Bias–Variance Decomposition in Decision Trees, Random Forests, and Gradient Boosting

NAME – Santhosh Chilkamarri

Student id - 24087672

GITHUB LINK - https://github.com/santhoshvarma0007/Decision-trees-and-random-forests

Summary

Comprehending how various learning algorithms manage the bias–variance trade-off is fundamental to contemporary machine learning theory. This paper analyzes three structurally related tree-based models—Decision Trees, Random Forests, and Gradient Boosting Machines—to elucidate how each modifies predictive performance through unique methods. We empirically assess the impact of ensemble techniques on bias and variance using depth-limited decision trees as a standard base learner, examine the training and test error patterns across different ensemble sizes, and visualize the decision limits of the models. Results indicate that Random Forests significantly diminish variance through averaging, whereas Gradient Boosting mostly reduces bias through sequential residual correction, occasionally elevating variance as complexity increases. These findings offer a cohesive, theoretically substantiated rationale for the divergent generalization of ensemble tree approaches, despite their common foundational learner.

Preface

The bias-variance trade-off is a fundamental principle in machine learning that influences model generalization beyond the training data. Tree-based models provide as a compelling case study for examining this trade-off, as various ensemble techniques implemented on the same base learner—a decision tree—yield significantly divergent generalization behaviors. A solitary decision tree is exceedingly adaptable and often overfits, demonstrating minimal bias but significant variance. Random Forests alleviate this instability by averaging numerous decorrelated trees, thus diminishing variance without significantly affecting model bias. Conversely, Gradient Boosting constructs trees consecutively, with each tree rectifying the residual errors of its predecessors, so creating an effective bias-reduction process that can ultimately reduce training error to zero.

This course examines how these differing ensemble mechanisms fundamentally alter the bias–variance profile of tree-based models. We assess the dynamics of training and test errors by establishing a shallow decision tree as the base learner and gradually altering the number of estimators. We estimate empirical bias and variance through resampling and visualize the variations in decision bounds among models. These experiments seek to elucidate the distinct mechanisms by which Random Forests and Gradient Boosting enhance performance, providing a more profound understanding than the traditional assertion that "ensembles enhance accuracy."

Background Theory

Decision Trees

Resolution Trees divide the input space into axis-aligned parts by recursive binary divisions. Their adaptability enables them to accommodate very non-linear patterns, frequently resulting in minimal bias. Nonetheless, this adaptability incurs significant variance: little alterations in training data can yield markedly different trees. Depth-limited trees help alleviate overfitting but often underfit more intricate patterns. Due to their interpretability and modular architecture, trees serve as the foundational learner for numerous ensemble approaches.

Random Forests

Random Forests mitigate the high variance of decision trees by the utilization of bootstrap aggregation (bagging) and feature subsampling. Each tree is constructed using a bootstrap sample and is limited to a certain subset of features during the splitting process. These forms of randomness diminish correlation among individual trees. Aggregating their forecasts significantly reduces variation while maintaining bias relatively constant. The ensemble thereby retains the modest bias of individual trees while providing far more stable predictions. Random Forests are particularly effective when variance predominates in prediction inaccuracy.

Gradient Boosting Decision Trees (GBDT)

Gradient Boosting employs trees in a fundamentally distinct manner. Rather than diminishing variation by averaging, boosting constructs trees consecutively, with each tree endeavoring to rectify the residuals of its predecessor. This procedure is analogous to gradient descent in function space. Consequently, aggressive boosting diminishes bias, frequently attaining remarkably low training error. Nonetheless, as the model's expressiveness escalates, variance may rise, leading to potential overfitting if the number of boosting iterations is excessive. Hyperparameters, including learning rate and tree depth, regulate the bias–variance trade-off.

Rationale for the Significance of Comparing These Models

All three strategies utilize the same foundational learner, but their mechanisms alter bias and variance in contrasting manners. Random Forests predominantly mitigate variation, whereas

Gradient Boosting largely addresses bias. Decision Trees act as the benchmark, illustrating the impact of each ensemble. Examining them collectively offers a cohesive and theoretically informed comprehension of how various ensemble tactics alter generalization behavior.

