

Educational Series

# Introduction To Ensemble Learning

Dr. Oliver Steinki, CFA, FRM

Ziad Mohammad

July 2015



## What Is Ensemble Learning?

In broad terms, ensemble learning is a procedure where multiple learner modules are applied on a dataset to extract multiple predictions, which are then combined into one composite prediction.

The ensemble learning process is commonly broken down into two tasks: First, constructing a set of base learners from the training data; second, combining some or all of these models to form a unified prediction model.

“Ensemble learning is a process that uses a set of models, each of them obtained by applying a learning process to a given problem. This set of models (ensemble) is integrated in some way to obtain the final prediction.” (Moreira, et al. 2012, 3)

Ensemble methods are mathematical procedures that start with a set of base learner models. Multiple forecasts based on the different base learners are constructed and combined into an enhanced composite model superior to the base individual models. The final composite model will provide a superior prediction accuracy than the average of all the individual base models predictions. This integration of all good individual models into one improved composite model generally leads to higher accuracy levels.

Ensemble learning provides a critical boost to forecasting abilities and decision-making accuracy. Ensemble methods attempt to improve forecasting bias while simultaneously increasing robustness

and reducing variance. Ensemble methods produce predictions according to a combination of all the individual base model forecasts to produce the final prediction. Ensemble methods are expected to be useful when there is uncertainty in choosing the best prediction model and when it is critical to avoid large prediction errors. These criteria clearly apply to our context of predicting returns of financial securities.

## The Rationale Behind Ensemble Methods

(Dietterich 2000b) Lists three fundamental reasons why ensembles are successful in machine learning applications. The first one is statistical. Models can be seen as searching a hypothesis space  $H$  to identify the best hypothesis. However, the statistical problem arises as we often have only limited datasets in practice. Hence, we can find many different hypotheses in  $H$  which fit reasonably well and we do not know which one of them has the best generalization performance. This makes it difficult to choose among them. Therefore, the use of ensemble methods can help to avoid this issue by averaging over several models to get a good approximation of the unknown true hypothesis.

The second reason is computational. Many models work by performing some form of local searches such as the gradient descent to minimize error functions that could get stuck in local optima. An ensemble constructed by starting the local

search from many different points may provide a better approximation to the true unknown function.

The third argument presented by Dietterich (2000b) is representational. In many situations, the unknown function we are looking for is not included in  $H$ . However, a combination of several hypotheses drawn from  $H$  can enlarge the space of representable functions, which could then also include the unknown true function. (Dietterich 2000b)

## Common Approaches To Ensemble Methods

The ensemble learning process can be broken into different stages depending on the application and the approach implemented. We choose to categorize the learning process into three steps following [Roli et al. 2001]: ensemble generation, ensemble pruning and ensemble integration (Moreira, et al. 2012). In the ensemble generation phase, a number of base learner models are generated according to a chosen learning procedure, to be used to predict the final output. In the ensemble pruning step, a number of base models are filtered out based on various mathematical procedures to improve the overall ensemble accuracy. In the ensemble integration phase, the filtered learner models are combined intelligently to form one unified prediction that is more accurate than the average of all the individuals' base models.

## Ensemble Generation

Ensemble Generation is the first step in the application of ensemble methods. The goal of this step is to obtain a set of calibrated models that have an individual prediction of the analyzed outcome.

An ensemble is called homogeneous if all base models belong to the same class of models in terms of their predictive function. If the base models are more diverse than the original set, the ensemble is called heterogeneous (Mendes-Moreira et al. 2012). The second approach is expected to obtain a more diverse ensemble with generally better performance (Wichard et al., 2003). Next to the accuracy of the base models, diversity is considered one of the key success factors of ensembles (Perrone and Cooper, 1993).

However, we do not have control over the diversity of the base models in the ensemble generation phase since the forecasting models used could have correlated forecasting errors. By calibrating a larger number of models from different classes of forecasting models, we increase the likelihood of having an accurate and diverse subset of base models however at the expense of computational requirements. This increased probability is the rationale for the introduction of a diverse range of base learner models.

