Does Ensemble Learning Always Lead to Better Forecasts?

Hitoshi Hamori[1] & Shigeyuki Hamori[2]

[1] Department of Aeronautics and Astronautics, University of Tokyo, 7-3-1, Hongo, Bunkyo-Ku, Tokyo 113-8654, Japan

[2] Graduate School of Economics, Kobe University, 2-1, Rokkodai, Nada-Ku, Kobe 657-8501, Japan

Correspondence: Shigeyuki Hamori, Graduate School of Economics, Kobe University, 2-1, Rokkodai, Nada-Ku, Kobe 657-8501, Japan.

Abstract

Ensemble learning is a common machine learning technique applied to business and economic analysis in which several classifiers are combined using majority voting for better forecasts as compared to those of individual classifier. This study presents a counterexample, which demonstrates that ensemble learning leads to worse classifications than those from individual classifiers, using two events and three classifiers. If there is an outstanding classifier, we should follow its forecast instead of using ensemble learning.

Keywords: ensemble learning, forecasts, classifier, majority voting

JEL Classification: C45; C58

1. Introduction

The impact of artificial intelligence on society has increased due to the availability of big data and rapid advances in computer technology. The application of machine learning, an aspect of artificial intelligence, in business and economic analysis has been explored in energy economics by Tso and Yau (2005), Weron (2014), Ziel and Steinert (2016), and Lago et al. (2018); stock price forecasting by Zhang et al. (1998), Hegazy et al. (2013), Rather et al. (2015), and Moghaddam et al. (2016); early warning systems by Tanaka et al. (2016); financial hazard map by Tanaka et al. (2018); and credit risk assessment by Angelini et al. (2008), Khashman (2009), Khashman (2010), Khemakhem and Boujelbene (2015), and Hamori et al. (2018).

Ensemble learning, among others, is one of the most useful machine learning techniques, in which multiple learners are trained to solve the same problem. Several classifiers are combined and majority voting is used, which leads to better forecasts than from an individual classifier. Bagging (Breiman, 1996), random forest (Breiman, 2001), and boosting (Schapire 1999; Shapire and Freund 2012) are well-known methods of ensemble learning. This idea has also been extended to Chen and Guestrin's (2016) XGboost and Ke et al.'s (2017) Light GBM.

As compared to deep learning, ensemble learning through methods, such as random forest has several merits. Deep learning requires tedious hyper parameter setting, while random forest requires simple hyper parameter setting; only the number of trees. Deep learning is a black box and provides results that are difficult to interpret, while random forest uncovers important variables. Furthermore, deep learning requires high machine power and computational cost to obtain high performance, while random forest does not. Thus, there are many applications of ensemble learning in economic and business data. For example, Tanaka et al. (2016) and Tanaka et al. (2018) use a random forest approach for risk management.

It is generally believed that ensemble learning leads to better classification than an individual classifier (Lantz, 2015). This study sets up a simple example model using two events and three classifiers to provide a counter example that demonstrates that ensemble learning leads to worse classification than that of individual classifier. To the best of our knowledge, this is the first study that emphasizes the problem of ensemble learning.

In section 2, we illustrate the model and counter example. Section 3 summarizes the conclusion.

2. Model and Counter Example

For simplicity, assume that there are two events, A or B to classify and three classifiers—M1, M2, and M3. Events A and B are assumed to be exclusive. For example, each classifier forecasts whether stock prices will rise or fall in the next period. Then, we consider three cases.

Case 1: Each classifier has the same accuracy rate with $(1/2 < p < 1)$.

Table 1 indicates the situations where the majority vote leads to answer A. For example, M1 and M2 forecast Event A but M3 forecasts Event B in Situation 1. According to the majority vote, we choose Event A in this situation. Similarly, we choose Event A for situations 2, 3, and 4 based on majority vote.

