TUTORIAL

A Step by Step ID3 Decision Tree Example

Step 1: Compute the Entropy for data set-Entropy(S)

Step 2: For Every Attribute/Feature:

      1. Calculate entropy for all other values Entropy (A)

      2. Take Average Information Entropy for the current Attribute

      3. Calculate Gain for the current attribute.

Step 3. Pick the highest gain attribute.

Step 4. Repeat until we get the tree we desired.

Make a Decision tree that predicts whether tennis will be played on the day?

Data set

For instance, the following table informs about decision making factors to play tennis at outside for previous 14 days.

Day	Outlook	Temp.	Humidity	Wind	Decision
1	Sunny	Hot	High	Weak	No
2	Sunny	Hot	High	Strong	No
3	Overcast	Hot	High	Weak	Yes
4	Rain	Mild	High	Weak	Yes
5	Rain	Cool	Normal	Weak	Yes
6	Rain	Cool	Normal	Strong	No
7	Overcast	Cool	Normal	Strong	Yes
8	Sunny	Mild	High	Weak	No
9	Sunny	Cool	Normal	Weak	Yes
10	Rain	Mild	Normal	Weak	Yes
11	Sunny	Mild	Normal	Strong	Yes
12	Overcast	Mild	High	Strong	Yes
13	Overcast	Hot	Normal	Weak	Yes
14	Rain	Mild	High	Strong	No

S = [9+, 5-]

OUTLOOK

SUNNY
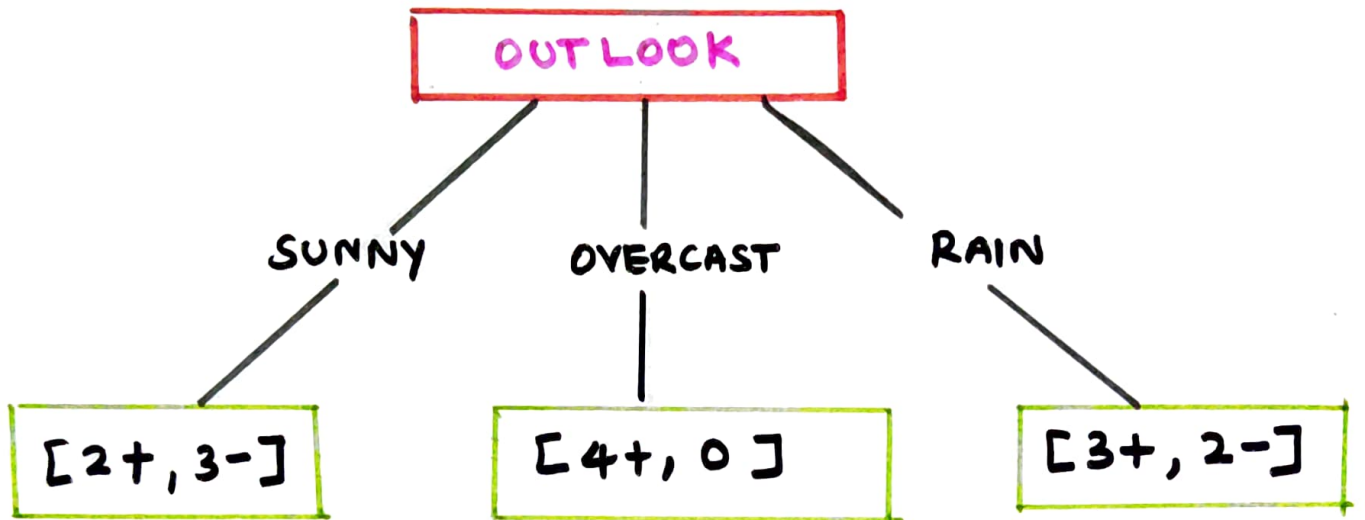
OVERCAST

RAIN

[2+, 3-]

[4+, 0]

