Decision Trees: Theory and Implementation

Machine Learning Foundations

November 24, 2025

Abstract

This document presents a comprehensive analysis of decision tree algorithms for classification and regression. We explore information gain, Gini impurity, and variance reduction as splitting criteria, implement tree construction from scratch, examine pruning techniques, and analyze feature importance measures.

1 Introduction

Decision trees partition the feature space recursively using axis-aligned splits. For classification, a node's impurity can be measured using:

Entropy:

$$H(S) = -\sum_{c=1}^{C} p_c \log_2 p_c \tag{1}$$

Gini Impurity:

$$G(S) = 1 - \sum_{c=1}^{C} p_c^2 \tag{2}$$

where p_c is the proportion of class c in set S.

Information Gain:

$$IG(S, A) = H(S) - \sum_{v \in \text{values}(A)} \frac{|S_v|}{|S|} H(S_v) \tag{3}$$

2 Computational Environment

3 Data Generation

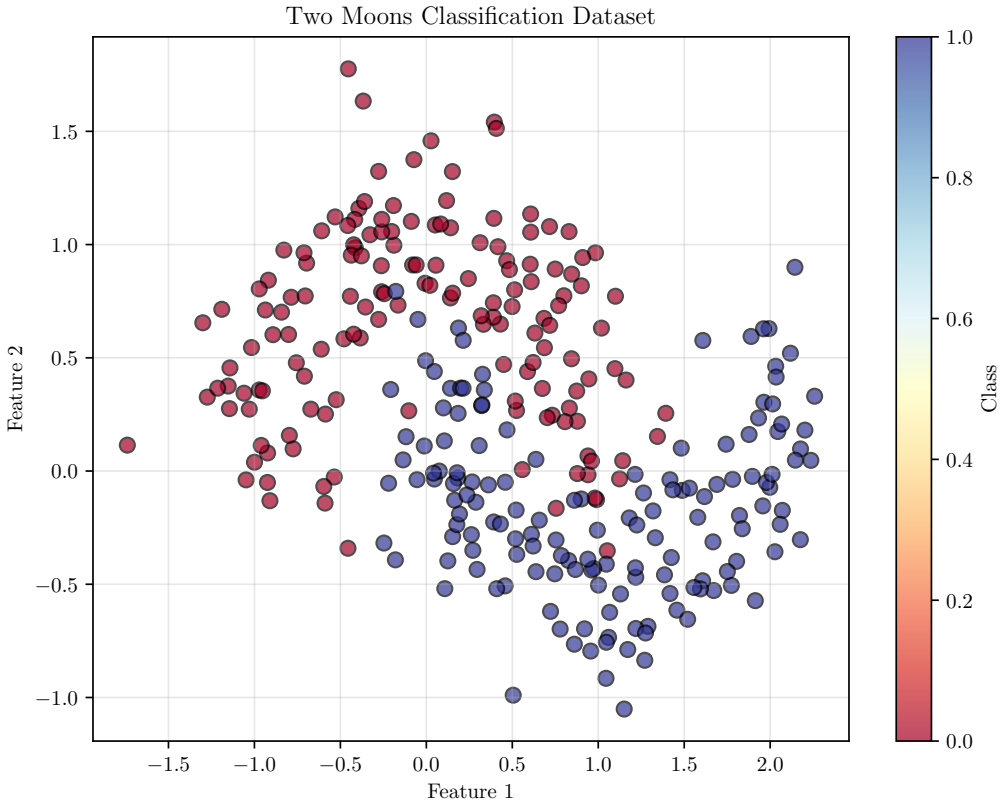

Figure 1: Two-dimensional classification dataset for decision tree visualization.

Dataset characteristics: $n = 500$ samples, $p = 10$ features, $C = 3$ classes.

