Random Forests and Gradient Boosting

Shreeyesh Menon

1 Introduction

Tree-based methods are among the most widely used tools in applied prediction problems. They are flexible, easy to interpret at a local level, and work well with heterogeneous economic data containing nonlinearities and interactions. This note introduces two powerful ensemble methods built upon decision trees: *Random Forests* and *Gradient Boosting*. Both techniques combine many trees to improve predictive accuracy, but they do so in very different ways.

2 From a Single Tree to Ensembles

A single decision tree partitions the feature space into regions, predicting the average outcome within each leaf. While interpretable, a single tree tends to have high variance. Small changes in the training sample can lead to very different splits and predictions.

Ensemble methods reduce variance by aggregating many trees. In economics, this is analogous to reducing sampling variability through repeated estimates from bootstrap samples or through model averaging across analysts.

3 Random Forests

Random Forests (RF) average the predictions of many *independent* trees, each grown on a perturbed version of the training set. The key ingredients are:

- **Bootstrap sampling:** Each tree is trained on a bootstrap sample of the data.
- **Random feature selection:** At each split, the tree only considers a random subset of predictors. This de-correlates trees and increases the diversity of the forest.
- **Deep trees:** Each tree is grown fully (or nearly fully), so the individual trees are high-variance but low-bias. Averaging reduces the variance.

The Random Forest prediction is

$$\hat{f}_{RF}(x) = \frac{1}{B} \sum_{b=1}^{B} \hat{f}^{(b)}(x),$$

where $\hat{f}^{(b)}(x)$ is the b-th tree's prediction.

3.1 Econometric Interpretation

Random Forests behave like a nonparametric estimator that automatically captures complex interactions between variables (e.g., income $\times$ education). The randomness helps avoid overfitting and yields strong out-of-sample performance—a desirable property in policy prediction and forecasting.

3.2 Out-of-Bag Error

Because each tree is trained on a bootstrap sample, roughly one-third of observations are left out ("out-of-bag"). These can be used as a built-in cross-validation mechanism, providing an unbiased estimate of predictive performance without needing a separate validation set.

4 Gradient Boosting

Gradient Boosting Machines (GBM) take a fundamentally different approach. Rather than averaging many independent trees, GBM builds trees *sequentially*, where each new tree attempts to correct the mistakes of the previous ones.

4.1 Additive Model Structure

The model takes the form

$$\hat{f}_M(x) = \sum_{m=1}^{M} \nu \, h_m(x),$$

where $h_m(x)$ is a small tree (often called a "weak learner"), and $\nu \in (0, 1)$ is a learning rate controlling how aggressively the model fits each stage.

4.2 Gradient Descent in Function Space

At each step, GBM fits a tree to the *negative gradient* of the loss function, evaluated at the current model. This can be seen as performing gradient descent, not on coefficients, but on entire functions. For squared-error regression:

$$\text{Residuals:} \quad r_i^{(m)} = y_i - \hat{f}_{m-1}(x_i),$$

and the next tree fits these residuals.

4.3 Interpretation

Boosting creates a model that gradually improves its fit, capturing complex nonlinear patterns. With small learning rates and many trees, GBM becomes a highly accurate predictor, often outperforming Random Forests when tuned carefully.

5 Comparison of Random Forests and Gradient Boosting

	Random Forests	**Gradient Boosting**
Tree dependence	Independent trees	Sequential dependence
Bias–variance	Reduces variance	Reduces bias gradually
Tuning	Fewer tuning parameters	More tuning (learning rate, depth, trees)
Risk of overfitting	Low (with enough trees)	Higher if not tuned properly
Interpretability	Moderate (e.g., variable importance)	Lower, but partial dependence plots help
Performance	Very robust	Often more accurate

For forecasting or prediction-heavy tasks in economics (e.g. credit scoring, policy targeting), boosting often achieves state-of-the-art accuracy, while Random Forests are preferred for baseline robustness and interpretability.

References

Friedman, J. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29(5).

Breiman, L. (2001). Random forests. *Machine Learning*, 45:5–32.