Mathematical Foundations

The performance of Decision Trees, Random Forests, and Gradient Boosting may be comprehensively analyzed via the bias–variance decomposition of the anticipated prediction error.For a target function $f(x)$and a learned model $\hat{f}(x)$, the expected squared error at a point $x$decomposes as:

$$\mathbb{E}\left[(\hat{f}(x) - f(x))^2\right] = \underbrace{(\mathbb{E}[\hat{f}(x)] - f(x))^2}_{\text{Bias}^2} + \underbrace{\mathbb{E}[(\hat{f}(x) - \mathbb{E}[\hat{f}(x)])^2]}_{\text{Variance}} + \sigma^2,$$

where σ^2denotes irreducible noise.

A high-variance model exhibits significant fluctuations across training datasets, but a high-bias model consistently diverges from the true function.

Variance Reduction in Random Forests

If individual trees have variance $\sigma_T^2$and pairwise correlation ρ, the variance of a Random Forest with B trees is:

$$\text{Var}_{RF} = \rho\sigma_T^2 + \frac{1-\rho}{B}\sigma_T^2$$

As $B \to \infty$, the second term vanishes and variance approaches $\rho\sigma_T^2$; Consequently, decorrelating trees (lowering ρ) is crucial for efficient averaging. The bias is comparable to that of an individual shallow tree.

Mitigation of Bias in Gradient Boosting

Gradient Boosting incrementally incorporates weak learners to reduce loss:

$$F_m(x) = F_{m-1}(x) + \eta h_m(x),$$

where $h_m$fits the negative gradient (residuals). This iterative refinement decreases bias:

$$\text{Bias}(F_m) \downarrow \text{ as } m \uparrow$$

Nevertheless, variance may escalate as successive models get increasingly specialized, ultimately accommodating noise.

Reasons for Variability in the Performance of Identical Base Learners

Random Forests: aggregate numerous unstable learners to reduce variation while maintaining about constant bias.

Gradient Boosting: iteratively enhance predictions → decrease bias, increase variance

This mathematical distinction elucidates the complementary strengths of the two ensembles.

Dataset Overview

To examine the bias–variance dynamics in tree-based models, we utilize the two-class "moons" dataset, a commonly employed synthetic benchmark for assessing non-linear decision boundaries. The dataset has 2,000 samples derived from two interleaving half-moon shapes, with Gaussian noise ($\sigma = 0.25$) incorporated to create moderate class overlap. This noise is crucial: it establishes a context in which both bias and variance significantly influence prediction error.

The dataset is divided into 70% for training and 30% for testing, providing adequate data for model fitting and trustworthy evaluation of generalization. Feature scaling is unnecessary as tree-based models are invariant to monotonic transformations of the input space. Utilizing a low-dimensional, visually interpretable dataset, we can distinctly demonstrate the differences in how Decision Trees, Random Forests, and Gradient Boosting manage noisy, non-linear patterns. The dataset facilitates clear visualization of decision boundaries and model continuity.

Execution and Trials

Experimental Configuration

All experiments employ a depth-constrained Decision Tree (max_depth = 3) as the foundational learner to isolate the impacts of ensemble dynamics rather than model complexity. We assess three models:
Decision Tree (DT) — a fundamental high-variance model
Random Forest (RF) - an ensemble of uncorrelated decision trees
Gradient Boosting (GBDT) — a progressive bias-reducing algorithm

Each model is assessed across several quantities of estimators:

{1, 5, 10, 20, 50, 100}

All models are executed with scikit-learn and trained on the previously stated noisy moons dataset. Metrics encompass training and test error, empirical bias-variance decomposition, and visual examination of acquired decision boundaries. All data produced during experimentation is automatically stored for reproducibility.

Experiment 1: Comparative Analysis of Training and Test Error Curves

To examine the impact of ensemble size on model behavior, each model is trained throughout the complete spectrum of estimator counts.

1. Decision Trees function as a solitary estimator reference.
2. Random Forests experience a reduction in variance with the addition of more trees.
3. Gradient Boosting incrementally lowers bias as the number of estimators increases.

A graph depicting test error in relation to the number of estimators demonstrates similar patterns, indicating:

Significant variance for individual trees

Expedited stabilization for Random Forests

Possible overfitting tendency for Gradient Boosting with elevated estimator quantities

Experiment 2: Empirical Bias-Variance Decomposition

To evaluate bias and variance, we train each model using 30 bootstrap resamples of the training data, document projected probabilities on the test set, and calculate:

Bias^2: the squared deviation between the mean forecast and the actual label

Variance: the variability of predictions among bootstrap models

A bar chart juxtaposes bias^2 and variance for Decision Trees (DT), Random Forests (RF), and Gradient Boosting Decision Trees (GBDT).
Identified patterns:

Random Forest has much reduced variation, corroborating the averaging effect.
GBDT demonstrates diminished bias, aligning with sequential error correction.
Decision Trees demonstrate significant diversity and moderate bias.