Table 1. Majority Rule

	M1	M2	M3
Situation 1	A	A	B
Situation 2	A	B	A
Situation 3	B	A	A
Situation 4	A	A	A

Then, the probability that majority rule will lead to each classification is given by Table 2.

Table 2. Majority Rule: Case 1

	Event A	Event B
Situation 1	$pp(1-p)$	$(1-p)(1-p)p$
Situation 2	$p(1-p)p$	$(1-p)p(1-p)$
Situation 3	$(1-p)pp$	$p(1-p)(1-p)$
Situation 4	ppp	$(1-p)(1-p)(1-p)$

In summary, if we use majority rule and forecast Event A using M1, M2, and M3, then the probability that the true event is Event A can be calculated as follows:

$$\frac{p^3 + 3p^2(1-p)}{p^3 + 3p^2(1-p) + 3p(1-p)^2 + (1-p)^3} = p^2(3-2p)$$

Let $\pi = p^3 + 3p^2(1-p)$. Then,

$$\frac{\pi}{p} = p^2 + 3p(1-p) = -2\left(p - \frac{3}{4}\right)^2 + \frac{9}{8} > 1 \text{ for } \frac{1}{2} < p < 1$$

Thus, $\pi > p$ for $1/2 < p < 1$. Therefore, the majority rule leads to a better forecast than from each classifier. As a numerical example, we have $\pi = 0.784$ for $p = 0.7$ and $\pi = 0.896$ for $p = 0.8$.

Case 2: One classifier has a low accuracy rate with p_2 and the others have a high accuracy rate with p_1 $(1/2 < p_2 < p_1 < 1)$.

Then, the probability that the majority rule leads to each classification is given in Table 3. In summary, if we use majority rule and forecast Event A using M1, M2, and M3, then the probability that the true event is Event A can be calculated as follows:

$$\frac{p_1^2(1-p_2) + 2p_1p_2(1-p_1) + p_1^2p_2}{p_1^2(1-p_2) + 2p_1p_2(1-p_1) + p_1^2p_2 + 2p_1(1-p_1)(1-p_2) + p_2(1-p_1)^2 + (1-p_1)^2(1-p_2))}$$
$$= 2p_1p_2(1-p_1) + p_1^2$$

Let $\pi = 2p_1p_2(1-p_1) + p_1^2$. Then, the π function can be drawn in the (p_2, π) plane, as shown in Figure 1.

Table 3. Majority Rule: Case 2

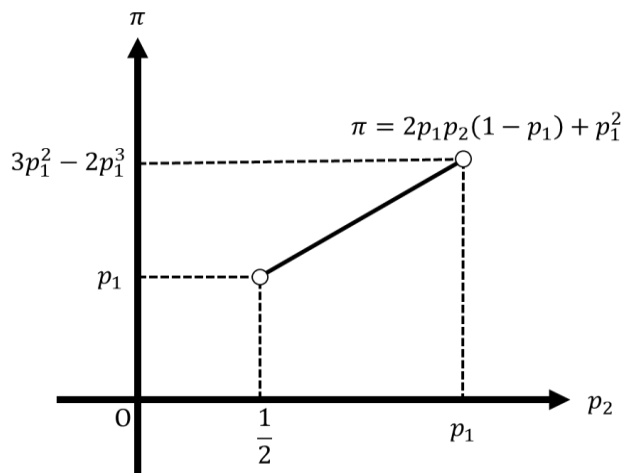|	Event A	Event B
Situation 1	$p_1 p_1 (1 - p_2)$	$(1 - p_1)(1 - p_1)p_2$
Situation 2	$p_1 (1 - p_1) p_2$	$(1 - p_1) p_1 (1 - p_2)$
Situation 3	$(1 - p_1) p_1 p_2$	$p_1 (1 - p_1)(1 - p_2)$
Situation 4	$p_1 p_1 p_2$	$(1 - p_1)(1 - p_1)(1 - p_2)$

Figure 1. $\pi = 2p_1 p_2 (1 - p_1) + p_1^2$ function

From Figure 1, $\pi > p_1$ is established for $1/2 < p_2 < p_1 < 1$. This example also indicates that majority rule leads to a better forecast than from each classifier. As a numerical example, we have $\pi = 0.936$ for $p_1 = 0.9$ and $p_2 = 0.7$, and $\pi = 0.864$ for $p_1 = 0.8$ and $p_2 = 0.7$.

