Published in final edited form as:

Proc Int Conf Mach Learn Appl. 2010 December 12; : 373–378. doi:10.1109/ICMLA.2010.62.

The Upper and Lower Bounds of the Prediction Accuracies of Ensemble Methods for Binary Classification

Xueyi Wang and

Department of Mathematics and Computer Science, Northwest Nazarene University, Nampa, ID, USA, xwang@nnu.edu

Nicholas J. Davidson

Department of Mathematics, Boise State University, Boise, ID, USA, njdavidson03@gmail.com

Abstract

Ensemble methods have been widely used to improve prediction accuracy over individual classifiers. In this paper, we achieve a few results about the prediction accuracies of ensemble methods for binary classification that are missed or misinterpreted in previous literature. First we show the upper and lower bounds of the prediction accuracies (i.e. the best and worst possible prediction accuracies) of ensemble methods. Next we show that an ensemble method can achieve > 0.5 prediction accuracy, while individual classifiers have < 0.5 prediction accuracies. Furthermore, for individual classifiers with different prediction accuracies, the average of the individual accuracies determines the upper and lower bounds. We perform two experiments to verify the results and show that it is hard to achieve the upper and lower bounds accuracies by random individual classifiers and better algorithms need to be developed.

Keywords

ensemble methods; binary classification; prediction accuracy; upper bound; lower bound

I. INTRODUCTION

Supervised learning learns a function $y = f(\mathbf{x})$ from a set of training examples in the form of $(\mathbf{x}, f(\mathbf{x}))$, where input $\mathbf{x}$ is usually a vector, and then predicts output y of any valid $\mathbf{x}$. The output y could be drawn from either a continuous value (regression) or from a discrete set of classes (classification). In this paper, we consider the binary classification only, where the y is drawn from $\{-1, 1\}$.

Using ensemble methods to improve the prediction accuracy of classification problems has become a well established research area. An ensemble method usually has an odd number of individual classifiers and uses a simple majority voting rule, where the prediction goes with the majority of the individual classifiers.

To build a good ensemble method, we need to find a set of individual classifiers so that the prediction accuracy of the ensemble method is better than that of each individual classifier or the average. Some ensemble methods use individual classifiers generated from a single machine learning algorithm with different subsets of training data. In these methods, empirical studies have demonstrated that bagging, boosting, and their variants can successfully improve the prediction accuracy [2,3,7]. Some other ensemble methods combine individual classifiers generated from two or more different machine learning algorithms, where the individual classifiers may still be generated from single machine learning algorithms via bagging, boosting, or other methods [4,11]. Although ensemble

methods have been successfully used to improve the prediction accuracies in some examples, it has been argued that they may not perform better than the best individual classifiers [6].

A lot of research has been done in understanding the performance of ensemble methods. It has been proposed that the key to build a good ensemble method is to find a set of accurate and diverse individual classifiers [8]. Accurate individual classifiers mean their prediction accuracies should be better than random, which is 0.5 for binary classifiers. While there is no well-accepted definition or measure for the diversity of individual classifiers [9], diverse classifiers usually mean that individual classifiers predict the output differently, so those outputs incorrectly predicted by a minority of individual classifiers may be correctly predicted by a majority of them and the accuracy of the ensemble method is improved. [5] proposed statistical, computational, and representational reasons why an ensemble method may perform better than its individual classifiers. Furthermore, some research proposed various methods for measuring the diversity of individual classifiers. [9] proposed ten statistics for measuring diversity but raised questions on the usefulness of measuring diversity for empirical problems, [1] proposed the Percentage Correct Diversity Measure to measure the diversity of individual classifiers, and [4] proposed another set of ten measures for demonstrating the goodness of ensembles.

