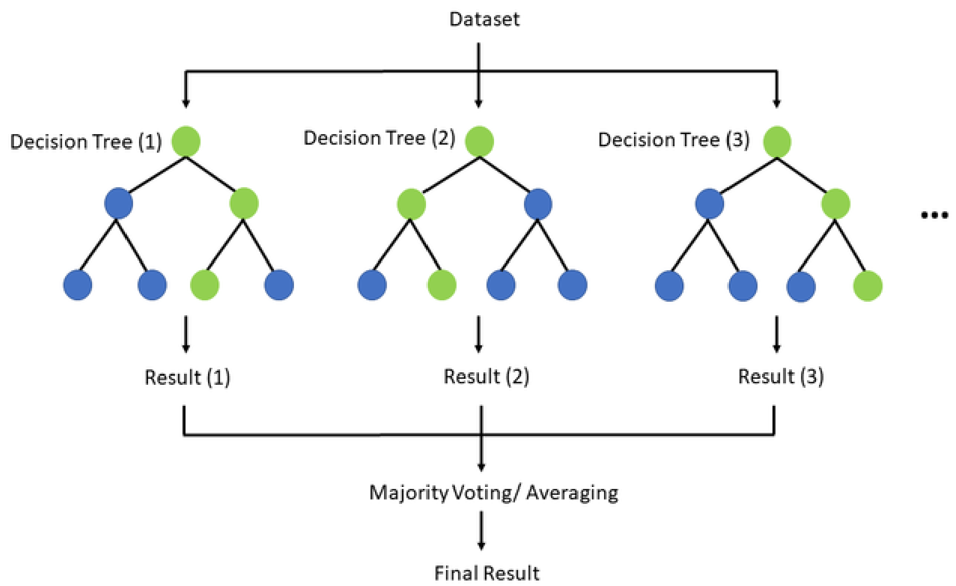
RANDOM FOREST

Introduction

Random Forest is an ensemble learning method that builds multiple decision trees and combines their predictions using averaging (regression) or majority voting (classification). It reduces variance and improves generalization.



Random Forest

1. Problem Formulation

Same input-output structure as decision tree.

Goal: $\min \ \mathbb{E}[\ell(y, \frac{1}{B}\sum_{b=1}^{B} f_b\,(x))]$, where B= number of trees.

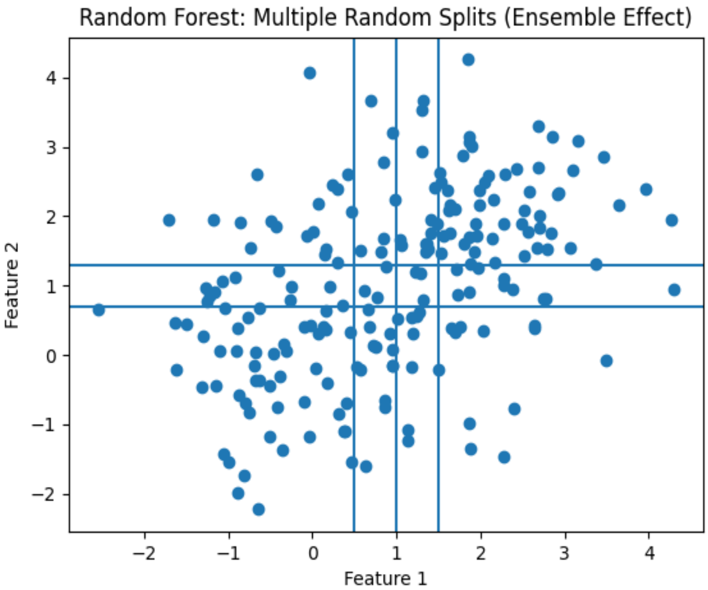
2. Model Specification

Random Forest builds:

- Multiple decision trees
- Each trained on bootstrap samples
- Each split uses random subset of features

Prediction:

- Classification → majority vote
- Regression → average prediction



3. Loss Function

Each tree uses:

- Gini or entropy (classification)
- MSE (regression)

Forest aggregates predictions.

4. Objective Function

Implicit objective:

Reduce variance through averaging: $Var(\bar{f}) = \frac{1}{B}Var(f) + \frac{B-1}{B}Cov(f_i, f_j)$

Random feature selection reduces correlation between trees.

Lower correlation → lower variance.

5. Optimization Method

Algorithm:

For each tree:

1. Draw bootstrap sample
2. At each split:
    - Randomly select subset of features

    - Choose best split among them
3. Grow tree fully (usually no pruning)

6. Statistical Interpretation

Random Forest approximates: $E[y \mid x]$

It is a variance reduction technique via:

- Bagging
- Feature randomness

7. Regularization & Generalization

Why it works:

- Individual trees → high variance
- Averaging → variance reduction

Generalization improves if trees are:

- Accurate
- Weakly correlated

8. Theoretical Properties

- Consistent under certain conditions
- Nonlinear decision boundaries
- Robust to overfitting compared to single tree
- Works well in high-dimensional spaces

9. Computational Complexity

Training: $O(Bpn\log n)$

Inference: $O(B \cdot \text{depth})$

Memory: $O(B \cdot \text{nodes})$

10. Limitations

1. Large memory usage
2. Slower inference than single tree

3. Less interpretable
4. Can still overfit noisy data
5. Poor extrapolation