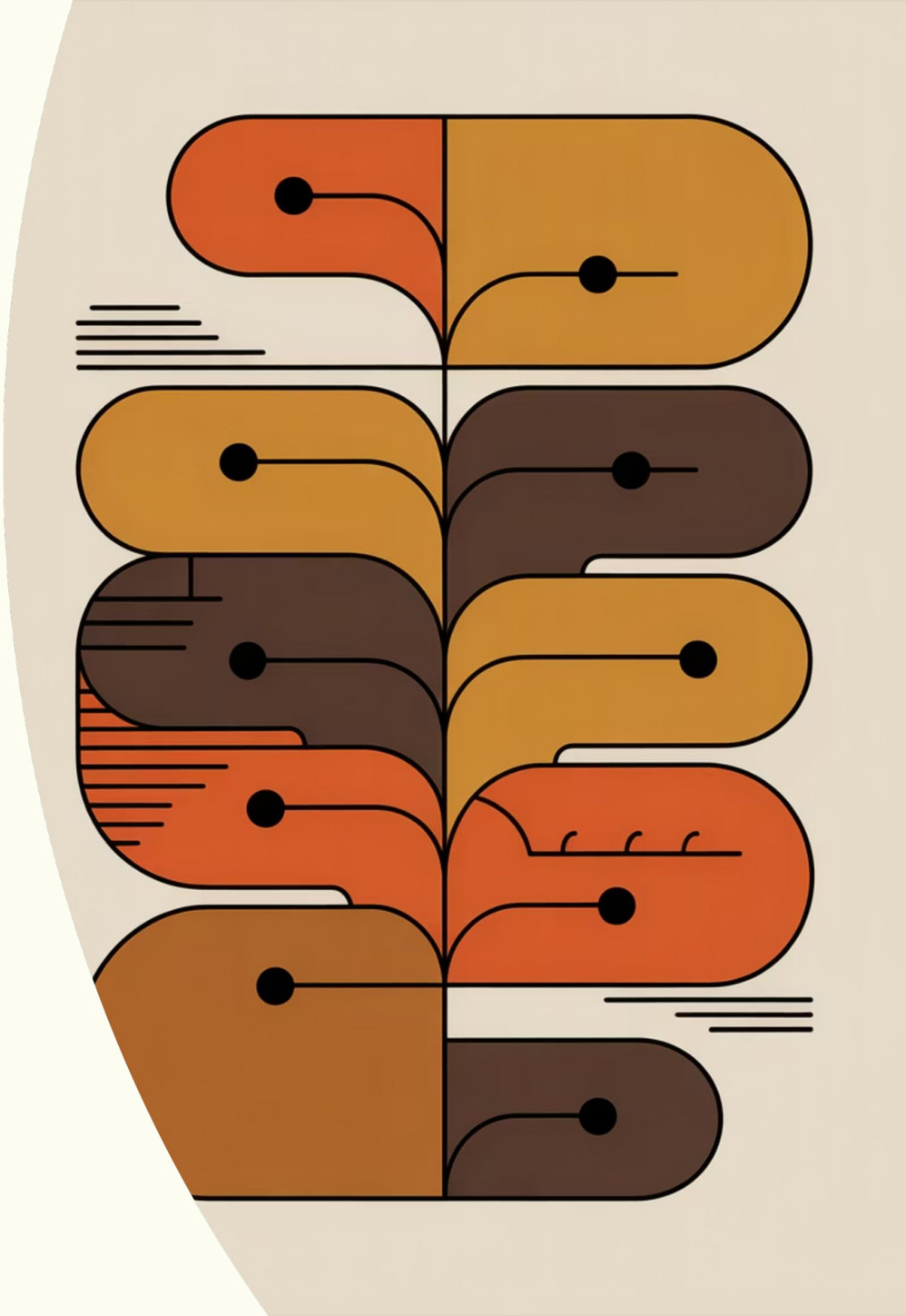
Entropy, Gini Index, and Information Gain

Module 11.2 – Entropy, Gini Index, and Information Gain



Why Do We Need Impurity Measures?

Decision trees work by asking questions to split data into groups. But not all questions are equally valuable. We need a systematic way to evaluate which question creates the most useful split.

This is where **impurity measures** come in. They quantify how "mixed" or "pure" a group of data points is after a split. The goal is simple: create groups that are as pure as possible, meaning most or all members belong to the same class.