[3+, 2-]

$$S = [9+, 5-]$$

```
          OUTLOOK
   /         |         \
 SUNNY    OVERCAST    RAIN
```
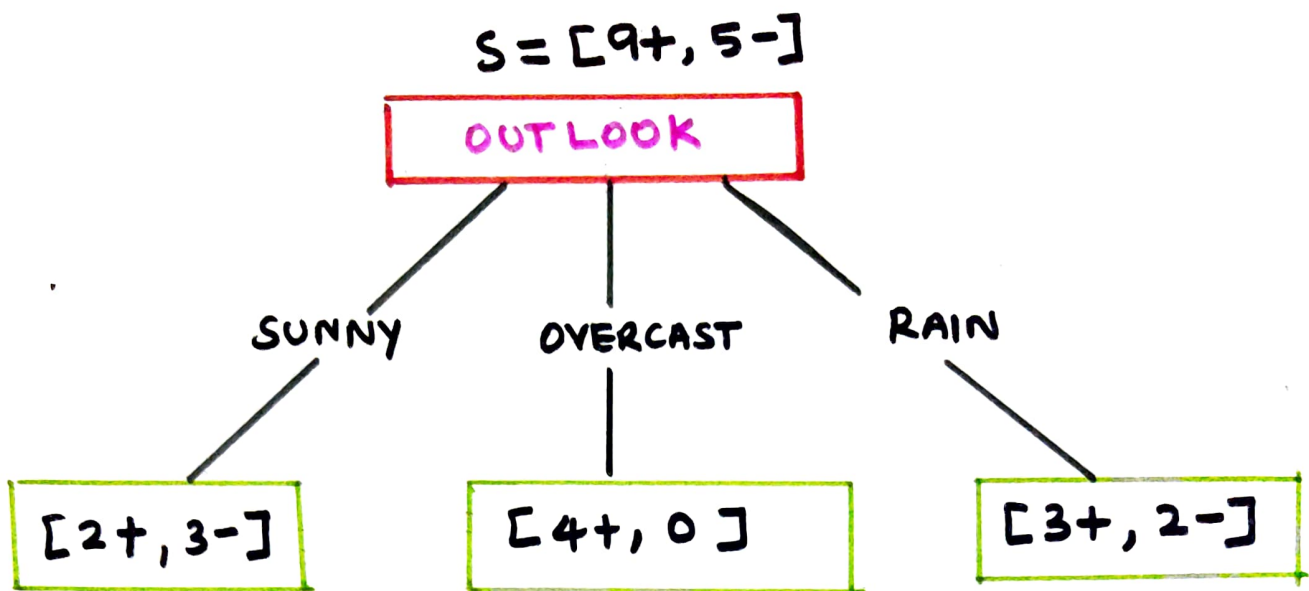
| [2+, 3-] | [4+, 0] | [3+, 2-] |

Attribute : OUTLOOK

Values (OUTLOOK) = Sunny, overcast, Rain

$$S = [9+, 5-]$$

```
            ┌─────────────┐
            │   OUTLOOK   │
            └─────────────┘
          /        |        \
      SUNNY    OVERCAST     RAIN
        /          |          \
  ┌─────────┐ ┌─────────┐ ┌─────────┐
  │ [2+,3-] │ │ [4+,0]  │ │ [3+,2-] │
  └─────────┘ └─────────┘ └─────────┘
```

Attribute : OUTLOOK

Values (OUTLOOK) = Sunny, overcast, Rain

$S \Rightarrow [9+, 5-]$

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

$$= 0.94$$

$S_{SUNNY} \Rightarrow [2+, 3-]$

$$\text{Entropy } (S_{Sunny}) = -\frac{2}{5} \log_2 \frac{2}{5} - \frac{3}{5} \log_2 \frac{3}{5}$$

$$= 0.971$$

$S_{overcast} \Rightarrow [3+, 2-]$

$$\text{Entropy } (S_{overcast}) = -\frac{4}{4} \log_2 \frac{4}{4} - \frac{0}{4} \log_2 \frac{0}{4}$$

$$= 0$$

$S_{Rain} \Rightarrow [3+, 2-]$

$$\text{Entropy } (S_{Rain}) = -\frac{3}{5} \log_2 \frac{3}{5} - \frac{2}{5} \log_2 \frac{2}{5}$$

$$= 0.971$$

$$\text{Gain (S, Outlook)} = \text{Entropy (S)} - \frac{5}{14} \text{Entropy } (S_{sunny})$$

$$- \frac{4}{14} \text{Entropy } (S_{overcast}) - \frac{5}{4} \text{Entropy } (S_{Rain})$$

