layers is two or more, the network is called a deep NN (DNN).

In Figure 1, units are arranged into three parts (input layer, middle layer, and output layer), and the outputs of each unit in the input and the middle layers are linked to all of the units in the next layer. This type of model is called a fully connected NN.

The activation function in an NN is very important, because it expresses the functional relationship between the input and output in each unit. In this study, we employed two types of activation functions—hyperbolic tangent (tanh) and rectified linear unit (ReLU). These functions are defined as follows:

$$\text{tanh:} \quad f(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}} \tag{1}$$

$$\text{ReLU:} \quad f(x) = \max(0, x) \tag{2}$$

Figures 2(a) and 2(b) illustrate the tanh and ReLU functions, respectively. The tanh function maps a real-valued number into the range [−1, 1]. Its activations saturate, and its output is zero-centered. The ReLU function is an alternative activation function in NNs.[1] One of its major benefits is the reduced likelihood of the gradient to vanish.

Although DNNs are powerful machine learning tools, they are susceptible to overfitting. This is addressed using a technique called dropout (Figure 3) in which units are randomly dropped (along with their incoming and outgoing connections) in the network. This prevents units from overly co-adapting (Srivastava et al., 2014).

The eXtreme Gradient Boosting (XGBoost**)** framework developed by Chen and Guestrin (2016), is a variant of gradient boosting and is another machine learning method. While the random forest method employs independent learning, boosting employs sequential learning (Schapire, 1999; Schapire and Freund, 2012). In boosting, on the basis of supervised learning, weights are successively adjusted, and multiple learning results are sought. These results are then combined and integrated to improve the overall accuracy (Figure 4). XGBoost not only uses the random sample of features but also the random samples of the entire dataset to create each decision tree.

[1] See LeCun et al. (2015).

2.2. Data

Sala-i-Martin (1997) identifies some variables to explain the economic growth and divides these variables into the following nine categories.[2]

1. Geography
2. Political institution
3. Religion
4. Market distortions and market performance
5. Investment and its composition
6. Dependence on primary products
7. Trade
8. Market orientation
9. Colonial history

His data consists of 134 countries and 61 variables. There are some missing observations in his data set. Fernández et al. (2001) eliminate all countries or features having missing observations, and therefore use the data of 72 countries and 42 variables. On the other hand, we eliminated those countries for which more than 50% of their features are missing. Thus, we eliminated 9 countries and imputed the missing observations for 125 countries. As a result, we use the data of 125 countries and 61 variables.

We split the countries into two groups—high growth and low growth countries. The growth rate of each country in the former group is higher than the median growth rate, and that in the latter group is lower than the median growth rate. We used the growth rate (low growth = 0, high growth = 1) as the explained variable and the other 60 variables as explanatory variables.

It is well known that data standardization can improve performance. Classifiers are required to calculate the objective function, which is the mean squared error between the

[2] Some examples of each category are as follows: Geography (Latitude, Total area, etc.), Political institution (Civil liberties index, War dummy, etc.), Religion (Fraction of Buddhist, Fraction of Catholic, etc.), Market distortions and market performance (Exchange rate distortions, Black market premium, etc.), Investment and its composition (Equipment investment, Non-equipment investment, etc.), Dependence on primary products (Fraction of GDP in mining, etc.), Trade (Terms of trade growth, etc.), Market orientation (Degree of capitalism), and Colonial history (Spanish colony, French colony, etc.).

predicted value and the observation. If some of the features have a broad range of values, the mean squared error may be governed by these particular features, and objective functions may not work properly. Therefore, it is desirable to standardize the ranges of all features so that each feature contributes equally to the cost function (Aksoy and Haralick, 2001). Sola and Sevilla (1997) pointed out that data standardization prior to NN training enables researchers to speed up the calculations and obtain good results.

We standardized the data using the following formula:

$$z_i = \frac{x_i - \bar{x}}{\sqrt{s^2}} \tag{3}$$

where z_i is the normalized data; x_i is each dataset; $\bar{x}$ is the sample average of x_i; and s^2 is the sample variance of x_i. This method rescales the range of features to a mean of zero and a standard deviation of one. We analyzed both the original and standardized data in order to evaluate the robustness of our experimental results.[3]

2.3. Performance Evaluation Metrics

We used accuracy to evaluate the performance of each machine learning method. For our two-class problem, the confusion matrix (Table 1) gives a summary of prediction results on the classification problem. The confusion matrix shows not only the errors of our classification method but also the different types of errors.