In this paper, we achieve a few results about the prediction accuracies of ensemble methods for binary classification that are missed or misinterpreted in previous literature. Given the prediction accuracies of individual classifiers, we show the upper and lower bounds of the prediction accuracies of ensemble methods that use the binary classification and simple majority voting rule. While previous research focuses on finding diverse classifiers to achieve better prediction accuracy, we show that the individual classifiers in an ensemble method that achieves the upper bound of the prediction accuracy are neither completely diverse nor completely converged. The upper bound, i.e. the best possible accuracy, is achieved when for each correctly predicted input by the ensemble method, the correct predictions of individual classifiers are just over majority (i.e. as diverse as possible) and for each incorrectly predicted input, all predictions of individual classifiers are incorrect (i.e. as converged as possible); the lower bound is achieved in exactly the reverse way. This result explains why sometimes a set of diverse classifiers cannot form a good ensemble method [6].

While it is well accepted that individual classifiers should have prediction accuracy > 0.5 in order to achieve better prediction accuracy for ensemble methods, we show that an ensemble method having all individual classifiers with prediction accuracy < 0.5 can achieve a better accuracy (e.g. > 0.5). Furthermore, we show that for individual classifiers with different prediction accuracies, the average of the individual accuracies determines the upper and lower bounds of the prediction accuracies of ensemble methods.

We perform two experiments for the same and different accuracies of individual classifiers, each using two different sizes of individual classifiers ($n = 3$ and 15). We show histograms of the prediction accuracies of ensemble methods and compare the distributions of the predictions by individual classifiers in the best and worst cases to the theoretical results. Finally, we perform an experiment showing that it is very hard to achieve the upper and lower bounds of the prediction accuracies of ensemble methods if we just random select a set of individual classifiers, so better algorithms must be developed to reach the theoretical upper bound.

The theoretical analysis performed in the paper is based on a training data set and any general machine learning algorithm for individual classifiers. As long as the testing data set

follows the same distribution and no overfitting, the training and testing data sets should generate similar prediction accuracies. While we do not propose any new algorithms in the paper, the theoretical results can be used as the guidelines for selecting a set of individual classifiers in developing new algorithms.

II. PREDICTION ACCURACY OF ENSEMBLE METHODS

We assume there are n individual binary classifiers BC_i for $(1 \leq i \leq n)$ in an ensemble method EM and there is a sufficient large and evenly distributed training set T with N inputs (e.g. vectors) $\mathbf{x}_j$ for $(1 \leq j \leq N)$. Given an input $\mathbf{x}_j$, each individual binary classifier BC_i predicts the output y_{ij} from $\{-1, 1\}$ separately.

We set $n = 2k+1$, where k is a natural number, and use a simple majority voting rule for the ensemble method EM . For a given input $\mathbf{x}_j$, the output y_j by EM is correct if and only if the outputs y_{ij} by at least $k+1$ individual classifiers are correct and the output y_j by EM is incorrect if and only if the outputs y_{ij} by at most k individual classifiers are correct. For the N inputs in the training set T , we assume each individual classifier BC_i for $(1 \leq i \leq n)$ can correctly predict N_i inputs and hence has a prediction accuracy $P(BC_i) = N_i / N = p_i$ or an error rate $e_i = 1 - P(BC_i) = 1 - p_i$.

We first discuss the prediction accuracy of ensemble methods with identical or independent individual classifiers. Then we show the upper and lower bounds of the prediction accuracies of ensemble methods, in which cases the individual classifiers are neither independent nor identical.

A. Identical or Independent Individual Classifiers

Identical and independent individual classifiers are two special cases and have been studied intensively [5,8]. When all individual classifiers are identical, each classifier predicts the same output for each input $\mathbf{x}_j$ ($1 \leq j \leq N$). So we have prediction accuracies $p_1 = p_2 = \dots = p_n = p$ and the ensemble method EM has the same prediction accuracy to the individual classifiers:

$$P(EM) = p \quad (1)$$

If all individual classifiers are independent and have the prediction accuracies $p_1 = p_2 = \dots = p_n = p$, then the classifiers follow binomial distribution and we can obtain the prediction accuracy of the ensemble method using the cumulative binomial distribution [10], where the ensemble method makes a correct prediction if and only if at least $k+1$ individual classifiers make correct predictions:

$$P(EM) = \sum_{i=k+1}^n \binom{n}{i} p^i (1-p)^{n-i} \quad (2)$$

It can be shown that when $p < 0.5$, $P(EM)$ is monotonically decreasing as n increases and when $p > 0.5$, $P(EM)$ is monotonically increasing as n increases. Figure 1 shows the $P(EM)$ for $n = 3, 5, 7$, and 9 .