$$= 0.94 - \frac{5}{14} \times 0.971 - \frac{4}{14} \times 0 - \frac{5}{14} \times 0.971 = 0.2464$$
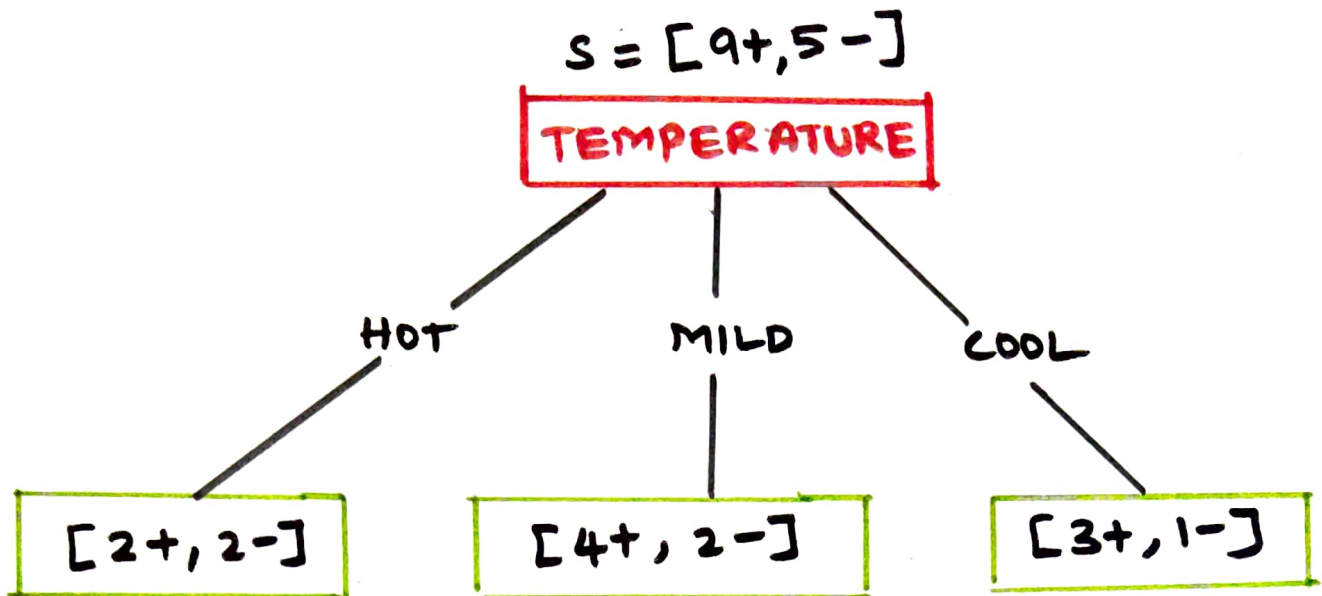
$$S = [9+, 5-]$$

TEMPERATURE

HOT — MILD — COOL

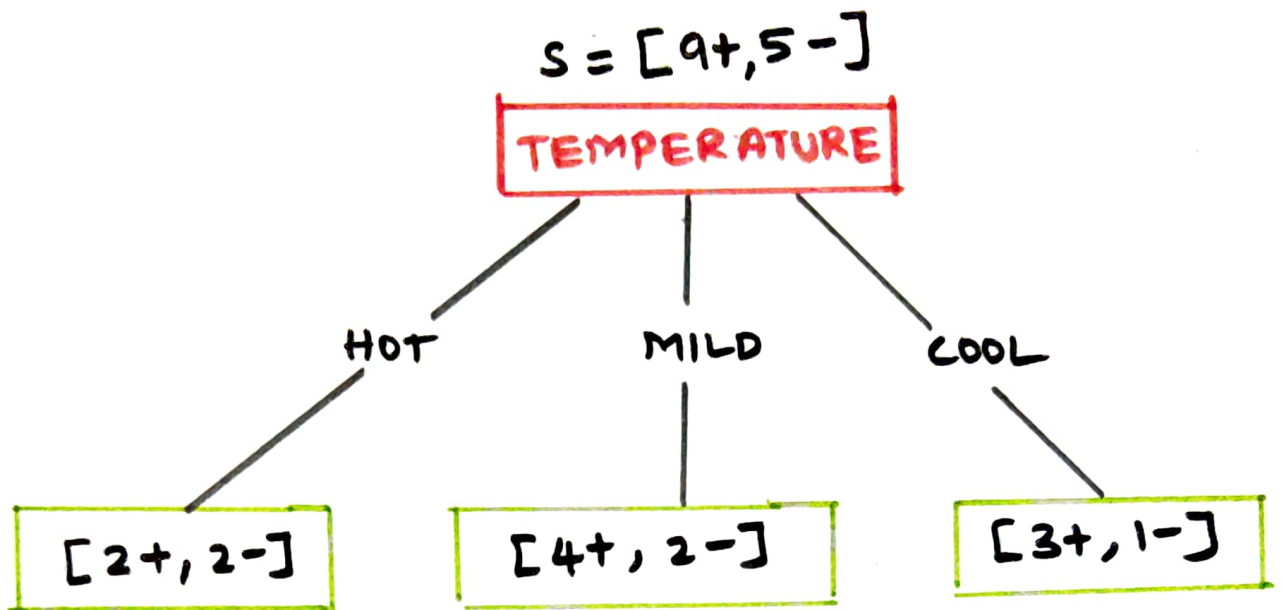[2+, 2-]   [4+, 2-]   [3+, 1-]