Note that true negative (TN) indicates the case where the actual class is no-event and we correctly predict the case to be no-event; false positive (FP) indicates the case where the actual class is no-event but we incorrectly predict the class to be event; false negative (FN) indicates the case where the actual class is event but we incorrectly predict the class to be no-event; and true positive (TP) indicates the case where the actual class is event and we correctly predict the class to be event. Then, the prediction accuracy rate is defined as

[3] Shanker et al. (1996) analyzed the effect of data standardization on the performance of neural networks. They found that neural networks yield better results in general if data are standardized but that the advantage diminishes as the network and sample size become large.

$$\text{prediction accuracy rate} = \frac{TP + TN}{TP + FP + FN + TN} \tag{4}$$

Next, we analyzed the classification ability of each method by examining the receiver operating characteristic (ROC) curve and the AUC value. When considering whether a model is appropriate, it is not sufficient to rely solely on the accuracy rate. The proportion of correctly identified instances in the given class is called the true positive rate; the proportion of incorrectly identified instances in the given class is called the false positive rate. When the false positive rate is plotted on the horizontal axis and the true positive rate on the vertical axis, the combination produces an ROC curve. A good model is one that shows a high true positive rate value and a low false positive value. AUC refers to the area under the ROC curve. A perfectly random prediction yields an AUC of 0.5, i.e., the ROC curve is a straight line connecting the origin (0, 0) and the point (1, 1).

We also report the F-score for each case, which is defined as follows:

$$\text{F-score} = \frac{2 \times (recall) \times (precision)}{(recall) + (precision)} \tag{5}$$

where *recall* is equal to TP/(TP + FN) and *precision* is equal to TP/(TP + FP). Thus, the F-score is the harmonic average of recall and precision.

2.4. Experiment Procedure

We implemented the methods in R—specifically, the ‘H2O’ package for NN, and XGBoost. The number of units in the middle layers of the NNs is determined using the Bayesian optimization method. The ratio of training data to test data was set as 70–30%.

We analyzed the prediction accuracy rate of each method for the two cases, that is, original and normalized data. Further, we examined the classification ability of each method based on the AUC value and F-score.

3. Experimental Results

We carried out each experiment 100 times using the original data series and summarized the results in Table 2. This table indicates the average and standard deviation of the prediction accuracy rate, average AUC value, and average F-score. The numbers of units in the middle layers of the NNs are determined using the Bayesian optimization method for each experiment.

As shown in the table, the prediction accuracy rate of the training data is close to 100% for all methods. However, the average prediction accuracy rate ranges from 71.842% to 76.754% for NNs, and is 77.105% for XGBoost. Also, the standard deviation of the prediction accuracy rate for XGBoost is smaller than that of NNs. Thus, it is clear that XGBoost achieves better prediction accuracy than NNs. The AUC value ranges from 0.835 to 0.847 for NNs and is 0.771 for XGBoost. Furthermore, the F-score ranges from 0.751 to 0.776 for NNs and is 0.759 for XGBoost. Thus, XGBoost shows better performance in terms of prediction accuracy than NNs[4].

We carried out each experiment 100 times using the standardized data series and summarized the results in Table 3. This table indicates the average and standard deviation of the prediction accuracy rate, average AUC value, and average F-score. The numbers of units in the middle layers of the NNs are determined using the Bayesian optimization method for each experiment.

Surprisingly enough, the results shown in Table 3 are slightly different from those in Table 2. As shown in the table, the prediction accuracy rate of the training data is close to 100% for all methods. However, the average prediction accuracy rate of the test data ranges from 74.473% to 77.719% for NNs and is 77.315% for XGBoost. Therefore, the prediction accuracy of NNs for the normalized data series is better than that for the original data series. Jayalakshmi and Santhakumaran (2011) point out that statistical standardization techniques enhance the reliability of feed-forward backpropagation NNs and the performance of the data classification model, which is consistent with our results.

The prediction accuracy of XGBoost for the normalized data series is also better than that for the original data series. Note that the prediction accuracy of XGBoost is still as good as that of NNs. Note that the standard deviation of prediction accuracy rate for XGBoost is

[4] However, since DNN records better values for AUC, caution is required depending on the threshold value to be used.

smaller than that of NNs. The AUC value ranges from 0.832 to 0.856 for NNs and is 0.775 for XGBoost. Furthermore, the F-score ranges from 0.765 to 0.781 for NNs and is 0.756 for XGBoost. Thus, these results indicate that the performance of NNs improves for the standardized data series.