Previous research has suggested that we should always use individual classifiers with $p > 0.5$ in an ensemble method in order to achieve $P(EM) > p$ [8]. But the equation (2) is obtained under the assumption that all individual classifiers are independent. As we will show in the next subsection, the prediction accuracy in the equation (2) is not the upper bound for an

ensemble method and it is possible for an ensemble method having all individual classifiers with $p < 0.5$ to achieve $P(EM) > 0.5$.

B. The Upper and Lower Bounds of the Prediction Accuracies of Ensemble Methods

First we assume that all individual classifiers BC_i ($1 \leq i \leq n$) have the same prediction accuracies: $p_1 = p_2 = \dots = p_n = p$ and discuss the case of different individual prediction accuracies later. Theorem 1 shows the upper and lower bounds of the prediction accuracies of an ensemble method.

Theorem 1. For an ensemble method having $n = 2k+1$ individual binary classifiers BC_i ($1 \leq i \leq n$) with the same prediction accuracy p , the upper bound of the prediction accuracy is

$$P(EM)_{\text{upper}} = \min(np/(k+1), 1) \quad (3)$$

and the lower bound of the prediction accuracy is

$$P(EM)_{\text{lower}} = \max((np - k)/(k+1), 0). \quad (4)$$

Proof. To obtain the upper bound of the prediction accuracy, i.e. the best possible prediction accuracy, we should maximize the correct prediction rate (and minimize the error rate). For a given input $\mathbf{x}_j$, the ensemble method predicts the output y_j correctly if and only if at least $k+1$ individual classifiers predict the output correctly. When each correctly predicted $\mathbf{x}_j$ by the ensemble method is correctly predicted by exactly $k+1$ individual classifiers, we maximize the correct rate for the ensemble method.

Given a total of N inputs, each individual classifier makes Np correct predictions and all individual classifiers make a total of nNp correct predictions. In maximizing the correct rate, the ensemble method correctly predicts $nNp/(k+1)$ inputs and we have the upper bound of the prediction accuracy $P(EM)_{\text{upper}} = np/(k+1)$.

Since $P(EM) \leq 1$, we have $P(EM)_{\text{upper}} = \min(np/(k+1), 1)$.

To obtain the lower bound of the prediction accuracy, i.e. the worst possible prediction accuracy, we should maximize the error rate (and minimize the correct rate). For a given input $\mathbf{x}_j$, the ensemble method predicts the output y_j incorrectly if and only if at least $k+1$ individual classifiers predict the output incorrectly. When each incorrectly predicted $\mathbf{x}_j$ by the ensemble method is incorrectly predicted by exactly $k+1$ individual classifiers (i.e. correctly predicted by k individual classifiers), we maximize the error rate for the ensemble method.

Follow the above reasoning by replacing the correct rate with the error rate in $P(EM)_{\text{upper}}$, we have $1 - P(EM)_{\text{lower}} = n(1-p)/(k+1)$.

Simplify it, we have $P(EM)_{\text{lower}} = (np - k)/(k+1)$.

Since $P(EM) \geq 0$, we have $P(EM)_{\text{lower}} = \max((np - k)/(k+1), 0)$.

Before the upper bound of the prediction accuracy $P(EM)_{\text{upper}}$ reaches 1, since no correct predictions by individual classifiers are wasted on incorrectly predicted inputs by the ensemble method, so we achieve both the maximum correct prediction rate and the minimum error rate. In the same reasoning, before the lower bound of the prediction accuracy $P(EM)_{\text{lower}}$ reaches 0, we achieve both the minimum correct prediction rate and the

maximum error rate. Figure 2 shows an example for the upper and lower bounds of the prediction accuracies achieved by an ensemble method with $n = 3$ and $p = 0.6$, where the upper bound accuracy is 0.9 and the lower bound accuracy is 0.4.

