Decision Tree (ID3): Mathematical Notes

This note demonstrates how to build an **ID3 Decision Tree** using a small **binary classification** dataset with **three features**. ID3 selects splits using **Information Gain** based on **Entropy**.

Dataset: Weather vs Play (Binary Class)

Features:

- X_1 = Outlook {Sunny, Overcast, Rainy}
- X_2 = Humidity {High, Normal}
- X_3 = Wind {Weak, Strong}

Class label:

$$y = \text{Play} \in \{\text{Yes}, \text{No}\}$$

ID	Outlook	Humidity	Wind	Play (y)
1	Sunny	High	Weak	No
2	Sunny	High	Strong	No
3	Sunny	Normal	Weak	Yes
4	Overcast	High	Weak	Yes
5	Overcast	Normal	Strong	Yes
6	Rainy	High	Weak	Yes
7	Rainy	Normal	Weak	Yes
8	Rainy	High	Strong	No

Goal: Build the ID3 decision tree and draw the final tree.

I. Entropy and Information Gain

1. Entropy

For a dataset S with two classes (Yes/No), entropy is:

$$H(S) = -P_{Yes}\log_2(P_{Yes}) - P_{No}\log_2(P_{No})$$

2. Information Gain (ID3 Criterion)

For a feature A:

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

where S_v is the subset of S where feature A takes value v.

II. Step-by-Step Tree Construction (ID3)

Step 1: Compute Entropy of the Full Dataset S

From the dataset:

$$|S| = 8, \quad \#Yes = 5, \quad \#No = 3$$

$$P_{Yes} = \frac{5}{8}, \quad P_{No} = \frac{3}{8}$$

$$H(S) = -\frac{5}{8}\log_2\left(\frac{5}{8}\right) - \frac{3}{8}\log_2\left(\frac{3}{8}\right)$$

Numerically:

$$\log_2\left(\frac{5}{8}\right) \approx -0.6781, \quad \log_2\left(\frac{3}{8}\right) \approx -1.4150$$

$$H(S) \approx -\frac{5}{8}(-0.6781) - \frac{3}{8}(-1.4150) \approx 0.9544$$

So,

$$\boxed{H(S) \approx 0.9544}$$

Visual 1: Class Distribution at the Root

Root Dataset S
Yes = 5, No = 3, Total = 8

Step 2: Compute Information Gain for Each Feature

A) Information Gain for Humidity Split by Humidity:

- Humidity = Normal: (IDs 3,5,7) ⇒ Yes=3, No=0
- Humidity = High: (IDs 1,2,4,6,8) ⇒ Yes=2, No=3

Entropy for Normal:

$$H(S_{Normal}) = 0$$

Entropy for High:

$$P_{Yes} = \frac{2}{5},\ P_{No} = \frac{3}{5}$$

$$H(S_{High}) = -\frac{2}{5}\log_2\left(\frac{2}{5}\right) - \frac{3}{5}\log_2\left(\frac{3}{5}\right) \approx 0.9710$$

Weighted entropy after split:

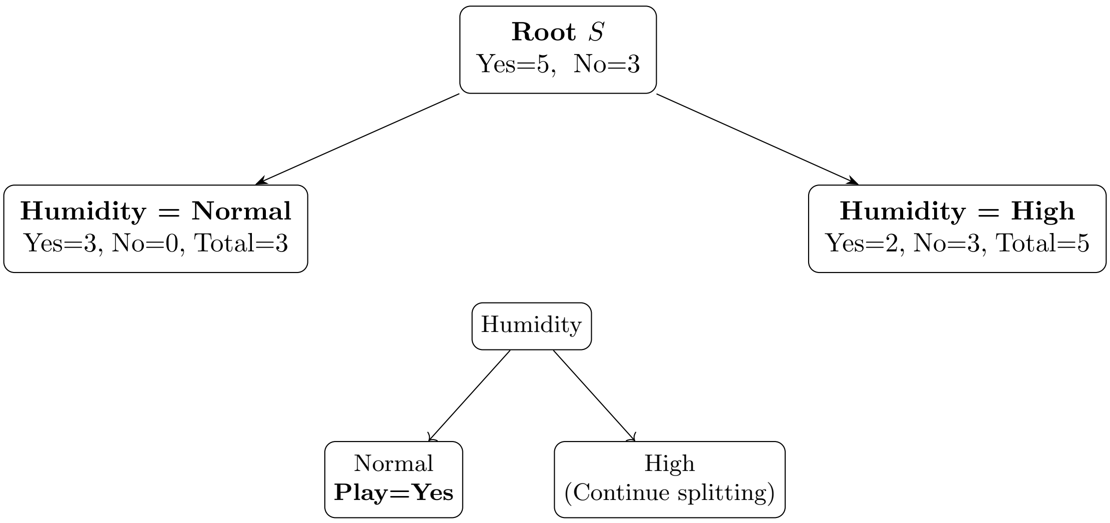
$$H(S|Humidity) = \frac{3}{8}(0) + \frac{5}{8}(0.9710) \approx 0.6069$$

