<u>Analysis: Random Forests</u> <u>- Dr. Leo Breiman</u>

Random Forests was originally developed *Dr. Leo Breiman* in a paper he published in 1999, building on the contributions including the CART decision tree. He continued his work with his former Ph.D. student *Dr. Adele Cutler* to develop the perfect form of Random Forests including the graphics permitting deeper data understanding. This article by Dr. Leo Breiman is a complete guide on the random forests. To be honest I read about this topic for the first time and unlike other papers and articles, it was easy to understand this article without any need to explore the internet. This article is a combination of experiments along with some good theory and real-life example to support their article.

Random Forests is a tool that leverages the power of many decision trees, judicious randomization, and ensemble learning to produce astonishingly accurate predictive models, insightful variable importance rankings, missing value imputations, novel segmentations, and laser-sharp reporting on a record-by-record basis for deep data understanding.

Random Forests are collections of decision trees that together produce predictions and deep insights into the structure of data. The core building block of a Random Forest is a CART inspired decision tree. The earliest version of the random forest was the *bagger.*

The bagger represented a distinct advance in machine learning when it was first introduced in 1994. Dr. Breiman discovered that the bagger was a good machine learning method but not as accurate as he had hoped for. Analyzing the details of many models he concluded that the trees in the bagger were like each other. Putting Randomness into Random Forests. The key idea was to introduce randomness not just into the training samples but also into the actual tree growing as well. This would guarantee that different trees would be quite dissimilar to each other.

Conclusion:

Random forests are an effective tool in prediction. Because of the Law of Large Numbers, they do not overfit. Injecting the right kind of randomness makes them accurate classifiers and regressors. Furthermore, the framework in terms of strength of the individual predictors and their correlations gives insight into the ability of the random forest to predict. Using out-of-bag estimation makes concrete the otherwise theoretical values of strength and correlation.

Likes:

1. This article is complete, it describes the random forests, shows some experiments and convincing results, has good theory, strengths and weaknesses, why random forests is an important tool and the applicability.
2. The article was so good that it never took me a second reading to understand what has been described. I was able to relate the article as an algorithm to use random forests.
3. This article covers the entire domain of the topic, I have come across many articles that look incomplete.
4. This was a very important introduction as random forests can help in several important predictions.

Dislikes:

Randomness produces good results in classification, less so in regression. The only types of randomness used in this study is bagging and random features. They say that other types of randomness can give better results, but they do not suggest or try that.

Inspirations:

It was an honor to read the work of one of my teachers and her major professor. Reading about them introducing a tool that is so good and accurate inspires me. There was so much more I wanted to include in this writeup but since it is a one-page analysis I restricted my content.