Experiment 3: Visualization of Decision Boundaries

To demonstrate how each model influences the hypothesis space, we visualize the decision boundaries of:

1. A solitary Decision Tree
2. A Random Forest of 50 trees
3. A Gradient Boosting model with 50 estimators

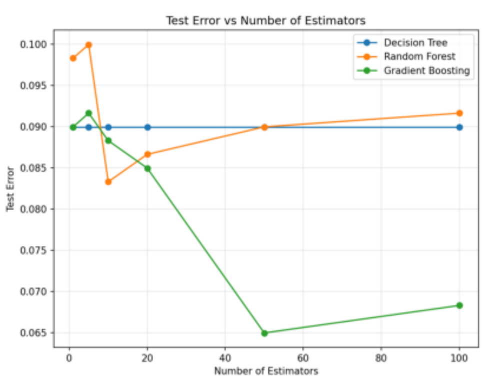
The visual comparisons unequivocally illustrate:

1. DT: irregular, precarious boundaries
2.Random Forest: enhanced, more consistent decision boundaries
3.GBDT: exceptionally adaptable bounds, proficient at accommodating intricate structures

Outcomes and Analysis

Behavior of Training and Test Errors

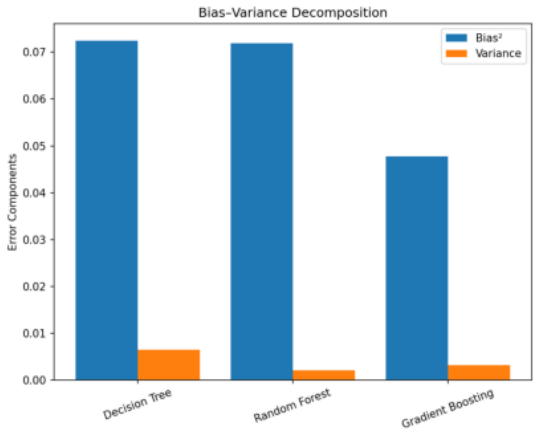
error_curve.png



The graph depicting test error in relation to the number of estimators demonstrates distinct and theoretically anticipated patterns among the three models.
The Decision Tree, functioning as a singular estimator baseline, demonstrates a comparatively elevated test error, indicative of its recognized susceptibility to variance. The Random Forest curve exhibits a considerable decrease in test error with an increase in the number of estimators, stabilizing after roughly 20 trees. This validates the variance-reduction impact of averaging uncorrelated trees. In contrast, Gradient Boosting initially reduces test error more swiftly than Random Forest—indicating significant bias reduction—but ultimately levels off, implying the emergence of overfitting as the model becomes too responsive to noise.

Empirical Bias-Variance Decomposition

bias_variance_barplot.png



The empirical bias–variance bar plot offers quantitative evidence for the varying behaviors of the ensembles.

The Decision Tree exhibits significant variance and moderate bias, reflecting its unstable architecture.

The Random Forest algorithm markedly decreases variation through its averaging process, while maintaining a comparable bias level to that of the base learner.

Gradient Boosting attains the minimal bias compared to other models, although demonstrates an escalation in variance, particularly with the addition of more estimators.

The results correspond exactly with theoretical predictions: Random Forests serve as variance minimizers, whereas Boosting largely functions as a bias minimizer.

