CS364: Machine Learning

Semester 2- year 1444

Level 6 BSc CS

Decision Tree Learning

Instructor:

Prof. Hachemi BinNaceur

Dr. Waad Alhoshan

Outline

- Decision tree learning
- Decision tree representation
- ID3 algorithm
- Entropy
- Classification example

Decision tree learning

- Trees
  - One of the most powerful Data Structure in CS.

- Decision tree learning
  - Decision tree learning is a method for approximating discrete-valued target functions, in which the learned function is represented by a decision tree.
  - Most widely used and practical methods for inductive inference.
  - Simple but powerful learning algorithm
  - Decision trees represent a disjunction of conjunctions of constraints on the attribute values of instances.

3

Decision tree learning

- Learning method
  - Approximating discrete-valued function
  - Learning disjunctive expressions

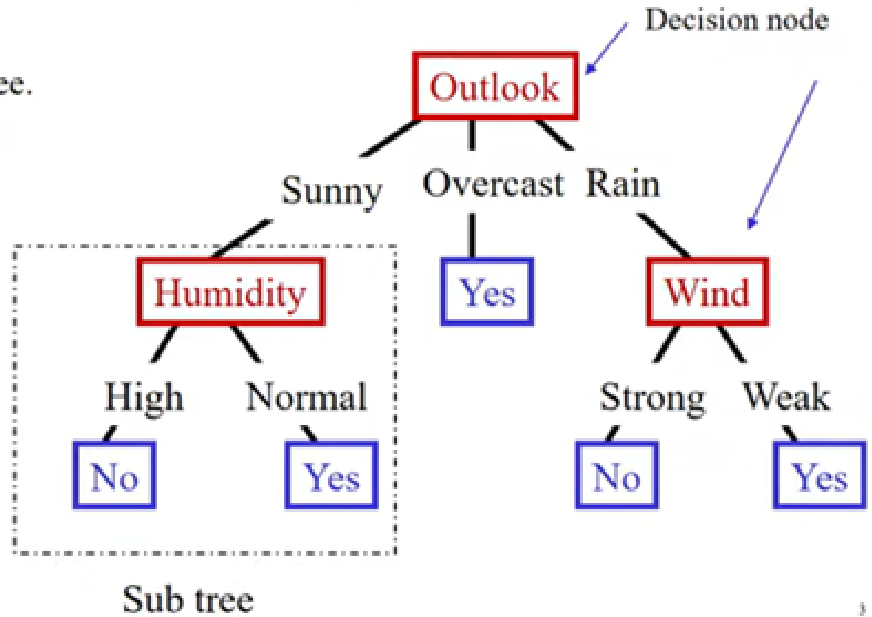
- Learned trees can also be represented by sets of if-then rules

- Decision trees classify instances or examples by starting at the root of the tree and moving through it until a leaf node.

4

Terminologies in Decision Tree

- **Root node** initiates the decision tree that represents the entire dataset D. Then D is divided into two or more homogeneous sets.
- **Internal** (Decision) nodes represent the features of a dataset that make the decision, **branch**es represent the decision rules to make decision, and each **leaf node** represents the decision (outcome).
- **Splitting** process divides the decision node/root node into sub-nodes according to the given conditions.
- **Branch/Sub Tree** is formed by splitting the tree.
- **Pruning** process removes the unwanted branches from the tree.



Decision tree examples

Day	Outlook	Temp	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

Target concept: “good days to play tennis”

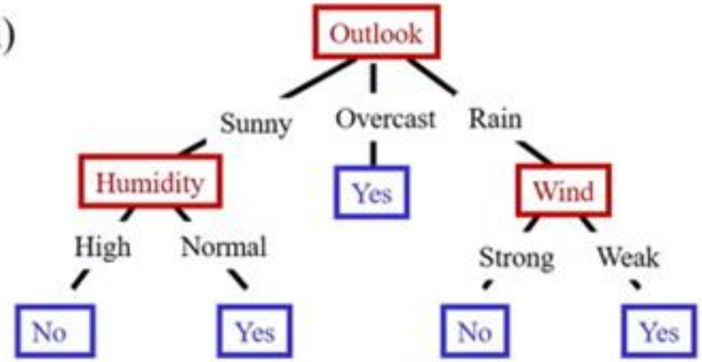
Decision Tree Representation

- Decision trees represent a **disjunction of conjunctions of constraints** on the attribute values of instances.
  - E.g. (Outlook = Sunny ∧ Humidity = Normal) ∨ (Outlook = Overcast) ∨ (Outlook = Rain ∧ Wind = Weak)
- Each path from the root to a leaf corresponds to a conjunction of attribute tests, and
- The tree itself is a disjunction of these conjunctions.
- Each node represents a **feature**, and each link represents a **decision**.
- Each leaf node represents an **outcome** (classification)

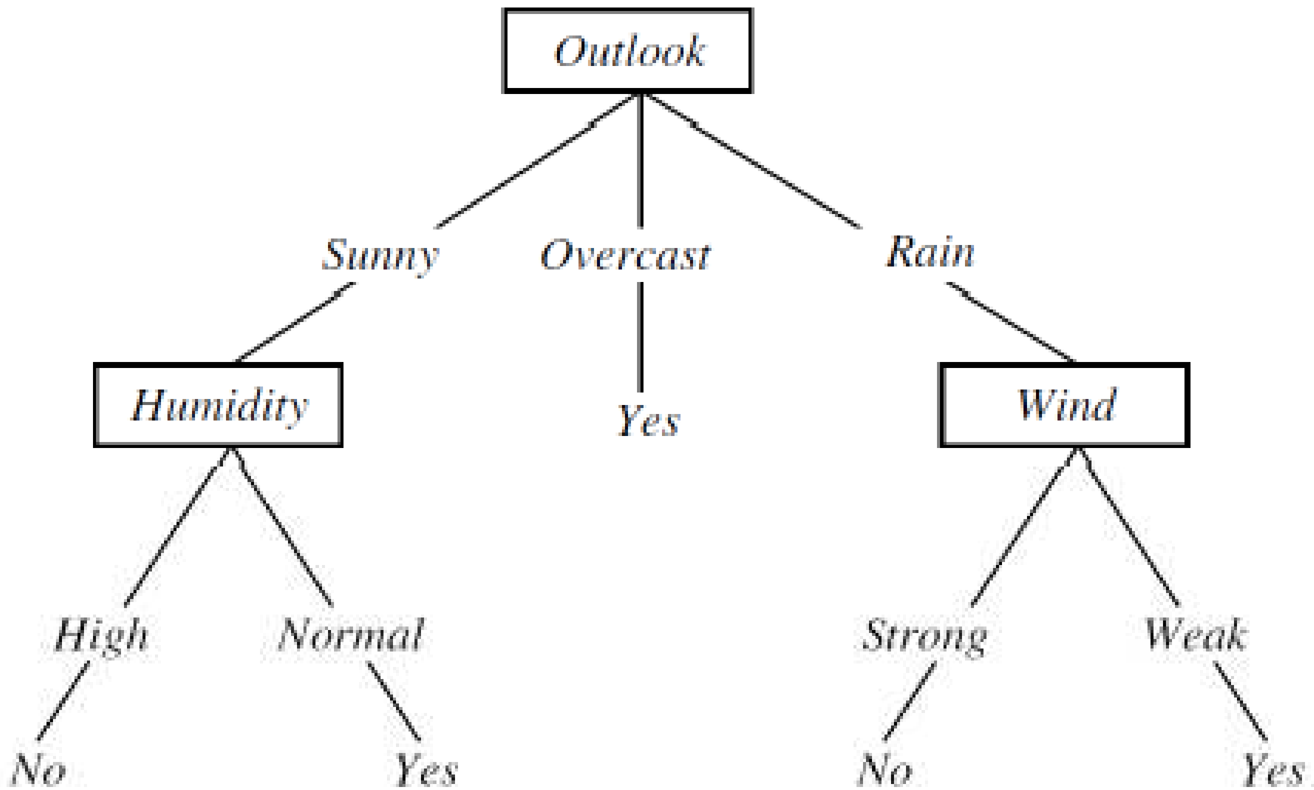
(Outlook = Sunny ∧ Humidity = Normal)

∨ (Outlook = Overcast)

∨ (Outlook = Rain ∧ Wind = Weak)



Decision tree examples (Cont.)



- Example:
  <Outlook = Sunny, Temperature = Hot, Humidity =High, Wind = Strong>
  Classification?

Building a Decision Tree

1. First test all attributes and select the **one attribute** that would function as the **best** root.
2. Break-up the training set into **subsets** based on the branches of the root node.
3. Test the **remaining attributes** to check which one fit best underneath the **branches** of the root node;
4. Continue this process for all other branches until
   a. all examples of a subset are of one type
   b. there are no examples left (return majority classification of the parent)
   c. there are **no more** attributes left (default value should be majority classification)

Decision tree representation

$$
\begin{aligned}
& (Outlook = Sunny \ \wedge \ Humidity = Normal) \\
\vee \quad & (Outlook = Overcast) \\
\vee \quad & (Outlook = Rain \ \wedge \ Wind = Weak)
\end{aligned}
$$

Decision tree representation

- A decision tree reaches its decision by performing a sequence of tests:
  - Each **internal leaf** corresponds to a test of some *attribute* of the instance
  - **Branches** labeled with the possible *values* of the test.
  - Each **leaf node** specifies the value to be returned if that leaf is reached (each leaf node assigns a *classification*).

When to consider decision trees?

