Decision Tree Classifier

(Introduction, ID3 Algorithm)

Dr. JASMEET SINGH

ASSISTANT PROFESSOR, CSED

TIET, PATIALA

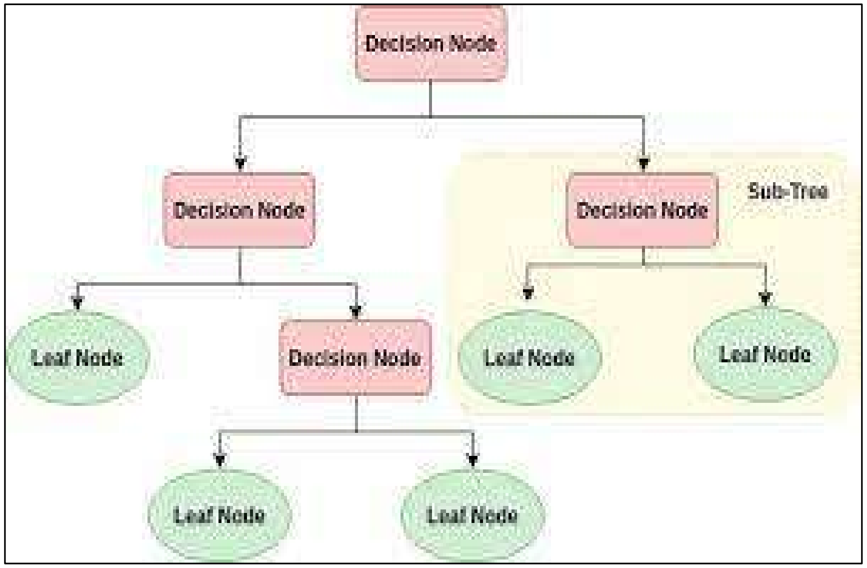
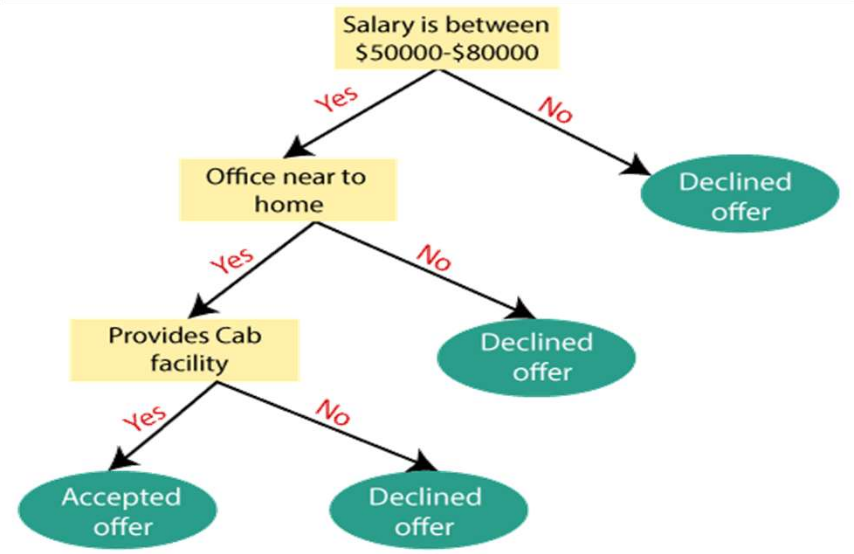
Decision Tree Classifier - Introduction

- Decision Tree Classifier is a supervised learning algorithm
- It is a distribution-free or non-parametric method, which does not depend upon probability distribution assumptions.
- Decision trees are one of the most important concepts in modern machine learning.
- Not only are they an effective approach for classification and regression problems, but they are also the building block for more sophisticated algorithms like random forests and gradient boosting.

Decision Tree Classifier - Introduction

- Decision trees can handle high dimensional data with good accuracy by constructing internal decision-making logic in a form of a decision tree.
- A decision tree is a flowchart-like tree structure where
  - an internal node represents feature(or attribute)-represented as square,
  - the branch represents a decision rule,
  - and each leaf node represents the outcome- represented as ovals.
- Conceptually, decision trees are quite simple. We split a dataset into smaller and smaller groups, attempting to make each one as “pure” or “homogenous” as possible.
- Once we finish splitting, we use the final groups to make predictions on unseen data.

