Decision Tree Algorithm

(Part-I)

BY:

DR. VIPIN KUMAR

DEPARTMENT OF COMPUTER SCIENCE &IT

MAHATMA GANDHI CENTRAL UNIVERSITY

MOTIHARI, BIHAR

Outline...

- *Basic decision tree algorithm*
- Suitable Problem for Decision Tree Learning
- *Method to find best the attributes*
- *Attribute selection mearures*
  - *Information gain*
  - *Rain ratio*
  - *Gini index*
- *Example of ID3 method*
- *Student assignment to construct the decision tree*

Basic Decision Tree Algorithms

- Algorithm employ a top-down, greedy search through the space of possible decision trees.
- ID3 and C4.5 are basic decision tree algorithm.
- For learning decision tree the question should ask

“which attribute should be tested at the root of the tree”

- ***Statistical property*** *is used to determine how well it alone classifies the training examples.*
- *Then the best attribute is selected for root node of the tree.*

Suitable Problem for Decision Tree Learning

- ***Dataset tuples must be in represented by attribute–value pairs.***
  - *e.g: Hot, Mild, Cold*
- ***The target function must have discrete output values.***
  - *e.g: Yes or No*
- ***Required disjunctive descriptions***
- ***Training data containing error***
  - *It is robust to classification error of training and error in attribute values of the dataset.*
- ***Training data containing missing attribute values***

How to Find Best Attribute

- Before defining the information gain, Entropy needs to define first that is commonly used in information theory.
- Let collection of samples S containing two type of target concepts (Positive and negative).
- Then, entropy corresponding target concept can be defined as:

$$Entropy(S) = -p_{+} \log_2 p_{+} - p_{-} \log_2 p_{-} \ldots\ldots\ldots(1)$$

How to Find Best Attribute

where p_+ and p_- are the proportion of positive and negative instances in S.

- Example:

$$Entropy[9+,5-] = -(9/14)\log_2(9/14) - (5/14)\log_2(5/14) = 0.940.....(2)$$

- In general, If target attribute have C different values, then C-wise classification is defined as:

$$Entropy(S) = \sum_{i=1}^{C} - p_i \log_2 p_i ....(3)$$

Attribute Selection Measures

- There are many method to measure best split of instances.
- The following measures are attribute selection measures:
  1) **Information Gain (IG)**
  2) **Gain Ratio (GR)**
  3) **Gini Index (GI)**

1). ID3 (Iterative Dichotomiser) (Information Gain Measure)

- Developed by J. Ross Quinlan in late 1970's and early 1980's
- Its measure the effectiveness of an attribute in classifying the training data.
- Information gain, Gain(S,A) can be defined as:

$$IG(S,A) = Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v) \quad \ldots(5)$$

1). ID3 (Iterative Dichotomiser) (Information Gain Measure)

- where $S_v \subseteq S$ for which attribute A has value

and $Values(A)$ is the set of all possible values of attribute A

Example:

Values(Wind)=Weak, Strong;

Number of positive and negative instances S=[9+,5-];