- Instances describable by attribute value pairs
  – Instances are described by a fixed set of attributes (e.g., Temperature) and their values (e.g., Hot).
- Discrete-valued target function
  – (e.g., yes or no)
- Disjunctive hypothesis may be required
- Possible noisy training data
  – May contain errors
  – May contain missing attribute values
- e.g.
  – Medical diagnosis
  – Credit risk analysis

Hypotheses spaces

- Would it be possible to use a "generate-and-test" strategy to find a correct decision tree?

  How many distinct decision trees with *n* Boolean attributes?

  = number of Boolean functions

  = number of distinct truth tables with 2^n rows

  $= 2^{2^n}$

- E.g., with 6 Boolean attributes, there are 18,446,744,073,709,551,616 trees

Decision tree induction

- Goal is, given set of training examples, construct decision tree that will classify those training examples correctly (and, hopefully, generalize).
- Original idea of decision trees developed in 1960s by psychologists Hunt, Marin, and Stone, as model of human concept learning. (CLS = "Concept Learning System")
- In 1970s, AI researcher Ross Quinlan used this idea for AI concept learning:
  - ID3 ("Itemized Dichotomizer 3"), 1979

Top-down decision tree induction (ID3)

1. Determine which attribute is, by itself, the most useful one for distinguishing the two classes over all the training data. Put it at the root of the tree.
2. Create branches from the root node for each possible value of this attribute. Sort training examples to the appropriate value.
3. At each descendant node, determine which attribute is, by itself, the most useful one for distinguishing the two classes for the corresponding training data. Put that attribute at that node.
4. Go to 2, but for the current node.

Choosing an attribute

- How to determine which attribute is the best classifier for a set of training examples?
  - If there are **some positive and some negative** examples, then choose **the best attribute** to split them.
  - If all the remaining examples are **positive** (**or all negative**), then we can answer **Yes** or **No**.
  - If there are **no examples left**, to represent the current combination of attribute values, then, we return a **default value** calculated from the **majority** classification at the node's parent.
  - If there are **no attributes**, but **both positive and negative examples**, we have a problem.
    - These examples have exactly the same description, but different classifications (incorrect data, incomplete attributes).

Which Attribute Is the Best Classifier?

- Information gain
  - Measures how well a given attribute separates the training examples according to their target classification.
  - This measure is used to select among the candidate attributes at each step while growing the tree.

Entropy

- Entropy Measures
  - A measure of homogeneity of the set of examples.
  - Given a set S of positive and negative examples of some target concept (a 2-class problem), the entropy of set **S** relative to this binary classification is:

$$\text{Entropy}(S) = -\,p(P).\log_2 p(P) - p(N).\log_2 p(N)$$

Entropy

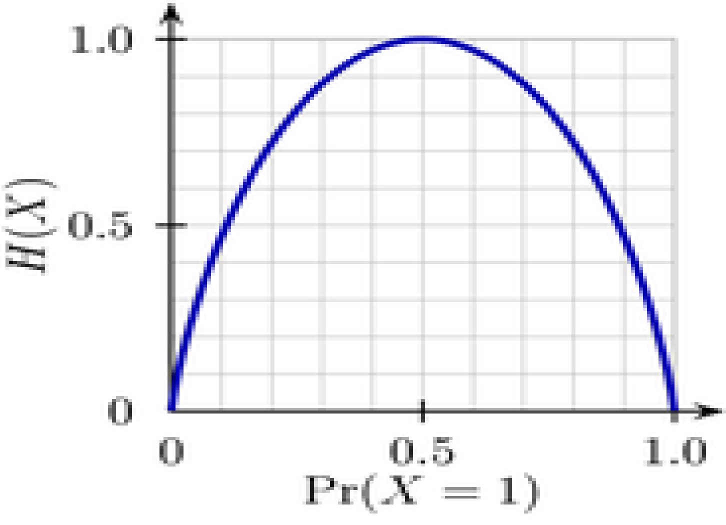
- ***Entropy*** characterizes (measures) the impurity in a given attribute of arbitrary training examples.
- It specifies degree of randomness (uncertainty) in data.
- Entropy values used in splitting i.e., which node is to be split first.
- Given a collection S, containing positive and negative examples of some target concept.
- The ***entropy of S*** relative to this Boolean classification is:

$$\textbf{Entropy(S)} = -p / (p + q) * \log_2(p / (p+q)) - q / (p+q) * \log_2 (q / (p + q))$$

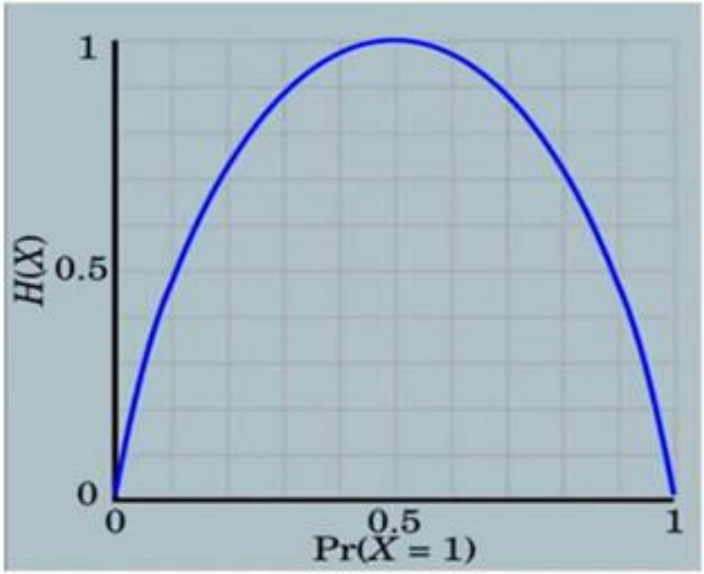
- **S** is a total number of training samples
- **p** is the proportion of positive classes
- **q** is the proportion of negative classes.

Entropy (Cont.)

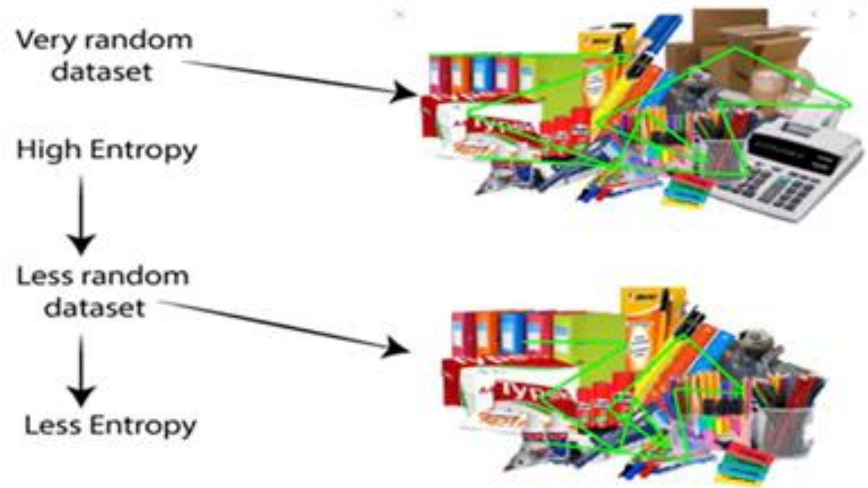
- Entropy Measures
  - The entropy is 0 if the outcome is ”certain”.
  - The entropy is maximum if we have no knowledge of the system (or any outcome is equally possible).



Entropy of a 2-class problem with regard to the portion of one of the two groups.



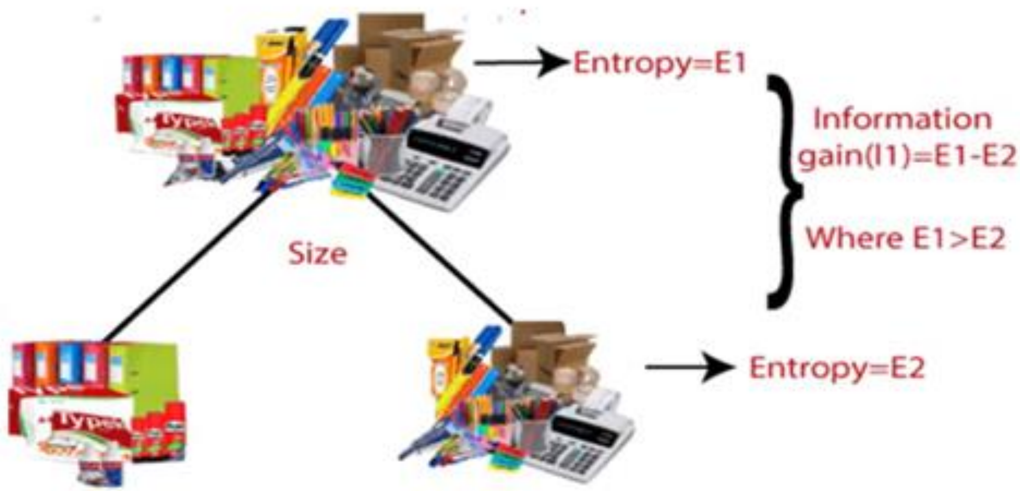
Entropy



- If data is completely (highly) pure / (highly) impure, randomness is 0.
- If impurity is 0.5, randomness (entropy) is 1.

Decision tree - ID3 algorithm

- **Iterative Dichotomiser 3 (**ID3) algorithm uses this *information gain* measure to select among the candidate attributes at each step that return the **highest** data gain while growing the tree.



Entropy (Cont.)

