**Assignment 3: Take Home Exam**

**31005 Machine Learning**

**Student name: Yujun Liang**

**Student ID: 12786277**

**1. Introduction**

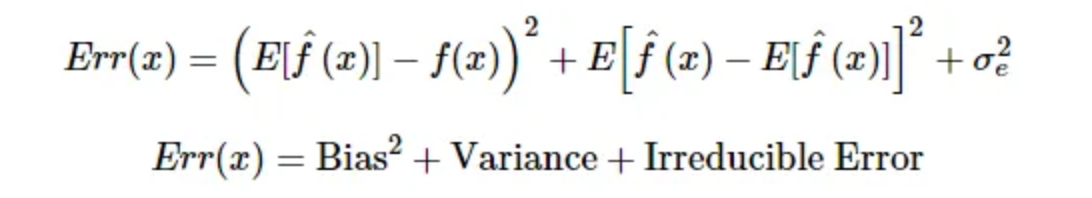
The question 6 has been chosen to answer in this assignment, the question is about the understanding of ensemble Methods. This paper will give a short description to ensemble learning, describes two existing approaches of fusing the decisions from individual classifiers with their advantages and disadvantages. Finally, discuss the diversity of ensemble methods.

1. **Ensemble methods**

Deploying multiple models is an efficient and rational way to solving specific a problem in human society, Helen (1950) states: ‘Alone we can do a little and together we can do much’. This view can be reflected in the machine learning filed. An ensemble is a group of learned classifiers and combines their predictions to get better predictive performance, comparing to predictions made by the single classifier, a group of classifiers have such high accuracy. According to Zhou (2012): ensemble methods have become a major learning paradigm since last 20 years.

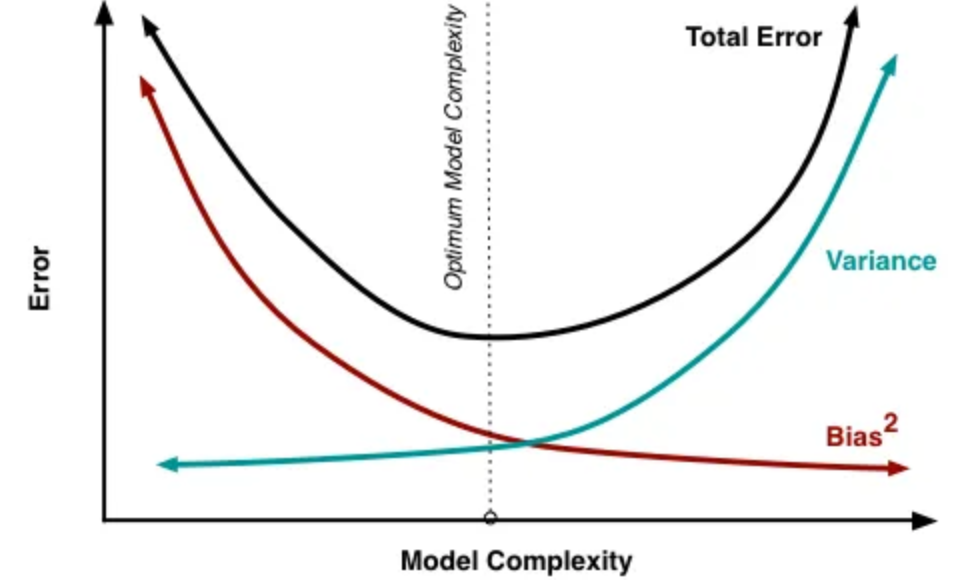
**2.1 Error in ensemble learning**

As the main purpose of ensemble learning is to increase the accuracy of the predictive result, figure 1(Tavish 2015) broken down three components of error.



In general, the algorithm that suffers from the statistical problem is called have a high ‘variance’, the algorithm that suffers from representational problem is called have a high ‘bias’. In detail, statistical problem refers to the large hypothesis space is too large to explore in limited training data, once the different hypotheses that giving the same accuracy is chosen by the algorithm, there is a risk that chosen could not predict the future data well. For representational problems, the true un-known hypothesis could not be represented by any hypothesis in the hypothesis (Zhou 2015). Conclusively, both two problems are main causes that lead to traditional learning approaches fail, through combination, the variance and bias can be significantly reduced (Opitz & Maclin 1999).