Pure Group

All members share the same label

✓ Good split

Mixed Group

Members have different labels

✗ Poor split

Think of it like sorting a deck of cards: a pile of all hearts is "pure," while a pile with mixed suits is "impure." Decision trees prefer splits that maximize purity.

Intuition for Purity

Let's develop intuition by examining three different groups with varying levels of purity. Notice how the proportion of "Yes" to "No" labels affects our sense of how well-organized each group is.

Group A: Perfectly Pure

10 Yes, 0 No

This group is completely homogeneous. Every single member belongs to the same class. This represents the ideal outcome of a split—maximum purity with zero uncertainty.

Group B: Maximum Impurity

5 Yes, 5 No

This group is evenly split between classes. It's the most uncertain scenario possible—if you randomly picked a member, you'd have exactly 50% chance of getting either class. Maximum disorder.

Group C: Moderate Purity

8 Yes, 2 No

This group leans heavily toward one class but isn't perfectly pure. It's better than Group B but not as ideal as Group A. Most members share the same label, creating moderate certainty.

Decision trees actively seek splits that move us from scenarios like Group B toward scenarios like Group A, progressively increasing purity at each level of the tree.

Entropy: Measuring Disorder

Entropy is a concept borrowed from physics and information theory that quantifies the amount of **disorder** or **uncertainty** in a dataset. In decision trees, it helps us measure how mixed our groups are.

The key intuition:

High entropy = messy, unpredictable groups with mixed classes
Low entropy = clean, organized groups with similar members
Zero entropy = perfectly pure group with only one class

1

50% Yes, 50% No

Maximum entropy—complete uncertainty about class membership

2

80% Yes, 20% No

Moderate entropy—some predictability emerging

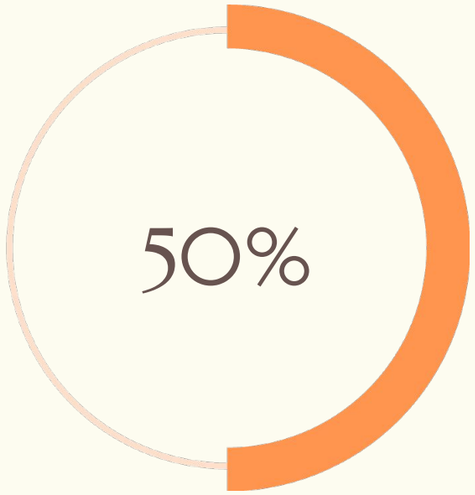
3

100% Yes, 0% No

Zero entropy—perfect certainty and complete order

Entropy Visualized

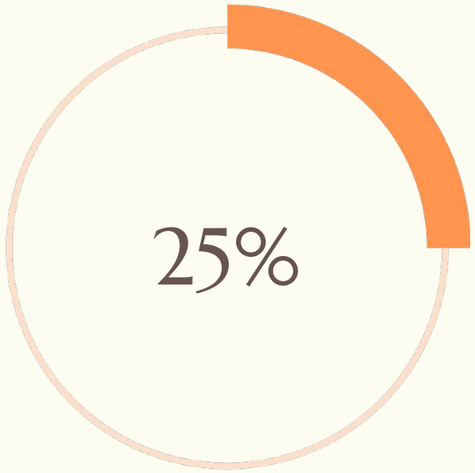
Let's see how entropy changes as groups become more pure. Each scenario shows a different mix of "Yes" and "No" labels, with corresponding entropy levels. Notice the pattern: as one class becomes more dominant, entropy decreases.



High Entropy

4 Yes, 4 No

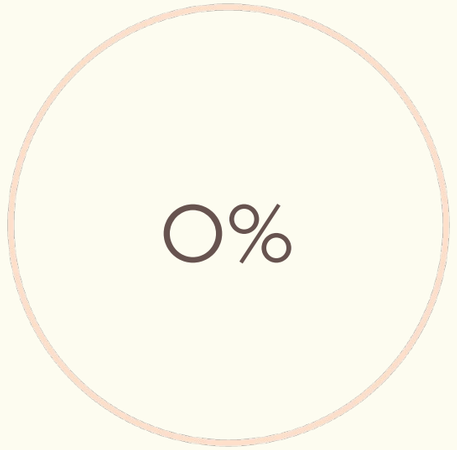
Maximum disorder—this group provides minimal information for classification



Medium Entropy

6 Yes, 2 No

Moderate disorder—the group leans toward one class but uncertainty remains



Low Entropy

8 Yes, 0 No

Minimal disorder—this pure group gives us maximum certainty

Key Insight: We don't need to calculate exact entropy values to understand the concept. The important takeaway is that entropy decreases as groups become more homogeneous, and decision trees aim to minimize entropy through strategic splits.