Example: Suppose S has 25 examples,

15 positive and

10 negatives

➔ **entropy of S relative to this classification is**

$$Entropy(S) = -(15/25)\log_2(15/25) - (10/25)\log_2(10/25)$$
$$= 0.97$$

Information gain

Information gain: measures the expected reduction in entropy, or uncertainty.

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

Entropy(S) is the entropy of the original collection **S**.

Values(A) is the set of all possible values for attribute A, and $\mathbf{S_v}$ the subset of **S** for which attribute **A** has value **v** ($\mathbf{S_v = \{s \text{ in } S \mid A(s) = v\}}$)

$\sum_{\boldsymbol{v \in Values(A)}} \frac{|\boldsymbol{S_v}|}{|\boldsymbol{S}|} \boldsymbol{Entropy(S_v)}$ is the expected value of the entropy after **S** is partitioned using attribute A.

Information gain (Cont.)

It is simply the expected reduction in entropy caused by partitioning the examples according to this attribute.

Examples: Entropy calculation

- **Entropy(S) = -p / (p + q) * $\log_2$(p / (p+q)) – q / (p+q) * $\log_2$ (q / (p + q))**
- 9 positive instances and 5 negative instances among 14 total instances.

- Entropy value range is from 0 to 1.
- **Leaf nodes** with **greater entropy** value is considered for further splitting.
- Entropy(S) = Entropy([9+, 5-]) = – (9/14) $\log_2$(9/14) – (5/14) $\log_2$(5/14) = 0.940

  (94% impure or non-homogeneous)
- In given dataset, If **50% is positive** and **50% is negative** after the split, the entropy value is 1 (worst case).

  Entropy([8+, 8-] = – (8/16) $\log_2$(8/16) – (8/16) $\log_2$(8/16) = 1.0
- In a given dataset, **All examples are positive.**

  Entropy([8+, 0-] = – (8/8) $\log_2$(8/8) – (0/8) $\log_2$(0/8) = 0.0
- In a given dataset, **All examples are Negative.**

  Entropy([0+, 8-] = – (0/8) $\log_2$(0/8) - (8/8) $\log_2$(8/8) = 0.0

Calculate Average Information Entropy of Attribute:

I(attribute) = (p_i+ q_i)/(p+q) Entropy(A)

- pi,q_i . +ve, -ve values of corresponding attribute (A) possibility value
- p, q – total +ve, -ve values of dataset

ID3 Decision tree example

Day	Outlook	Temp	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

Target concept: "good days to play tennis"

Day	Outlook	Temp	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

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

$$\begin{aligned} Entropy(S) &= -\,\mathrm{p(P)}.\log_2 \mathrm{p(P)} - \mathrm{p(N)}.\log_2 \mathrm{p(N)} \\ &= -\,(9/14)\log_2 (9/14) - (5/14)\log_2 (5/14) \\ &= 0.94 \end{aligned}$$

Day	Outlook	Temp	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

Attribute: Outlook

$$Entropy(S) = 0.94$$

$$S_{sunny} \rightarrow [2+,\ 3-]$$

$$Entropy(S_{Sunny}) = -\ (2/5).\log_2\ (2/5) - (3/5).\log_2\ (3/5) = 0.971$$

Day	Outlook	Temp	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

Attribute: Outlook

$$Entropy(S) = 0.94$$

$$S_{Sunny} \rightarrow [2+,\ 3-]$$

$$S_{Overcast} \rightarrow [4+,\ 0-]$$

$$Entropy(S_{Sunny}) = -(2/5).\log_2 p(2/5) - (3/5).\log_2 (3/5) = 0.971$$

$$Entropy(S_{Overcast}) = -(4/4).\log_2 p(4/4) - (0/4).\log_2 (0/4) = 0$$

Day	Outlook	Temp	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

Attribute: Outlook

$$Entropy(S) = 0.94$$

$$S_{Sunny} \rightarrow [2+,\ 3-]$$

$$S_{Overcast} \rightarrow [4+,\ 0-]$$

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

$$Entropy(S_{Sunny}) = -(2/5).\log_2(2/5) - (3/5).\log_2(3/5) = 0.971$$

$$Entropy(S_{Overcast}) = -(4/4).\log_2(4/4) - (0/4).\log_2(0/4) = 0$$

$$Entropy(S_{Rain}) = -(3/5).\log_2(3/5) - (2/5).\log_2(2/5) = 0.971$$

Day	Outlook	Temp	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

Attribute: Outlook

$$Entropy(S) = 0.94$$

$$S_{Sunny} \rightarrow [2+, 3-]$$

$$S_{Overcast} \rightarrow [4+, 0-]$$

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

$$Entropy(S_{Sunny}) = -(2/5).\log_2(2/5) - (3/5).\log_2(3/5) = 0.971$$

$$Entropy(S_{Overcast}) = -(4/4).\log_2(4/4) - (0/4).\log_2(0/4) = 0$$

$$Entropy(S_{Rain}) = -(3/5).\log_2(3/5) - (2/5).\log_2(2/5) = 0.971$$

$$Gain(S, Outlook) = Entropy(S) - \sum_{v \in Values(Outlook)} \frac{|S_v|}{|S|} Entropy(S_v)$$

$$Gain(S, Outlook) = 0.94 - \frac{5}{14} Entropy(S_{Sunny}) - \frac{4}{14} Entropy(S_{Overcast}) - \frac{5}{14} Entropy(S_{Rain})$$

$$= 0.94 - \frac{5}{14}(0.971) - \frac{4}{14}(0) - \frac{5}{14}(0.971)$$

$$= 0.2464$$

Day	Outlook	Temp	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

$$Entropy(S_{Hot}) = -(2/4).\log_2 (2/4) - (2/4).\log_2 (2/4) = 1$$

Attribute: Temp

$$Entropy(S) = 0.94$$

$$S_{Hot} \rightarrow [2+,\ 2-]$$

Day	Outlook	Temp	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

Attribute: Temp

$$Entropy(S) = 0.94$$

$$S_{Hot} \rightarrow [2+,\ 2-]$$

$$S_{Mild} \rightarrow [4+,\ 2-]$$

$$Entropy(S_{Hot}) = -(2/4).\log_2 (2/4) - (2/4).\log_2 (2/4) = 1$$

$$Entropy(S_{Mild}) = -(4/6).\log_2 (4/6) - (2/6).\log_2 (2/6) = 0.9183$$

Day	Outlook	Temp	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

Attribute: Temp

$$Entropy(S) = 0.94$$

$$S_{Hot} \rightarrow [2+,\ 2-]$$

$$S_{Mild} \rightarrow [4+,\ 2-]$$

$$S_{Cool} \rightarrow [3+,\ 1-]$$

$$Entropy(S_{Hot}) = -(2/4).\log_2(2/4) - (2/4).\log_2(2/4) = 1$$

$$Entropy(S_{Mild}) = -(4/6).\log_2(4/6) - (2/6).\log_2(2/6) = 0.9183$$

$$Entropy(S_{Cool}) = -(3/4).\log_2(3/4) - (1/4).\log_2(1/4) = 0.8113$$

Day	Outlook	Temp	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

Attribute: Temp

$$Entropy(S) = 0.94$$

$S_{Hot} \rightarrow [2+,\ 2-]$

$S_{Mild} \rightarrow [4+,\ 2-]$

$S_{Cool} \rightarrow [3+,\ 1-]$

$$Entropy(S_{Hot}) = -(2/4).\log_2(2/4) - (2/4).\log_2(2/4) = 1$$

$$Entropy(S_{Mild}) = -(4/6).\log_2(4/6) - (2/6).\log_2(2/6) = 0.9183$$

$$Entropy(S_{Cool}) = -(3/4).\log_2(3/4) - (1/4).\log_2(1/4) = 0.8113$$

$$Gain(S, Temp) = Entropy(S) - \sum_{v \in Values(Temp)} \frac{|S_v|}{|S|} Entropy(S_v)$$

$$= 0.94 - \frac{4}{14} Entropy(S_{Hot}) - \frac{6}{14} Entropy(S_{Mild}) - \frac{4}{14} Entropy(S_{Cool})$$

$$= 0.94 - \frac{4}{14}(1) - \frac{6}{14}(0.9183) - \frac{4}{14}(0.8113)$$

$$= 0.0289$$

Day	Outlook	Temp	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

Attribute: Humidity

$$Entropy(S) = 0.94$$

$$S_{High} \rightarrow [3+,\ 4-]$$

$$Entropy(S_{High}) = -\ (3/7).\log_2 (3/7) - (4/7).\log_2 (4/7) = 0.9852$$

Day	Outlook	Temp	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

Attribute: Humidity

$$Entropy(S) = 0.94$$

$$S_{High} \rightarrow [3+,\ 4-]$$

$$S_{Normal} \rightarrow [6+,\ 1-]$$

$$Entropy(S_{High}) = -\ (3/7).\log_2 (3/7) - (4/7).\log_2 (4/7) = 0.9852$$

$$Entropy(S_{Normal}) = -\ (6/7).\log_2 (6/7) - (1/7).\log_2 (1/7) = 0.5916$$

Day	Outlook	Temp	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

Attribute: Humidity

$$Entropy(S) = 0.94$$

$$S_{High} \rightarrow [3+,\ 4-]$$

$$S_{Normal} \rightarrow [6+,\ 1-]$$

$$Entropy(S_{High}) = -\ (3/7).\log_2 (3/7) - (4/7).\log_2 (4/7) = 0.9852$$

$$Entropy(S_{Normal}) = -\ (6/7).\log_2 (6/7) - (1/7).\log_2 (1/7) = 0.5916$$

$$Gain(S, Humidity) = Entropy(S) - \sum_{v \in Values(Humidity)} \frac{|S_v|}{|S|} Entropy(S_v)$$