The figure 2(Tavish 2015) indicates the ensemble methods are a way to execute the complexity of the model, resulting in the improvement of predictive performance.

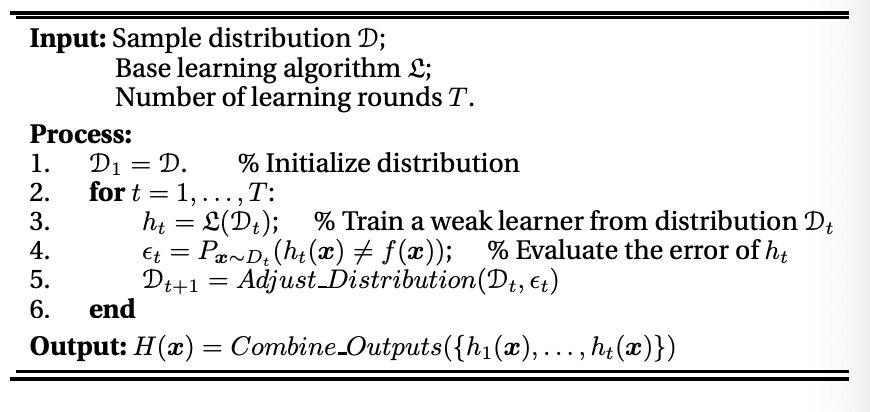


**2.2 Main Ensemble learning techniques**

According to Breiman(1996): two main paradigms of ensemble methods is sequential ensemble method and parallel ensemble methods. The sequential ensemble methods so call boosting techniques, with Adaboost and GBDT as representatives. The parallel ensemble methods refer to the base learners are generated in parallel, with Bagging (Bootstrap Aggregating) as representative.

**Boosting:**

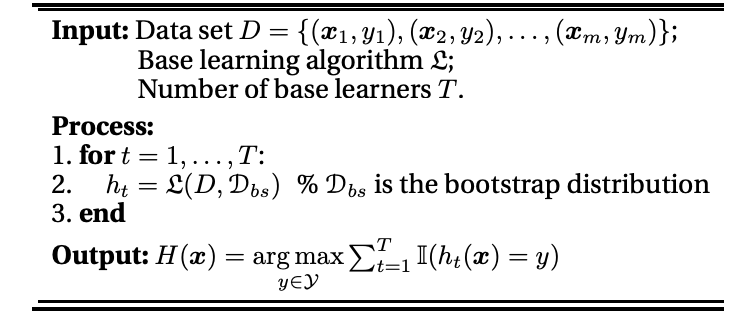
The term of boosting can be defined as a method to convert weak learners to strong learners. According to Zhou (2012): The favor just a slightly better than random guess, the latter is close to the perfect performance. The improvement is made by adjusting the weight of an observation base on the last classification. In the beginning, equal weights are given in each observation, when the first learner predicted incorrectly in the classes, then the higher weight is given to the missed classified observation. In my opinion, the second learner focuses more on the error the last learner faced. Boosting produce better predictive accuracy in general, yet it may overfit on the training data occasionally (Sunil 2015)



(figure 3: procedure of boosting techniques (Zhou 2012)

**Bagging:**

Firstly, the random sample of training data set is created. Secondly, build the classifier for each sample is created. Finally, combine all results of these classifiers using average or voting, which will discuss in the next section. The binary classification and multi-class classification can be deal with Bagging, the following figure summarizes the Bagging algorithm (Zhou 2012)

(Figure 4)

1. **Fusing the decisions from individual classifiers**

After generating a group of classifiers, rather than find the best single classifier, ensemble methods aim to combine them in order to achieve the generalization ability (Zhou 2012). How the individual classifier is weighted for the ensemble’s result? The following section discussed two main combination methods with their advantages and disadvantages.

**3.1 Averaging:**

For the numeric outputs, the most common combination strategy is to use averaging method. There are two main averaging methods: Sample and weighted. For the sample averaging, equal weight is given to all models and the average of predictions is taken. Weighted averaging combines output by different weights implying different importance (Zhou 2012)

The training samples in the actual task are usually insufficient and noisy, which will make the weight of the learning not completely reliable. Especially for the larger-scale combination, it is easier to lead to over-fitting. In general, the weighted average method should be used when the performance of individual learners differs greatly, and the simple average method should be used when the performance of individual learners is similar.