$$S = [9+, 5-]$$

TEMPERATURE

HOT — MILD — COOL

[2+, 2-]   [4+, 2-]   [3+, 1-]

Attribute : TEMPERATURE

Values (Temp) = Hot, Mild, Cool

$$S = [9+, 5-]$$

$$\boxed{\text{TEMPERATURE}}$$

HOT   MILD   COOL

$$[2+, 2-] \qquad [4+, 2-] \qquad [3+, 1-]$$

Attribute : TEMPERATURE

Values (Temp) = Hot, Mild, Cool

$S = [9+, 5-]$

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

$$= 0.94$$

$S_{HOT} = [2+, 2-]$

$$\text{Entropy}(S_{HOT}) = -\frac{2}{4} \log_2 \frac{2}{4} - \frac{2}{4} \log_2 \frac{2}{4}$$
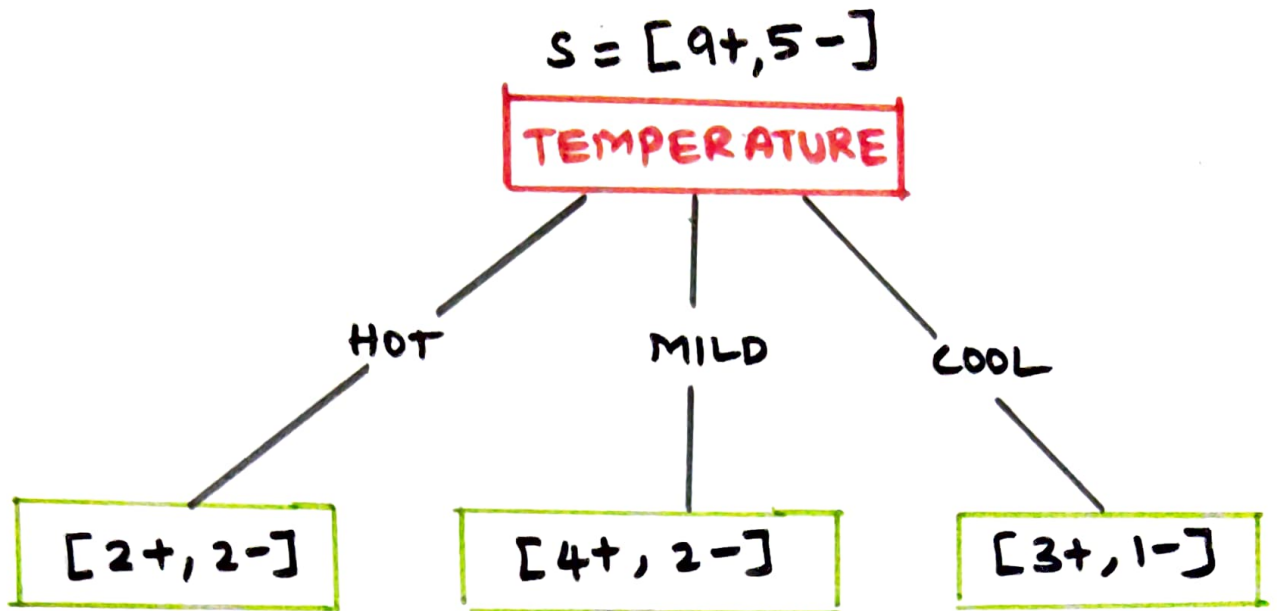
$$= 1.0$$