Next, we analyze the relative importance of explanatory variables to forecast the growth rate. Table 4 indicates the experimental results. In this table, we drop each category one by one and see how the result affects the prediction accuracy rate of the test data. We use XGBoost in this experiment. We use both original data and standardized data. Among them, we choose the variable whose correct answer rate is lower for both data as an important variable. It is clear from this table that if we drop "Political institution", "Investment and its composition", "Trade", and "Colonial history", then the prediction accuracy rate decreases for both data. Thus, we may say that "Political institution", "Investment and its composition," "Trade", and "Colonial history" are important factors in analyzing the economic growth.

"Investment and its composition" are found to be important variables in Sala-i-Martin (1997) and Fernández (2001), and machine learning has also supports this. "Political institution" and "Colonial history" are also chosen as important variable in Sala-i-Martin (1997), which is consistent with our results. It is interesting that "Trade" is considered to be an important factor to forecast the cross-country growth in our results. This variable is emphasized neither in Sala-i-Martin (1997) nor in Fernández (2001), and thus is something new as a result of using the machine learning technique.

4. Conclusion

Because of the rapid development of AI, machine learning methods have been widely applied to various research activities in economics and finance. There is a high degree of machine learning applied to econometrics and microeconomics, but this is not the case for macroeconomics (Chow, 2018). This paper uses machine learning methods for macroeconomic analysis. More specifically, cross-country growth patterns are analyzed using machine learning methods.

Five machine learning methods are used to predict economic growth based on a country's attributes, and their prediction accuracy is compared. The methods used are four neural network methods with different activation functions, and the XGBoost method. We compare their performance in terms of their ability to predict economic growth rate using three measures (prediction accuracy rate, AUC-value, and F-score).

The main results are summarized as follows:

1) For the original data, XGBoost outperforms NNs in terms of prediction accuracy.

2) Data standardization enhances the reliability of NNs. Further, it improves the prediction accuracy, AUC-value, and F-score of NNs.

3) XGBoost has a smaller standard deviation of prediction accuracy rate than that of NNs.

4) “Political institution”, “Investment and its composition”, “Trade”, and “Colonial history” are important factors for cross-country economic growth.

Sala-i-Martin (1997) and Fernández (2001) find that “Investment and its composition” is an important variable, which is consistent with our result. Sala-i-Martin (1997) also find that “Political institution” and “Colonial history” are important variables, which is also consistent with our result. Although Sala-i-Martin (1997) and Fernández (2001) did not find “Trade” to be an important growth factor, our results indicate that “Trade” is also an important factor in forecasting cross-country growth.

References

Ahmad, M. W., Mourshed, M., and Rezgui, Y. 2017. Trees vs. Neurons: Comparison between random forest and ANN for high-resolution prediction of building energy consumption. Energy and Buildings 147: 77–89.

Aksoy, Selim, and Robert M. Haralick. 2001. Feature normalization and likelihood-based similarity measures for image retrieval. Pattern Recognition. Letters 22: 563–82.

Angelini, Eliana, Giacomo di Tollo, and Andrea Roli. 2008. A neural network approach for credit risk evaluation. Quarterly Review of Economics and Finance 48: 733–55.

Austin, P.C., Tu, J.V., Ho, J. E., Levy, D., and Lee, D. S. 2013. Using methods from the data-mining and machine-learning literature for disease classification and prediction: A case study examining classification of heart failure subtypes. Journal of Clinical Epidemiology. Volume 66, Issue 4, pp. 398–407.

Breiman, Leo. 2001. Random forests. Machine Learning 45: 5–32.

Chen, Tianqi, and Carlos Guestrin. 2016. XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16). ACM, pp. 785–94.

Chow W. 2018. Prospects for economics in the machine learning and big data era. Keynote Speech. SIBR 2018 Conference on Interdisciplinary Business & Economics Research.

Ding, Xiao, Yue Zhang, Ting Liu, and Junwen Duan. 2014. Using structured events to predict stock price movement: An empirical investigation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). Association for Computational Linguistics, pp. 1415–25.

Fernández, Carmen, Eduard Ley, and Mark F. J. Steel. 2001. Model uncertainty in cross-country growth regressions. Journal of Applied Econometrics 16: 563–76.

Hamori, Shigeyuki, Minami Kawai, Takahiro Kume, Yuji Murakami, and Chikara Watanabe. 2018. Ensemble learning or deep learning? Application to default risk analysis. Journal of Risk and Financial Management 11: 12.