**3.2 Voting:**

Comparing averaging, voting is the most common combination strategy for nominal outputs. There are three main averaging methods: majority voting, plurality voting and weighted voting. In majority voting, every classifier vote for one label, the label with more than half of all votes is the final output. If there is no class label receives more than half votes, there will be no prediction and a rejection option is given. In contrast to majority voting, plurality voting chose the labels more with the highest number of votes (no requiring more than 50 percentage) as final result, if more than two labels have the same vote, the result is chosen randomly in these labels. Furthermore, if individual classifiers performed differently, weighted voting gives more power to stronger classifiers in voting.

1. **Diversity:**

The diversity of ensemble methods refers to the difference with each individual classifier. If individual classifiers are same, there would be no performance improvement when they combined (Tumer & Ghosh. 1995). One of the hot topics in ensemble methods is how to create diverse classifiers, I believe the main idea of creating diverse classifiers is to implement the randomness. According to Zhou (2012): manipulating the data sample, input features, learning parameters and output representations are main popular ideas for ensemble diversity.

In detail, data sample manipulation is the individual classifiers are trained from different data samples that come from generating the parent data set. Although this mechanism significantly works well in unstable classifier like decision tree, some insensitive classifiers like support vector machine, k-nearest neighbors are not performed well on this mechanism, so-called stable base classifier.

For input features manipulation, the training data is described by a set of features normally, subspace is different subsets of feature, providing different views on the data. Therefore, when different individual classifiers trained from different subspaces are usually diverse (Ho. 1998).

For learning parameter manipulation, individual classifiers are being generated by using different parameter settings. In particular, different initial weights can be assigned to individual neural networks (Kolen & Pollack 1991), individual decision trees can apply different split sections (Kwok & Carter 1998)

For output representation manipulation. The diversity is achieved by using different output representations. For example, the output smearing method converts multi-class outputs to multivariate regression outputs to generating different individual classifiers (Breiman.2000)

In addition, the mechanisms can be used together for diversity generation. Random forest (Breiman,2001) implements both the data sample manipulation and input feature manipulation.

1. **Reference:**

Breiman. L. 1996. *Bagging predictors. Machine Learning*, Vol. 24(2). pp. 123–140.

Breiman. L. 2000. *Randomizing outputs to increase prediction accuracy.* Machine

Learning. Vol. 40(3). pp.113–120

Breiman. L 2001. *Random forests. Machine Learning*, Vol.45(1) pp. 5–32.

Kwok. S.W. & Carter. C. *Multiple decision trees*. International Conference on Uncertainty in Artificial Intelligence, pp. 327–338, New York.

Kolen. J. F & Pollack. J. B. 1991. *Back propagation is sensitive to initial conditions*. Advanced Neural Information Processing Systems 3, pp. 860–867. San Francisco.

Opitz.D & Maclin. R. 1991.*Popular ensemble methods: An empirical study*. Journal of Artificial Intelligence Research. Vol.11. pp. 169–198.

Sunil. R. 2015, *5 Easy questions on ensemble modeling everyone should know,* Analytics Vidhya, viewed 08 September 2019,

< <https://www.analyticsvidhya.com/blog/2015/09/questions-ensemble-modeling>>

Tumer. K. & Ghosh. J. 1995. *Theoretical foundations of linear and order statistics combiners for neural pattern classifiers*. Computer and Vision Research Center, University of Texa.

Tavish. S. 2015, *Basics of ensemble Learning Explained in simple English*, Analytics Vidhya, viewed 08 September 2019,

<<https://www.analyticsvidhya.com/blog/2015/08/introduction-ensemble-learning/>>

Valiance Solutions 2015, *Improving Predictions with Ensemble Model,* Data secience central, viewed 08 September 2019, <<https://www.datasciencecentral.com/profiles/blogs/improving-predictions-with-ensemble-model>>

Zhou. Z. H. 2012, *Ensemble Methods: Foundations and Algorithms*, CRC Press. China.