$S_{mild} = [4+, 2-]$

$$\text{Entropy}(S_{mild}) = -\frac{4}{6} \log_2 \frac{4}{6} - \frac{2}{6} \log_2 \frac{2}{6}$$

$$= 0.9183$$

$S_{cool} = [3+, 1-]$

$$\text{Entropy}(S_{cool}) = -\frac{3}{4} \log_2 \frac{3}{4} - \frac{1}{4} \log_2 \frac{1}{4}$$

$$= 0.8113$$

$$\text{Gain}(S, Temp) = \text{Entropy}(S) - \frac{4}{14} \text{Entropy}(S_{HOT})$$

$$- \frac{6}{14} \text{Entropy}(S_{mild}) - \frac{4}{14} \text{Entropy}(S_{cool})$$

$$\text{Gain}(S, Temp) = 0.94 - \frac{4}{14} \times 1 - \frac{6}{14} \times 0.9183 - \frac{4}{14} \times 0.8113$$

$$= 0.0289$$

$$S = [9+, 5-]$$

$$\boxed{\text{TEMPERATURE}}$$

HOT   MILD   COOL

$$[2+, 2-] \qquad [4+, 2-] \qquad [3+, 1-]$$

Attribute : TEMPERATURE

Values (Temp) = Hot, Mild, Cool

$S = [9+, 5-]$
$$\text{Entropy}(S) = -\frac{9}{14} \log_2 \frac{9}{14} - \frac{5}{14} \log_2 \frac{5}{14} =$$
$$= 0.94$$

$S_{HOT} = [2+, 2-]$
$$\text{Entropy}(S_{Hot}) = -\frac{2}{4} \log_2 \frac{2}{4} - \frac{2}{4} \log_2 \frac{2}{4}$$
$$= 1.0$$

$S_{mild} = [4+, 2-]$
$$\text{Entropy}(S_{mild}) = -\frac{4}{6} \log_2 \frac{4}{6} - \frac{2}{6} \log_2 \frac{2}{6}$$
$$= 0.9183$$
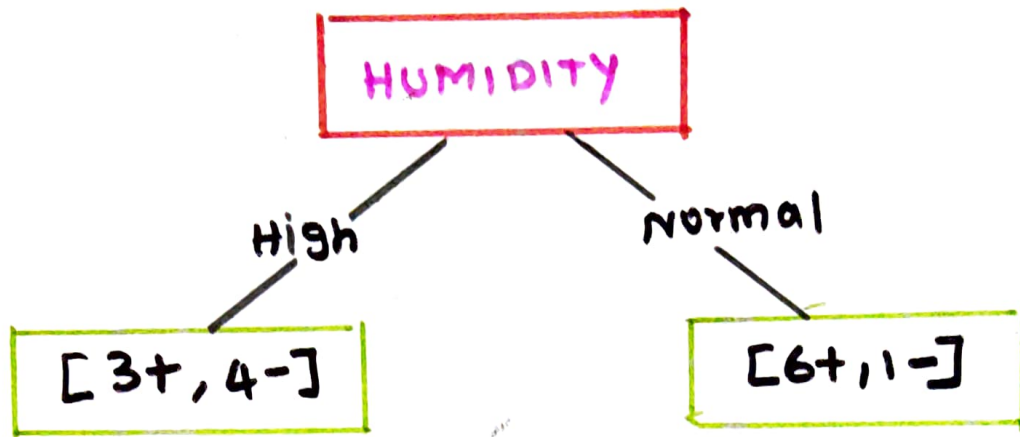
$S_{cool} = [3+, 1-]$
$$\text{Entropy}(S_{cool}) = -\frac{3}{4} \log_2 \frac{3}{4} - \frac{1}{4} \log_2 \frac{1}{4}$$
$$= 0.8113$$