While previous research focuses on finding diverse individual classifiers but yet to reach an agreement on how to measure the diversity of individual classifiers [9], we have three observations from our theorem, showing that we cannot simply find a diverse set of individual classifiers. First, the upper bound of the prediction accuracy cannot be achieved if the individual classifiers are independent. By comparing the Figure 1 and Figure 3, we can see that the theoretical upper bound accuracy of an ensemble method is much better than the best accuracy with independent classifiers. Second, the upper bound accuracy is achieved when for each correctly predicted input by the ensemble method, the predictions of individual classifiers are as dispersed as possible and for each incorrectly predicted input, the predictions of individual classifiers are as concentrated as possible. In this case, the individual classifiers are diverse on some predictions but converged on other predictions. Third, if we focus on finding a diverse set of individual classifiers, it is equally likely that the prediction accuracy of an ensemble method goes better or worse, where for the upper bound accuracy, all correct predictions are dispersed and all incorrect predictions are concentrated.

The Boost by Majority algorithm [12] uses a similar way to obtain a good ensemble method, where in each iteration of the algorithm, it concentrates the input data near the decision threshold to generate a new individual classifier and improves the prediction accuracy. It differs from the way of achieving the upper bound accuracy, where we need to concentrate on all the input data below the decision threshold and add multiple individual classifiers that make dispersed correct prediction on these input data.

Figure 3 shows $P(EM)_{\text{upper}}$ and $P(EM)_{\text{lower}}$ for $n = 3, 5, 7$, and 9. Both $P(EM)_{\text{upper}}$ and $P(EM)_{\text{lower}}$ are two straight lines with the same slope of $n/(k+1)$. When $n = 1$, we have $P(EM)_{\text{upper}} = P(EM)_{\text{lower}} = p$, which is the same to the prediction accuracy of an individual classifier. Since we have $P(EM)_{\text{upper}} > p > P(EM)_{\text{lower}}$, it shows that $P(EM)$ could be better or worse than p for any given p .

While individual classifiers with $p < 0.5$ are below random prediction rate and are generally not considered for ensemble methods (apparently if an individual classifier BC_i has $P(BC_i) = p < 0.5$, we can just use the reverse classifier $\neg BC_i$ that has $P(\neg BC_i) = 1 - p > 0.5$), we may still achieve $P(EM) \geq 0.5$ by these individual classifiers. We have the following corollary.

Corollary 1. For an ensemble method having $n = 2k+1$ individual binary classifiers BC_i ($1 \leq i \leq n$) with the same prediction accuracy p , if $p \geq (k+1)/2n$, then $P(EM)_{\text{upper}} \geq 0.5$.

Proof. This is obtained immediately by letting $P(EM)_{\text{upper}} = np/(k+1) \geq 0.5$.

The Corollary 1 shows the minimum requirement of the prediction accuracy of individual classifiers for achieving $P(EM) \geq 0.5$. For example, if $n = 3$, we just need $p \geq 1/3$ to make $P(EM)_{\text{upper}} \geq 0.5$.

One interesting result is, if we allow to use $\neg BC_i$ instead of BC_i when $p < 0.5$ (so we consider only the right half of the Figure 3), then when $n \rightarrow \infty$, we always have $P(EM)_{\text{upper}} = 1$ for any given p . It shows that when n is sufficiently large, it is possible to always achieve near 100% accuracy if we can find a good set of individual classifiers with any prediction accuracies.

For the case that individual classifiers have different prediction accuracies, we have the following theorem.