Ensemble generation methods can be classified according to how they attempt to generate different base models: either through manipulating the data or through manipulating the modeling process (Mendes-Moreira et al., 2012).

Data manipulation can be further broken down into subsampling from the training data and manipulating input features or output variables. The manipulation process can also be subdivided further: It can be achieved by using different parameter sets or manipulating the induction algorithm or the resulting model.

### Ensemble Pruning

The methods introduced for ensemble generation create a diverse set of models. However, the resulting set of predictor models do not ensure the best accuracy possible.

Ensemble pruning describes the process of choosing the appropriate subset from the candidate pool of base models. Ensemble pruning methods try to improve ensemble accuracy and/or to reduce computational cost. They can be divided into partitioning-based and search based methods.

Partitioning based approaches split the base models into subgroups based on a predetermined criteria. Search based approaches try to find a subset of models with improved ensemble accuracy by either adding or removing models from the initial candidate pool.

Furthermore, the different pruning approaches could be classified according to their stopping criterion: Direct ensemble pruning methods are approaches where the number of models used is determined ex-ante, whereas evaluation ensemble methods determine the number of models used according to the ensemble accuracy (Mendes-Moreira et al., 2012).

### Ensemble Integration

Following the ensemble generation and ensemble pruning step, the last step of the ensemble learning process is called ensemble integration. It describes how the remaining calibrated models are combined into a single composite model.

Ensemble integration methods vary in approach and classification. Ensemble integration approaches could be broadly classified as 'combination' or 'selection'. In the combination approach, all learner models are combined into one composite model, in the selection approach only the most promising model(s) are used to construct the final composite model.

A common challenge in the integration phase is multicollinearity, the correlation between the base learner models predictions, which could lower the accuracy of the final ensemble prediction.

Suggestions to avoid or reduce the existence of multicollinearity include several methods applied during the ensemble generation or ensemble pruning step to guarantee an accurate, yet diverse (and hence not perfectly correlated) set of base models. (Steinki 2014, 109). A detailed review of ensemble methods can be found in chapter 4 of Oliver's [doctoral thesis](#).

## Success Factors Of Ensemble Methods

A successful ensemble could be described as having accurate predictors and commits errors in the different parts of the input space. An important factor in measuring the performance of an ensemble lies in the generalization error. Generalization error measures how a learning module performs on out of sample data. It is measured as the difference between the prediction of the module and the actual results. Analyzing the generalization errors allows us to understand the source of the error and the correct technique to minimize it.

Understanding the generation error also allows to probe the base predictors underlying characteristics causing this error.

To improve the forecasting accuracy of an ensemble, the generalization error should be minimized by increasing the ambiguity yet without increasing the bias. In practice, such an approach could be challenging to achieve.

Ambiguity is improved by increasing the diversity of the base learners where a more diverse set of parameters is used to induce the learning process. As the diversity increases, the space for prediction function also increases. A larger space for prediction improves the accuracy of the prediction function given the more diverse set of parameters used to induce learning. The larger space of input given for the prediction models improves the

accuracy on the cost of a larger generalization error.

Brown provides a good discussion on the relation between ambiguity and co-variance (Brown 2004). An important result obtained from the study of this relation is the confirmation that it is not possible to maximize the ensemble ambiguity without affecting the ensemble bias component as well, i.e., it is not possible to maximize the ambiguity component and minimize the bias component simultaneously (Moreira, et al. 2012, 8).

Dietterich (2000b) states an important criteria for successful ensemble methods is to construct individual learning algorithms with prediction accuracy above 50% whose errors are at least somewhat uncorrelated.

## Proven Applications of Ensemble Methods

Numerous academic studies analyzed the success of ensemble methods in diverse application fields such as medicine (Polikar et al., 2008), climate forecasting (Stott and Forest, 2007), image retrieval (Tsoumakas et al., 2005) and astrophysics (Bazell and Aha, 2001). Several academic studies have shown that ensemble predictions can often be much more accurate than the forecasts of the base learners (Freund and Schapire, 1996; Bauer, 1999; Dietterich, 2000a), reduce variance (Breiman, 1996; Lam and Suen, 1997) or bias and variance (Breiman, 1998).