Decision Tree Classifier - Introduction

GENERAL DECISION TREE STRUCTURE	DECISION TREE- THAT ACCEPTS OR REJECTS A JOB OFFER ON THE BASIS OF SALARY, DISTANCE AND CAB FACILITY
	

Variants of Decision Tree Classifier

- A successful decision tree is one that does a good job of "splitting" data into homogeneous groups.
- Therefore, in order to build a good decision tree algorithm, we'll need a method for evaluating splits.
- There are several different algorithm used to generate trees such as,
- **ID3**
  - Uses information gain, to decide the partition feature,
  - not designed to deal with continuous features
- **CART** (Classification and regression trees)
  - Uses Gini coefficient to decide partition feature
- **C4.5**
  - Works similar to ID3 by using information gain to split data.
  - However C4.5 can handle continuous features, as well as can work with missing data

ID3 (Iterative Dichotomiser 3) Algorithm

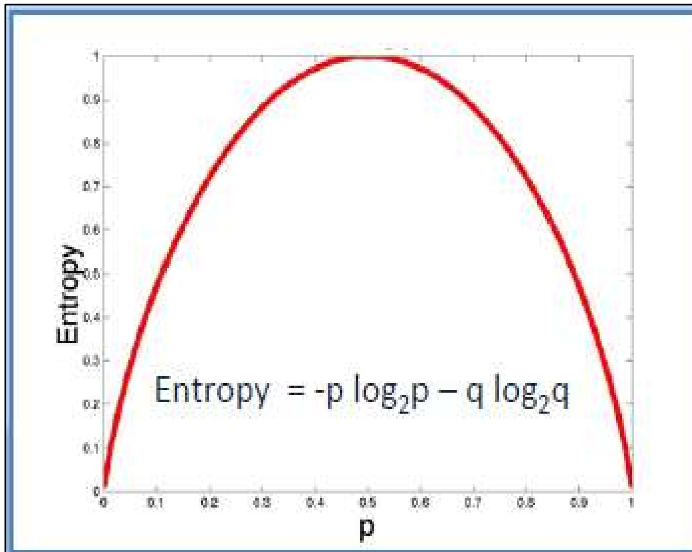
- ID3 stands for ***Iterative Dichotomiser 3*** and is named such because the algorithm iteratively (repeatedly) dichotomizes(divides) features into two or more groups at each step.
- Invented by Ross Quinlan, ID3 uses a **top-down greedy** approach to build a decision tree.
- In simple words, the **top-down** approach means that we start building the tree from the top and the **greedy** approach means that at each iteration we select the best feature at the present moment to create a node.

ID3 (Iterative Dichotomiser 3) Algorithm

- ID3 generates a tree by considering the whole set S as the root node.
- It then iterates on every attribute and splits the data into fragments known as subsets to calculate the entropy or the information gain of that attribute.
- After splitting, the algorithm recourses on every subset by taking those attributes which were not taken before into the iterated ones.
- The best feature in ID3 is selected using *Entropy and Information Gain* metrics.

Entropy

- It is defined as a measure of impurity present in the data.
- Entropy calculates the homogeneity of a sample.
- If the sample is completely homogeneous the entropy is zero and if the sample is equally divided it has entropy of one (as shown in figure for binary classes).
- Entropy with the lowest value makes a model better in terms of prediction as it segregates the classes better.



$$\text{Entropy} = -0.5\log_2 0.5 - 0.5\log_2 0.5 = 1$$

Entropy (Contd…..)

- Entropy of dataset (S) is computed as follows:

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

Where n is the total number of classes in the target column (in our case n = 2 i.e YES and NO)

p_i is the **probability of class 'i'** or the ratio of "*number of rows with class i in the target column*" to the "*total number of rows*" in the dataset.

Information Gain

- In a decision tree building process, two important decisions are to be made
  - what is the best split(s)
  - and which is the best variable to split a node.
- Information Gain criteria helps in making these decisions.
- We need to calculate Entropy of Parent and Child Nodes for calculating the information gain due to the split.
- The concept of Information Gain is based on:

The more we know about a topic, the less new information you are apt to get about it. To be more concise: If you know an event is very probable, it is no surprise when it happens, that is, it gives us little information that it actually happened.

- The amount of information gained is inversely proportional to the probability of an event happening.
- We can also say that as the Entropy increases the information gain decreases. This is because Entropy refers to the probability of an event.

