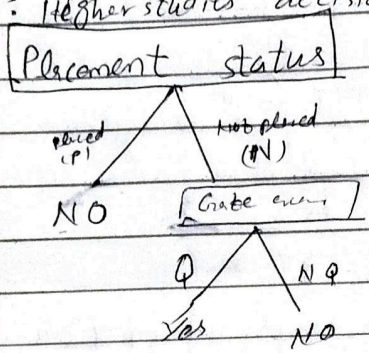
Abdul Raffy
W25-PAK-INP-AI-16

Decision tree (classifier)

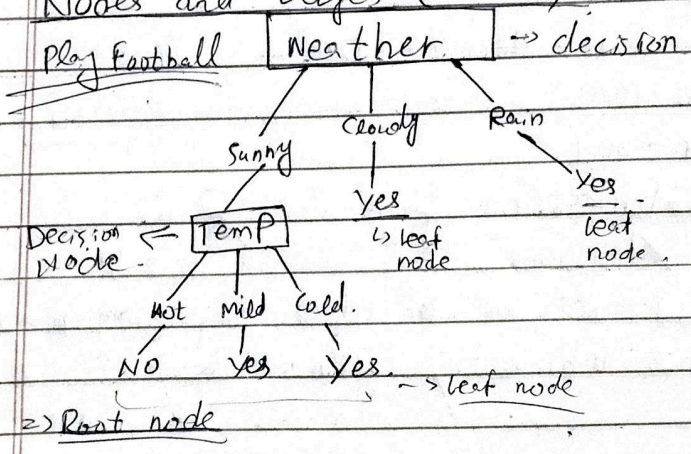
Can be used for both classification and regression but is mainly used for classification.

Email, Spam or Not Spam.

: Higher studies decision for student.



Nodes and Edges (Vertex)



⇒ Root node

Splitting / selection = Entropy / → information Gain.

Data impure or pure.

If data is pure information gain is more, if data is impure information gain is low.

Pruning

Removing unnecessary edges is called pruning.

- Tree structure
- Decision Nodes
- Leaf Nodes
- Splitting
- Entropy and information Gain -
- Pruning.

ID3 Algorithm.

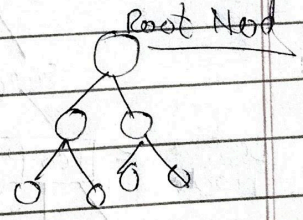
Iterative decomizer.

> Calculate entropy and information gain.

L> For decision tree.

Two steps

- Entropy
- Information gain.



The dataset with the lowest impurity means highest information gain will become root Nod.

Foot ball play classification

Day, weather, Temperature, humidity, wind, Play football

(The Information gain of all attributes will be checked whose is high will be the root node

Calculate Info gain of Weather.

Step 1 : Entropy of entire dataset.

Number of Yes in dataset (+)

$$S\{+9, -5\} = -\frac{9}{14}\log_2\frac{9}{14} - \frac{5}{14}\log_2\frac{5}{14} = 0.94$$

- S: Entropy sign
- 5: Number of NO in datasets
- 14: Number of rows in dataset

Step 2

Entropy of all attributes.

- Entropy of sunny $\{+2, -3\} = -\frac{2}{5}\log_2\frac{2}{5} - \frac{3}{5}\log_2\frac{3}{5} = 0.97$.
  - 2: Number of times when sunny is yes.
  - 3: Number of times when sunny is NO
  - 5: Number of rows of sunny.
- Entropy of Cloudy $\{+4, -0\}$ = " " " "
- Entropy of Rain $\{+3, -2\}$ = " " " "

Entropy

- Information Gain = Entropy (whole data) $*_{-1}$

$$= \frac{5}{14} \times Ent(S) - \frac{4}{14} Ent(C) - \frac{5}{14} Ent(R)$$

- $\frac{5}{14}$: Total Number of rows of sunny / Total number of rows of entire dataset
- Cloudy
- Rain.

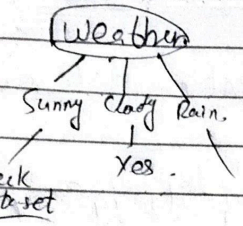
> Calculate Information Gain of other attributes with same process as above

- Gain (S, Weather) = 0.246
- Gain (S, Temp) = 0.029
- Gain (S, humidity) = 0.15
- Gain (S, Wind) = 0.0478

Now that, take attribute with max gain which is

weather.

