Ensemble Methods 101

Ensemble Learning

- In statistics and machine learning, ensemble methods use multiple learning algorithms to obtain better predictive performance than could be obtained from any of the learning algorithms alone.
- The Wisdom of Crowd (book): "In 1906, the great statistician Francis Galton observed a competition to guess the weight of an ox at a country fair. 800 people entered. He discovered that the average guess of all competitors (1,197lb) was extremely close to the actual weight (1,198lb) of the ox."

Common Types of Ensemble

- Bayes Optimal Classifier
- **Bootstrap Aggregating (Bagging)**
- **Boosting**
- Bayesian Parameter Averaging
- Bayesian Model Combination
- Bucket of Models
- **Stacking**

Pros and Cons

Advantages of ensemble methods

- Intuitively, ensembles allow the different needs of a difficult problem to be handled by hypotheses suited to those particular needs.
- Mathematically, ensembles provide an extra degree of freedom in the classical bias/variance tradeoff, allowing solutions that would be difficult (if not impossible) to reach with only a single hypothesis.
- They’re unlikely to overfit.

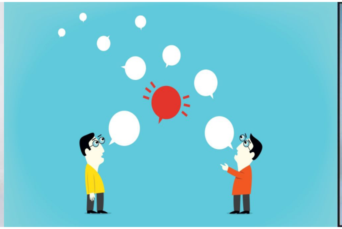
Disadvantages of ensemble methods

- The model that is closest to the true data generating process will always be best and will beat most ensemble methods. So if the data come from a linear process, linear models will be much superior to ensemble models.
- Ensemble models suffer from lack of interpretability. Sometimes we need predictions and explanations of the predictions. It is hard to convince people to act on predictions when the methods are too complex for their comfort level.
- Ensemble methods are usually computationally expensive. Therefore, they add learning time and memory constraints to the problem.

How to Make an Ensemble

Here are the top 4 reasons for a model to be different. They can be different because of a mix of these factors as well:

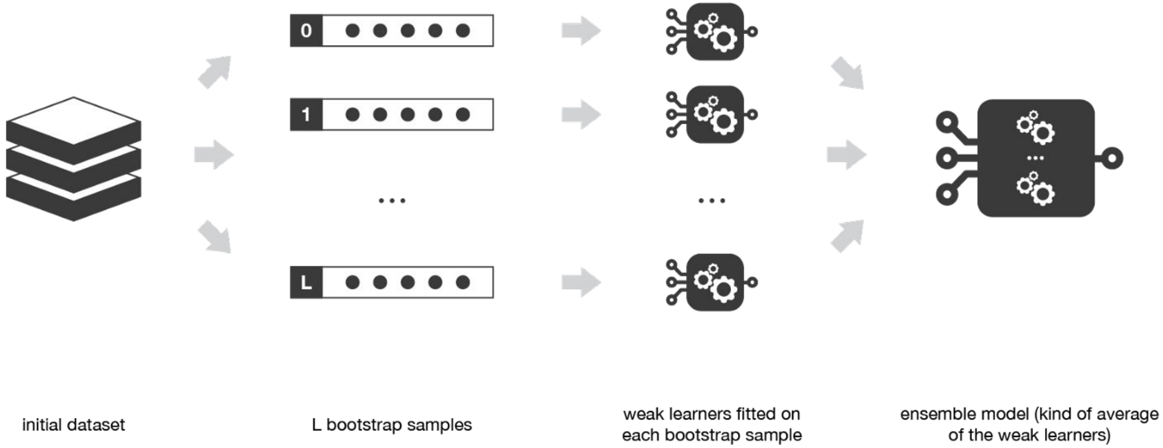
- 1) Difference in Population
- 2) Difference in Hypothesis
- 3) Difference in Modeling Technique
- 4) Difference in Initial Seed

Bootstrap Aggregating (Bagging)

Bagging tries to implement similar learners on small sample populations and then takes a mean of all the predictions.

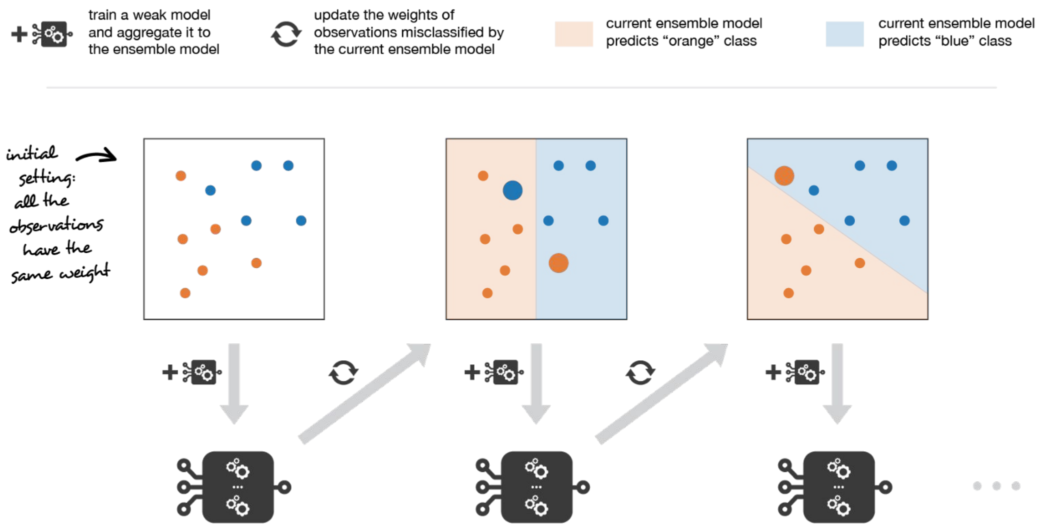
Example: Random Forests



Boosting

It focus on adjusting the weight of an observation based on the last classification. If an observation was classified incorrectly, it tries to increase the weight of this observation and vice versa.

Example: Gradient Boosting



Stacking

Involves training a learning algorithm to combine the predictions of several other learning algorithms. First, all of the other algorithms are trained using the available data, then a combiner algorithm is trained to make a final prediction using all the predictions of the other algorithms as additional inputs.

Why Stacking?

- Dataset (Feature Engineering)
- Training Algorithm
- Approach (Team Up!)
- Target Transformation
- Stacking Architectures

Let’s Code!