$$\text{Gain}(S, Temp) = \text{Entropy}(S) - \frac{4}{14} \text{Entropy}(S_{Hot})$$
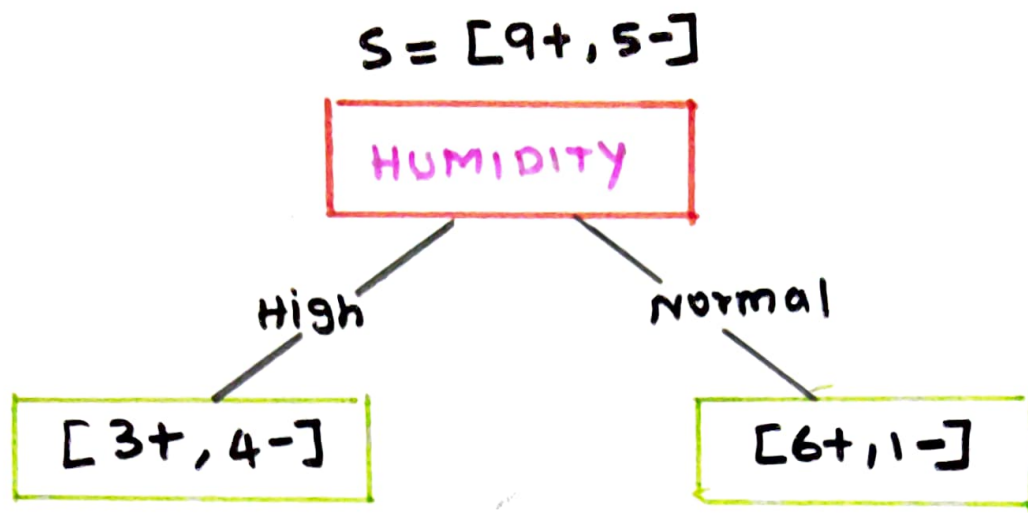$$- \frac{6}{14} \text{Entropy}(S_{Mild}) - \frac{4}{14} \text{Entropy}(S_{cool})$$

$$\text{Gain}(S, Temp) = 0.94 - \frac{4}{14} \times 1 - \frac{6}{14} \times 0.9183 - \frac{4}{14} \times 0.8113$$
$$= 0.0289$$

$$S = [9+, 5-]$$

```
            ┌─────────────┐
            │  HUMIDITY   │
            └─────────────┘
           High         Normal
      ┌──────────┐      ┌──────────┐
      │ [3+, 4-] │      │ [6+, 1-] │
      └──────────┘      └──────────┘
```

Attribute : HUMIDITY

Values (Humidity) = High, Normal)

$$S = [9+, 5-]$$

HUMIDITY

High   Normal

$$[3+, 4-]$$ $$[6+, 1-]$$

Attribute: HUMIDITY

Values (Humidity) = High, Normal)

$S = [9+, 5-]$ Entropy(S) $= -\frac{9}{14} \log_2 \frac{9}{14} - \frac{5}{14} \log_2 \frac{5}{14}$

$$= 0.94$$

$S_{High} = [3+, 4-]$ Entropy(S_{High}) $= -\frac{3}{7} \log_2 \frac{3}{7} - \frac{4}{7} \log_2 \frac{4}{7}$

$$= 0.9852$$

$S_{normal} = [6+, 1-]$ Entropy(S_{normal}) $= -\frac{6}{7} \log_2 \frac{6}{7} - \frac{1}{7} \log_2 \frac{1}{7}$

$$= 0.5916.$$

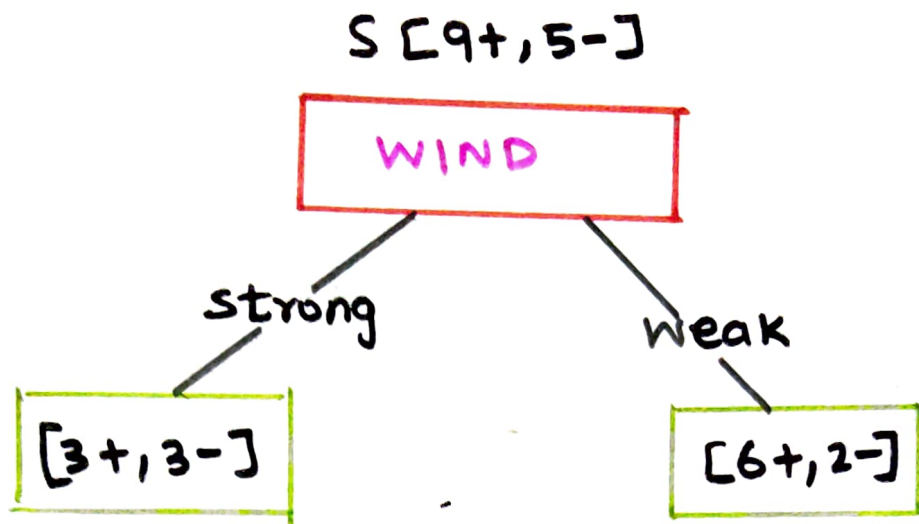Gain(S, Humidity) $=$ Entropy(S) $- \frac{7}{14}$ Entropy(S_{High})