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

$$S_{Strong} \leftarrow [3+,3-]$$

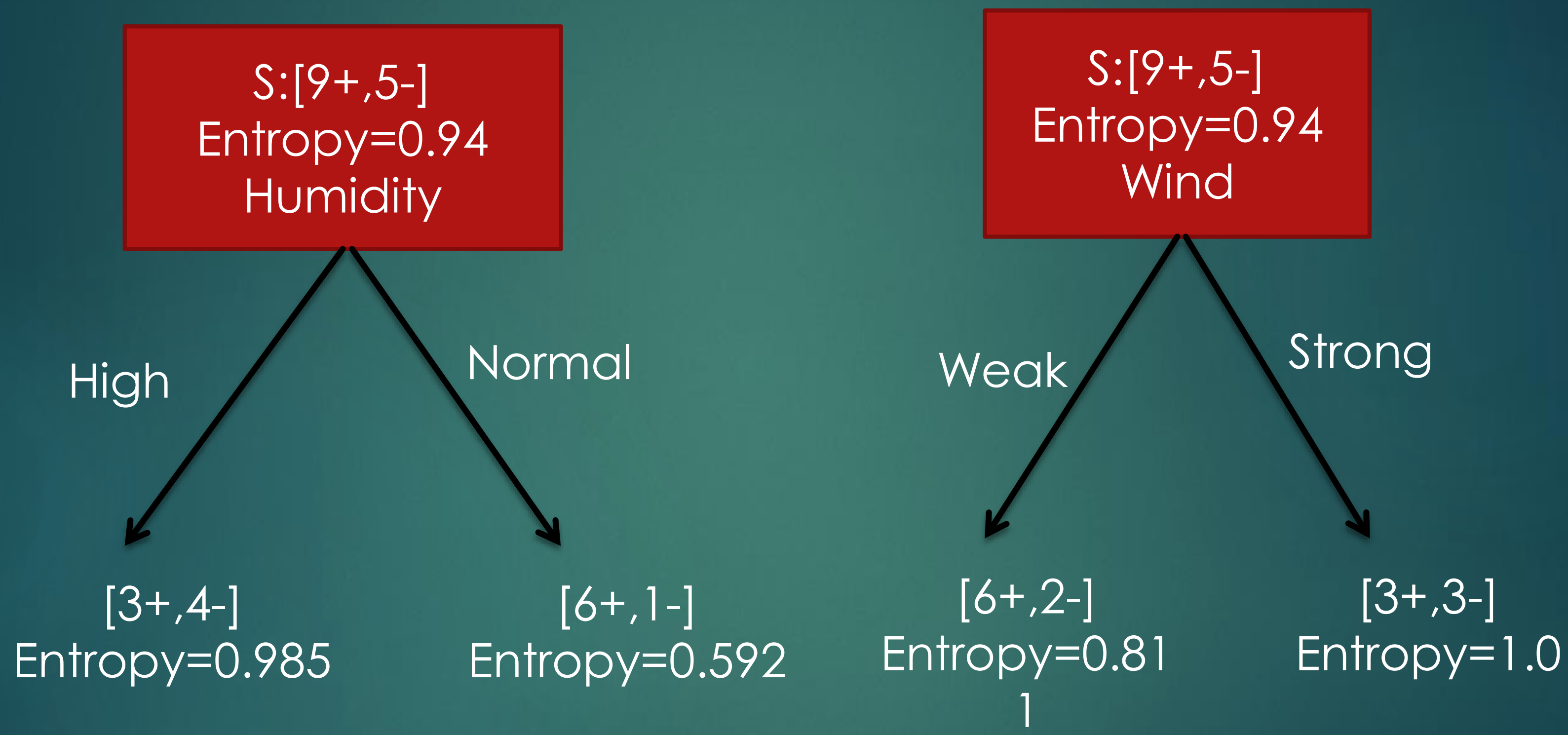
$$then$$

$$IG(S, wind) = Entropy(S) - (8/14)Entropy(S_{weak}) - (6/14)Entropy(S_{strong})$$

$$= 0.94 - (8/14) \times 0.811 - (6/14) \times 1.0$$

$$= 0.048$$

Information gain of attributes for weather data



$$IG(S,Humidity)=0.940-(7/14)*0.985-(7/14)*0.592=0.151$$

$$IG(S,Wind)=0.940-(8/14)*0.811-(6/14)*1.0=0.048$$

Information gain of attributes for weather data

- Information gain of the all attributes:

IG(S,Outlook)=0.246

IG(S,humidity)=0.151

IG(S,Wind)=0.048

IG(S,Temperature)=0.029

Outlook

Sunny

Overcast

Rain

S1

S.N.	Temperature	Humidity	Windy	Play
D1	**Hot**	**High**	**Weak**	**No**
D2	Hot	High	Strong	No
D8	**Mild**	**High**	**Weak**	**No**
D9	Cool	Normal	Weak	Yes
D11	**Mild**	**Normal**	**Strong**	**Yes**