Information Gain (Contd…)

- Information Gain = Entropy of Parent – sum (weighted % * Entropy of Child)

Weighted % = Number of observations in particular child/sum (observations in all child nodes)

- In particular, Information Gain for a feature column **A** is calculated as:

$$Information\ Gain(S, A) = Entropy(S) - \sum_{v=1}^{|v|} \frac{|S_v|}{|S|} Entropy(S_v)$$

where $\mathbf{S_v}$ is the set of rows in **S** for which the feature column **A** has value **v**, $|\mathbf{S_v}|$ is the number of rows in $\mathbf{S_v}$ and likewise |**S**| is the number of rows in **S.**

- **Information Gain calculates the reduction in the entropy and measures how well a given feature separates or classifies the target classes.**
- **The feature with the highest Information Gain is selected as the best one.**

ID3 Algorithm-Pseudocode

- **ID3**(*instances*, *target_attribute*, *attributes*)
  - Create a new *root* node to the tree.
  - **If** all instances have the target_attribute belonging to the same class *c*,
    - **Return** the tree with single *root* node with label *c*.
  - **If** *attributes* is empty, then
    - **Return** the tree with single root node with the most common label of the *target_attribute* in *instances*.
  - **Else**
    - A ← the attribute in *attributes* which best classifies *instances*
    - root decision attribute ← A
    - **Foreach** possible value v_i **of** A,
      - Add a new ramification below root, corresponding to the test A = v_i
      - Let $instances_{vi}$ be the subset of instances with the value v_i for A
      - **If** $instances_{vi}$ is empty then
        - Below this ramification, add a new leaf node with the most common value of *target_attribute* in *instances*.
      - **Else** below this ramification, add the subtree given by the recursion:
        ID3($instances_{vi}$, *target_attribute*, *attributes* – { A })
  - **End**

Numerical Example 1

Consider the **weather dataset** in which we have to decide that whether the player should play golf or not on the basis of weather conditions (shown in figure).

Train a decision tree classifier (using ID3 algorithm) that classifies any new test case according to given weather conditions.

S. No.	Outlook	Temperature	Humidity	Windy	PlayTennis
1	Sunny	Hot	High	Weak	No
2	Sunny	Hot	High	Strong	No
3	Overcast	Hot	High	Weak	Yes
4	Rainy	Mild	High	Weak	Yes
5	Rainy	Cool	Normal	Weak	Yes
6	Rainy	Cool	Normal	Strong	No
7	Overcast	Cool	Normal	Strong	Yes
8	Sunny	Mild	High	Weak	No
9	Sunny	Cool	Normal	Weak	Yes
10	Rainy	Mild	Normal	Weak	Yes
11	Sunny	Mild	Normal	Strong	Yes
12	Overcast	Mild	High	Strong	Yes
13	Overcast	Hot	Normal	Weak	Yes
14	Rainy	Mild	High	Strong	No

Solution- Example 1

Compute Entropy of the entire dataset:

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

Positive Examples = 9

Negative Examples =5

Total= 14

$$Entropy(S) = -\frac{9}{9+5} log_2\left(\frac{9}{9+5}\right) - \frac{5}{9+5} log_2\left(\frac{5}{9+5}\right) = 0.940$$

Solution- Example 1 (Contd…)

- For each attribute: (let say Outlook)
  - Calculate Entropy of each values, i.e., 'Sunny', 'Rainy', 'Overcast'

$$Entropy(Outlook\ = Sunny) = -\frac{2}{2+3}log_2\left(\frac{2}{2+3}\right) - \frac{3}{2+3}log_2\left(\frac{3}{2+3}\right) = 0.971$$

$$Entropy(Outlook\ = Rainy) = -\frac{3}{2+3}log_2\left(\frac{3}{2+3}\right) - \frac{2}{2+3}log_2\left(\frac{2}{2+3}\right) = 0.971$$

$$Entropy(Outlook\ = Overcast) = -\frac{4}{4+0}log_2\left(\frac{4}{4+0}\right) - \frac{0}{4+0}log_2\left(\frac{0}{4+0}\right) = 0$$

$$Average\ Information\ Entropy = \sum_{v=1}^{|v|}\frac{|S_v|}{|S|}Entropy(S_v)$$

$$Average\ Information\ Entropy =$$

$$I(S, Outlook) = \frac{2+3}{9+5}\times 0.971 + \frac{3+2}{9+5}\times 0.971 + \frac{4+0}{9+5}\times 0 = 0.693$$

