An Overview of Ensemble Learning

Zhaokun Zhang

May 5, 2019

1 Definition

When you want to purchase a new car, will you just walk up to the first car shop and purchase one based on the advice of the dealer? For most people it's highly unlikely. On the contrary, you would likely browser some websites where a large number of people have posted their reviews and compare different cars carefully, checking for their features and prices. It's also possible for you to ask your friends and colleagues for their advice. That is, you wouldn't directly reach a conclusion but will instead make a decision considering the opinions of others as well.

Ensemble models in machine learning operate on a similar idea. They combine the decisions from multiple models to improve the overall performance. In statistics and machine learning, ensemble methods use multiple learning algorithms to obtain better predictive performance than could be obtained from any of the constituent learning algorithms alone. Besides, a machine learning ensemble consists of only a concrete finite set of alternative models, but typically allows for much flexible structure to exist among those alternatives, compared to ensemble in statistical mechanics which is usually infinite.

Machine learning algorithms are most commonly described as preforming the task of searching through a hypothesis space to find a suitable hypothesis that will make better predictions for one particular problem. Instead of learning a single hypothesis from a hypothesis set, ensemble methods select a collection of hypotheses and combine their predictions, that is, ensembles combine multiple hypotheses to form a hopefully better hypothesis. With a collection of hypothesis, it typically requires more computation than evaluating the prediction of a single model, then ensembles sometimes are thought of as a way to compensate for poor learning algorithms by performing a lot of extra computation. Ensembles can be shown to have more flexibility in the functions they represent. This flexibility can in theory, enable them to over-fit the training data more than a single model would, but in reality, many ensemble techniques like bagging tend to reduce problems related to over-fitting of the training data.

Let's understand the concept of ensemble learning with an example. Suppose you are a movie director and you've created a short movie on a very interesting topic. Now, you want to take preliminary feedback on the movie before making it public. What are the possible ways by which you can do that?

A: You may ask one of your best friends or family members to rate the movie. But it's entirely possible that the person you have chosen loves you very much and doesn't want to break your heart by providing a 1-star rating to the horrible work you have created.

B: Another way could be by asking 5 colleagues or friends of yours to rate the movie. This should provide a better idea of the movie. This method may provide honest ratings for your movie. But there exists a problem that these 5 people may not be "Subject Matter Experts" on the topic of your movie.

C: The last way is to ask more than 50 people to rate the movie. Some of which can be your friends, some of them can be your colleagues and some may even be total strangers. The responses, in this case, would be more generalized and diversified since now you have people with different sets of skills. And as it turns out – this is a better approach to get honest ratings than the previous cases we assumed.

With the single example, you can infer that a diverse group of people are likely to make better decisions as compared to individuals. Similar is true for a diverse set of models in comparison to single models. This diversification in Machine Learning is achieved by a technique called Ensemble Learning.

2 Advantages

As mentioned above, ensemble learning tends to perform better compared to single learning algorithm. And there are some advantages for the ensemble methods:

- We 've already known that for a given set of N independent random variables each with variance σ^2, the variance of the mean is $\frac{\sigma^2}{N}$. Then for a given set of hypothesis set $\mathcal{H}$, when sample size is small, many different hypotheses in $\mathcal{H}$ can all give the same accuracy on the training data. Combining their predictions helps lower the variance and improve prediction accuracy;
- When we use the ensemble methods, all hypotheses are locally optimal based on local search, then combine all of them help improve the computational accuracy;
- It is possible to expand $\mathcal{H}$ and produce a better approximation of target function f since it forms weighted sums of hypotheses drawn from $\mathcal{H}$.

3 Simple types

Since the advanced methods of ensemble learning like bagging and boosting have been already taught in our class, in this section I want to talk a few simple techniques of ensemble methods.

3.1 Max voting

The max voting method is generally used for classification problems. For this technique, multiple models are used to make predictions for each data point, which is considered as a 'vote'. The predictions which we get from the majority of the models are used as the final prediction.

For example, in the example we mentioned above, when you ask 5 of your colleagues to rate your movie (out of 5), we assume three of them rate it as 4 while two of them give it a 5. Since the majority give a rating of 4, the final rating will be taken as 4. You can consider this as taking the mode of all the predictions.

3.2 Averaging

Similar to the max voting technique, multiple predictions are made for each data point in averaging. In this method, we take an average of predictions from all the models and use it to make the final prediction. Averaging can be used for making predictions in regression problems or while calculating probabilities for classification problems.

Similarly, the average method would give you the value of: $\frac{1}{5} \times (4 \times 3 + 5 \times 2) = 4.4$.

3.3 Weighted Average

This is an extension of the averaging method. All models are assigned different weights defining the importance of each model for prediction. That is, those results or predictions with more importance will have more weights in calculation of the final prediction.

4 Summary

The goal of any machine learning problem is to find a single model that will best predict our wanted outcome. Rather than making one model and hoping this model is the best or most accurate predictor we can make, ensemble methods take a myriad of models into account, and average those models to produce one final model.

References

[1] Wikipedia, "Ensemble learning", https://en.wikipedia.org/wiki/Ensemble_learning

[2] Aishwarya Singh(2018). "A Comprehensive Guide to Ensemble Learning (with Python codes)"

[3] Evan Lutins(2017). "Ensemble Methods in Machine Learning: What are They and Why Use Them?"

[4] Jiaming Mao, "Decision Trees and Ensemble Methods", https://jiamingmao.github.io/data-analysis/assets/Lectures/Decision_Trees_and_Ensemble_Methods.pdf