S2

S.N.	Temperature	Humidity	Windy	Play
D4	**Mild**	**High**	**Weak**	**Yes**
D5	Cool	Normal	Weak	Yes
D6	**Cool**	**Normal**	**Strong**	**No**
D10	**Mild**	**Normal**	**Weak**	**Yes**
D14	**Mild**	**high**	**Strong**	**No**

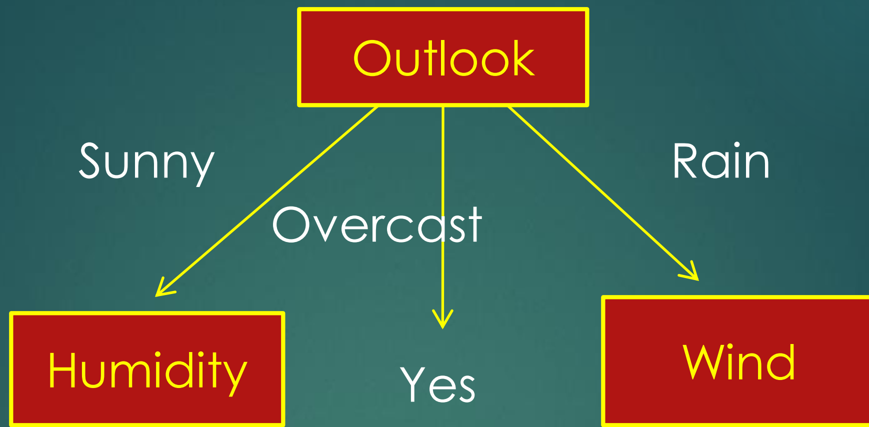
Yes

S3

S.N.	Temperature	Humidity	Windy	Play
D3	**Hot**	**High**	**Weak**	**Yes**
D7	**Cool**	**Normal**	**Strong**	**Yes**
D12	**Mild**	**High**	**Strong**	**Yes**
D13	Hot	Normal	Weak	Yes

Entropy(S1)=0.971
IG(S1,Temperature)=0.571
IG(S1,Humidity)=0.971
IG(S1, Windy)=0.02

Entropy(S2)=0.971
IG(S2,Temperature)=0.02
IG(S2,Humidity)=0.02
IG(S2, Windy)=0.971



S1

S.N.	Temperature	Humidity	Windy	Play
D1	**Hot**	**High**	**Weak**	**No**
D2	Hot	High	Strong	No
D8	**Mild**	**High**	**Weak**	**No**
D9	Cool	Normal	Weak	Yes
D11	**Mild**	**Normal**	**Strong**	**Yes**

S2

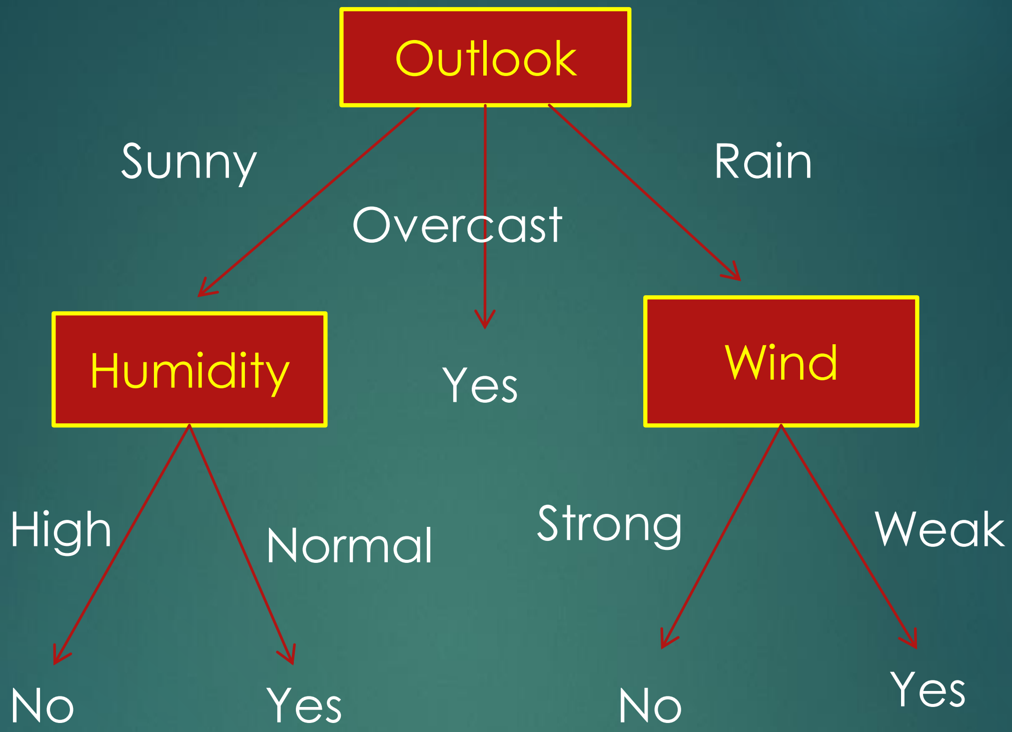
S.N.	Temperature	Humidity	Windy	Play
D4	**Mild**	**High**	**Weak**	**Yes**
D5	Cool	Normal	Weak	Yes
D6	**Cool**	**Normal**	**Strong**	**No**
D10	**Mild**	**Normal**	**Weak**	**Yes**
D14	**Mild**	**high**	**Strong**	**No**

