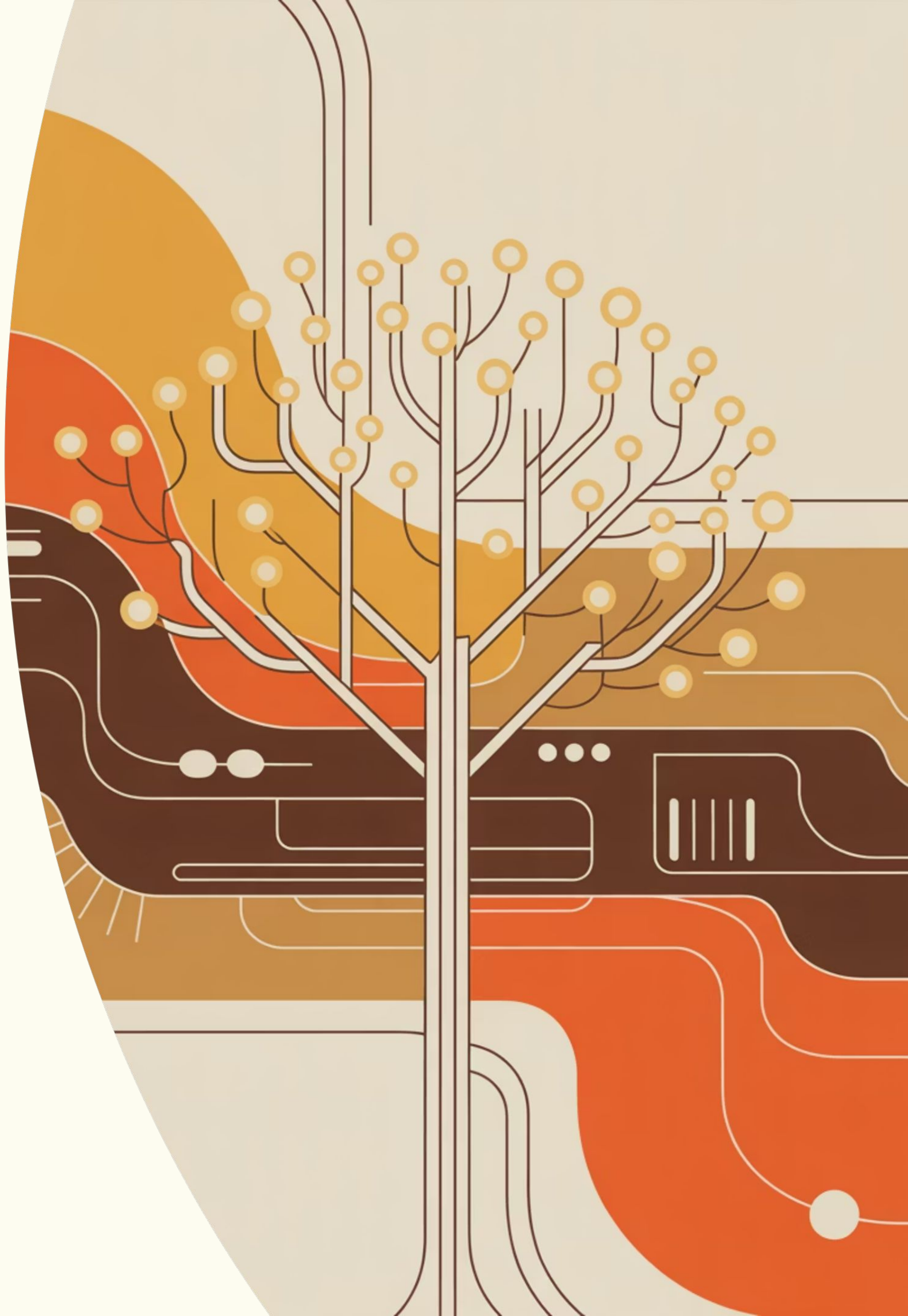
Introduction to Decision Trees

Module 11.1 — Machine Learning Fundamentals



Why Decision Trees?