$$\boldsymbol{Information\ Gain(S, Outlook) = Entropy(S) - \mathrm{I}(S, Outlook)}$$

$$\boldsymbol{= 0.940 - 0.693 = 0.247}$$

Outlook	PlayTennis
Sunny	No
Sunny	No
Sunny	No
Sunny	Yes
Sunny	Yes

Outlook	PlayTennis
Rainy	Yes
Rainy	Yes
Rainy	No
Rainy	Yes
Rainy	No

Outlook	PlayTennis
Overcast	Yes
Overcast	Yes
Overcast	Yes
Overcast	Yes

Outlook	p	n	Entropy
Sunny	2	3	0.971
Rainy	3	2	0.971
Overcast	4	0	0

Solution- Example 1 (Contd...)

- For each attribute: (let say Temperature)
  - Calculate Entropy of each values, i.e., 'Hot', 'Mild', 'Cool'

$$Entropy(Temp = Hot) = -\frac{2}{2+2}log_2\left(\frac{2}{2+2}\right) - \frac{2}{2+2}log_2\left(\frac{2}{2+2}\right) = 0$$

$$Entropy(Temp = Mild) = -\frac{4}{4+2}log_2\left(\frac{4}{4+2}\right) - \frac{2}{4+2}log_2\left(\frac{2}{4+2}\right) = 0.918$$

$$Entropy(Temp\ = Cool) = -\frac{3}{3+1}log_2\left(\frac{3}{3+1}\right) - \frac{1}{3+1}log_2\left(\frac{1}{3+1}\right) = 0.811$$

$$Average\ Information\ Entropy = \sum_{v=1}^{|v|}\frac{|S_v|}{|S|}Entropy(S_v)$$

$$Average\ Information\ Entropy =$$

$$I(S,Temp) = \frac{2+2}{9+5}\times 1 + \frac{4+2}{9+5}\times 0.918 + \frac{3+1}{9+5}\times 0.811 = 0.911$$

$$\boldsymbol{Information\ Gain(S,Temp) = Entropy(S) - \mathrm{I}(S,Temp)}$$

$$\mathbf{= 0.940 - 0.911 = 0.029}$$

Temperature	PlayTennis
Hot	No
Hot	No
Hot	Yes
Hot	Yes

Temperature	PlayTennis
Mild	Yes
Mild	No
Mild	Yes
Mild	Yes
Mild	Yes
Mild	No

Temperature	PlayTennis
Cool	Yes
Cool	No
Cool	Yes
Cool	Yes

Temperature	p	n	Entropy
Hot	2	2	1
Mild	4	2	0.918
Cool	3	1	0.811

Solution- Example 1 (Contd…)

- For each attribute: (let say Humidity)
  - Calculate Entropy of each values, i.e., ‘High’, ‘Normal’,

$$Entropy(Humidity = High) = -\frac{3}{3+4}log_2\left(\frac{3}{3+4}\right) - \frac{4}{3+4}log_2\left(\frac{4}{3+4}\right) = 0.985$$

$$Entropy(Hunidity = Normal) = -\frac{6}{6+1}log_2\left(\frac{6}{6+1}\right) - \frac{1}{6+1}log_2\left(\frac{1}{6+1}\right) = 0.591$$

$$Average\ Information\ Entropy = \sum_{v=1}^{|v|}\frac{|S_v|}{|S|}Entropy(S_v)$$

$$Average\ Information\ Entropy =$$

$$I(S, Humidity) = \frac{3+4}{9+5} \times 0.985 + \frac{6+1}{9+5} \times 0.591 = 0.788$$

$$\boldsymbol{Information\ Gain(S, Humdity)}$$
$$\boldsymbol{= Entropy(S) - I(S, Humidity)}$$
$$\boldsymbol{= 0.940 - 0.788 = 0.152}$$

Humidity	PlayTennis
Normal	Yes
Normal	No
Normal	Yes
Normal	Yes
Normal	Yes
Normal	Yes
Normal	Yes

Humidity	PlayTennis
High	No
High	No
High	Yes
High	Yes
High	No
High	Yes
High	No

Humidity	p	n	Entropy
High	3	4	0.985
Normal	6	1	0.591

Solution- Example 1 (Contd...)