Case 3: One classifier has a high accuracy rate with p_1 and the others have a low accuracy rate with p_2 ($1/2 < p_2 < p_1 < 1$).

Then, the probability that the majority rule leads to each classification is given in Table 4. In summary, if we use majority rule and forecast Event A using M1, M2, and M3, then the probability that the true event is Event A can be calculated as follows:

$$\frac{2p_1 p_2 (1 - p_2) + (1 - p_1)p_2^2 + p_1 p_2^2}{2p_1 p_2 (1 - p_2) + (1 - p_1)p_2^2 + p_1 p_2^2 + 2(1 - p_1)p_2 (1 - p_2) + p_1 (1 - p_2)^2 + (1 - p_1)(1 - p_2)^2}$$
$$= 2p_1 p_2 (1 - p_2) + p_2^2$$

Let $\pi = 2p_1 p_2 (1 - p_2) + p_2^2$. Then, π function can be drawn in the (p_1, π) plane, as shown in Figure 2.

Table 4. Majority Rule: Case 3

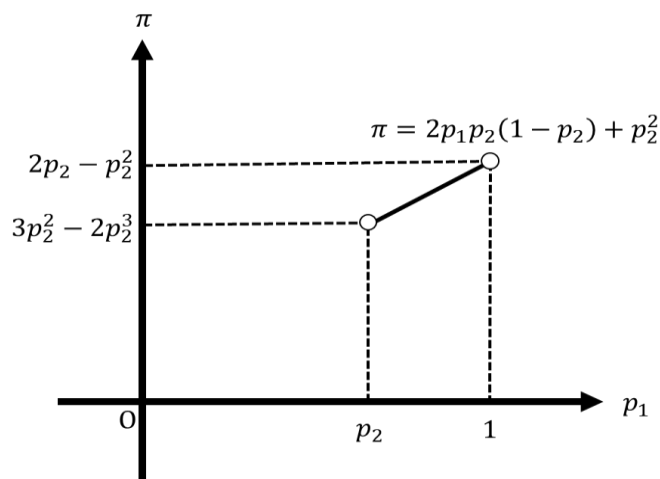|	Event A	Event B
Situation 1	$p_1 p_2 (1 - p_2)$	$(1 - p_1)(1 - p_2)p_2$
Situation 2	$p_1 (1 - p_2) p_2$	$(1 - p_1) p_2 (1 - p_2)$
Situation 3	$(1 - p_1) p_2 p_2$	$p_1 (1 - p_2)(1 - p_2)$
Situation 4	$p_1 p_2 p_2$	$(1 - p_1)(1 - p_2)(1 - p_2)$

Figure 2. $\pi = 2p_1p_2(1-p_2) + p_2^2$ function

For $1/2 < p_2 < 1$, we have $3p_2^2 - 2p_2^3 > p_2$ (See case 2). We also have

$$2p_2 - p_2^2 = -(p_2 - 1)^2 + 1 < 1$$

Figure 3 illustrates $\pi = p_1$ and $\pi = 2p_1p_2(1-p_2) + p_2^2$ in the (p_1, π) plane. As is clear from Figure 3, it does not necessarily hold true that $\pi > p_1$ for any p_1. Figure 3 indicates that $\pi > p_1$ for $p_2 < p_1 < p_1^*$ (majority rule leads to a better forecast than p_1), whereas $\pi < p_1$ for $p_1^* < p_1 < 1$ (majority rule leads to a worse forecast than p_1).[1] As a numerical example, we have $\pi = 0.868$ for $p_1 = 0.9$ and $p_2 = 0.7$, and $\pi = 0.826$ for $p_1 = 0.8$ and $p_2 = 0.7$.