$$= 0.94 - \frac{7}{14} Entropy(S_{High}) - \frac{7}{14} Entropy(S_{Normal})$$

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

$$= 0.1516$$

Day	Outlook	Temp	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

Attribute: Wind

$$Entropy(S) = 0.94$$

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

$$Entropy(S_{Strong}) = -\ (3/6).\log_2\ (3/6) - (3/6).\log_2\ (3/6) = 1$$

Day	Outlook	Temp	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

Attribute: Wind

$$Entropy(S) = 0.94$$

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

$$S_{Weak} \rightarrow [6+,\ 2-]$$

$$Entropy(S_{Strong}) = -(3/6).\log_2(3/6) - (3/6).\log_2(3/6) = 1$$

$$Entropy(S_{Weak}) = -(6/8).\log_2(6/8) - (2/8).\log_2(2/8) = 0.8113$$

Day	Outlook	Temp	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

Attribute: Wind

$Entropy(S) = 0.94$

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

$S_{Weak} \rightarrow [6+, 2-]$

$Entropy(S_{Strong}) = -(3/6).\log_2(3/6) - (3/6).\log_2(3/6) = 1$

$Entropy(S_{Weak}) = -(6/8).\log_2(6/8) - (2/8).\log_2(2/8) = 0.8113$

$$Gain(S, Wind) = Entropy(S) - \sum_{v \in Values(Wind)} \frac{|S_v|}{|S|} Entropy(S_v)$$

$$= 0.94 - \frac{6}{14} Entropy(S_{Strong}) - \frac{8}{14} Entropy(S_{Weak})$$

$$= 0.94 - \frac{6}{14}(1) - \frac{8}{14}(0.8113)$$

$$= 0.0478$$

Day	Outlook	Temp	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

Which Attribute Is the Best Classifier?

$$Gain(S, Outlook) = 0.246$$

$$Gain(S, Humidity) = 0.151$$

$$Gain(S, Wind) = 0.048$$

$$Gain(S, Temperature) = 0.029$$

Day	Outlook	Temp	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

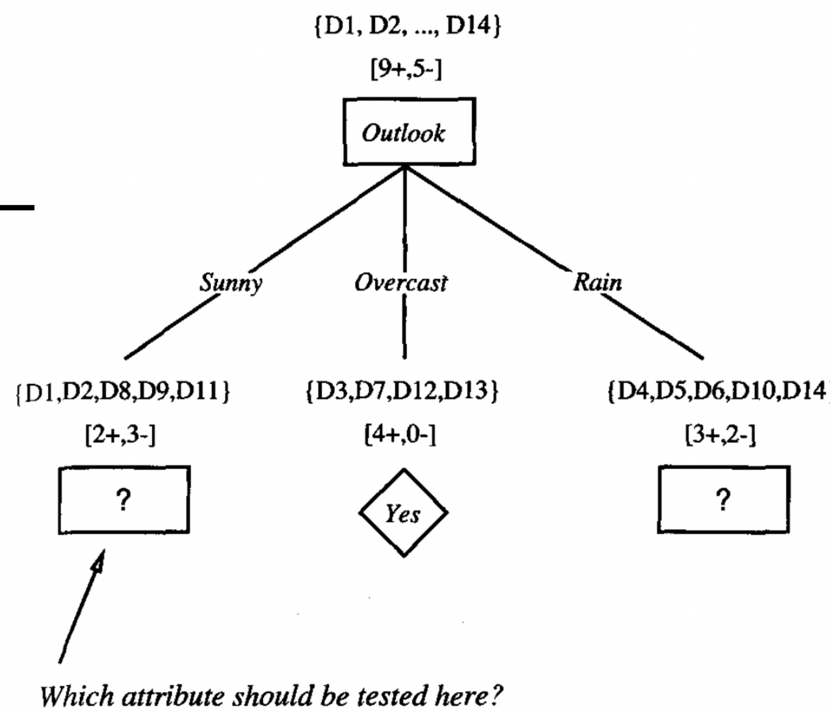
Which Attribute Is the Best Classifier?

$$Gain(S, Outlook) = 0.246$$

$$Gain(S, Humidity) = 0.151$$

$$Gain(S, Wind) = 0.048$$

$$Gain(S, Temperature) = 0.029$$



Day	Outlook	Temp	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes

$$S_{Sunny} = [2+, 3-]$$

$$\begin{aligned} Entropy(S_{Sunny}) &= -\,(P).\log_2 (P) - (N).\log_2 (N) \\ &= -\,(2/5) \log_2 (2/5) - (3/5) \log_2 (3/5) \\ &= 0.97 \end{aligned}$$

Attribute: Temp

Day	Outlook	Temp	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes

$$Entropy(S_{Sunny}) = 0.97$$

$$S_{Hot} \rightarrow [0+, 2-]$$

$$Entropy(S_{Hot}) = 0$$

Attribute: Temp

Day	Outlook	Temp	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes

$$Entropy(S_{Sunny}) = 0.97$$

$$S_{Hot} \rightarrow [0+,\ 2-]$$

$$S_{Mild} \rightarrow [1+,\ 1-]$$

$$Entropy(S_{Hot}) = 0$$

$$Entropy(S_{Mild}) = 1$$

Attribute: Temp

Day	Outlook	Temp	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes

$$Entropy(S_{Sunny}) = 0.97$$

$$S_{Hot} \rightarrow [0+,\ 2-]$$

$$S_{Mild} \rightarrow [1+,\ 1-]$$

$$S_{Cool} \rightarrow [1+,\ 0-]$$

$$Entropy(S_{Hot}) = 0$$

$$Entropy(S_{Mild}) = 1$$

$$Entropy(S_{Cool}) = 0$$

Attribute: Temp

Day	Outlook	Temp	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes

$$Entropy(S_{Sunny}) = 0.97$$

$$S_{Hot} \rightarrow [0+, 2-]$$

$$S_{Mild} \rightarrow [1+, 1-]$$

$$S_{Cool} \rightarrow [1+, 0-]$$

$$Entropy(S_{Hot}) = 0$$

$$Entropy(S_{Mild}) = 1$$

$$Entropy(S_{Cool}) = 0$$

$$Gain(SSunny, Temp) = Entropy(S) - \sum_{v \in Values(Temp)} \frac{|S_v|}{|S|} Entropy(S_v)$$

$$= 0.97 - \frac{2}{5} Entropy(S_{Hot}) - \frac{2}{5} Entropy(S_{Mild}) - \frac{1}{5} Entropy(S_{Cool})$$

$$= 0.97 - \frac{2}{5}(0) - \frac{2}{5}(1) - \frac{1}{5} Entropy(0)$$

$$= 0.570$$

Attribute: Humidity

Day	Outlook	Temp	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes

$$Entropy(S_{Sunny}) = 0.97$$

$$S_{High} \rightarrow [0+,\ 3-]$$

$$Entropy(S_{High}) = 0$$

Attribute: Humidity

Day	Outlook	Temp	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes

$$Entropy(S_{Sunny}) = 0.97$$

$$S_{High} \rightarrow [0+, 3-]$$

$$S_{Normal} \rightarrow [2+, 0-]$$

$$Entropy(S_{High}) = 0$$

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

Attribute: Humidity

Day	Outlook	Temp	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes

$$Entropy(S_{Sunny}) = 0.97$$

$$S_{High} \rightarrow [0+,\ 3-]$$

$$S_{Normal} \rightarrow [2+,\ 0-]$$

$$Entropy(S_{High}) = 0$$

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

$$Gain(SSunny, Humidity) = Entropy(S) - \sum_{v \in Values(Humidity)} \frac{|S_v|}{|S|} Entropy(S_v)$$

$$= 0.97 - \frac{3}{5} Entropy(S_{High}) - \frac{2}{5} Entropy(S_{Normal})$$

$$= 0.97 - \frac{3}{5}(0) - \frac{2}{5}(0)$$

$$= 0.97$$

Attribute: Wind

Day	Outlook	Temp	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes

$$Entropy(S_{Sunny}) = 0.97$$

$$S_{Strong} \rightarrow [1+,\ 1-]$$

$$Entropy(S_{Strong}) = 1$$

Attribute: Wind

Day	Outlook	Temp	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes

$$Entropy(S_{Sunny}) = 0.97$$

$$S_{Strong} \rightarrow [1+,\ 1-]$$

$$S_{Weak} \rightarrow [1+,\ 2-]$$

$$Entropy(S_{Strong}) = 1$$

$$Entropy(S_{Weak}) = -(1/3).\log_2(1/3) - (2/3).\log_2(2/3) = 0.9183$$

Attribute: Wind

Day	Outlook	Temp	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes

$Entropy(S_{Sunny}) = 0.97$

$S_{Strong} \rightarrow [1+,\ 1-]$

$S_{Weak} \rightarrow [1+,\ 2-]$

$Entropy(S_{Strong}) = 1$

$Entropy(S_{Weak}) = -(1/3).\log_2(1/3) - (2/3).\log_2(2/3) = 0.9183$

$$Gain(SSunny, Wind) = Entropy(S) - \sum_{v \in Values(Wind)} \frac{|S_v|}{|S|} Entropy(S_v)$$