- For each attribute: (let say Windy)
  - Calculate Entropy of each values, i.e., 'Strong', 'Weak',

$$Entropy(Windy = Strong) = -\frac{3}{3+3}log_2\left(\frac{3}{3+3}\right) - \frac{3}{3+3}log_2\left(\frac{3}{3+3}\right) = 1$$

$$Entropy(Windy = Weak) = -\frac{6}{6+2}log_2\left(\frac{6}{6+2}\right) - \frac{2}{6+1}log_2\left(\frac{2}{6+2}\right) = 0.811$$

$$Average\ Information\ Entropy = \sum_{v=1}^{|v|}\frac{|S_v|}{|S|}Entropy(S_v)$$

$$Average\ Information\ Entropy =$$

$$I(S, Windy) = \frac{3+3}{9+5}\times 1 + \frac{6+2}{9+5}\times 0.811 = 0.892$$

$$\boldsymbol{Information\ Gain(S, Windy) = Entropy(S) - \mathrm{I}(S, Windy)}$$

$$\boldsymbol{= 0.940 - 0.892 = 0.048}$$

Since Information Gain for Outlook is max, therefore root is *Outlook*

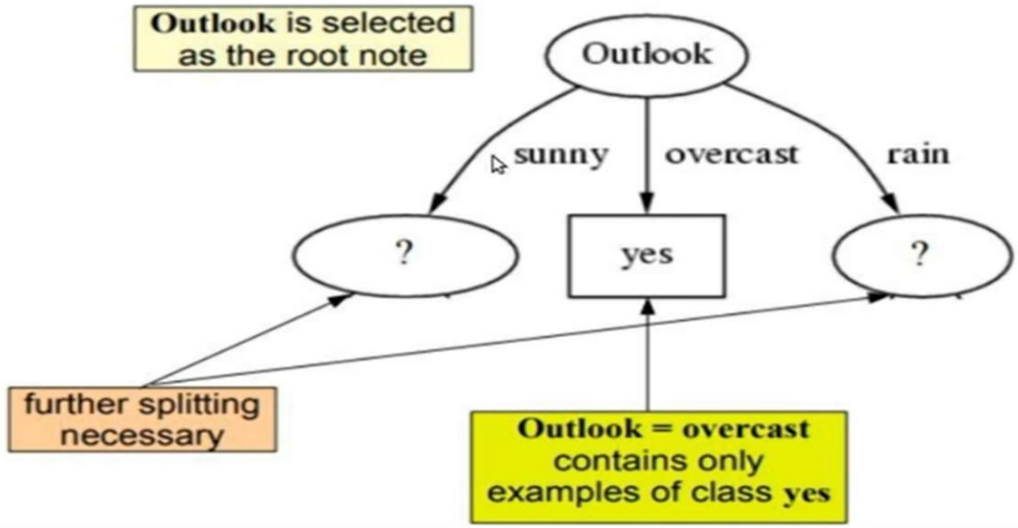
Windy	PlayTennis
Weak	No
Weak	Yes
Weak	Yes
Weak	Yes
Weak	No
Weak	Yes
Weak	Yes
Weak	Yes

Windy	PlayTennis
Strong	No
Strong	No
Strong	Yes
Strong	Yes
Strong	Yes
Strong	No

Windy	p	n	Entropy
Strong	3	3	1
Weak	6	2	0.811

Solution- Example 1 (Contd…)

Outlook	Temperature	Humidity	Windy	PlayTennis
Overcast	Hot	High	Weak	Yes
Overcast	Cool	Normal	Strong	Yes
Overcast	Mild	High	Strong	Yes
Overcast	Hot	Normal	Weak	Yes



Solution- Example 1 (Contd...)

- Repeat the same thing for sub-trees till we get the tree.

Outlook	Temperature	Humidity	Windy	PlayTennis
Sunny	Hot	High	Weak	No
Sunny	Hot	High	Strong	No
Sunny	Mild	High	Weak	No
Sunny	Cool	Normal	Weak	Yes
Sunny	Mild	Normal	Strong	Yes

Outlook = "Sunny"

Outlook	Temperature	Humidity	Windy	PlayTennis
Rainy	Mild	High	Weak	Yes
Rainy	Cool	Normal	Weak	Yes
Rainy	Cool	Normal	Strong	No
Rainy	Mild	Normal	Weak	Yes
Rainy	Mild	High	Strong	No

Outlook = "Rainy"

Solution- Example 1 (Contd...)