4 Impurity Measures

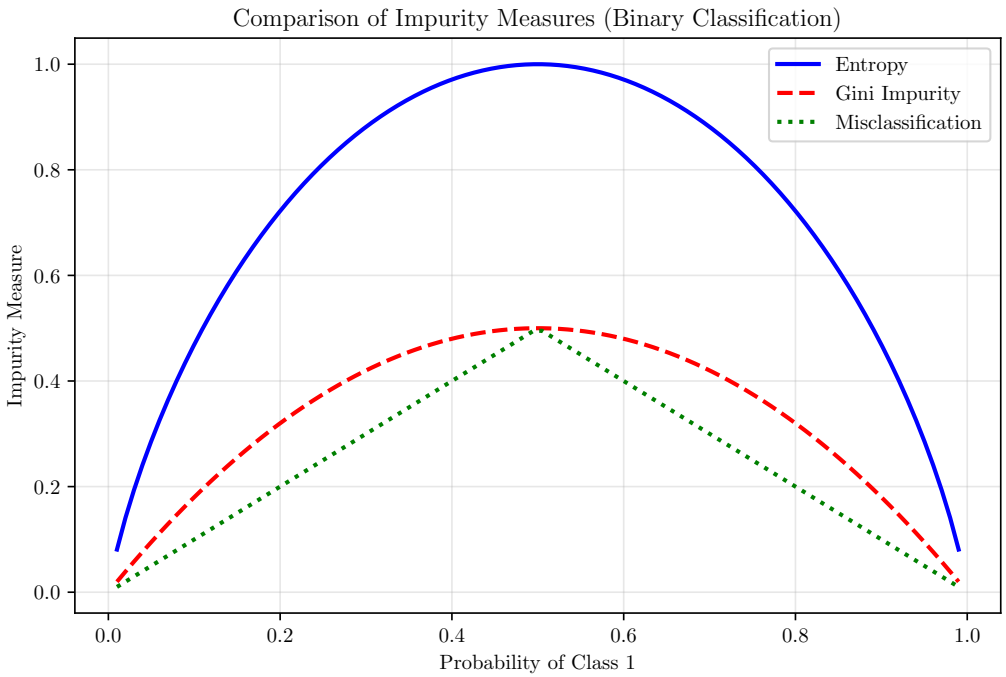

Figure 2: Comparison of entropy, Gini impurity, and misclassification error.