Hegazy, Osman, Omar S. Soliman, and Mustafa Abdul Salam. 2013. A machine learning model for stock market prediction. International Journal of Computer Science and Telecommunications 4(12): 17–23.

Jayalakshmi, T., and A. Santhakumaran. 2011. Statistical Normalization and Back Propagation for Classification. International Journal of Computer Theory and Engineering 3: 83–93.

Khashman, Adnan. 2009. A neural network model for credit risk evaluation. International Journal of Neural Systems 19: 285–94.

Khashman, Adnan. 2010. Neural networks for credit risk evaluation: Investigation of different neural models and learning schemes. Expert Systems with Applications 37: 6233–39.

Khemakhem, Sihem, and Younés Boujelbène. 2015. Credit risk prediction: A comparative study between discriminant analysis and the neural network approach. Accounting and Management Information Systems 14: 60–78.

Kurt, I., Ture, M., and Kurum, A. T. 2008. Comparing performances of logistic regression, classification and regression tree, and neural networks for predicting coronary artery disease. Expert Systems with Applications. Volume 34, Issue 1, pp. 366–374.

LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. 2015. Deep learning. Nature 521: 436–44.

Lucas Jr., Robert E. 1988. On the mechanics of economic development. Journal of Monetary Economics 22: 3–42.

Oberlechner, Thomas, and Sam Hocking. 2004. Information sources, news, and rumors in financial markets: Insights into the foreign exchange market. Journal of Economic Psychology 25: 407–24.

Oreski, Stjepan, Dijana Oreski, and Goran Oreski. 2012. Hybrid system with genetic algorithm and artificial neural networks and its application to retail credit risk assessment. Expert Systems with Applications 39: 12605–17.

Razi M. A. and Athappilly, K. 2005. A comparative predictive analysis of neural networks (NNs), nonlinear regression and classification and regression tree (CART) models, Expert Systems with Applications, Volume 29, Issue 1, pp. 65–74.

Romer, Paul M. 1986. Increasing returns and long-run growth. Journal of Political Economy 94: 1002–37.

Romer, Paul M. 1990. Endogenous technological change. Journal of Political Economy 98: S71–S102.

Roondiwala, Murtaza, Harshal Patel, and Sharaddha Varma. 2017. Predicting stock prices using LSTM. International Journal of Science and Research 6(4): 1754–56.

Sala-i-Martin, Xavier X. 1997. I just ran two million regressions. American Economic Review 87(2): 178–83.

Santos Pinheiro, Leonardo dos, and Mark Dras. 2017. Stock market prediction with deep learning: A character-based neural language model for event-based trading. In Proceedings of the Australasian Language Technology Association Workshop 2017. Association for Computational Linguistics, pp. 6–15.

Schapire, Robert E. 1999. A brief introduction to boosting. In Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence (IJCAI-99) (II). IJCAI, pp. 1401–06.

Schapire, Robert E., and Yoav Freund. 2012. Boosting: Foundations and Algorithms. Cambridge, MA, MIT Press.

Shanker, M., M. Y. Hu, and M. S. Hung. 1996. Effect of data standardization on neural network training. Omega 24: 385–97.

Solow, Robert M. 1956. A contribution to the theory of economic growth. Quarterly Journal of Economics 70: 65–94.

Sola, J., and Joaquin Sevilla. 1997. Importance of input data normalization for the application of neural networks to complex industrial problems. IEEE Transactions on Nuclear Science 44: 1464–68.

Srivastava, Nitish, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. 2014. Dropout: A simple way to prevent neural networks from overfitting. Journal of Machine Learning Research 15: 1929–58.

Swan, T. W. 1956. Economic growth and capital accumulation. Economic Record 32: 334–61.

Tso, G. K. F. and Yau, K. K. W. 2005. Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks. Energy, Volume 32, Issue 9, pp. 1761–1768

Yeh, I-Cheng, and Che-hui Lien. 2009. The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients. Expert Systems with Applications 36: 2473–80.

Zhuge, Qun, Lingyu Xu, and Gaowei Zhang. 2017. LSTM neural network with emotional analysis for prediction of stock price. Engineering Letters 25: 167–75.