Theorem 2. For an ensemble method having $n = 2k+1$ individual binary classifiers BC_i with prediction accuracies p_i for $(1 \leq i \leq n)$, the upper bound of the prediction accuracy is

$$P(EM)_{\text{upper}} = \min(n\bar{p}/(k+1), 1) \quad (5)$$

and the lower bound of the prediction accuracy is

$$P(EM)_{\text{lower}} = \max((n\bar{p} - k)/(k+1), 0), \quad (6)$$

where $\bar{p} = \sum_{i=1}^n p_i/n$.

Proof. We show only the upper bound of the prediction accuracy here.

From the proof of Theorem 1, to maximize the correct rate and obtain the upper bound of the prediction accuracy of the ensemble method, each correctly predicted input $\mathbf{x}_j$ by the ensemble method should be correctly predicted by exactly $k+1$ individual classifiers. Given a total of N inputs, each individual classifier makes Np_i correct predictions and all individual classifiers make a total of $N \sum_{i=1}^n p_i = Nn\bar{p}$ correct predictions. In maximizing the correct rate, the ensemble method correctly predicts $Nn\bar{p}/(k+1)$ inputs and we have the upper bound of the prediction accuracy $P(EM)_{\text{upper}} = n\bar{p}/(k+1)$.

Since $P(EM) \leq 1$, we have $P(EM)_{\text{upper}} = \min(n\bar{p}/(k+1), 1)$.

Theorem 2 extends Theorem 1 to allow different prediction accuracies for different individual classifiers. By replacing p with $\bar{p}$, all the analysis for the Theorem 1 holds for the Theorem 2.

While Theorem 2 shows that the higher the $\bar{p}$, the better the $P(EM)_{\text{upper}}$, it should not be explained as just finding a few individual classifiers with best accuracies for an ensemble method. Theorem 2 shows the upper bound accuracy, but to achieve it, the correct predictions of individual classifiers should be as dispersed as possible and the incorrect predictions should be as concentrated as possible. A few best individual classifiers may not have this property. Also as we will see from Corollary 2 below, as long as $\bar{p} \geq (k+1)/n$, it is always possible to achieve the upper bound accuracy of 1, so finding a good distribution of individual classifiers is more important than finding a few best individual classifiers.

We end the subsection with Corollary 2, which shows when an ensemble method achieves the upper bound accuracy of 1 or the lower bound accuracy of 0, given the average prediction accuracy of individual classifiers $\bar{p}$.

Corollary 2. The upper bound of the prediction accuracy of an ensemble method $P(EM)_{\text{upper}} = 1$ if and only if

$$\bar{p} \geq (k+1)/n \quad (7)$$

and the lower bound of the prediction accuracy $P(EM)_{\text{lower}} = 0$ if and only if

$$\bar{p} \leq k/n \quad (8)$$

Proof. We can immediately obtain (7) and (8) by solving equations $P(EM)_{\text{upper}} = n\bar{p}/(k+1) \geq 1$ and $P(EM)_{\text{lower}} = (n\bar{p} - k)/(k+1) \leq 0$.

III. EXPERIMENTS AND DISCUSSIONS

We perform two experiments for the prediction accuracies of ensemble methods, where the individual classifiers have either the same or different prediction accuracies, and compare the distributions of the individual classifiers in the best and worst cases to the distribution analysis in Theorems 1 and 2. We create a testing data set containing 100 artificial inputs: half outputs “1” and half outputs “-1”. In the first experiment, all individual classifiers have the same prediction accuracy $p = 0.6$ and in the second experiment, all individual classifiers have prediction accuracies evenly distributed between 0.4–0.8 with the average $\bar{p} = 0.6$. The predictions of individual classifiers in both experiments are uniformly distributed. We run both experiments with different values of n and repeat 100,000 times. The experiments are done on MATLAB.

Figures 4 shows the results of the first experiments with $n = 3$ and 15. Figures 4a and 4b show the histograms of the prediction accuracies of the ensemble method in 100,000 runs. The histograms show that the individual classifiers used in the experiments that make uniformly distributed predictions are independent: the average prediction accuracies follow the cumulative binomial distribution with expected values of 0.65 for $n = 3$ (Figure 4a) and 0.79 for $n = 15$ (Figure 4b). The histograms follow normal distributions, which can be expected from the central limit theorem [10].