Ensemble methods have been successful in solving numerous statistical problems. Applications of

ensemble methods have been used in a broad range of industries; Air traffic controllers utilize ensembles to minimize airplanes arrival delay time, numerous weather forecast agencies implement ensemble learning to improve weather forecasting accuracy.

A recent public competition by Netflix offered a monetary reward for any contesteer that could improve its film-rating prediction algorithm. After many proposed solutions, the winning team that finally sealed the competition implemented an approach based on ensemble methods.

### The Netflix Competition

The Netflix Competition 2009 was a public competition with a grand prize of US\$1,000,000 to be given for any contesteer that can develop a collaborative filtering algorithm that would predict user rating for films with a RMSE (root-mean-squared error) score lower than 0.8563. The contesters were given a dataset consisting of seven years of past film rating data without any further information on the users or the films. The winning team approach was based on gradient boosted decision trees; a technique applied to regression problems to produce predictions. The prediction was based on an ensemble of 500 decision trees, which were used as base learners and combined to formulate the final prediction of film ratings. In 2009, BellKor's Pragmatic Chaos won the competition and provided a solution that resulted in the lowest RMSE score among the contesters and had better prediction capabilities than the prevailing Netflix algorithm.

EVOLUTIQ's systematic multi-asset class strategy, the *Pred - X Model*, is based on the application of ensemble methods using Lévy based market models to predict daily market moves. The investment strategy is built upon scholarly research on the applicability of ensemble methods to enhance option pricing models based on Lévy processes conducted by Dr. Oliver Steinki.



Dr. Oliver Steinki, CFA, FRM  
*CEO & Co-Founder of EVOLUTIQ*

Responsible for the entrepreneurial success of EVOLUTIQ. He combines his expertise in statistical learning techniques, ensemble methods and quantitative trading strategies with his fundamental research skills to maximize investment profitability.

Oliver started working in the financial industry in 2003. Previous roles include multi-asset-class derivatives trading at Stigma Partners, a systematic global macro house in Geneva, research at MSCI (Morgan Stanley Capital Intl.) and corporate banking with Commerzbank. From an entrepreneurial perspective, Oliver has co-founded and invested in several successful startups in Switzerland, Germany and the UK. Some of these startups received awards from the FT Germany, McKinsey, Ernst & Young and the German federal ministry of economics. Oliver is also an adjunct professor teaching algorithmic trading and portfolio management courses at IE Business School in Madrid and on the Hong Kong campus of Manchester Business School.

Oliver completed his doctoral degree in financial mathematics at the University of Manchester and graduated as a top 3 student from the Master in Financial Management at IE Business School in Madrid. His doctoral research investigated ensemble methods to improve the performance of derivatives pricing models based on Lévy processes. Oliver is also a CFA and FRM charter holder



Ziad Mohammad  
*Sales & Research Analyst*

As part of his role, Ziad splits his time between the research and sales departments. On one hand, he focuses on researching fundamental market strategies and portfolio optimization techniques. On the other hand, he participates in the fundraising & marketing efforts for EVOLUTIQ's recently launched multi asset class strategy.

In his past role as a 'Financial Analyst' at McKinsey & Company, he applied statistical and data mining techniques on data pools to extract intelligence to aid in the decision making process.

Ziad recently completed his Masters degree in Advanced Finance from IE Business School, where he focused his research on emerging markets and wrote his master's final thesis focusing on bubble formations in frontier markets. He completed his bachelors degree in Industrial Engineering from Purdue University and a diploma in Investment Banking from the Swiss Finance Academy.