Table 1

Confusion Matrix

		Predicted Class	
		No-Event	Event
Actual Class	No-Event	TN (True Negative)	FP (False Positive)
	Event	FN (False Negative)	TP (True Positive)

Table 2

Experimental results for original data

	Training Data	Test Data		
	Prediction Accuracy Rate (%)	Prediction Accuracy Rate (%)	AUC Value	F-score
NN (Tanh)	100.000 (0.000)	74.298 (5.810)	0.847	0.770
NN (Tanh, Dropout)	98.421 (3.49)	76.754 (6.486)	0.8412	0.776
NN (ReLU)	100.000 (0.000)	71.842 (6.987)	0.835	0.751
NN (ReLU, Dropout)	99.298 (2.065)	76.052 (6.641)	0.843	0.764
XGBoost	100.000 (0.000)	77.105 (5.528)	0.771	0.759

Note: Numbers in parentheses indicate standard deviation.

Table 3

Experimental results for standardized data

	Training Data	Test Data		
	Prediction Accuracy Rate (%)	Prediction Accuracy Rate (%)	AUC Value	F-score
NN (Tanh)	99.912 (0.480)	74.649 (6.823)	0.844	0.771
NN (Tanh, Dropout)	98.421 (3.564)	76.052 (5.998)	0.832	0.765
NN (ReLU)	99.473 (1.274)	74.473 (6.451)	0.844	0.774
NN (ReLU, Dropout)	100.000 (0.000)	77.719 (6.726)	0.856	0.781
XGBboost	100.000 (0.000)	77.315 (5.705)	0.775	0.756

Note: Numbers in parentheses indicate standard deviation.

Table 4

Relative Importance of Each Category

Dropped Category	Original data Prediction Accuracy Rate (%)	Standardized data Prediction Accuracy Rate (%)
No	77.105	77.315
Geography	77.315	77.315
Political institution	**76.947**	**76.526**
Religion	77.473	77.052
Market distortions and market performance	76.736	78.684
Investment and its composition	**76.842**	**75.578**
Dependence on primary products	76.947	78.210
Trade	**76.263**	**77.052**
Market orientation	77.421	76.526
Colonial history	**76.789**	**77.157**

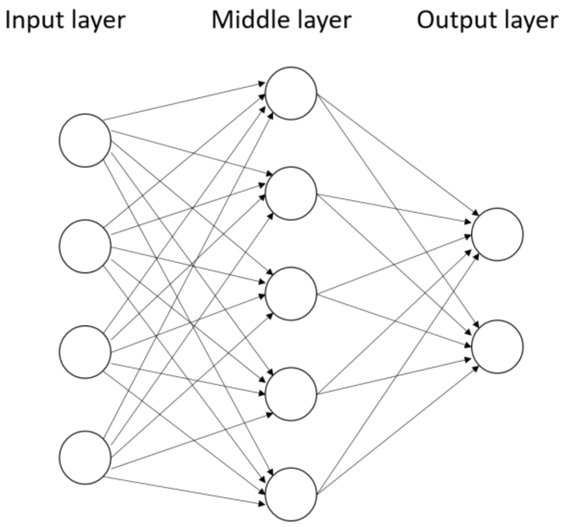
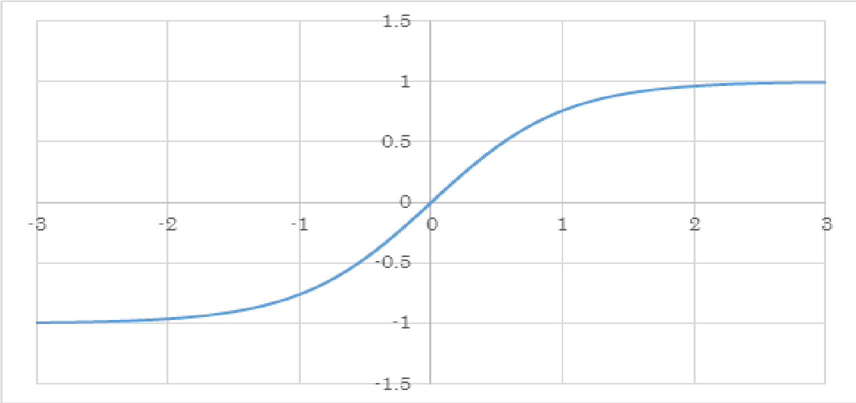
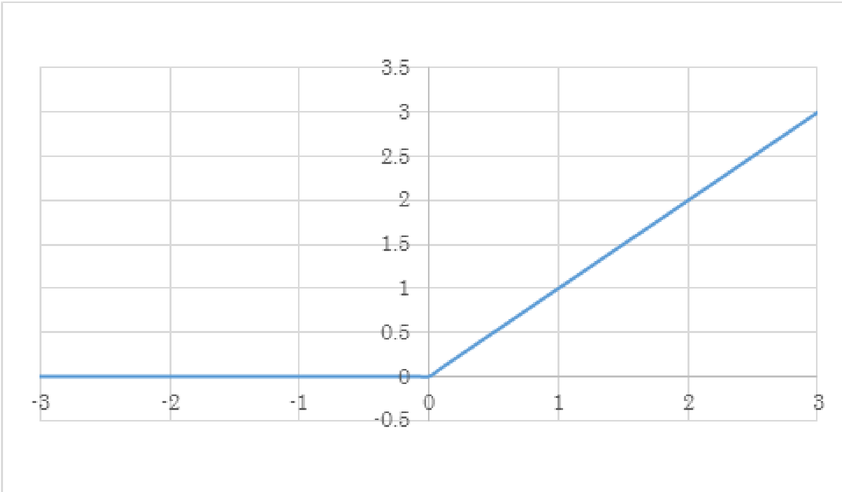