The histograms of correct predictions made by individual classifiers in the best case found, as shown in Figure 4c and 4d, and the histograms in the worst case found, as shown in Figure 4e and 4f, match the distribution analysis in Theorem 1. From the Theorems, before the upper bound of the prediction accuracy reaches 1, the upper bound accuracy is achieved when all correctly predicted outputs by the ensemble method are correctly predicted by exactly $k+1$ individual classifiers and no incorrectly predicted outputs are correctly predicted by any individual classifiers (the lower bound accuracy is achieved in the reverse way). By comparing Figure 4c to Figure 4e, we can see that in 4c, more outputs are correctly predicted by two or more individual classifiers, while in 4e, more outputs are correctly predicted by one or less individual classifier (i.e. these outputs are incorrectly predicted by the ensemble method), so the accuracy of the ensemble method in 4c is much better than 4e. Figures 4d and 4f show the same for $n = 15$, where 4d has more outputs agreed by eight or more individual classifiers and 4f has more agreed by only seven or less individual classifiers.

The histograms in Figure 5 are almost the same to those in Figure 4. It matches Theorem 2 that an ensemble method having individual classifiers with the average accuracy of $\bar{p}$ performs the same to those having individual classifiers with the same prediction accuracy p .

Although based on Theorems 1 and 2, the range of prediction accuracies of an ensemble method is large, $P(EM)_{\text{upper}} - P(EM)_{\text{lower}} = k/(k+1)$, it is unlikely to achieve either the upper bound or the lower bound if we just randomly pickup individual classifiers to form an ensemble method. Figure 6 shows the best and worst prediction accuracies with $p = 0.5$ and $n = 3, 5, \dots, 51$ in 1,000,000 runs and the upper and lower bounds of the prediction accuracies. The theoretically upper bound accuracies are mostly over 0.9, but all best

accuracies from the experiment are below 0.8, which are far from the upper bound. A better algorithm should choose a set of individual classifiers so that the correct predictions are as dispersed as possible and the incorrect predictions are as concentrated as possible. Theorems 1 and 2 and the distribution analysis in the paper can be used as the guidelines for developing new algorithms.

IV. CONCLUSION

We show two theorems about the upper and lower bounds of the prediction accuracies of ensemble methods using binary classification and a simple majority voting rule, which are missed in previous literature. To achieve the upper bound accuracy for the ensemble method, instead of finding a set of diverse individual classifiers, for each correctly predicted input by the ensemble method, the predictions of individual classifiers should be as dispersed as possible and for each incorrectly predicted input, the predictions of individual classifiers should be as concentrated as possible. For any prediction accuracies of individual classifiers, the ensemble method may achieve a better or worse accuracy depends on the distribution of correct and incorrect predictions by individual classifiers. We also show that an ensemble method can achieve > 0.5 accuracy if all individual classifiers have < 0.5 accuracies. Finally we perform several experiments showing that the experiments match the theorems well.

The theoretical analysis in the paper helps us to better understand the prediction accuracies of ensemble methods and clarify some misunderstandings. The distribution analysis of individual classifiers can be used as the guidelines for selecting individual classifiers in developing new algorithms with better performance, that is, making the correct predictions as dispersed as possible and incorrect predictions as concentrated as possible. It is also interesting to analyze existing ensemble methods such as bagging to see how the distribution of individual classifiers correlates with the prediction accuracy of an ensemble method.

Acknowledgments

This research is supported by NIH Grant # P20 RR016454 from the INBRE Program of the National Center for Research Resources.