$$- \frac{7}{14} \text{ Entropy}(S_{normal}).$$

$$= 0.94 - \frac{7}{14} \times 0.9852 - \frac{7}{14} \times 0.5916$$

$$= 0.1516.$$
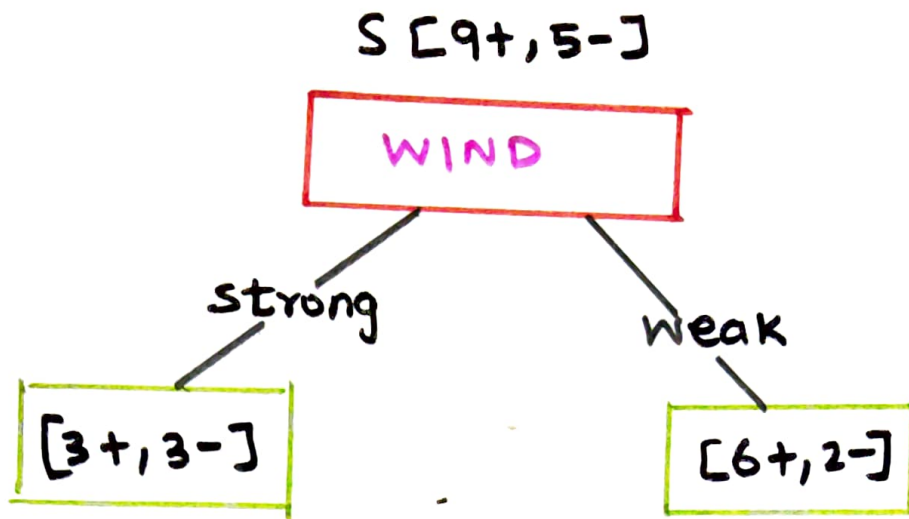
S [9+, 5-]


```
        S [9+,5-]
        ┌─────────┐
        │  WIND   │
        └─────────┘
       strong     weak
    [3+,3-]        [6+,2-]
```

Attribute: wind

Values(wind) = Strong, wind.

$$S\ [9+,5-]$$

WIND

strong — [3+, 3-]

weak — [6+, 2-]

Attribute: Wind

Values(wind) = strong, wind.

$S = [9+, 5-]$

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

$$= 0.94$$

$S_{strong} = [3+, 3-]$

$$\text{Entropy}\ (S_{strong}) = 1.0$$

$S_{weak} = [6+, 2-]$

$$\text{Entropy}\ (S_{weak}) = -\frac{6}{8} \log_2 \frac{6}{8} - \frac{2}{8} \log_2 \frac{2}{8}$$

$$= 0.8113$$

$$\text{Gain}\ (S, wind) = \text{Entropy}\ (S) - \frac{6}{14}\ \text{Entropy}\ (S_{strong})$$

$$- \frac{8}{14}\ \text{Entropy}\ (S_{weak})$$

$$= 0.0478$$

$$\text{Gain}(S, \text{outlook}) = 0.2464$$

$$\text{Gain}(S, \text{Temp}) = 0.0289$$

$$\text{Gain}(S, \text{Humidity}) = 0.1516$$

$$\text{Gain}(S, \text{wind}) = 0.0478$$

Gain(S, Outlook) = 0.2464

Gain(S, Temp) = 0.0289

Gain(S, Humidity) = 0.1516

Gain(S, wind) = 0.0478

Root Node: Outlook.

$\{D1, D2, \ldots D_{14}\}$

$[9+, 5-]$

OUTLOOK

SUNNY

OVERCAST

RAIN

$\{D1, D2, D8, D9, D11\}$

$[2+, 3-]$

YES

$\{D3, D7, D12, D13\}$

$[4+, 0-]$

$\{D4, D5, D6, D10, D14\}$

$[3+, 2-]$