$$= 0.97 - \frac{2}{5} Entropy(S_{Strong}) - \frac{3}{5} Entropy(S_{Weak})$$

$$= 0.97 - \frac{2}{5}(1) - \frac{3}{5}(0.9183)$$

$$= 0.0192$$

7

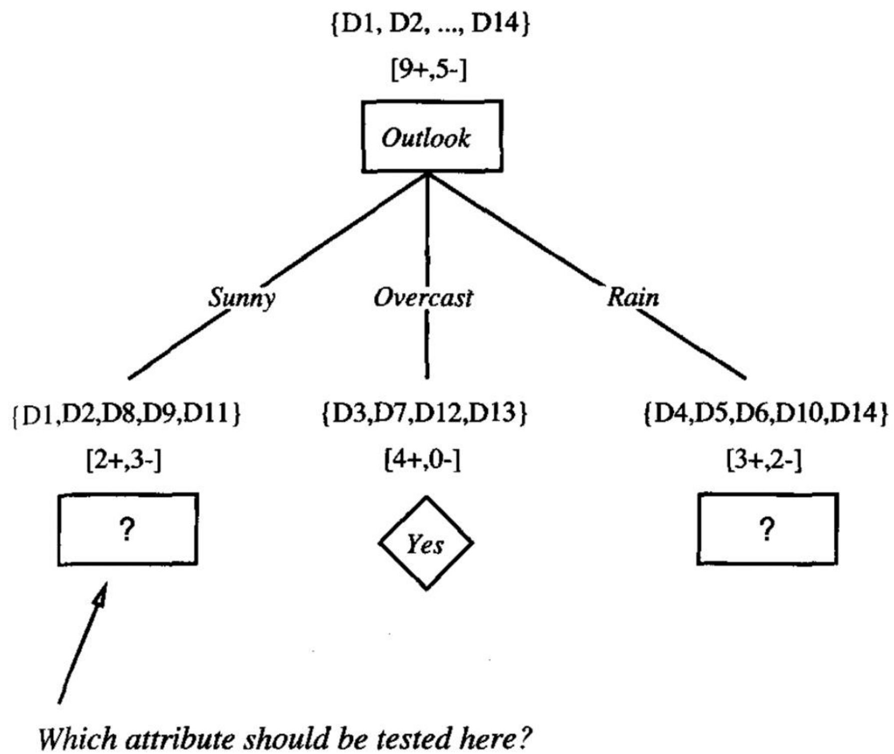
Which Attribute Is the Best Classifier?

Day	Outlook	Temp	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	
D11	Sunny	Mild	Normal	Strong	

$$Gain(S_{Sunny}, Temp) = 0.570$$

$$Gain(SSunny, Humidity) = 0.97$$

$$Gain(SSunny, Wind) = 0.0192$$



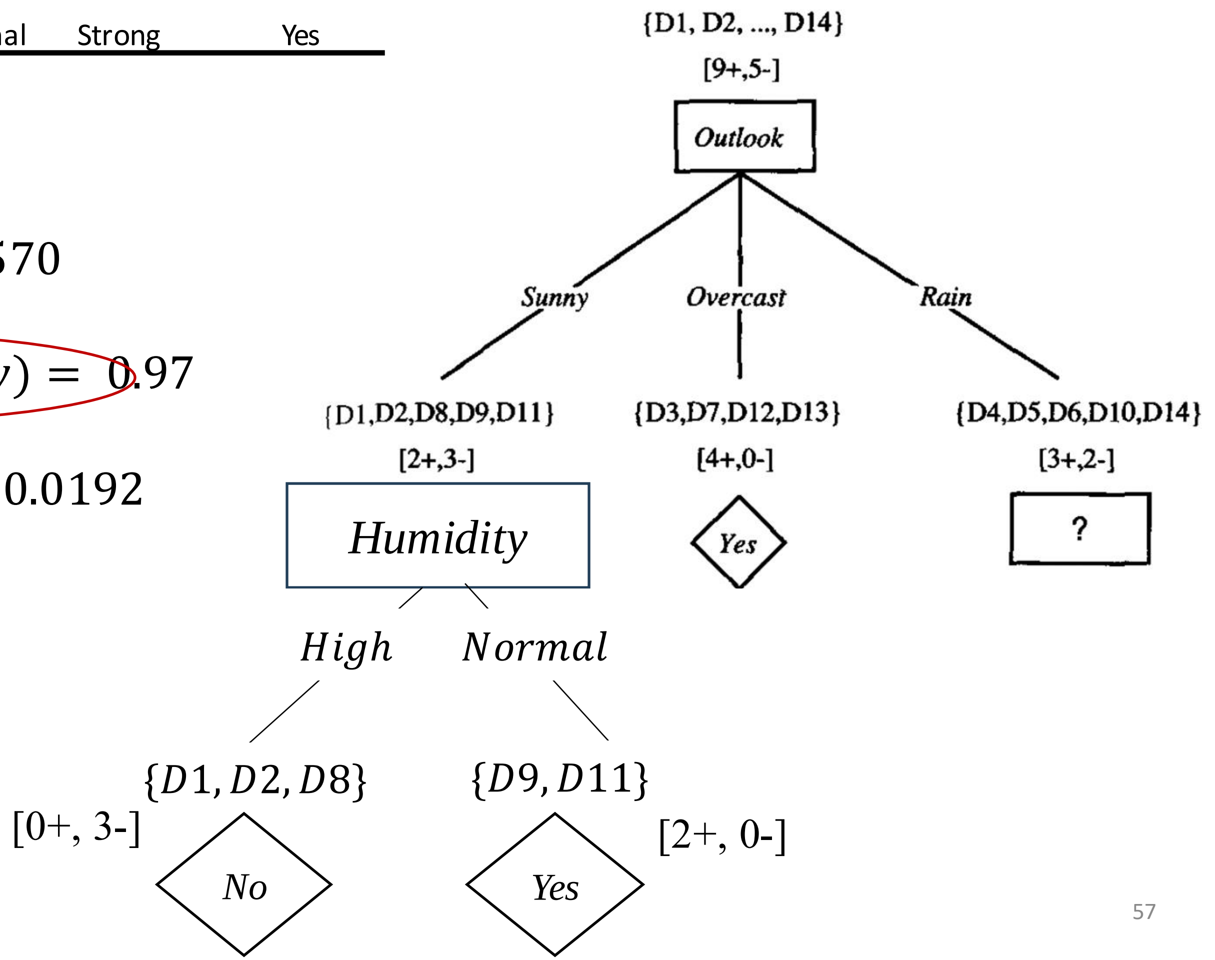
Which Attribute Is the Best Classifier?

Day	Outlook	Temp	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes

$Gain(S_{Sunny}, Temp) = 0.570$

$Gain(SSunny, Humidity) = 0.97$

$Gain(SSunny, Wind) = 0.0192$



Day	Outlook	Temp	Humidity	Wind	PlayTennis
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D10	Rain	Mild	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

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

$$\begin{aligned} Entropy(S_{Rain}) &= -\,\mathrm{p(P).log_2\,p(P)} - \mathrm{p(N).log_2\,p(N)} \\ &= -\,(3/5)\log_2(3/5) - (2/5)\log_2(2/5) \\ &= 0.97 \end{aligned}$$

Attribute: Temp

Day	Outlook	Temp	Humidity	Wind	PlayTennis
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D10	Rain	Mild	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

$$Entropy(S_{Rain}) = 0.97$$

$$S_{Hot} \rightarrow [0+,\ 0-]$$

$$Entropy(S_{Hot}) = 0$$

Attribute: Temp

Day	Outlook	Temp	Humidity	Wind	PlayTennis
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D10	Rain	Mild	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

$$Entropy(S_{Rain}) = 0.97$$

$$S_{Hot} \rightarrow [0+, 0-]$$

$$S_{Mild} \rightarrow [2+, 1-]$$

$$Entropy(S_{Hot}) = 0$$

$$Entropy(S_{Mild}) = -\,\mathrm{p}(2/3).\log_2 \mathrm{p}(2/3) - \mathrm{p}(1/3).\log_2 \mathrm{p}(1/3) = 0.9183$$

Attribute: Temp

Day	Outlook	Temp	Humidity	Wind	PlayTennis
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D10	Rain	Mild	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

$Entropy(S_{Rain}) = 0.97$

$S_{Hot} \rightarrow [0+, 0-]$

$S_{Mild} \rightarrow [2+, 1-]$

$S_{Cool} \rightarrow [1+, 1-]$

$$Entropy(S_{Hot}) = 0$$

$$Entropy(S_{Mild}) = -\mathrm{p}(2/3).\log_2 \mathrm{p}(2/3) - \mathrm{p}(1/3).\log_2 \mathrm{p}(1/3) = 0.9183$$

$$Entropy(S_{Cool}) = 1$$

Attribute: Temp

Day	Outlook	Temp	Humidity	Wind	PlayTennis
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D10	Rain	Mild	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