Thus, we show that majority rule does not necessarily lead to a better classification than from an individual classifier. Note that if there is an outstanding classifier, we should not choose majority rule but follow its forecast.
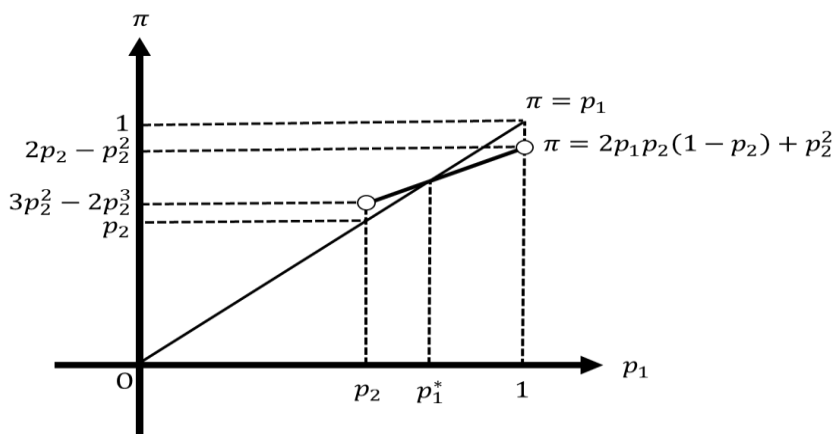

Figure 3. $\pi = p_1$ and $\pi = 2p_1p_2(1-p_2) + p_2^2$ functions

3. Conclusions

The impact of artificial intelligence on society has increased significantly due to the availability of big data and rapid advances in computer technology. Ensemble learning, among others, is one of the most useful machine learning

---

[1] The intersection of the two linear lines $\pi = p_1$ and $\pi = 2p_1p_2(1-p_2) + p_2^2$ leads to $p_1 = p_1^*$. Thus,

$$2p_1^*p_2(1-p_2) + p_2^2 = p_1^*$$

$$p_1^* = \frac{p_2^2}{1 - 2p_2 + 2p_2^2}$$

techniques, in which multiple learners are trained to solve the same problem. It is generally believed that ensemble learning leads to better classification than that of each classifier alone (Lantz 2015).

We demonstrate that ensemble learning does not necessarily lead to better classification than from each classifier using two events and three classifiers. The main results are as follows:

(1) When all the classifiers have an equal accuracy rate, ensemble learning always leads to better forecasts.

(2) When one classifier has a low accuracy rate and the others have high accuracy rates, ensemble learning always leads to better forecasts.

(3) When one classifier has a high accuracy rate and the others have low accuracy rates, ensemble learning does not necessarily lead to better forecasts. We should follow the forecasts of an outstanding classifier (if it exists) instead of using ensemble learning.

Thus, this study presents a simple counter example that demonstrates that ensemble learning leads to worse classification than from an individual classifier. We should adopt the forecast of an outstanding classifier instead of using ensemble learning.

Acknowledgements

References

Angelini, E., di Tollo, G., & Roli, A. (2008). A neural network approach for credit risk evaluation. *Quarterly Review of Economics and Finance, 48,* 733-755. https://doi.org/10.1016/j.qref.2007.04.001

Breiman, L. (1996). Bagging predictors. *Machine Learning, 24,* 123-140. https://doi.org/10.1007/BF00058655

Breiman, L. (2001). Random forests. *Machine Learning, 45,* 5-32. https://doi.org/10.1023/A:1010933404324

Chen, T., & Guestrin, C. (2016). *XGBoost: A Scalable Tree Boosting System.* KDD '16, August 13-17, 2016, San Francisco, CA, USA. http://dx.doi.org/10.1145/2939672.2939785