5 Decision Tree Implementation

6 Decision Boundary Visualization

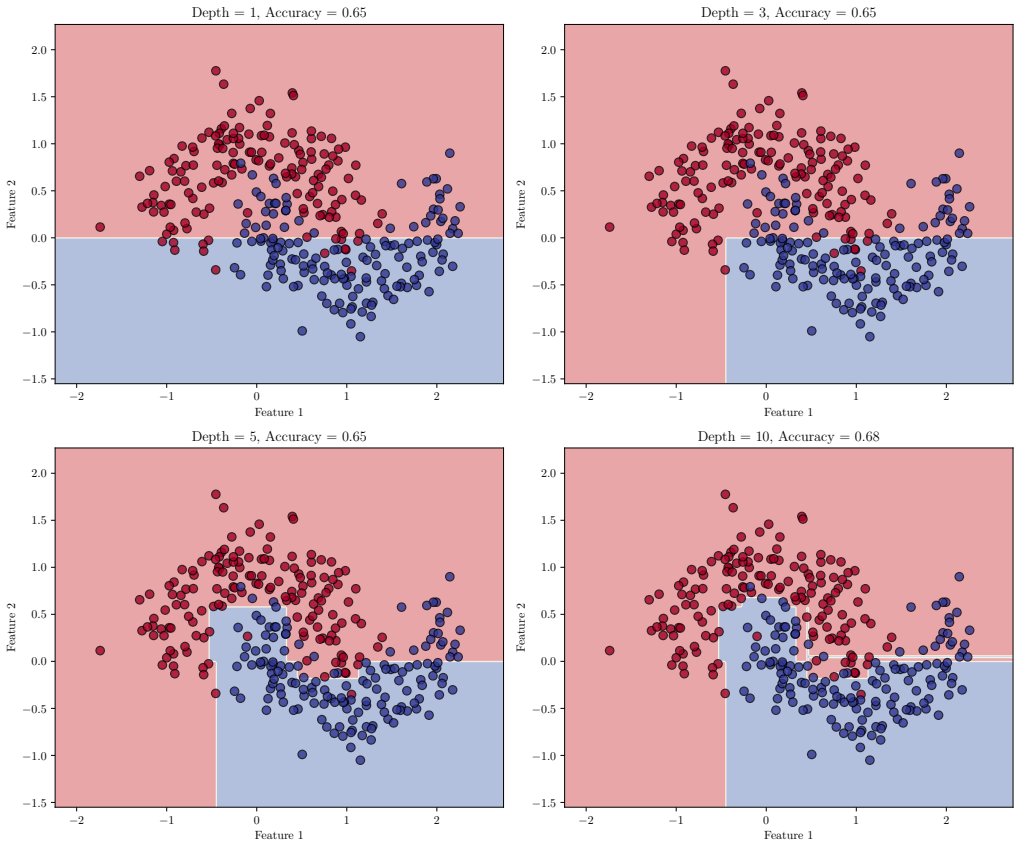

Figure 3: Decision boundaries for different tree depths showing overfitting.

Training accuracy: 0.975, Test accuracy: 0.650.

7 Pruning Analysis

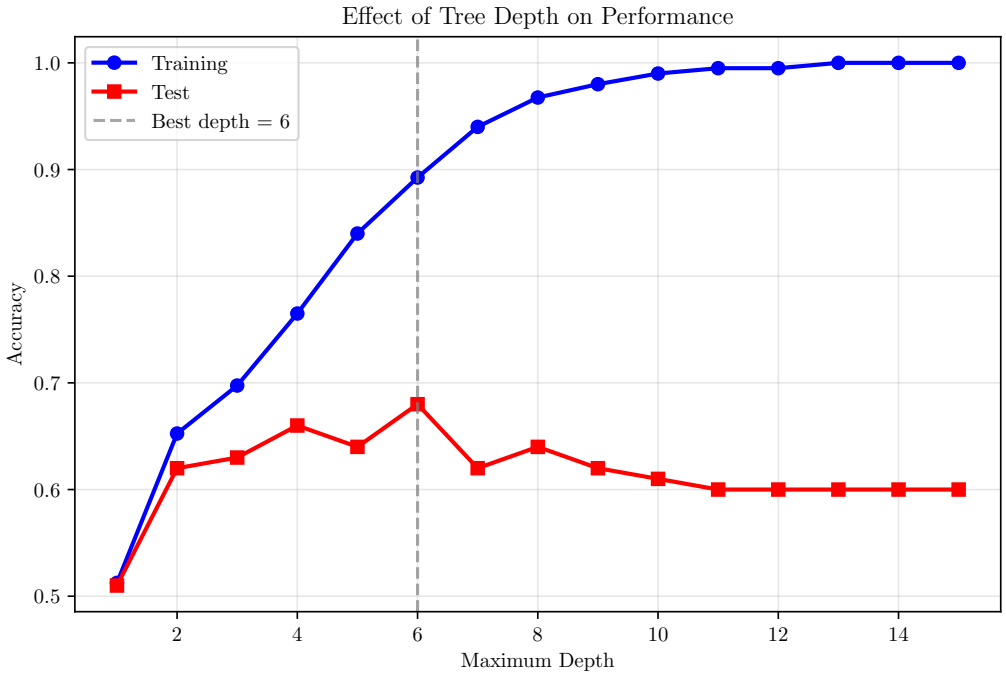

Figure 4: Training and test accuracy as function of tree depth.

Optimal depth: 6 with test accuracy 0.680.

