Zero Theorem Literature Review

"Ensemble Methods: Bagging, Boosting and Stacking, J. Rocca, B. Rocca, 2019"

By: SoReZ

https://www.zerotheorem.com

Highlights

- Understanding the basics of Ensemble Model that integrates the numerous weak models to yield a powerful model.
- Explore the types of Ensemble Methods through which one can yield maximum performance.

Background

Have you ever bought a new vehicle? Will you go straight to the auto dealership to get a new car based on the dealer's recommendation? It's exceedingly improbable. You'll probably go to a few websites where people have left reviews and compare several automobile models, looking at their features and prices. You'll almost certainly seek feedback from your friends and coworkers as well. In other words, you won't come to a decision based just on your own viewpoint, but rather on the opinions of others.

Furthermore, consider yourself a film director who has produced a short film about a topic that is both significant and fascinating. Now you want to get some early feedback on the film before releasing it to the whole public. What are your options for accomplishing this? You may ask one of your friends to review the film for you. It's also possible that the person you've selected adores you and doesn't want to break your heart by giving a one-star review to the dreadful job you've produced. Another option is to ask five of your coworkers to rate the film. This should give you a better idea of what to expect from the film. This method may generate accurate movie ratings. However, there is still a problem. These five individuals may or may not be "Subject Matter Experts" on the subject of your film. Sure, they may be able to appreciate the cinematography, images, and sound, but they may not be the finest judges of dark humour. How about interviewing 50 people for their opinions on the film?

Some of these people may be your friends, colleagues, or complete strangers. Because you now have people with different sets of expertise, the responses will be more broad and diverse. And, as it turns out, this is a superior method for obtaining honest feedback than the prior examples we've seen. With these two instances in mind, it's easy to conclude that a varied group of people is more likely to make better decisions than individuals. This is where ensemble learning comes into existence.

Introduction to Ensemble Methods

As Joseph (2019) stated that the theory behind ensemble learning is that when numerous weak models are integrated, they may typically yield a far more powerful model. It's basically a process that contain multiple models, such as classifiers or experts, and are strategically generated and combined to solve a particular computational intelligence problem. Moreover, ensemble learning is primarily used to improve the (classification, prediction, function approximation, etc.) performance of a model, or reduce the likelihood of an unfortunate selection of a poor models. Hence, Robert (2018) concludes that the major ways to carry out this combination include boosting (to decrease the model's bias), stacking (to increase the predictive force of the classifier), and bagging (to reduce model's variance) .

Types of Ensemble Methods

Now that one have got a gist of what ensemble learning is – let's have a look at the various types of Ensemble method.

1. Bagging
2. Boosting
3. Stacking

What is Bagging?

The term "bagging" refers to "bootstrap aggregation." It entails fitting numerous base models to various bootstrap samples and then constructing an ensemble model that "averages" the results of these weak models. It's basically a sampling approach that allows you to produce subsets of observations from an original dataset employing substitution, with the subsets being the same size as the original set. Furthermore, this technique use subsets (bags) to obtain a reasonable estimate of the distribution (complete set). The size of the bagging subsets may be smaller than the original set.

How bagging works?

The main principle of bagging, according to Joseph (2019), is to fit several separate models and "average" their predictions to create a model with a smaller variance. However, fully fitting independent models is not possible due to the large amount of data required. As a result, bootstrap samples with good "approximate characteristics" will be used to fit models that are almost independent. To have a better understanding have a look at Figure 1

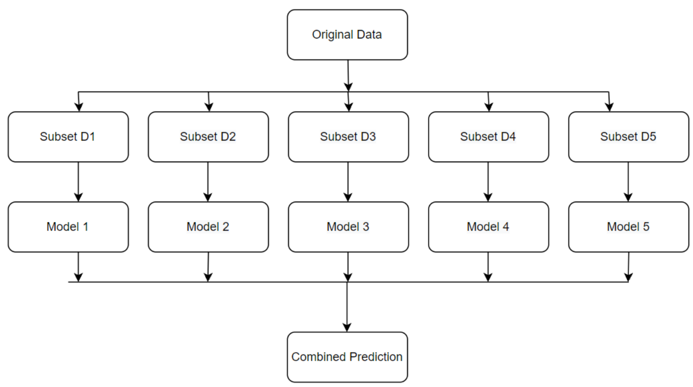

Figure 1: An Overview of the Bagging Process by Singh (2018)

What is Boosting?

After exploring the basics of Bagging, here's another question for you that needs to be addresses: If a data point is incorrectly predicted by the first model, and then the next (probably all models), will combining the predictions provide better results? To solve this issue, a sequential method that fits models iteratively known as Boosting comes into existence that is basically a sequential process Graczyk et al. (2010). In Boosting process, each subsequent model attempts to correct the errors of the previous model. Hence, the succeeding models are dependent on the previous model

How boosting works?

Joseph (2019) stated in his study that bagging is done in such a way that the training of a model at any given step depends on the models fitted at the previous steps. It results in an ensemble model that is, in general, less biased than the weak learners that compose it. Because this strategy is primarily concerned with lowering bias, models with a high bias but low variance are frequently considered for boosting. After the weak learners have been identified, it is necessary to determine how they will be aggregated and sequentially fitted.

Stacking

According to Joseph (2019), stacking differs from boosting and bagging in two ways. First, unlike boosting and bagging, which primarily consider homogeneous weak learners, stacking frequently considers heterogeneous weak learners (combines multiple learning methods). Second, stacking learns to combine base models using a meta-model, whereas boosting and bagging employ deterministic methods to combine weak learners.

How stacking works?

To have a basic understanding assume there is a need to fit a stacking ensemble which has L weak learners. This will be done in three steps. First, splitting the training data into two parts. After this choose L weak learners and fit them to data of the 1st part. Once this learner is selected third step is to perform, for each of the L weak learners, that makes predictions for observations in the second part. To validate the method, fit the meta-model using predictions made by the weak learners as inputs.

Conclusion

This led to the conclusion that ensemble learning is all about combining some base models to obtain an ensemble model with better performances. Hence, it is evident that bagging focuses on producing an ensemble model with less variance than its components, whereas stacking and boosting focuses on producing strong models that are less biased than their components.

References

Graczyk, M., Lasota, T., Trawinski, B., and Trawiński, K. (2010). Comparison of bagging, boosting and stacking ensembles applied to real estate appraisal. In *ACIIDS*.

Joseph, R. (2019). Ensemble methods: bagging, boosting and stacking. *medium-towards data science*.

Robert, D. (2018). Boosting, bagging, and stacking — ensemble methods with sklearn and mlens. *medium-towards data science*.

Singh, A. (2018). A comprehensive guide to ensemble learning (with python codes). *Analytics Vidhya*.