Ensemble Based Methods and Bagging

Tree ensembles have been found to generalize well when scoring new data. Some useful and popular tree ensembles are bagging, boosting, and random forests. Bagging, which combines decision trees by using bootstrap aggregated samples. An advantage specific to bagging is that this method can be multithreaded or computed in parallel. Most of these ensembles are assessed using out-of-bag error.

Random Forest

Random forest is a tree ensemble that has a similar approach to bagging. Their main characteristic is that they add randomness by only using a subset of features to train each split of the trees it trains. Extra Random Trees is an implementation that adds randomness by creating splits at random, instead of using a greedy search to find split variables and split points.

Boosting

Boosting methods are additive in the sense that they sequentially retrain decision trees using the observations with the highest residuals on the previous tree. To do so, observations with a high residual are assigned a higher weight.

Gradient Boosting

The main loss functions for boosting algorithms are:

- *0-1 loss function*, which ignores observations that were correctly classified. The shape of this loss function makes it difficult to optimize.
- *Adaptive boosting loss function*, which has an exponential nature. The shape of this function is more sensitive to outliers.
- *Gradient boosting loss function*. The most common gradient boosting implementation uses a binomial log-likelihood loss function called deviance. It tends to be more robust to outliers than AdaBoost.

The additive nature of gradient boosting makes it prone to overfitting. This can be addressed using cross validation or fine tuning the number of boosting iterations. Other hyperparameters to fine tune are:

- learning rate (shrinkage)
- subsample
- number of features.

Stacking

Stacking is an ensemble method that combines any type of model by combining the predicted probabilities of classes. In that sense, it is a generalized case of bagging. The two most common ways to combine the predicted probabilities in stacking are: using a majority vote or using weights for each predicted probability.