Hamori, S., Kawai, M., Kume, T., Murakami, Y., & Watanabe, C. (2018). Ensemble learning or deep learning? Application to default risk analysis. *Journal of Risk and Financial Management, 11,* 12. https://doi.org/10.3390/jrfm11010012

Hegazy, O., Soliman, O. S., & Salam, M. A. (2013). A machine learning model for stock market prediction. *International Journal of Computer Science and Telecommunications, 4,* 17-23.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., … Liu, T. Y. (2017). LightGBM: A Highly Efficient Gradient Boosting Decision Tree. *31st Conference on Neural Information Processing Systems* (NIPS 2017), Long Beach, CA, USA.

Khashman, A. (2009). A neural network model for credit risk evaluation. *International Journal of Neural Systems, 19,* 285-294. https://doi.org/10.1142/S0129065709002014

Khashman, A. (2010). Neural networks for credit risk evaluation: Investigation of different neural models and learning schemes. *Expert Systems with Applications, 37,* 6233-6239. https://doi.org/10.1016/j.eswa.2010.02.101

Khemakhem, S., & Boujelbene, Y. (2015). Credit risk prediction: A comparative study between discriminant analysis and the neural network approach. *Accounting and Management Information Systems, 14,* 60-78.

Lago, J., De Ridder, F., Vrancx, P., & De Schutter, B. (2018). Forecasting day-ahead electricity prices in Europe: The importance of considering market integration. *Applied Energy, 11,* 890-903. https://doi.org/10.1016/j.apenergy.2017.11.098

Lantz, B. (2015). *Machine Learning with R* (2nd ed.). Birmingham: Packt Publishing Ltd.

Moghaddam, A. H., Moghaddam, M. H., & Esfandyari, M. (2016). Stock market index prediction using artificial neural network. *Journal of Economics, Finance and Administrative Science, 21,* 89-93. https://doi.org/10.1016/j.jefas.2016.07.002

Rather, A. M., Agarwal, A., & Sastry, V. N. (2015). Recurrent neural network and a hybrid model for prediction of stock returns. *Expert Systems with Applications, 42,* 3234-3241. https://doi.org/10.1016/j.eswa.2014.12.003

Schapire, R. E. (1999). *A brief introduction to boosting.* Paper presented at the Sixteenth International Joint Conference on Artificial Intelligence, Stockholm, Sweden, July 31–August 6, pps. 1–6.

Shapire, R. E., & Freund, Y. (2012). *Boosting: Foundations and Algorithms.* Cambridge: The MIT Press.

Tanaka, K., Kinkyo, T., & Hamori, S. (2016). Random Forests-based Early Warning System for Bank Failures. *Economics Letters, 148,* 118-121. https://doi.org/10.1016/j.econlet.2016.09.024

Tanaka, K., Kinkyo, T., & Hamori, S. (2018). Financial Hazard Map: Financial Vulnerability Predicted by a Random Forests Classification Model. *Sustainability, 10,* 1530. https://doi.org/10.3390/su10051530

Tso, G. K. F., & Yau, K. K. W. (2005). Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks. *Energy, 32,* 1761-1768. https://doi.org/10.1016/j.energy.2006.11.010

Weron, R. (2014). Electricity price forecasting: A review of the state-of-the-art with a look into the future. *International Journal of Forecasting, 30,* 1030-1081. https://doi.org/10.1016/j.ijforecast.2014.08.008

Zhang, G., Patuwo, B. E., & Hu, M. Y. (1998). Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting, 14,* 35-62. https://doi.org/10.1016/S0169-2070(97)00044-7

Ziel, F., & Steinert, R. (2016). Electricity price forecasting using sale and purchase curves: The X-Model. *Energy Economics, 59,* 435-454. https://doi.org/10.1016/j.eneco.2016.08.008