Outlook	Temperature	Humidity	Windy	PlayTennis
Sunny	Hot	High	Weak	No
Sunny	Hot	High	Strong	No
Sunny	Mild	High	Weak	No
Sunny	Cool	Normal	Weak	Yes
Sunny	Mild	Normal	Strong	Yes

P= 2 N= 3
Total= 5

Compute Entropy of the Sunny:

$$Entropy\left(S_{sunny}\right) = -\sum_{i=1}^{n} p_i log_2(p_i)$$

$$Entropy(S_{sunny}) = -\frac{2}{2+3} log_2\left(\frac{2}{2+3}\right) - \frac{3}{2+3} log_2\left(\frac{3}{2+3}\right) = 0.971$$

Solution- Example 1 (Contd…)

- For each attribute: (let say Temperatre)
  - Calculate Entropy of each values, i.e., ‘Hot’, ‘Mild’, ‘Cool’

$$Entropy(Temp = Hot) = -\frac{0}{0+2} log_2\left(\frac{0}{0+2}\right) - \frac{2}{2+0} log_2\left(\frac{2}{2+0}\right) = 0$$

$$Entropy(Temp = Mild) = -\frac{1}{1+1} log_2\left(\frac{1}{1+1}\right) - \frac{1}{1+1} log_2\left(\frac{1}{1+1}\right) = 1$$

$$Entropy(Temp = Cool) = -\frac{1}{1+0} log_2\left(\frac{1}{1+0}\right) - \frac{0}{1+0} log_2\left(\frac{0}{1+0}\right) = 0$$

$$Average\ Information\ Entropy = \sum_{v=1}^{|v|} \frac{|S_v|}{|S|} Entropy(S_v)$$

$$Average\ Information\ Entropy =$$

$$I(S_{sunny}, Temperature) = \frac{0+2}{2+3} \times 0 + \frac{1+1}{2+3} \times 1 + \frac{1+0}{2+3} \times 0 = 0.4$$

$$\boldsymbol{Information\ Gain(S_{sunny}, Temp)}$$

$$\boldsymbol{= Entropy(S_{sunny}) - I(S_{sunny}, Temp)}$$

$$\boldsymbol{= 0.971 - 0.4 = 0.571}$$

Outlook	Temperature	PlayTennis
Sunny	Cool	Yes
Sunny	Hot	No
Sunny	Hot	No
Sunny	Mild	No
Sunny	Mild	Yes

Temperature	p	n	Entropy
Cool	1	0	0
Hot	0	2	0
Mild	1	1	1

Solution- Example 1 (Contd…)

- For each attribute: (let say Humidity)
  - Calculate Entropy of each values, i.e., ‘High’, ‘Normal’

$$Entropy(Hmidity = High) = -\frac{0}{0+3}log_2\left(\frac{0}{0+3}\right) - \frac{3}{3+0}log_2\left(\frac{3}{3+0}\right) = 0$$

$$Entropy(Humidity = Normal) = -\frac{2}{2+0}log_2\left(\frac{2}{2+0}\right) - \frac{0}{0+2}log_2\left(\frac{0}{0+2}\right) = 0$$

$$Average\ Information\ Entropy = \sum_{v=1}^{|v|}\frac{|S_v|}{|S|}Entropy(S_v)$$

$$Average\ Information\ Entropy =$$

$$I(S_{sunny}, Humidity) = \frac{0+3}{2+3}\times 0 + \frac{2+0}{2+3}\times 0 = 0$$

$$\boldsymbol{Information\ Gain}(S_{sunny}, Humidity)$$

$$= \boldsymbol{Entropy}(S_{sunny}) - \mathbf{I}(S_{sunny}, Humidity)$$

$$= \mathbf{0.971 - 0 = 0.971}$$

Outlook	Humidity	PlayTennis
Sunny	High	No
Sunny	High	No
Sunny	High	No
Sunny	Normal	Yes
Sunny	Normal	Yes

Humidity	p	n	Entropy
high	0	3	0
normal	2	0	0

Solution- Example 1 (Contd…)

- For each attribute: (let say Windy)
  - Calculate Entropy of each values, i.e., ‘Strong’, ‘False’

$$EntropyWindy = Strong) = -\frac{1}{1+1}log_2\left(\frac{1}{1+1}\right) - \frac{1}{1+1}log_2\left(\frac{1}{1+1}\right) = 1$$

$$Entropy(Windy = Weak) = -\frac{1}{1+2}log_2\left(\frac{1}{1+2}\right) - \frac{2}{1+2}log_2\left(\frac{2}{1+2}\right) = 0.918$$