$Entropy(S_{Rain}) = 0.97$

$S_{Hot} \rightarrow [0+, 0-]$

$S_{Mild} \rightarrow [2+, 1-]$

$S_{Cool} \rightarrow [1+, 1-]$

$Entropy(S_{Hot}) = 0$

$Entropy(S_{Mild}) = -(2/3).\log_2 (2/3) - (1/3).\log_2 (1/3) = 0.9183$

$Entropy(S_{Cool}) = 1$

$$Gain(SRain, Temp) = Entropy(S) - \sum_{v \in Values(Temp)} \frac{|S_v|}{|S|} Entropy(S_v)$$

$$= 0.97 - \frac{0}{5} Entropy(S_{Hot}) - \frac{3}{5} Entropy(S_{Mild}) - \frac{2}{5} Entropy(S_{Cool})$$

$$= 0.97 - \frac{0}{5}(0) - \frac{3}{5}(0.9183) - \frac{2}{5}(1)$$

$$= 0.0192$$

Attribute: Humidity

Day	Outlook	Temp	Humidity	Wind	PlayTennis
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D10	Rain	Mild	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

$$Entropy(S_{Rain}) = 0.97$$

$$S_{High} \rightarrow [1+,\ 1-]$$

$$Entropy(S_{High}) = 1$$

Attribute: Humidity

Day	Outlook	Temp	Humidity	Wind	PlayTennis
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D10	Rain	Mild	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

$$Entropy(S_{Rain}) = 0.97$$

$$S_{High} \rightarrow [1+, 1-]$$

$$S_{Normal} \rightarrow [2+, 1-]$$

$$Entropy(S_{High}) = 1$$

$$Entropy(S_{Normal}) = -(2/3).\log_2(2/3) - (1/3).\log_2(1/3) = 0.9183$$

Attribute: Humidity

Day	Outlook	Temp	Humidity	Wind	PlayTennis
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D10	Rain	Mild	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

$$Entropy(S_{Rain}) = 0.97$$

$$S_{High} \rightarrow [1+, 1-]$$

$$S_{Normal} \rightarrow [2+, 1-]$$

$$Entropy(S_{High}) = 1$$

$$Entropy(S_{Normal}) = -(2/3).\log_2 (2/3) - (1/3).\log_2 (1/3) = 0.9183$$

$$Gain(SRain, Humidity) = Entropy(S) - \sum_{v \in Values(Humidity)} \frac{|S_v|}{|S|} Entropy(S_v)$$

$$= 0.97 - \frac{2}{5} Entropy(S_{High}) - \frac{3}{5} Entropy(S_{Normal})$$

$$= 0.97 - \frac{2}{5}(1) - \frac{3}{5}(0.918)$$

$$= 0.0192$$

Attribute: Wind

Day	Outlook	Temp	Humidity	Wind	PlayTennis
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D10	Rain	Mild	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

$$Entropy(S_{Rain}) = 0.97$$

$$S_{Strong} \rightarrow [0+,\ 2-]$$

$$Entropy(S_{Strong}) = 0$$

Attribute: Wind

Day	Outlook	Temp	Humidity	Wind	PlayTennis
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D10	Rain	Mild	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

$$Entropy(S_{Rain}) = 0.97$$

$$S_{Strong} \rightarrow [0+,\ 2-]$$

$$S_{Weak} \rightarrow [3+,\ 0-]$$

$$Entropy(S_{Strong}) = 0$$

$$Entropy(S_{Weak}) = 0$$

Attribute: Wind

Day	Outlook	Temp	Humidity	Wind	PlayTennis
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D10	Rain	Mild	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

$$Entropy(S_{Rain}) = 0.97$$

$$S_{Strong} \rightarrow [0+,\ 2-]$$

$$S_{Weak} \rightarrow [3+,\ 0-]$$

$$Entropy(S_{Strong}) = 0$$

$$Entropy(S_{Weak}) = 0$$

$$Gain(SRain, Wind) = Entropy(S) - \sum_{v \in Values(Wind)} \frac{|S_v|}{|S|} Entropy(S_v)$$

$$= 0.97 - \frac{2}{5} Entropy(S_{strong}) - \frac{3}{5} Entropy(S_{Weak})$$

$$= 0.97 - \frac{2}{5}(0) - \frac{3}{5}(0)$$

$$= 0.97$$

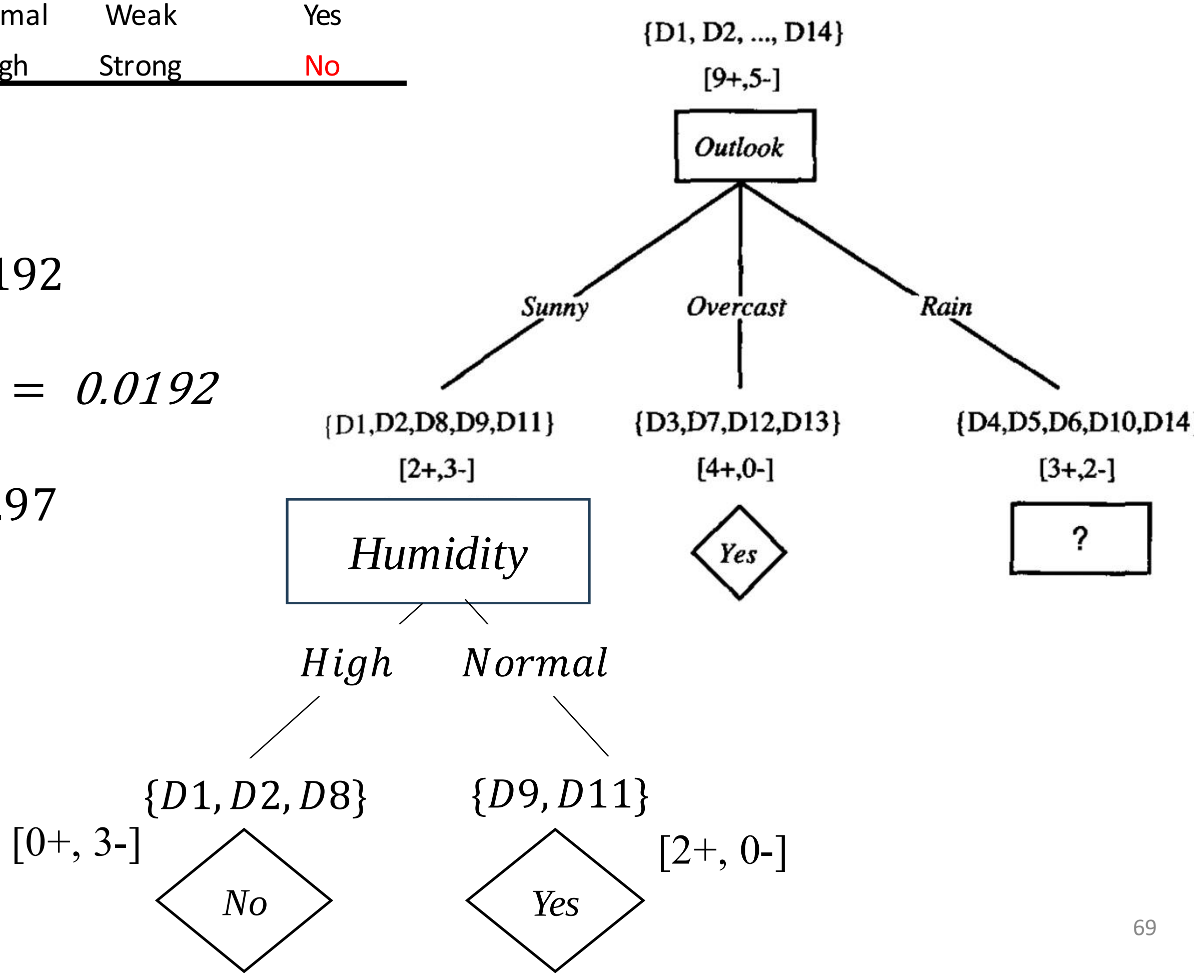
Which Attribute Is the Best Classifier?

Day	Outlook	Temp	Humidity	Wind	PlayTennis
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D10	Rain	Mild	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

$$Gain(S_{Rain}, Temp) = 0.0192$$

$$Gain(SRain, Humidity) = 0.0192$$

$$Gain(SRain, Wind) = 0.97$$



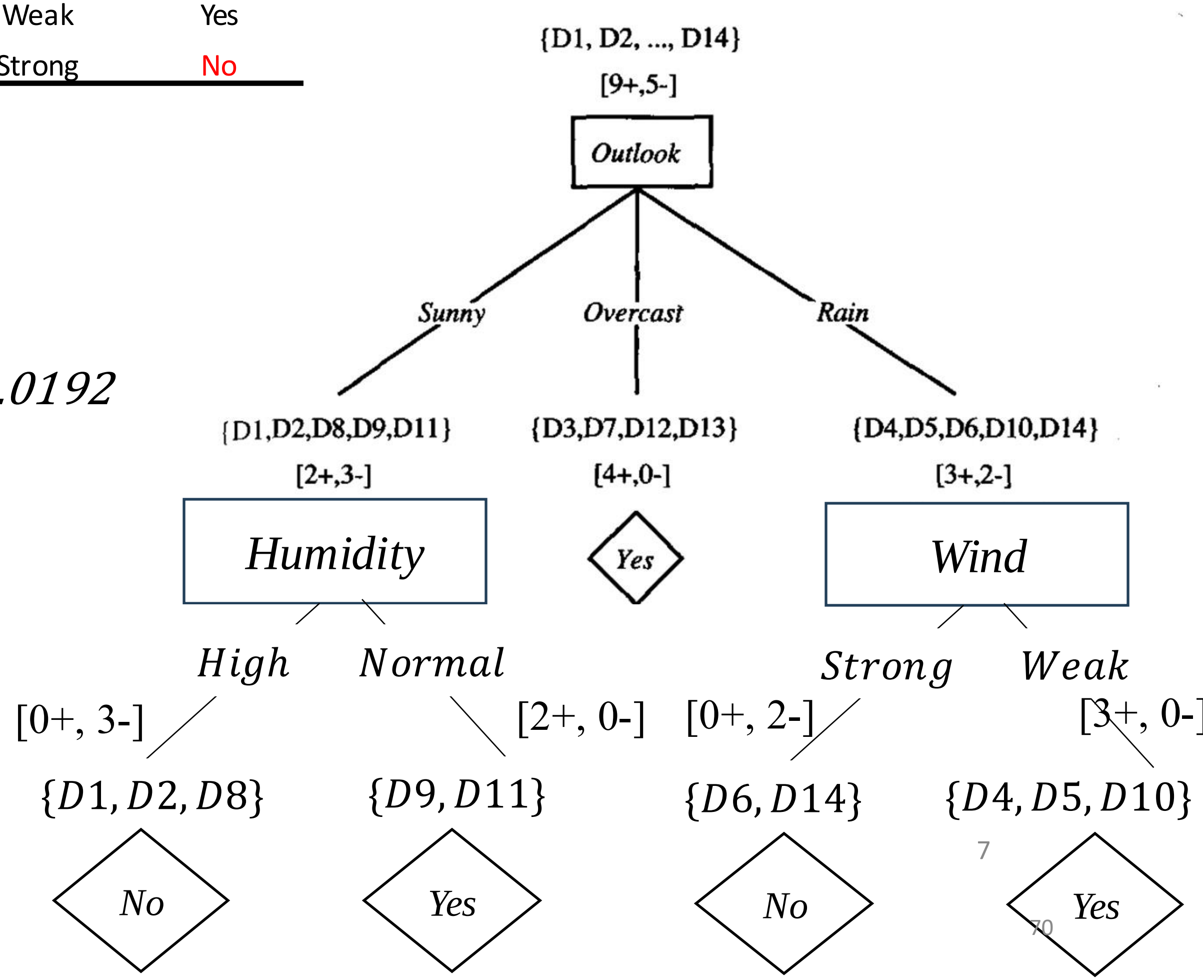
Which Attribute Is the Best Classifier?

