DECISION TREE : ID3

📘 Decision Tree Basics

A **Decision Tree** is a **supervised learning algorithm** used for both **classification** and **regression**. It uses a **tree-like structure** where:

- Internal nodes represent tests on features.
- Branches represent outcomes of the test.
- Leaf nodes represent class labels (in classification) or numeric values (in regression).

⚙️ Key Terminologies:

- **Root Node**: The top-most node (starting point).
- **Splitting**: Dividing the dataset based on an attribute.
- **Leaf/Terminal Node**: Final output label.
- **Information Gain**: A measure to select the best attribute to split the data.
- **Entropy**: A measure of impurity or randomness.

🌳 ID3 Algorithm (Iterative Dichotomiser 3)

ID3 builds the decision tree using:

- **Entropy** to measure impurity.

- **Information Gain (IG)** to choose the best feature for splitting.

🔸 Entropy Formula:

$$Entropy(S) = -\sum_{i=1}^{c} p_i \log_2(p_i)$$

Where p_i is the proportion of class i in dataset S.

🔸 Information Gain Formula:

$$IG(S, A) = Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} \cdot Entropy(S_v)$$

🧮 Numerical Example using ID3

Dataset: "Play Tennis"

Day	Outlook	Temperature	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes

Day	Outlook	Temperature	Humidity	Wind	PlayTennis
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

Step 1: Compute Entropy of Target (PlayTennis)

- Total = 14
- Yes = 9, No = 5

$$Entropy(S) = -\frac{9}{14}\log_2(\frac{9}{14}) - \frac{5}{14}\log_2(\frac{5}{14}) = 0.940$$

Step 2: Compute Information Gain for "Outlook"

Outlook	Count	Yes	No	Entropy
Sunny	5	2	3	0.971
Overcast	4	4	0	0.000
Rain	5	3	2	0.971

$$IG(S, Outlook) = 0.940 - (\frac{5}{14} \cdot 0.971 + \frac{4}{14} \cdot 0.000 + \frac{5}{14} \cdot 0.971)$$

$$= 0.940 - (0.347 + 0 + 0.347) = 0.940 - 0.694 = \boxed{0.246}$$

Repeat for other attributes and choose the one with the **highest IG**.

Step 3: Choose Attribute with Highest IG

Assume:

- IG(Outlook) = 0.246
- IG(Humidity) = 0.151
- IG(Wind) = 0.048
- IG(Temperature) = 0.029

🔹 So, **Outlook** is selected as root.

🌲 Partial Decision Tree

```plaintext
          Outlook
         /   |   \
     Sunny Overcast Rain
      /      |     \
    ...     Yes     ...
```

Each branch continues recursively, using remaining attributes on the subset.

✅ Summary

- **ID3** is simple and intuitive.
- Uses **entropy** and **information gain** to split nodes.
- Doesn't support continuous values or pruning (improved in **C4.5**).

SEE BELOW A COMPLETE EXAMPLE OF ID3 DECISION TREE

||

🧮 Dataset Recap

Day	Outlook	Temperature	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

- Total: 14 samples
- PlayTennis: Yes = 9, No = 5

🔹 Step 1: Calculate Entropy of Full Dataset

$$Entropy(S) = -\frac{9}{14}\log_2\left(\frac{9}{14}\right) - \frac{5}{14}\log_2\left(\frac{5}{14}\right)$$

$$= -0.643 \cdot \log_2(0.643) - 0.357 \cdot \log_2(0.357)$$

$$= -0.643 \cdot (-0.643) - 0.357 \cdot (-1.485) = 0.940$$

🔹 Step 2: Calculate IG for all attributes

A. Attribute: Outlook

Outlook	Total	Yes	No	Entropy
Sunny	5	2	3	$-\frac{2}{5}\log_2\frac{2}{5} - \frac{3}{5}\log_2\frac{3}{5} = 0.971$
Overcast	4	4	0	0.0
Rain	5	3	2	same as Sunny = 0.971

$$IG(S, Outlook) = 0.940 - (\frac{5}{14} \cdot 0.971 + \frac{4}{14} \cdot 0 + \frac{5}{14} \cdot 0.971)$$

$$= 0.940 - (0.347 + 0 + 0.347) = 0.940 - 0.694 = \boxed{0.246}$$

B. Attribute: Humidity

Humidity	Total	Yes	No	Entropy
High	7	3	4	0.985
Normal	7	6	1	0.591

$$IG(S, Humidity) = 0.940 - (\frac{7}{14} \cdot 0.985 + \frac{7}{14} \cdot 0.591)$$

$$= 0.940 - (0.493 + 0.296) = 0.940 - 0.789 = \boxed{0.151}$$

C. Attribute: Wind

Wind	Total	Yes	No	Entropy
Weak	8	6	2	0.811
Strong	6	3	3	1.0

$$IG(S, Wind) = 0.940 - (\frac{8}{14} \cdot 0.811 + \frac{6}{14} \cdot 1.0)$$

$$= 0.940 - (0.463 + 0.429) = 0.940 - 0.892 = \boxed{0.048}$$

D. Attribute: Temperature

Temperature	Total	Yes	No	Entropy
Hot	4	2	2	1.0
Mild	6	4	2	0.918
Cool	4	3	1	0.811

$$IG(S, Temp) = 0.940 - (\frac{4}{14} \cdot 1.0 + \frac{6}{14} \cdot 0.918 + \frac{4}{14} \cdot 0.811)$$

$$= 0.940 - (0.286 + 0.393 + 0.232) = 0.940 - 0.911 = \boxed{0.029}$$

✅ Select the Attribute with Highest Gain:

- **Outlook = 0.246 (highest)** → Chosen as the root node.

🌲 Step 3: Build the Tree Recursively

➤ Branch: Outlook = Overcast

- 4 samples → All **Yes** ⇒ Leaf = **Yes**

➤ Branch: Outlook = Rain

Subset: D4, D5, D6, D10, D14

(PlayTennis: Yes = 3, No = 2)

Entropy = 0.971

Now calculate **IG for Rain subset**:

i. Attribute: Wind

Wind	Total	Yes	No	Entropy
Weak	3	3	0	0.0
Strong	2	0	2	0.0

$$IG = 0.971 - (\frac{3}{5} \cdot 0 + \frac{2}{5} \cdot 0) = 0.971$$

✅ Max gain → Split on **Wind**

- Wind = Weak → All Yes ⇒ Leaf = Yes
- Wind = Strong → All No ⇒ Leaf = No

➤ Branch: Outlook = Sunny

Subset: D1, D2, D8, D9, D11

(Yes = 2, No = 3)

Entropy = 0.971

Try splitting:

i. Attribute: Humidity

Humidity	Total	Yes	No	Entropy
High	3	0	3	0.0

Humidity	Total	Yes	No	Entropy
Normal	2	2	0	0.0

$$IG = 0.971 - (\frac{3}{5} \cdot 0 + \frac{2}{5} \cdot 0) = 0.971$$

✅ Max gain → Split on **Humidity**

- High → All No ⇒ Leaf = No
- Normal → All Yes ⇒ Leaf = Yes

✅ Final Decision Tree:

plaintext

```
            Outlook
          /   |    \
       Sunny Overcast  Rain
       /      |      \
   Humidity    Yes    Wind
    /  \              /  \
 High Normal       Weak   Strong
  No    Yes          Yes     No
```

📝 Summary

- Used **ID3** with **entropy** and **information gain**.
- Chose attributes recursively with highest IG.
- Constructed a complete decision tree.
- Tree is **perfectly consistent** with training data.