

Module 11 – Decision Trees

From basic structure to real-world evaluation



What This Module Covers

Decision trees are intuitive and powerful machine learning algorithms. They split data into branches, like a flowchart, to make predictions.



How Decision Trees Work

Learn the mechanics of recursive partitioning and how trees make predictions through a series of questions.



Entropy & Information

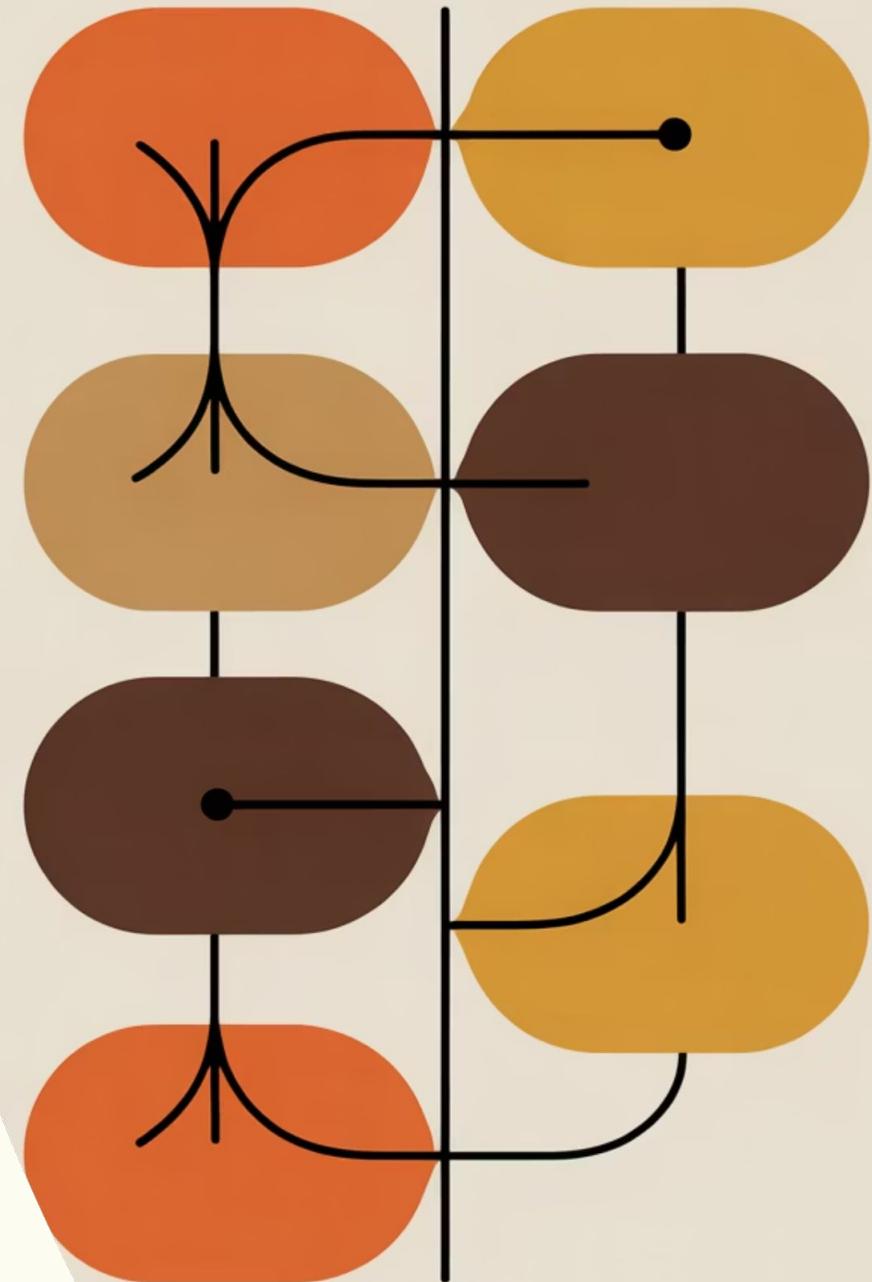
Gain

Understand mathematical concepts like Gini Index and entropy for optimal feature splitting.



Building & Visualizing Trees

Master techniques for constructing and visualizing trees to illuminate model logic.



Making Trees Smarter

Decision trees can overfit training data if too complex, performing poorly on new information. This section explores strategies to balance complexity with generalization ability.

The Overfitting Challenge

Deep, unpruned trees often learn noise instead of signal, leading to poor real-world performance. Addressing this is essential for reliable models.

- Trees grow too deep without constraints
- Model memorizes training data noise
- Poor generalization to new examples

Control Techniques

Several parameters help prevent overfitting by controlling tree growth. These include maximum depth, minimum samples per split, and minimum samples per leaf.

- Pruning removes unnecessary branches
- Depth limits prevent excessive growth
- Split thresholds ensure simplicity

Model Evaluation Metrics

Effective model evaluation is crucial. Choosing the right metrics depends on your problem and priorities.

1

Accuracy & Error Rates

Percentage of correct predictions. Can be misleading with imbalanced datasets.

2

Precision & Recall

Precision (correct positive predictions) and Recall (actual positives caught). F1 score balances both.

3

Confusion Matrix

Table of true/false positives and negatives. Provides a complete picture of classification performance.

4

ROC Curve & AUC

Plots true positive rate vs. false positive rate. AUC summarizes overall discriminative ability.

Ready to Build Your First Tree

Master the fundamentals

Understand how trees split data and make predictions.

Control complexity

Apply pruning and hyperparameter tuning to prevent overfitting.

Evaluate performance

Use metrics to assess model quality.