Day	Outlook	Temp	Humidity	Wind	PlayTennis
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D10	Rain	Mild	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

$$Gain(S_{Rain}, Temp) = 0.0192$$

$$Gain(SRain, Humidity) = \ 0.0192$$

$$Gain(SRain, Wind) = 0.97$$



Hypothesis space search by ID3 (Cont.)

- ID3 maintains only a **single** current hypothesis
  - Contrasts with Candidate-Elimination method which maintains the set of all hypotheses consistent with the available training examples.
- It does not have the ability to determine how many alternative decision trees are consistent with the available training data.

Hypothesis space search by ID3 (Cont.)

- No backtracking
  - Local minima
- All training examples are used at each step in the search
  - Statistically-base decision to refine the current hypothesis.
  - Contrasts with Find-S and Candidate-Elimination (based on individual-training examples)
  - Robust to noisy data

ID3's inductive bias

- What is the inductive bias?
  - Policy by which ID3 generalizes from observed training examples to classify unseen instances.
- Given a set of training examples, there are typically many decision trees consistent with that set.
  - e.g., what would be another decision tree consistent with the example training data?
- Of all these, which one does ID3 construct?
  - First acceptable tree found in greedy search

ID3's inductive bias (Cont.)

- ID3 search strategy:
  - Selects in favor of shorter trees over long ones
  - Selects trees that place the attributes with highest information gain closest to the root.

ID3’s inductive bias (Cont.)

- Approximate inductive bias of ID3
  – Shorter trees are preferred over larger trees
- Bias is a preference for some hypotheses, rather than a restriction of hypothesis space H

ID3’s inductive bias (Cont.)

- A closer approximation to the inductive bias of ID3
  – Shorter trees are preferred over longer trees
  – Trees that place high information gain attributes close to the root are preferred over those that do not.

ID3's inductive bias (Cont.)

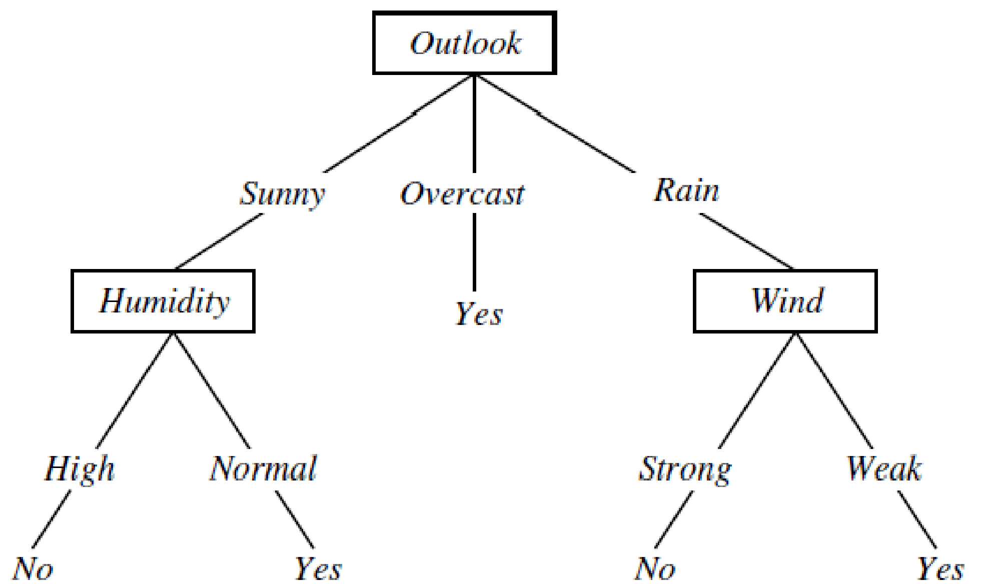
- Occam's razor: prefer the simplest hypothesis that fits the data
- Why prefer short hypotheses?
  - Fewer short hypotheses, than long hypotheses
    - Computational efficiency (avoid redundant, spurious attributes)
    - Avoid over-fitting training examples

Overfitting in decision trees

Consider adding noisy training example #15

Sunny, Hot, Normal, Strong, PlayTennis = No

What effect on earlier tree?



Overfitting in decision trees (Cont.)

Formal definition of **overfittin*g:***

Given a hypothesis space *H, a hypothesis h of H is said* to **overfit the training data if there exists some**