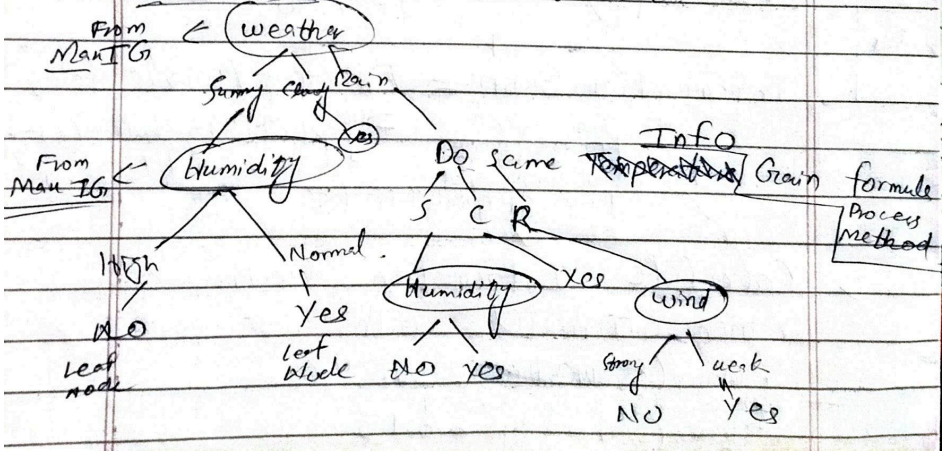
So, we take weather as root node



> Calculate information gain of ~~Sun~~ Temperature according
> Calculate information gain of Humidity to sunny (according to sunny)
> " Ig of Wind accordingly to sunny.

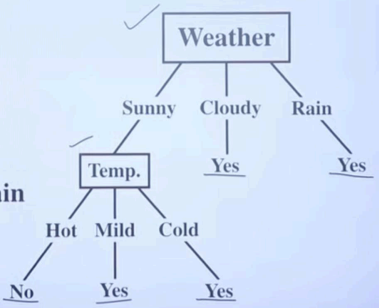
•) Gain (S_{sunny}, Temp) = 0.57.
•) Gain (S_{sunny}, Humidity) = 0.97
•) Gain (S_{sunny}, Wind) = 0.019.

weather was fixed as root node.



Decision Tree

- Machine learning algorithm used for both classification and regression tasks.
- Tree Structure
- Decision Nodes
- Leaf Nodes
- Splitting
- Entropy and Information Gain
- Pruning



Day	Weather	Temperature	Humidity	Wind	Play Football?
Day 1	Sunny	Hot	High	Weak	No
Day 2	Sunny	Hot	High	Strong	No
Day 3	Cloudy	Hot	High	Weak	Yes
Day 4	Rain	Mild	High	Weak	Yes
Day 5	Rain	Cool	Normal	Weak	Yes
Day 6	Rain	Cool	Normal	Strong	No
Day 7	Cloudy	Cool	Normal	Strong	Yes
Day 8	Sunny	Mild	High	Weak	No
Day 9	Sunny	Cool	Normal	Weak	Yes
Day 10	Rain	Mild	Normal	Weak	Yes
Day 11	Sunny	Mild	Normal	Strong	Yes
Day 12	Cloudy	Mild	High	Strong	Yes
Day 13	Cloudy	Hot	Normal	Weak	Yes
Day 14	Rain	Mild	High	Strong	No

Calculate IG of Weather

- **Step1: Entropy of entire dataset**

$$S\{+9,-5\} = -\frac{9}{14}\log_2\frac{9}{14} - \frac{5}{14}\log_2\frac{5}{14} = 0.94$$

- **Step2: Entropy of all attributes:**
- **Entropy of Sunny** $\{+2,-3\} = -\frac{2}{5}\log_2\frac{2}{5} - \frac{3}{5}\log_2\frac{3}{5} = 0.97$
- **Entropy of Cloudy** $\{+4,-0\} = -\frac{4}{4}\log_2\frac{4}{4} - \frac{0}{4}\log_2\frac{0}{4} = 0$
- **Entropy of Rain** $\{+3,-2\} = -\frac{3}{5}\log_2\frac{3}{5} - \frac{2}{5}\log_2\frac{2}{5} = 0.97$
- **Information Gain** $= \text{Entropy(whole data)} - \frac{5}{14}\text{Ent(S)} - \frac{4}{14}\text{Ent(C)} - \frac{5}{14}\text{Ent(R)}$

$= 0.246$

Day	Weather	Temperature	Humidity	Wind	Play Football?
Day 1	Sunny	Hot	High	Weak	No
Day 2	Sunny	Hot	High	Strong	No
Day 8	Sunny	Mild	High	Weak	No
Day 9	Sunny	Cool	Normal	Weak	Yes
Day 11	Sunny	Mild	Normal	Strong	Yes