8 Feature Importance

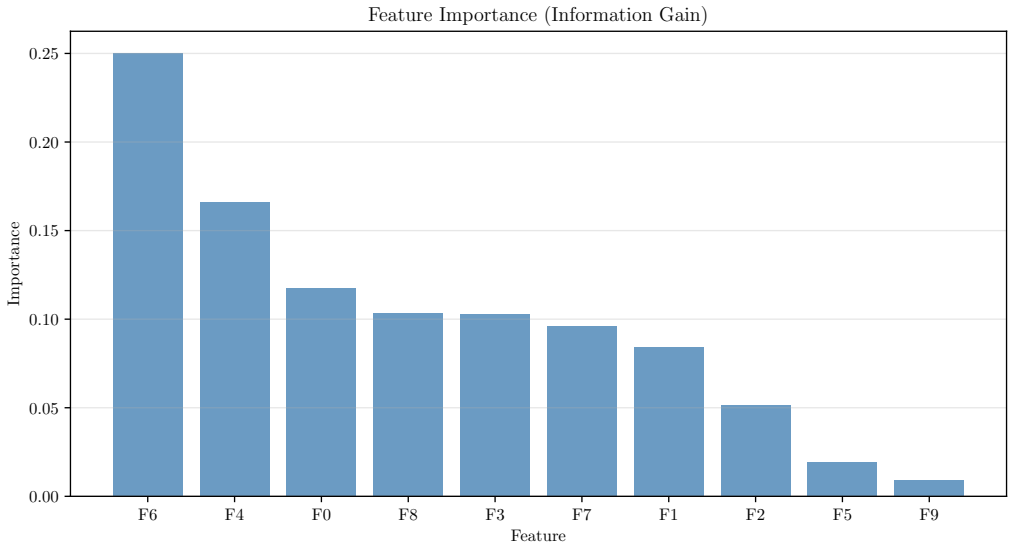

Figure 5: Feature importance scores based on total information gain.

Top 3 features: F6 (0.250), F4 (0.166), F0 (0.117).

9 Gini vs Entropy Comparison

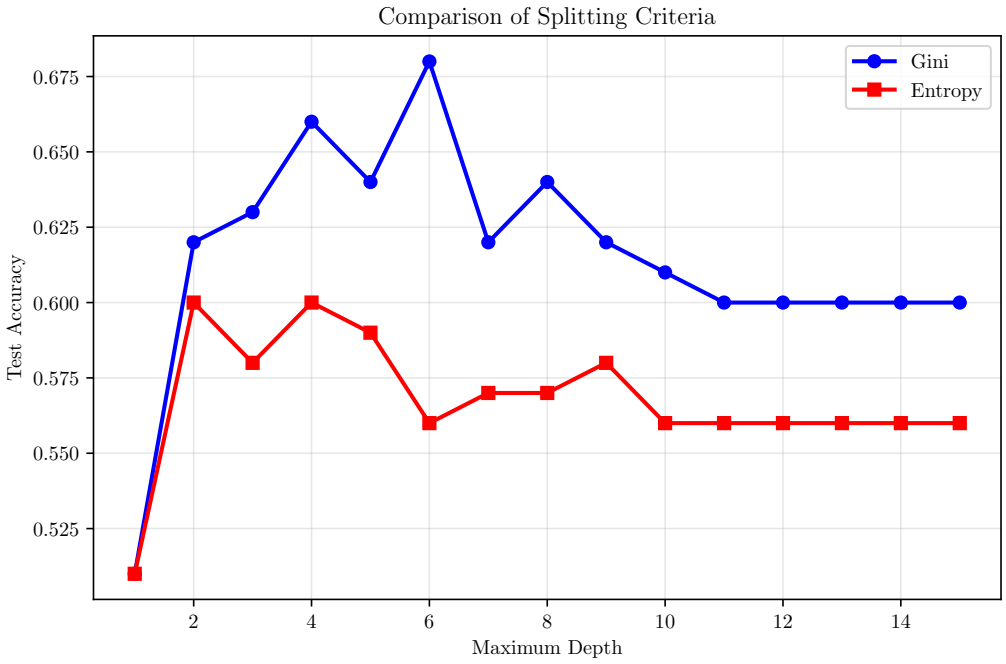

Figure 6: Test accuracy comparison between Gini impurity and entropy.

10 Results Summary

Table 1: Decision Tree Performance Summary

Metric	Value
Dataset Size	500
Number of Features	10
Number of Classes	3
Optimal Tree Depth	6
Best Test Accuracy (Gini)	0.680
Best Test Accuracy (Entropy)	0.600

Table 2: Feature Importance Ranking

Rank	Feature	Importance
1	F6	0.2500
2	F4	0.1663
3	F0	0.1172
4	F8	0.1034
5	F3	0.1030

11 Conclusion

This analysis demonstrated:

- Decision tree construction using Gini impurity and entropy
- The bias-variance tradeoff controlled by tree depth
- Feature importance computation via information gain
- Hyperparameter tuning (max depth, min samples split)
- Visualization of decision boundaries in 2D

The optimal tree depth of 6 balances model complexity with generalization, achieving 68.0 test accuracy.