REFERENCES

1. Banfield, BE.; Hall, LO.; Bowyer, KW.; Kegelmeyer, WP, Jr. Proceedings of the 4th International Workshop on Multiple Classifier Systems. Guildford, UK: 2003. A New Ensemble Diversity Measure Applied to Thinning Ensembles; p. 306-316.
2. Bauer E, Kohavi R. An Empirical Comparison of Voting Classification Algorithms: Bagging, Boosting, and Variants. *Machine Learning*. 1999; 36(1-2):105-139.
3. Breiman L. Bagging Predictors. *Machine Learning*. 1996; 24(2):123-140.
4. Caruana, R.; Niculescu-Mizil, A.; Crew, G.; Ksikes, A. Ensemble Selection from Libraries of Models. Proceedings of the 31st International Conference on Machine Learning; 2004.
5. Dietterich TG. Ensemble Methods in Machine Learning. Proceedings of the First International Workshop on Multiple Classifier Systems. Lecture Notes in Computer Science. 2001; 1857:1-15.
6. Dzeroski, S.; Zenko, B. Is Combining Classifiers Better than Selecting the Best One. Proceedings of the 19th International Conference on Machine Learning; 2002. p. 123-130.
7. Freund, Y.; Schapire, RE. Experiments with a New Boosting Algorithm. Proceedings of the 13th International Conference on Machine Learning; 1996. p. 325-332.
8. Hansen LK, Salamon P. Neural Network Ensembles. *IEEE Trans. Pattern Analysis and Machine Intell*. 1990; 12(10):993-1001.
9. Kuncheva LI, Whitaker CJ. Measures of Diversity in Classifier Ensembles and Their Relationship with the Ensemble Accuracy. *Machine Learning*. 2003; 51(2):181-207.

10. Spiegel, MR.; Stephens, LJ. Schaum's Outline of Statistics. 4th edition. McGraw-Hill; 2007.
11. Valentini G, Dietterich TG. Bias-Variance Analysis of Support Vector Machines for the Development of SVM-Based Ensemble Methods. *The Journal of Machine Learning Research*. 2004; 5:725–775.
12. Freund, Y. An Adaptive Version of the Boost by Majority Algorithm. *Proceedings of the Twelfth Annual Conference on Computational Learning Theory*; 2000. p. 102-113.

Figure 1.
The relationship of the prediction accuracy of an ensemble method $P(EM)$ having n independent classifiers and the prediction accuracy of an individual classifier p .

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
BC_1	Blue	Blue	Blue	Blue	Blue	Blue	White	White	White	White
BC_2	Blue	Blue	Blue	White	White	White	Blue	Blue	Blue	White
BC_3	White	White	White	Blue	Blue	Blue	Blue	Blue	Blue	White

a. The case of achieving the upper bound of the prediction accuracy

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
BC_1	Blue	Blue	Blue	Blue	Blue	Blue	White	White	White	White
BC_2	Blue	Blue	Blue	Blue	White	White	Blue	Blue	White	White
BC_3	Blue	Blue	Blue	Blue	White	White	White	White	Blue	Blue

b. The case of achieving the lower bound of the prediction accuracy

Figure 2.

Example of an ensemble method with $n = 3$ and $p = 0.6$. Total there are 10 inputs. The blue block denotes a correct prediction and white block denotes an incorrect prediction. 2a shows the upper bound accuracy of 0.9 and 2b shows the lower bound accuracy of 0.4.

Figure 3.
The upper and lower bounds of the prediction accuracies of an ensemble method.

a. histogram of prediction accuracies when $n=3$

b. histogram of prediction accuracies when $n=15$

c. histogram of correct predictions in best case ($n=3$)

d. histogram of correct predictions in best case ($n=15$)

e. histogram of correct predictions in worst case ($n=3$)

f. histogram of correct predictions in worst case ($n=15$)

Figure 4.
Results for an ensemble method with individual classifiers having the same $p = 0.6$.

Figure 5.
Results for an ensemble method with individual classifiers having p evenly distributed between 0.4–0.8 and $\bar{p}=0.6$.

Figure 6.

The upper and lower bounds of the prediction accuracies vs. the best and worst prediction accuracies found with $p = 0.5$ and $n = 3, 5, \dots, 51$ in 1,000,000 runs.