Gini Index: Probability of Mistakes

The Gini Index offers an alternative way to measure impurity, based on a simple thought experiment: *If you randomly picked a data point from this group and randomly assigned it a class label based on the group's distribution, how often would you be wrong?*

This probabilistic interpretation makes Gini intuitive:

High Gini = High chance of misclassification
Low Gini = Low chance of mistakes
Zero Gini = Pure group, impossible to be wrong

While entropy measures disorder, Gini measures expected error. Both metrics tell the decision tree the same fundamental story: **"How clean is this group?"**

In practice, entropy and Gini often lead to similar splitting decisions. The choice between them is usually based on computational efficiency or personal preference rather than dramatic performance differences.

Gini Index Examples

Let's examine the same scenarios we used for entropy, but through the lens of misclassification probability. Notice how Gini values parallel entropy—both decrease as groups become purer.

50%

High Gini

5 Yes, 5 No — Maximum misclassification risk

Mixed Group

Picking blindly would yield unpredictable results—high Gini reflects this uncertainty

35%

Medium Gini

7 Yes, 3 No — Moderate misclassification risk

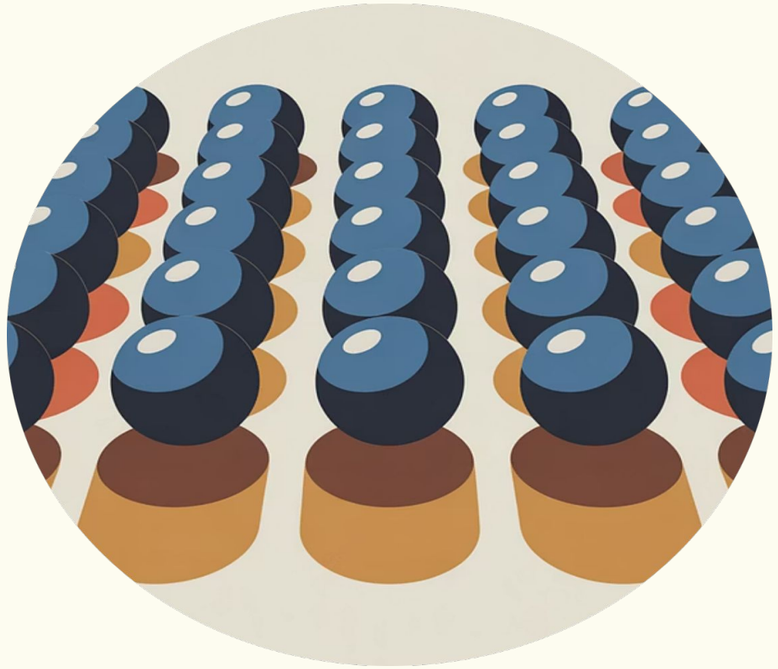
Dominated Group

Most picks would be correct, with occasional mistakes—moderate Gini

0%

Low Gini

10 Yes, 0 No — Zero misclassification risk



Pure Group

Every pick guarantees the same result—Gini reaches zero

Information Gain: Quantifying Improvement

Now we understand how to measure impurity in a single group. But decision trees need to evaluate *splits*—comparing the parent group's impurity to the children groups' impurity after asking a question.

Information Gain captures exactly this concept. It measures how much impurity decreases when we split a group using a particular question. The formula is conceptually simple:

> **Information Gain = Parent Impurity − Weighted Average of Children Impurity**

01

Start with parent group

Calculate impurity (entropy or Gini) of the original mixed group

02

Apply a split

Ask a question that divides data into two or more child groups

03

Measure children impurity

Calculate impurity for each resulting child group

04

Compute information gain

Subtract weighted children impurity from parent impurity

Trees systematically try different questions and choose the one that maximizes information gain—the split that creates the biggest jump toward purity.

A Simple Split Example

Let's walk through a concrete example to see information gain in action. We'll keep the numbers simple to focus on the concept rather than calculations.

1 Before Split

Parent Group: 6 Yes, 4 No

This group is fairly mixed—there's uncertainty about which class a randomly selected member belongs to. We need a better split.

2 After Split

Left Child: 5 Yes, 0 No (pure!)

Right Child: 1 Yes, 4 No (cleaner)

Both children are more organized than the parent. Purity has increased dramatically.