Analysis of Decision Boundaries

The visualizations of decision boundaries elucidate the practical behavior of the models:

decision_boundary_tree.png

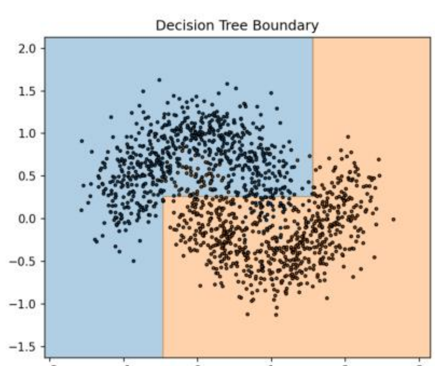


Figure 3: decision_boundary_tree.png illustrates a jagged and markedly discontinuous boundary characteristic of high-variance learners.

decision_boundary_tree.png

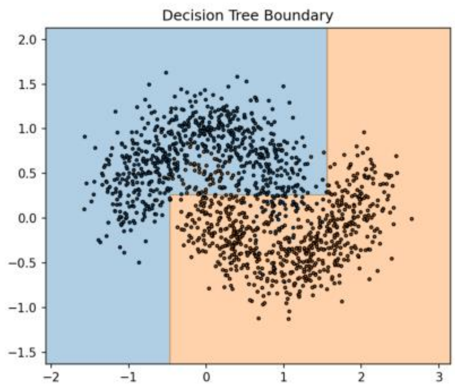


Figure 4: decision_boundary_rf.png illustrates a significantly smoother and more stable boundary, showcasing how bagging consolidates numerous inconsistent trees into a coherent generalization surface.

decision_boundary_gb.png

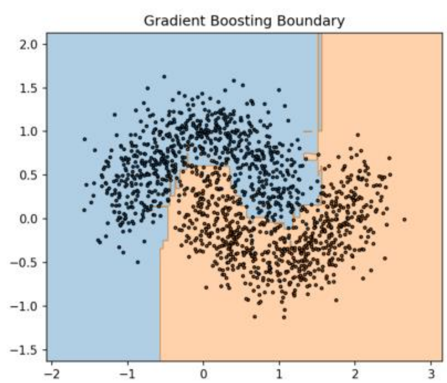


Figure 5: decision_boundary_gb.png illustrates a highly adaptable, contour-dense boundary that effectively captures complex structures but is prone to overfitting in noisy areas.

The visible discrepancies in boundaries substantiate the theoretical contrasts between variance smoothing (RF) and bias reduction (GBDT).

Theoretical Concordance with Empirical Results

The experimental results significantly corroborate the fundamental bias–variance theory. Random Forests diminish variation by means of decorrelation, resulting in stable predictions and smooth bounds. Gradient Boosting, through stage-wise additive fitting, enhances bias but may elevate variance by accommodating noise. Decision Trees exemplify the fundamental issue:

elevated variation coupled with inadequate stabilization. Collectively, these findings illustrate that ensemble dynamics, rather than the individual learner, dictate generalization behavior.

Conclusion

This study revealed that Decision Trees, Random Forests, and Gradient Boosting offer fundamentally distinct approaches to the bias–variance trade-off, even if they utilize the same underlying learner. Our controlled experiments and visualizations shown that Decision Trees exhibit considerable variance, Random Forests significantly mitigate this volatility through averaging, and Gradient Boosting effectively reduces bias by iteratively correcting residuals. The differing choice bounds and empirical bias-variance assessments demonstrate how each ensemble uniquely alters the hypothesis space. These data highlight that enhancements in ensemble performance are inconsistent: Random Forests stabilize predictions, but Gradient Boosting increases expressiveness, potentially leading to overfitting. Comprehending these dynamics is crucial for choosing models appropriate for particular data complexity and noise situations.

References

Breiman, L. (2001). *Random forests*. Machine Learning, 45(1), 5–32.

Breiman, L., Friedman, J., Olshen, R., & Stone, C. (1984). *Classification and Regression Trees*. Chapman & Hall/CRC.

Friedman, J. H. (2001). *Greedy function approximation: A gradient boosting machine*. Annals of Statistics, 29(5), 1189–1232.

Geman, S., Bienenstock, E., & Doursat, R. (1992). *Neural networks and the bias/variance dilemma*. Neural Computation, 4(1), 1–58.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2nd ed.). Springer.

Krogh, A., & Vedelsby, J. (1995). *Neural network ensembles, cross validation, and active learning*. Advances in Neural Information Processing Systems (NIPS).

Opitz, D., & Maclin, R. (1999). *Popular ensemble methods: An empirical study*. Journal of Artificial Intelligence Research, 11, 169–198.

Schapire, R. E. (1990). *The strength of weak learnability*. Machine Learning, 5(2), 197–227.

Zhou, Z.-H. (2012). *Ensemble Methods: Foundations and Algorithms*. Chapman & Hall/CRC.