Figure 1. Structure of a fully connected neural network

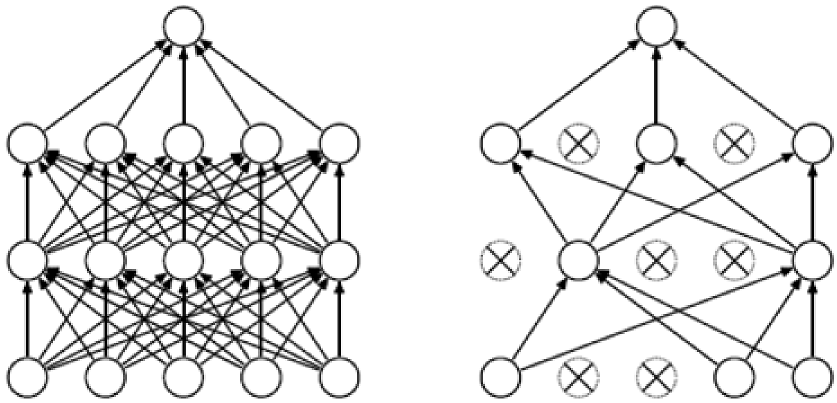


(a)



(b)

Figure 2. (a) Tanh function, and (b) ReLU function



(a) **(b)**

Figure 3. Illustration of the dropout technique. (a) Standard neural net, and (b) neural net after applying dropout. (Source: Srivastava et al. (2014))

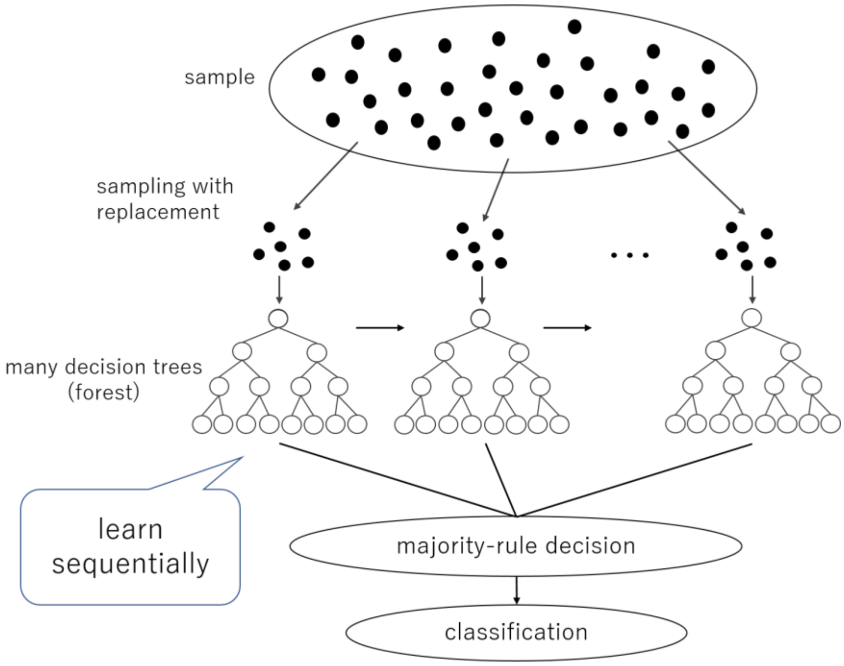


Figure 4. Basic concept of boosting