3 Result

High Information Gain

The split successfully reduced impurity. The tree would choose this question because it creates meaningful separation between classes.

Why This Split Works

- Left child is perfectly pure—no uncertainty remains
- Right child strongly favors one class—minimal uncertainty
- Together, they're much more organized than the original mixed group
- The question that created this split effectively separates the classes

A Simple Split Example

The diagram shows how a decision tree makes one split to organize data better:

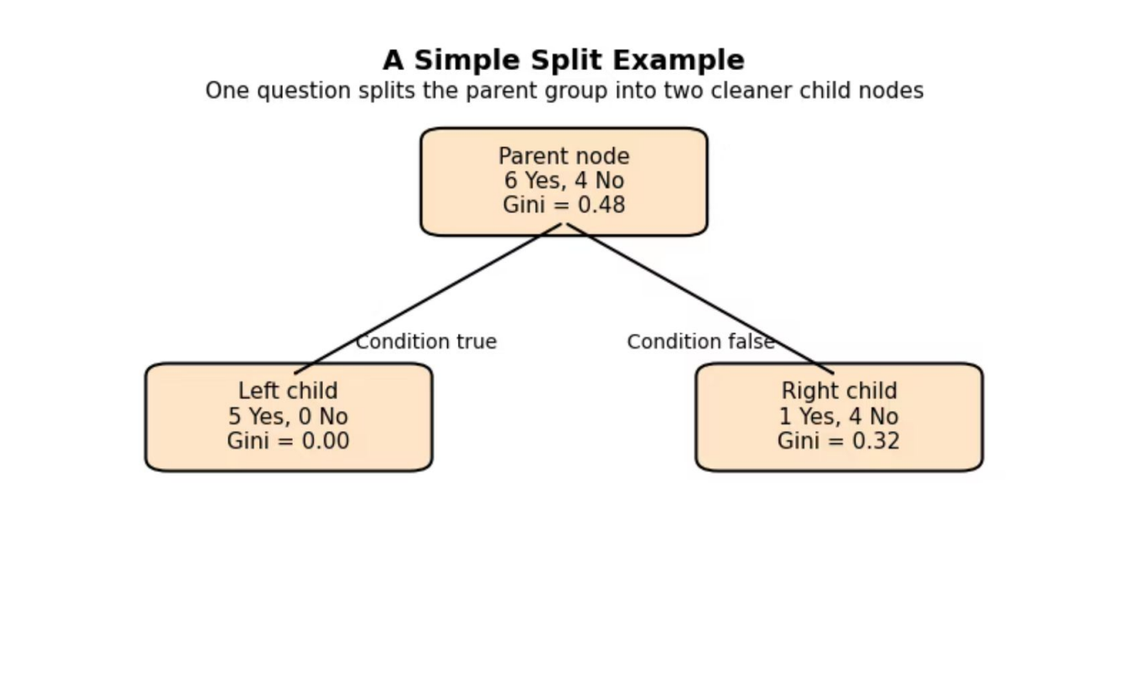
Parent Node (Top): We start with a mixed group of 6 Yes and 4 No. The Gini index is 0.48, which means this group is fairly impure - there's a lot of mixing between the two classes.

The Split: The tree asks one question (a condition) that divides the data into two groups.

Left Child (Condition True): This group has 5 Yes and 0 No. The Gini index is 0.00 - this is a perfectly pure group! Everyone belongs to the same class.

Right Child (Condition False): This group has 1 Yes and 4 No. The Gini index is 0.32 - this is much cleaner than the parent, with most members belonging to the "No" class.

The Result: By asking just one question, we went from a messy mixed group to two much cleaner groups. The left side is perfectly pure, and the right side is mostly pure. This is exactly what decision trees try to do - find questions that create the cleanest possible groups.



Key Takeaways

Entropy

Measures disorder and uncertainty in a group. High entropy means messy, mixed classes. Low entropy means clean groups.

Gini Index

Measures misclassification risk. Asks: "How often would we be wrong if we guessed randomly?" Pure groups have zero Gini.

Information Gain

Measures improvement from a split. Quantifies how much impurity decreases when we ask a question. Trees choose splits with maximum information gain.

What's Next?

Now that we understand the theory behind split selection, we're ready to put it into practice. In the next module, we'll build our first decision tree using synthetic data, watching these impurity measures guide the tree's construction step by step.

You'll see how information gain drives every decision, creating a hierarchy of questions that efficiently separates classes.