Simple & Intuitive

Mirror the way humans naturally make decisions through logical sequences

Easy to Visualize

Can be drawn as flowcharts that anyone can understand and explain

Versatile Data Handling

Works seamlessly with both numeric values and categorical variables

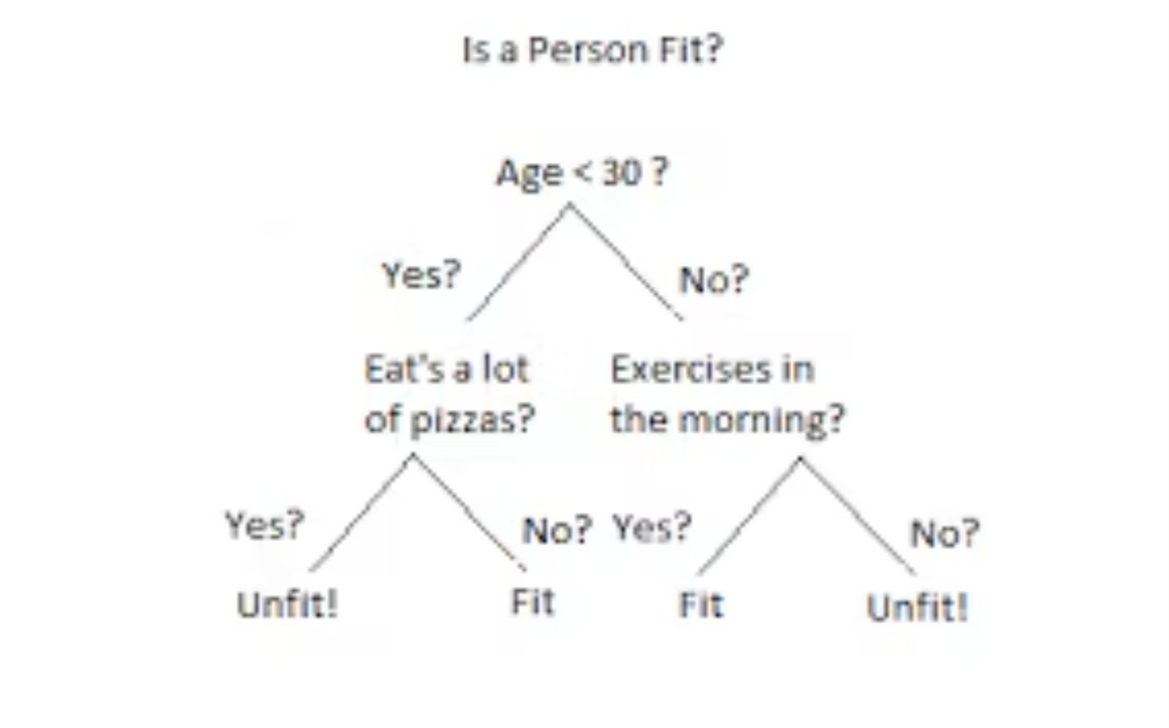
Minimal Preprocessing

Requires little feature scaling or data transformation before training

What Is a Decision Tree?

A decision tree is a predictive model that makes classifications or predictions by asking a sequence of **yes/no** or **threshold-based** questions about the data.

Think of it as an intelligent flowchart that guides you from a starting point to a final answer.



The Journey Through the Tree

Begin at the **root node** (top)

Follow a **branch** based on data

Arrive at a **leaf** (prediction)

Real-Life Example

Deciding Whether to Play Cricket

Let's examine a simple scenario where we predict whether to play cricket based on weather conditions.

Weather	Windy	Play?
Sunny	No	Yes
Sunny	Yes	Yes
Rainy	No	No
Rainy	Yes	No

The Resulting Decision Rule

If *Weather = Sunny* → **Play Cricket**

If *Weather = Rainy* → **Don't Play**

The tree learns this simple pattern from the training data!

Tree Structure Overview

Understanding the anatomy of a decision tree helps us build and interpret models effectively.

01

Root Node

The very first decision point at the top of the tree where all data begins. This split has the most impact on the model.

02

Internal Nodes

Intermediate decision points that further divide the data into more refined groups based on additional features.