Information Gain:

$$IG(S, Humidity) = 0.9544 - 0.6069 \approx 0.3475$$

$$\boxed{IG(S, Humidity) \approx 0.3476}$$

Visual 2: Split on Humidity (Root Split)



B) Information Gain for Outlook Split by Outlook:

- Sunny (IDs 1,2,3): Yes=1, No=2
- Overcast (IDs 4,5): Yes=2, No=0
- Rainy (IDs 6,7,8): Yes=2, No=1

Entropy:

$$H(S_{Sunny}) \approx 0.9183, \quad H(S_{Overcast}) = 0, \quad H(S_{Rainy}) \approx 0.9183$$

Weighted entropy:

$$H(S|Outlook) = \frac{3}{8}(0.9183) + \frac{2}{8}(0) + \frac{3}{8}(0.9183) \approx 0.6887$$

Information Gain:

$$IG(S, Outlook) = 0.9544 - 0.6887 \approx 0.2657$$

$$\boxed{IG(S, Outlook) \approx 0.2657}$$

C) Information Gain for Wind Split by Wind:

- Weak (IDs 1,3,4,6,7): Yes=4, No=1
- Strong (IDs 2,5,8): Yes=1, No=2

Entropy:

$$H(S_{Weak}) \approx 0.7219, \quad H(S_{Strong}) \approx 0.9183$$

Weighted entropy:

$$H(S|Wind) = \frac{5}{8}(0.7219) + \frac{3}{8}(0.9183) \approx 0.7956$$

Information Gain:

$$IG(S, Wind) = 0.9544 - 0.7956 \approx 0.1588$$

$$\boxed{IG(S, Wind) \approx 0.1589}$$

Step 3: Select the Root Node

Choose the feature with the highest Information Gain:

$$IG(Humidity) \approx 0.3476,\ IG(Outlook) \approx 0.2657,\ IG(Wind) \approx 0.1589$$

Thus,

$$\boxed{\text{Root Feature} = \text{Humidity}}$$

Step 4: Expand Each Branch

Branch 1: Humidity = Normal Subset (IDs 3,5,7): all are **Yes**. So it is a leaf node:

$$\boxed{\text{If Humidity} = \text{Normal} \Rightarrow \text{Play} = \text{Yes}}$$

Branch 2: Humidity = High Subset (IDs 1,2,4,6,8): Yes=2, No=3 (not pure). Now compute gains within this subset using remaining features {Outlook, Wind}.

From this subset (computed):

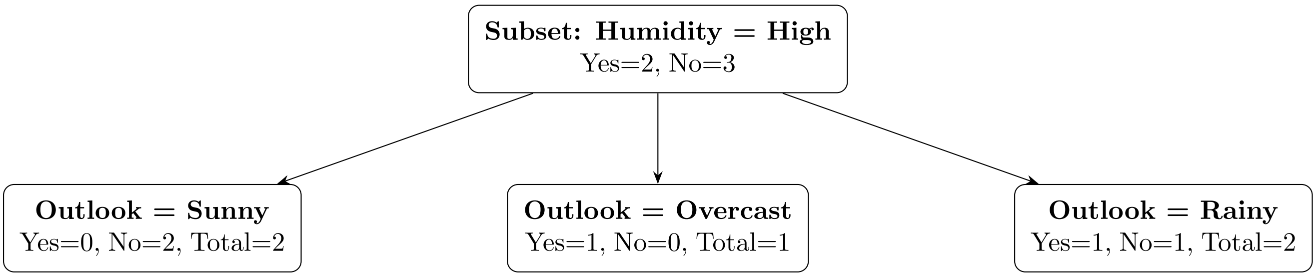
- $IG(Outlook \mid Humidity = High) \approx 0.5710$
- $IG(Wind \mid Humidity = High) \approx 0.4200$

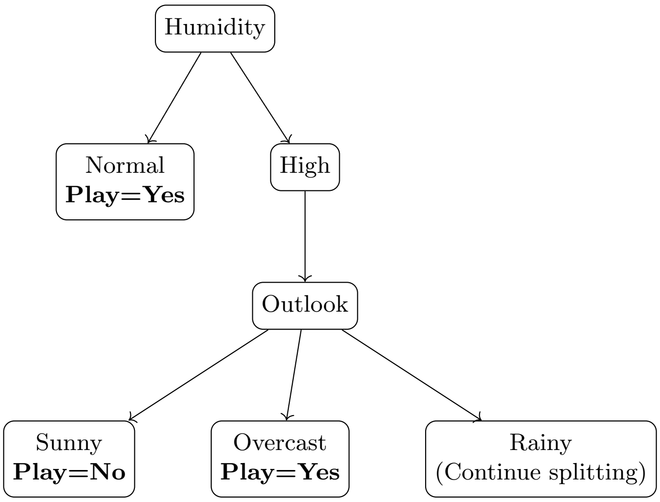
So next split:

$$\boxed{\text{Next Feature} = \text{Outlook (within Humidity=High)}}$$

Visual 3: Split (Humidity = High) on Outlook

Within the subset {Humidity=High} we have: Yes=2, No=3 (Total=5).