S3

S.N.	Temperature	Humidity	Windy	Play
D3	**Hot**	**High**	**Weak**	**Yes**
D7	**Cool**	**Normal**	**Strong**	**Yes**
D12	**Mild**	**High**	**Strong**	**Yes**
D13	Hot	Normal	Weak	Yes

Entropy(S1)=0.971
IG(S1,Temperature)=0.571
IG(S1,Humidity)=0.971
IG(S1, Windy)=0.02

Entropy(S2)=0.971
IG(S2,Temperature)=0.02
IG(S2,Humidity)=0.02
IG(S2, Windy)=0.971

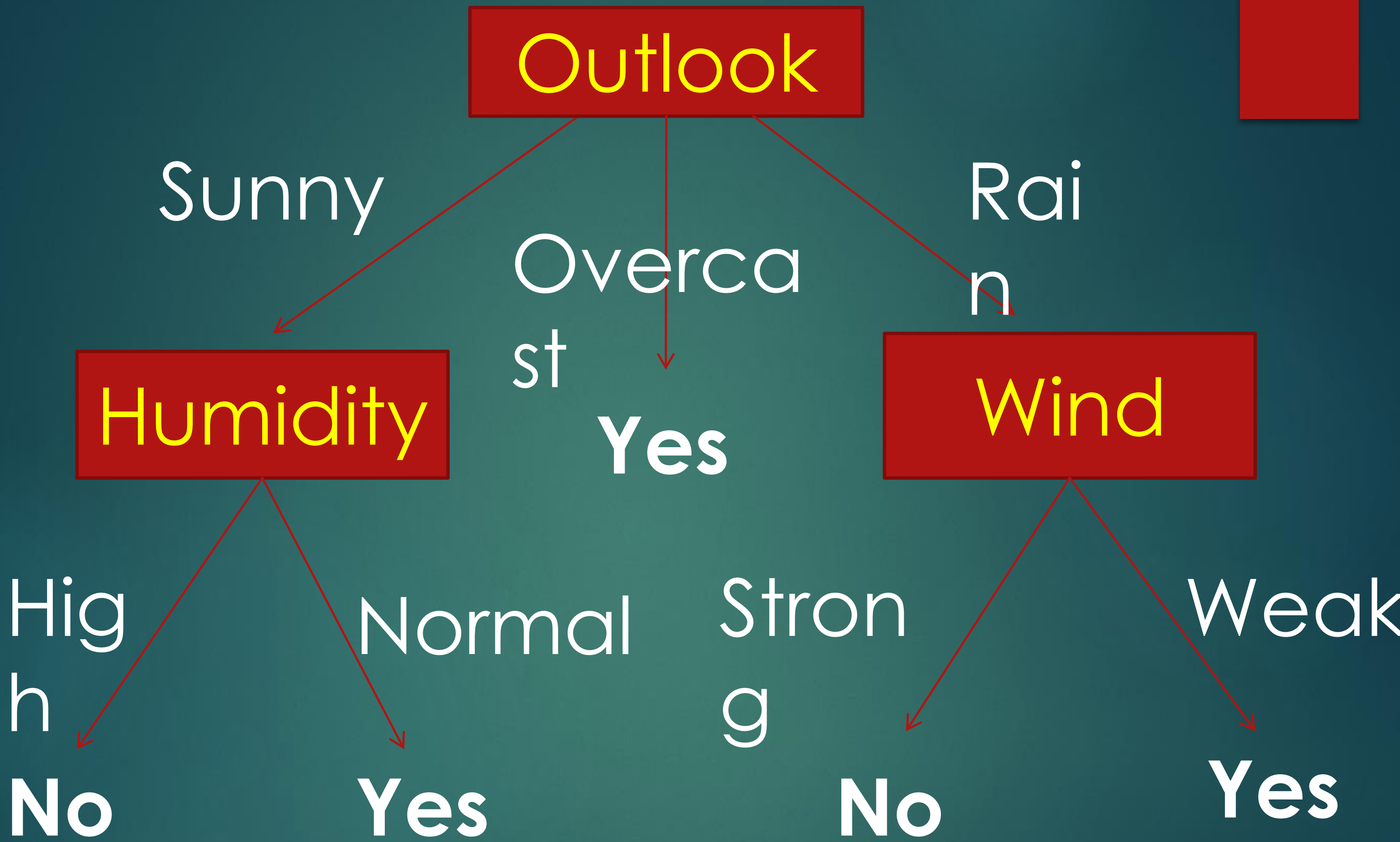


S.N.	Temperature	Windy	Play
D1	Hot	Weak	No
D2	Hot	Strong	No
D8	Mild	Weak	No

S.N.	Temperature	Windy	Play
D9	Cool	Weak	Yes
D11	Mild	Strong	Yes

S.N.	Temperature	Humidity	Play
D6	Cool	Normal	No
D14	Mild	high	No

S.N.	Temperature	Humidity	Play
D4	Mild	High	Yes
D5	Cool	Normal	Yes
D10	Mild	Normal	Yes



Final Decision Tree

Assignment-1: Construct decision tree using ID3 algorithm

RID	Age	Income	Student	Credit-rating	Class: buys computer
1	youth	High	No	Fair	No
2	Youth	High	No	Excellent	No
3	Middle_aged	High	No	Fair	Yes
4	Senior	Medium	No	Fair	Yes