03

Branches

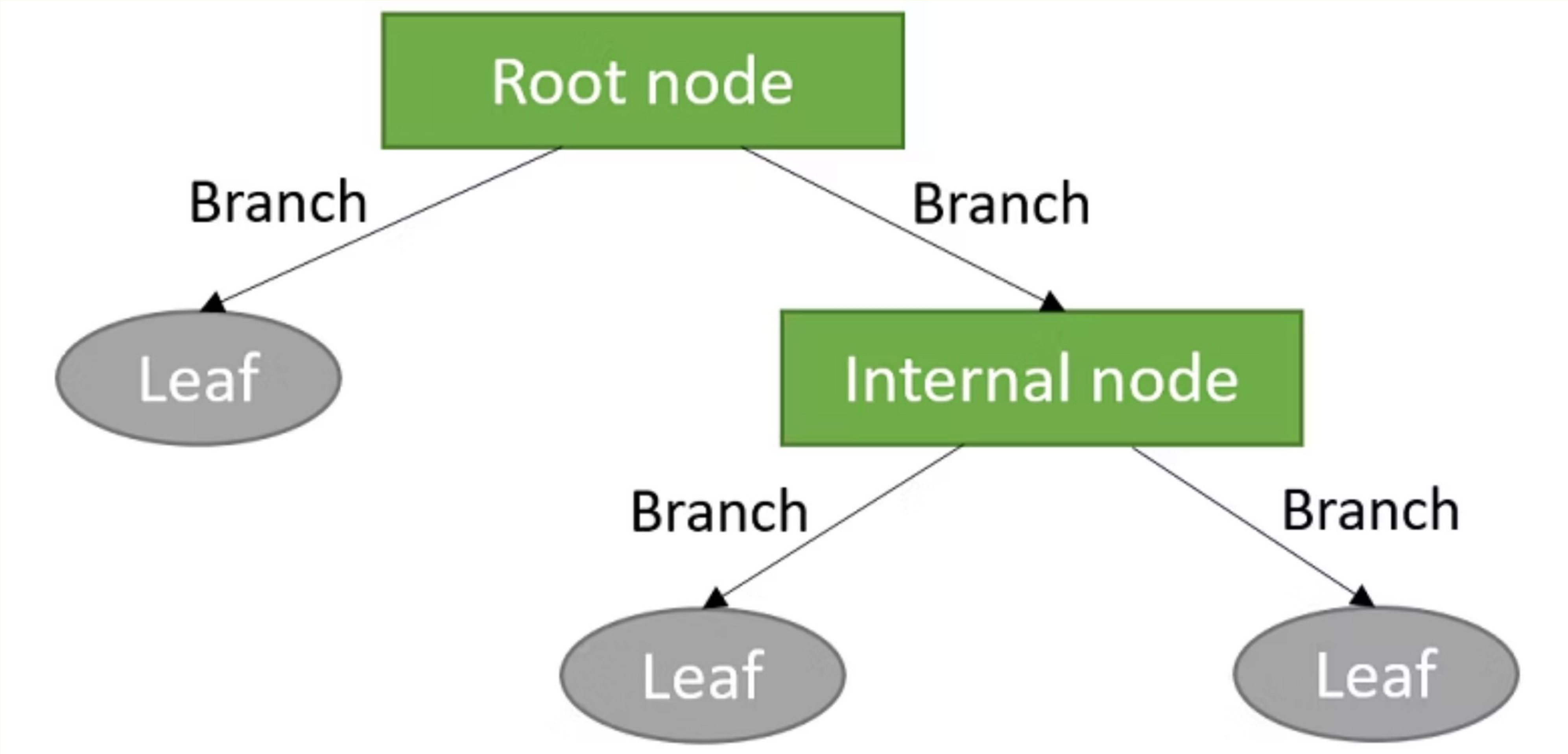
The pathways connecting nodes, representing the outcomes of decisions (yes/no, true/false, or threshold comparisons).

04

Leaf Nodes

Terminal nodes at the bottom of the tree that contain the final predictions or classifications for each path.

Decision Tree Anatomy



This diagram illustrates the complete structure of a decision tree. At the very top, the **root node** initiates the process by asking the first crucial question. Following this, **internal nodes** represent subsequent decision points, further segmenting the data based on various conditions. The connections between these nodes are known as **branches**, which indicate the paths taken depending on the outcome of each decision. Finally, the tree culminates in **leaf nodes** at the bottom, which hold the final predictions or classifications derived from the path traversed.

How Does a Tree Choose a Split?

At each node, the algorithm evaluates many possible questions to ask about the data. It systematically tests different features and thresholds.

Example Questions

- Is age > 45?
- Is cholesterol > 220?
- Is weather = sunny?
- Is income > $50,000?

Key Insight: The tree selects the question that **best reduces impurity** — creating the most homogeneous groups possible after the split.

Understanding Impurity

Building Intuition (Math Comes Later)

Impurity measures how mixed or uncertain a group is. The goal of each split is to create groups that are as *pure* as possible.

Pure Group

10 Yes, 0 No

Perfect! All examples belong to the same class.

Impure Group

5 Yes, 5 No

Highly mixed — the tree can't confidently predict.

In upcoming modules, we'll learn the mathematical measures: **Entropy** and **Gini Index** — which quantify this concept precisely.

Strengths of Decision Trees

Highly Explainable

Every prediction can be traced back through clear, logical rules. Stakeholders can understand exactly why a decision was made.

No Feature Scaling

Unlike many algorithms, decision trees don't require normalization or standardization of features before training.

Nonlinear Patterns

Can capture complex, nonlinear relationships between features without manual feature engineering.

Excellent Baselines

Quick to train and interpret, making them perfect starting points for any machine learning project.

Limitations to Consider

While powerful, decision trees have weaknesses we must understand and address.

Overfitting Tendency

Trees can grow too complex, memorizing training data noise rather than learning true patterns. This hurts performance on new data.

High Variance

Small changes in training data can produce completely different tree structures, making them unstable without proper regularization.

Requires Pruning

Without careful pruning strategies, trees become unwieldy and lose their interpretability advantage.

Struggles with High Dimensions

Performance degrades when dealing with datasets containing hundreds or thousands of features.

What's Next?

You've learned the fundamentals! Now we'll dive deeper into the mathematics and practical implementation of decision trees.

Entropy & Gini

Mathematical measures of impurity

Information Gain

How to select optimal splits

Build Your Tree

Hands-on implementation

Pruning

Preventing overfitting

Evaluation

ROC curves and AUC