Now check each Outlook branch under Humidity=High:

- Outlook = Sunny: (IDs 1,2) ⇒ all No
- Outlook = Overcast: (ID 4) ⇒ Yes
- Outlook = Rainy: (IDs 6,8) ⇒ Yes/No (needs one more split)

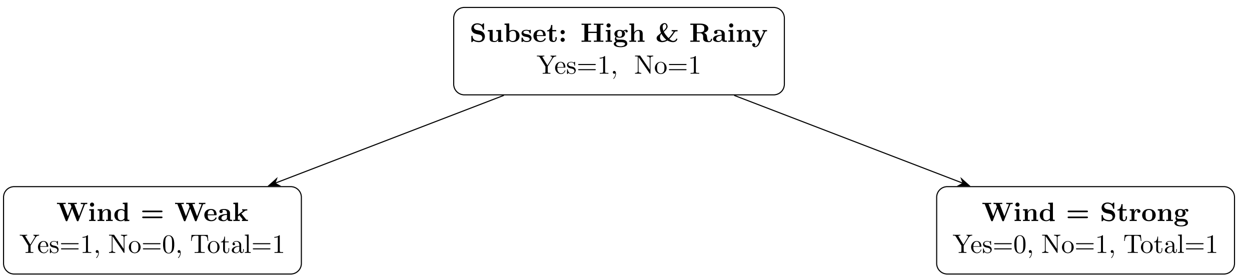
For Outlook = Rainy under Humidity=High:

- Wind = Weak (ID 6) ⇒ Yes
- Wind = Strong (ID 8) ⇒ No

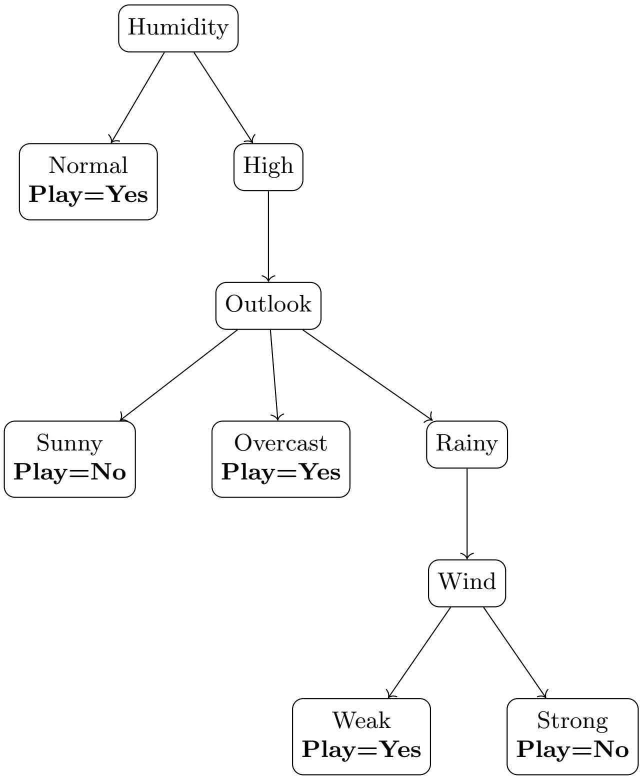
So we split by Wind and get pure leaves.

Visual 4: Split (Humidity=High, Outlook=Rainy) on Wind

For the subset {Humidity=High, Outlook=Rainy}: Yes=1, No=1 (Total=2).



III. Final ID3 Decision Tree (Graph)



Discussion

- **ID3 Selection Rule:** ID3 selects splits that maximize Information Gain, which reduces entropy (uncertainty) the most at each step.
- **Interpretability:** Decision Trees are highly interpretable. The decision-making process is transparent and can be visualized easily.
- **Non-Parametric Nature:** They make no assumptions about linearity or data distribution.
- **Greedy Algorithm:** Tree construction follows a greedy strategy by selecting locally optimal splits, which may not produce a globally optimal tree.
- **Overfitting Tendency:** Deep trees may overfit the training data. Pruning techniques (pre-pruning or post-pruning) are used to reduce overfitting.
- **Computational Complexity:** Training complexity is approximately:

  $$O(nd \log n)$$

  where n is number of samples and d is number of features.
- **Instability:** Small variations in data can result in significantly different tree structures.
- **Extension to Ensembles:** Decision Trees form the base learner for powerful ensemble models such as Random Forest and Gradient Boosting.

Prepared By:
Md. Atikuzzaman
Lecturer
Department of Computer Science and Engineering
Green University of Bangladesh
Email: atik@cse.green.edu.bd