### References

- Allwein, E. L., R. E. Schapire, and Y. Singer. 2000. *Reducing Multiclass to Binary: A Unifying Approach for Margin Classifiers*. The Journal of Machine Learning 1, 113-141.
- Bauer, E. 1999. *An Empirical Comparison of Voting Classification Algorithms : Bagging, Boosting, and Variants*. Machine Learning 36, 105-142.
- Bazell, D., and D. W. W Aha. 2001. *Ensembles of Classifiers for Morphological Galaxy Classification*. The Astrophysical Journal, 219-223.
- Breiman, L. 1998. *Arcing Classifiers*. The Annals of Statistics, 801-849.
- Breiman, L. 1996. *Bagging Predictors*. Machine Learning 24, 123-140.
- Brown, G. 2004. *Diversity In Neural Network Ensembles*. Ph.d. thesis, University of Birmingham.
- Dietterich, T. G. 2000a. *An Experimental Comparison of Three Methods for Constructing Ensembles of Decision Trees: Bagging, Boosting, and Randomization*. Machine Learning, 139-158.
- Dietterich, T. G. 2000b. *Ensemble Methods in Machine Learning*. In J. Kittler and F. Roli (Ed.), Multiple Classifier Systems, Springer-Verlag, 1-15.
- Freund, Y., and R. E. Schapire. 1996. *Experiments with a New Boosting Algorithm*. Morgan Kaufmann, 148-156.
- Kittler, J., M. Hatef, R. P. Duin, and J. Matas. 1998. *On Combining Classifiers*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 226-239.
- Kleinberg, E. M. 2000. *A Mathematically Rigorous Foundation for Supervised Learning*. In F. Roli and J. Kittler (Eds), Multiple Classifier Systems. Springer.
- Kleinberg, E. M. 1996. *An Overtraining-Resistant Stochastic Modeling Method for Pattern Recognition*. The Annals of Statistics, 2319-2349.
- Kong, E., and T. Dietterich. 1995. *Error-Correcting Output Coding Corrects Bias and Variance*. In The XII International Conference on Machine Learning, San Francisco, Morgan Kaufmann, 313-321.
- Lam, L., and C. Y. Suen. 1995. *Optimal Combinations Of Pattern Classifiers*. Pattern Recognition Letters, 945-954.
- Moreira, J., C. Soares, A. Jorge, and J. De Sousa. 2012. *Ensemble Approaches for Regression: A Survey*. Faculty of Economics, University of Porto, ACM Computing Surveys, 1-40.
- Perrone, M. P., and L. N. Cooper. 1993. *When Networks Disagree: Ensemble Methods For Hybrid Neural Networks*. Brown University.
- Polikar, R. A., D. Topalis, D. Parikh, D. Green, J. Frymiare, J. Kounios, and C. Clark. 2008. *An Ensemble Based Data Fusion Approach For Early Diagnosis Of Alzheimer's Disease*. Information Fusion 9, 83-95.
- Roli, F. 2001. *Methods for Designing Multiple Classifier Systems*. F. Roli and J. Kittler (Eds.), Multiple Classifier Systems, Springer, 78-87.
- Scott, P. A.; Forest, C. E. 2007. *Ensemble Climate Predictions Using Climate Models and Observational Constraints*. Mathematical, Physical, and Engineering Sciences 365 (1857), 52.
- Steinki, O. 2014. *An Investigation Of Ensemble Methods To Improve The Bias And/Or Variance Of Option Pricing Models Based On Levy Processes*. Doctoral Thesis, University of Manchester, 213.
- Tsoumakas, G., L. Angelis, and I. Vlahavas. 2005. *Selective Fusion Of Heterogeneous Classifiers*. Intelligent Data Analysis, 511-525.
- Wichard, J., C. Merkwirth, and M. Ogorzalek. 2003. *Building Ensembles With Heterogeneous Models*. Lecture Notes, AGH University of Science and Technology.

---

*EVOLUTIQ GmbH is issuing a series of white papers on the subject of systematic trading. These papers will discuss different approaches to systematic trading as well as present specific trading strategies and associated risk management techniques. This is the second paper of the EVOLUTIQ educational series.*



EVOLUTIQ GmbH

Schwerzistr.6 8807 Freienbach – Switzerland

Telephone: +41 55 410 7373

Website: [www.evolutiq.com](http://www.evolutiq.com)

Email: [sales@evolutiq.com](mailto:sales@evolutiq.com)