Outlook	Windy	PlayTennis
Sunny	Strong	No
Sunny	Strong	Yes
Sunny	Weak	No
Sunny	Weak	No
Sunny	Weak	Yes

Windy	p	n	Entropy
Strong	1	1	1
Weak	1	2	0.918

$$Average\ Information\ Entropy = \sum_{v=1}^{|v|}\frac{|S_v|}{|S|}Entropy(S_v)$$

$$Average\ Information\ Entropy =$$

$$I(S_{sunny}, Windy) = \frac{1+1}{2+3}\times 1 + \frac{1+2}{2+3}\times .918 = 0.951$$

$$\boldsymbol{Information\ Gain(S_{sunny}, Windy)}$$

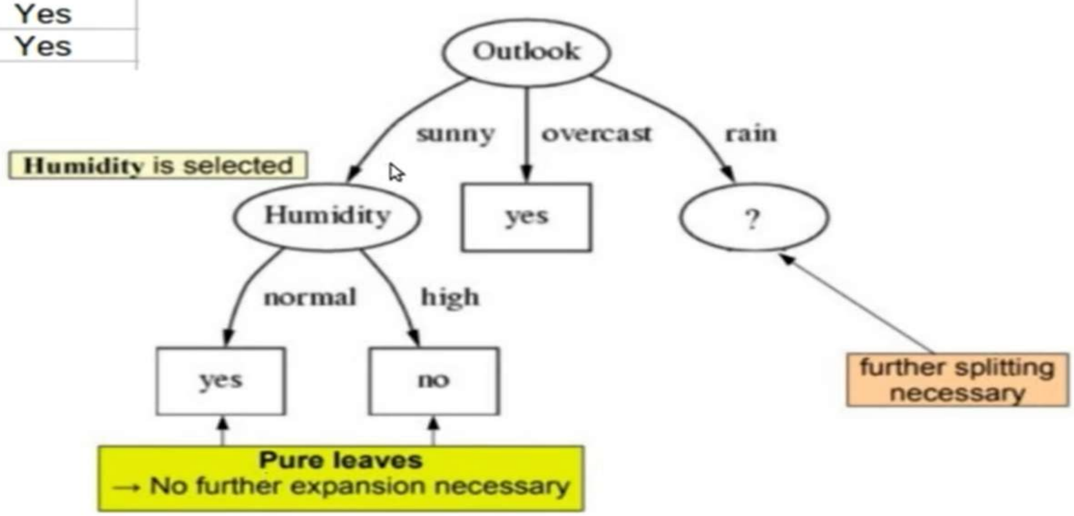
$$= \boldsymbol{Entropy(S_{sunny}) - \mathrm{I}(S_{sunny}, Windy)}$$

$$= \mathbf{0.971 - 0.951 = 0.020}$$

Since Information Gain for root Sunny is maximum for Humidity,
So, the node for Split under Outlook=Sunny is Humidity

Solution- Example 1 (Contd...)

Outlook	Humidity	PlayTennis
Sunny	High	No
Sunny	High	No
Sunny	High	No
Sunny	Normal	Yes
Sunny	Normal	Yes



Solution- Example 1 (Contd…)

Outlook	Temperature	Humidity	Windy	PlayTennis
Rainy	Mild	High	Weak	Yes
Rainy	Cool	Normal	Weak	Yes
Rainy	Cool	Normal	Strong	No
Rainy	Mild	Normal	Weak	Yes
Rainy	Mild	High	Strong	No

P= 3 N= 2

Total = 5

Compute Entropy of the Outlook=Rainy:

$$Entropy\left(S_{rainy}\right) = -\sum_{i=1}^{n} p_i log_2(p_i)$$

$$Entropy\left(S_{rainy}\right) = -\frac{3}{2+3} log_2\left(\frac{3}{2+3}\right) - \frac{2}{2+3} log_2\left(\frac{2}{2+3}\right) = 0.971$$

Solution- Example 1 (Contd...)

