BOOTSTRAP AGGREGATION

Piyush and Neil

IIT Hyderabad

March 4, 2019

Definition

Bootstrap aggregating, also called **bagging**, is a machine learning ensemble **meta-algorithm** designed to improve the stability and accuracy of machine learning algorithms used in statistical classification and regression. **It also reduces variance and helps to avoid overfitting**. Although it is usually applied to decision tree methods, it can be used with any type of method. Bagging is a special case of the model averaging approach.

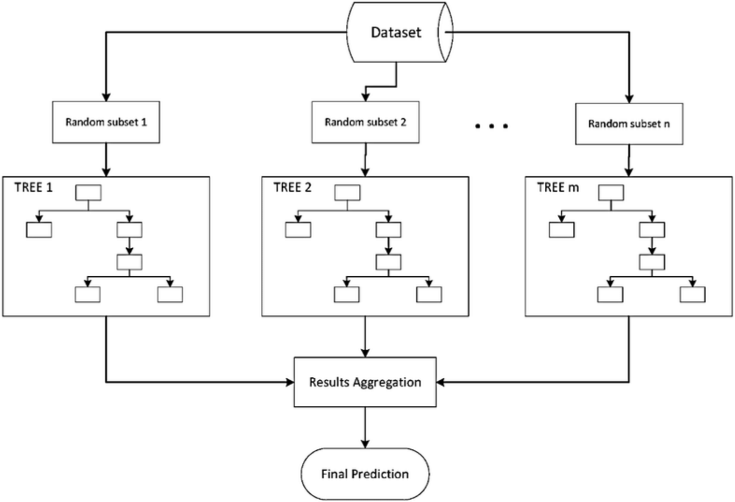
A brief introduction on **Ensemble Method**

Ensemble methods are machine learning techniques that combines several base models in order to produce one optimal predictive model. Eg: Bagging and random forests. We will talk about bagging here.

Difference between Bagging and Bootstrap

Bootstrapping is a sampling technique and Bagging is an machine learning ensemble based on bootstrapped sample.

Diagrammatic Representation

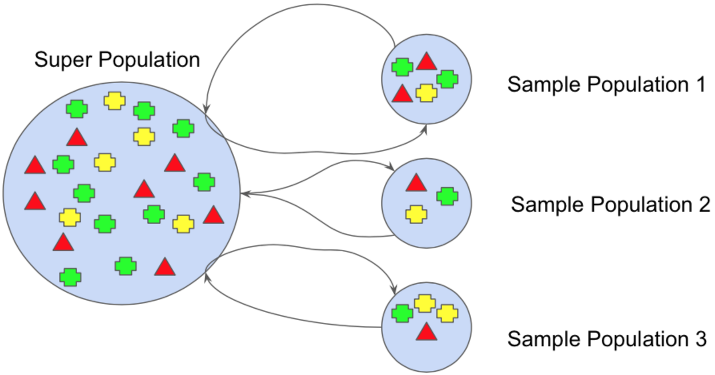


Technique Description

Given a standard training set D of size n, bagging generates m new training sets D_i, each of size $n^{'}$, by sampling from D uniformly and with replacement. By sampling with replacement, some observations may be repeated in each D_i. If $n^{'} = n$, then for large n the set D_i is expected to have the fraction $(1 - 1/e)$ (approx 63.2%) of the unique examples of D, the rest being duplicates. This kind of sample is known as a bootstrap sample. The m models are fitted using the above m bootstrap samples and combined by averaging the output (for regression) or voting (for classification).

Why Bagging?

Bagging leads to <u>improvements for unstable procedures</u>, which include, for example, artificial neural networks, classification and regression trees, and subset selection in linear regression.



Bagging in Regression

Consider first the regression problem. Suppose we fit a model to our training data $Z = (x_1, y_1), (x_2, y_2), ..., (x_N, y_N)$, obtaining the predic- tion $f(x)$ at input x. Bootstrap aggregation or bagging averages this prediction over a collection of bootstrap samples, thereby reducing its variance. For each bootstrap sample Z^{*b}, $b = 1, 2, ..., B$, we fit our model, giving prediction $f^{*b}(x)$.
The bagging estimate is defined by:

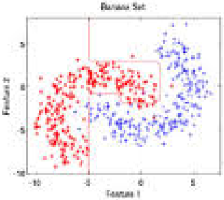
$$f_{bag}(x) = \frac{1}{B} \sum_{b=1}^{B} f^{*b}(x)$$

Example solved using Bagging

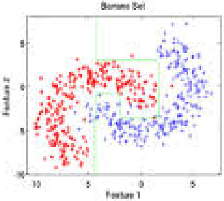
Algorithm

- Step1: From training dataset D(having n_1 elements), m distinct random sub datasets were chosen with elements $n_2 < n_1$.
- Step2: Now, m distinct decision trees were trained using these m distinct sub data sets.
- Step3: after training the decision trees were fed input and response was recorded and the final response was calculated as mean($f(x)$). This final output had low variance. This algorithm can be used to reduce variance and train unstable base learning algorithms.

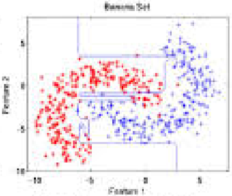
Bagging Example



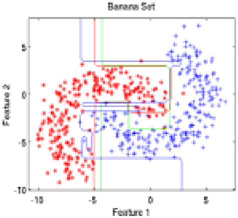
Decision boundary produced by one tree



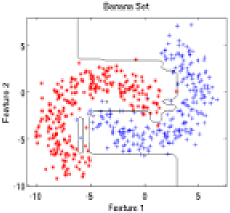
Decision boundary produced by a second tree



Decision boundary produced by a third tree



Three trees and final boundary overlaid



Final result from bagging all trees.