5	Senior	Low	Yes	Fair	Yes
6	Senior	Low	Yes	Excellent	No
7	Middle_aged	Low	Yes	Excellent	Yes
8	Youth	Medium	No	Fair	No
9	Youth	Low	Yes	Fair	Yes
10	Senior	Medium	Yes	Fair	Yes
11	Youth	Medium	Yes	Excellent	Yes
12	Middle_aged	Medium	No	Excellent	Yes
13	Middle_aged	High	Yes	Fair	Yes
14	Senior	Medium	No	Excellent	No

Bibliography

- Tan, Pang-Ning, Michael Steinbach, and Vipin Kumar. *Introduction to data mining*. Pearson Education India, 2016.
- Han, Jiawei, Jian Pei, and Micheline Kamber. *Data mining: concepts and techniques*. Elsevier, 2011.
- Mitchell, Tom M. "Machine learning." (1997).
- Alpaydin, Ethem. *Introduction to machine learning*. MIT press, 2020.
- Murphy, Kevin P. *Machine learning: a probabilistic perspective*. MIT press, 2012.
- Burkov, Andriy. *The hundred-page machine learning book*. Quebec City, Can.: Andriy Burkov, 2019.

- Burkov, Andriy. *The hundred-page machine learning book*. Quebec City, Can.: Andriy Burkov, 2019.

Thank You