- For each attribute: (let say Temperature)
  - Calculate Entropy of each values, i.e., 'Mild', 'Cool'

$$Entropy(Temp = Mild) = -\frac{2}{2+1}log_2\left(\frac{2}{2+1}\right) - \frac{1}{2+1}log_2\left(\frac{1}{2+1}\right) = 0.918$$

$$Entropy(Temp = Cool) = -\frac{1}{1+1}log_2\left(\frac{1}{1+1}\right) - \frac{1}{1+1}log_2\left(\frac{1}{1+1}\right) = 1$$

$$Average\ Information\ Entropy = \sum_{v=1}^{|v|}\frac{|S_v|}{|S|}Entropy(S_v)$$

$$Average\ Information\ Entropy =$$

$$I(S_{rainy}, Temperature) = \frac{2+1}{3+2}\times 0.918 + \frac{1+1}{3+2}\times 1 = 0.951$$

$$\boldsymbol{Information\ Gain(S_{rainy}\ Temp)}$$

$$\boldsymbol{= Entropy(S_{rainy}) - I(S_{rainy}, Temp)}$$

$$\boldsymbol{= 0.971 - 0.951 = 0.020}$$

Outlook	Temperature	PlayTennis
Rainy	Mild	Yes
Rainy	Cool	Yes
Rainy	Cool	No
Rainy	Mild	Yes
Rainy	Mild	No

Attribute	p	n	Entropy
Cool	1	1	1
Mild	2	1	0.918

Solution- Example 1 (Contd...)

- For each attribute: (let say Windy)
  - Calculate Entropy of each values, i.e., 'Strong', 'Weak'

$$Entropy(Windy = Strong) = -\frac{0}{0+2} log_2\left(\frac{0}{0+2}\right) - \frac{2}{0+2} log_2\left(\frac{2}{0+2}\right) = 0$$

$$Entropy(Windy = Weak) = -\frac{3}{3+0} log_2\left(\frac{3}{3+0}\right) - \frac{0}{0+3} log_2\left(\frac{0}{0+3}\right) = 0$$

$$Average\ Information\ Entropy = \sum_{v=1}^{|v|} \frac{|S_v|}{|S|} Entropy(S_v)$$

$$Average\ Information\ Entropy =$$

$$I(S_{rainy}, Windy) = \frac{0+2}{3+2} \times 0 + \frac{3+0}{3+2} \times 0 = 0$$

$$\boldsymbol{Information\ Gain}(S_{rainy}, Windy)$$
$$= \boldsymbol{Entropy}(S_{rainy}) - \mathbf{I}(S_{rainy}, \boldsymbol{Windy})$$
$$= \mathbf{0.971 - 0 = 0.971}$$

Outlook	Windy	PlayTennis
Rainy	Strong	No
Rainy	Strong	No
Rainy	Weak	Yes
Rainy	Weak	Yes
Rainy	Weak	Yes

Attribute	p	n	Entropy
Strong	0	2	0
Weak	3	0	0

Solution- Example 1 (Contd…)

- For each attribute: (let say Humidity)
  - Calculate Entropy of each values, i.e., ‘High’, Normal’

$$Entropy(Humidity = High) = -\frac{1}{1+1} log_2\left(\frac{1}{1+1}\right) - \frac{1}{1+1} log_2\left(\frac{1}{1+1}\right) = 1$$

$$Entropy(Humidity = Normal) = -\frac{2}{2+1} log_2\left(\frac{2}{2+1}\right) - \frac{1}{2+1} log_2\left(\frac{1}{1+2}\right) = 0.918$$

$$Average\ Information\ Entropy = \sum_{v=1}^{|v|} \frac{|S_v|}{|S|} Entropy(S_v)$$

$$Average\ Information\ Entropy =$$

$$I(S_{rainy}, Humidity) = \frac{1+1}{3+2} \times 1 + \frac{2+1}{3+2} \times 0.918 = 0.951$$

$$\boldsymbol{Information\ Gain(S_{rainy}, Humidity)}$$
$$\boldsymbol{= Entropy(S_{rainy}) - \mathrm{I}(S_{rainy}, Humidity)}$$
$$\boldsymbol{= 0.971 - .951} = \mathbf{0.020}$$

Outlook	Humidity	PlayTennis
Rainy	High	Yes
Rainy	High	No
Rainy	Normal	Yes
Rainy	Normal	No
Rainy	Normal	Yes

Attribute	p	n	Entropy
High	1	1	1
Normal	2	1	0.918

Since Information Gain for root Outlook=Rainy is maximum for Windy, So, the node for Split under Outlook=Runny is Windy

Solution- Example 1 (Contd...)

Outlook	Windy	PlayTennis
Rainy	Weak	Yes
Rainy	Weak	Yes
Rainy	Strong	No
Rainy	Weak	Yes
Rainy	Strong	No