alternative *h' in H, such that*

TrainingError(h) < TrainingError(h'),

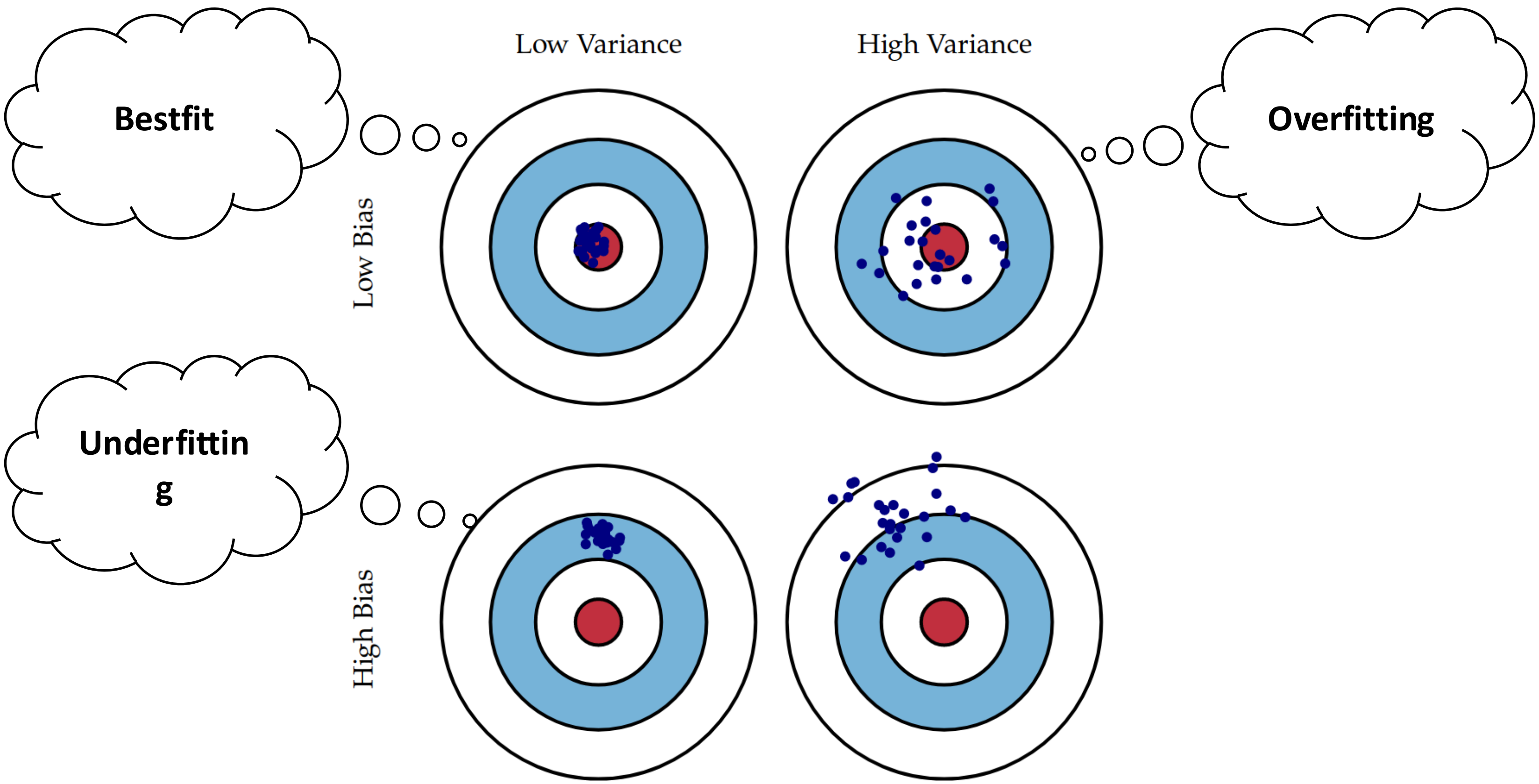
but

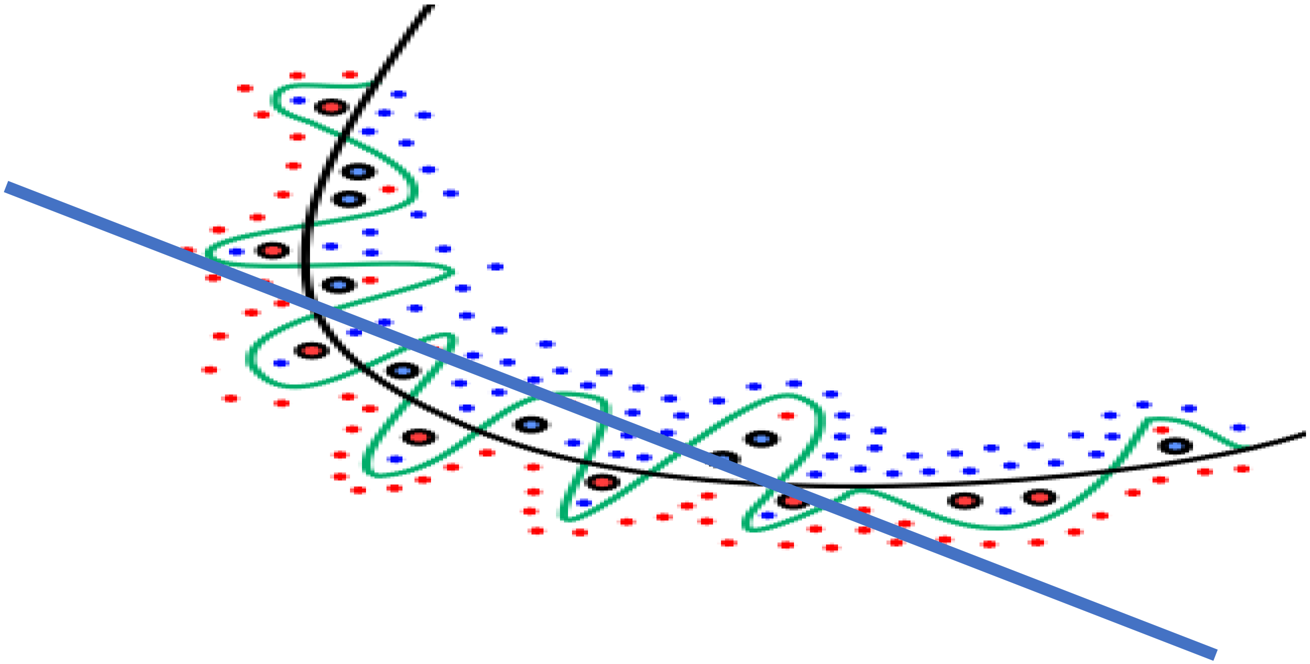
TestError(h') < TestError(h).

Overfitting

- In machine learning, overfitting occurs when an algorithm fits too closely or even exactly to its training data, resulting in a model that can't make accurate predictions or conclusions from any data other than the training data.

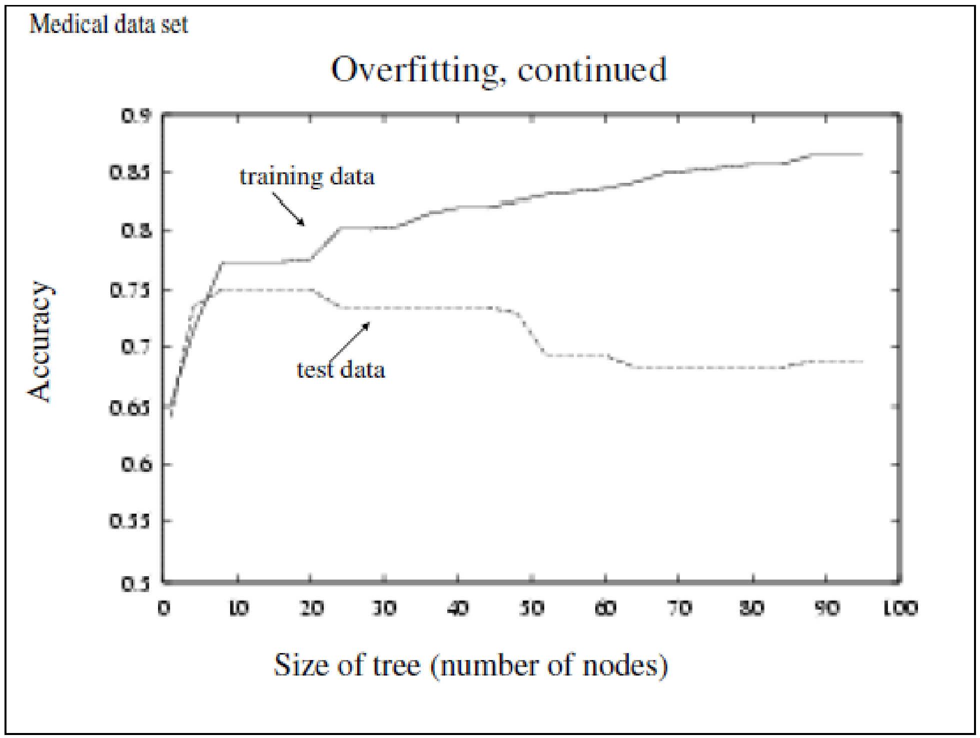
Bias and Varaince





Green model : overfitting
Black model generalize better than the green one
Blue model : underfitting

Overfitting in decision trees (Cont.)



Overfitting in decision trees (Cont.)

- How to avoid overfitting:
  - approaches that stop growing the tree earlier, before it reaches the point where it perfectly classifies the training data,
  - approaches that allow the tree to overfit the data, and then post-prune the tree.
  - a key question is what criterion is to be used to determine the correct final tree size

Overfitting in decision trees (Cont.)

- How to select best tree:
  - Measure performance over training data
  - Measure performance over separate validation data set
  - Minimum Description Length MDL:

  Minimize *size(tree)+size(misclassifications(tree))*

Overfitting in decision trees (Cont.)

- Interesting observation about noise:

  Quinlan found that adding a low level of noise to a training set resulted in trees with higher classification errors on training set, but with lower classification errors on new objects (not in training set).

Overfitting in decision trees (Cont.)

- Moral: “It is counter-productive to eliminate noise from the attribute information in the training set if these same attributes will be subject to high noise levels when the induced decision tree is put to use.”

Pruning the tree

- Pruning:
  - Remove subtree below a decision node.
  - Create a leaf node there and assign most common classification of the training examples affiliated with that node.
  - Helps get rid of nodes due to overfitting.

Pruning the tree (Cont.)

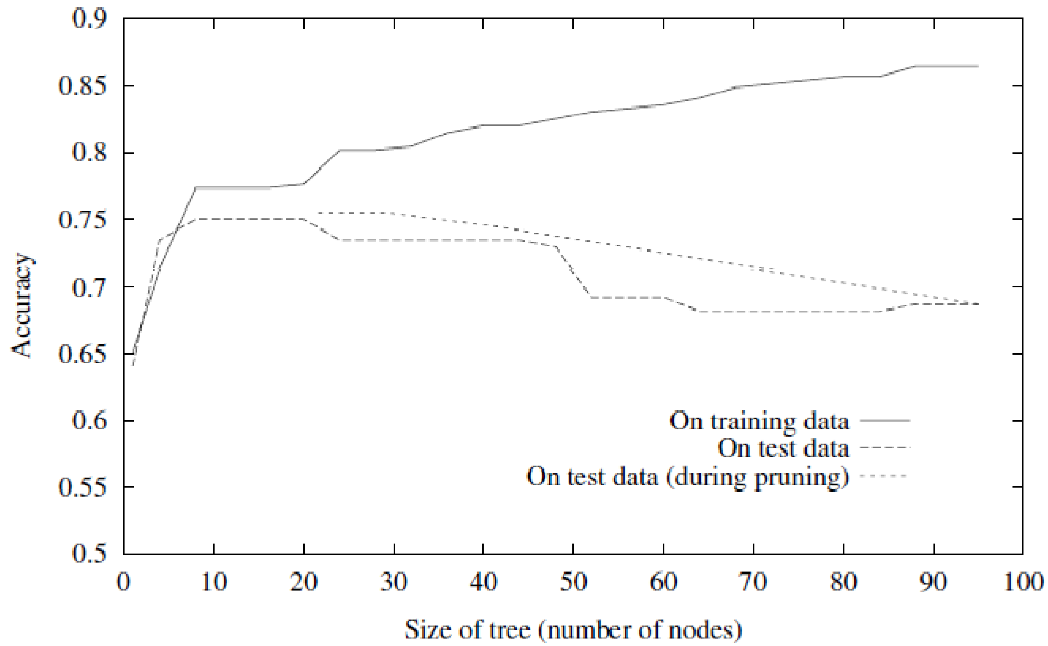
- Reduced-error pruning:
  - Consider each decision node as candidate for pruning.
  - For each node, try pruning node. Measure accuracy of pruned tree over validation set.
  - Select single-node pruning that yields best increase in accuracy over validation set.

Pruning the tree (Cont.)

- Reduced-error pruning:
  – If no increase, select one of the single-node prunings that does not decrease accuracy.
  – If all prunings decrease accuracy, then don't prune.

    Otherwise, continue this process until further pruning is harmful.

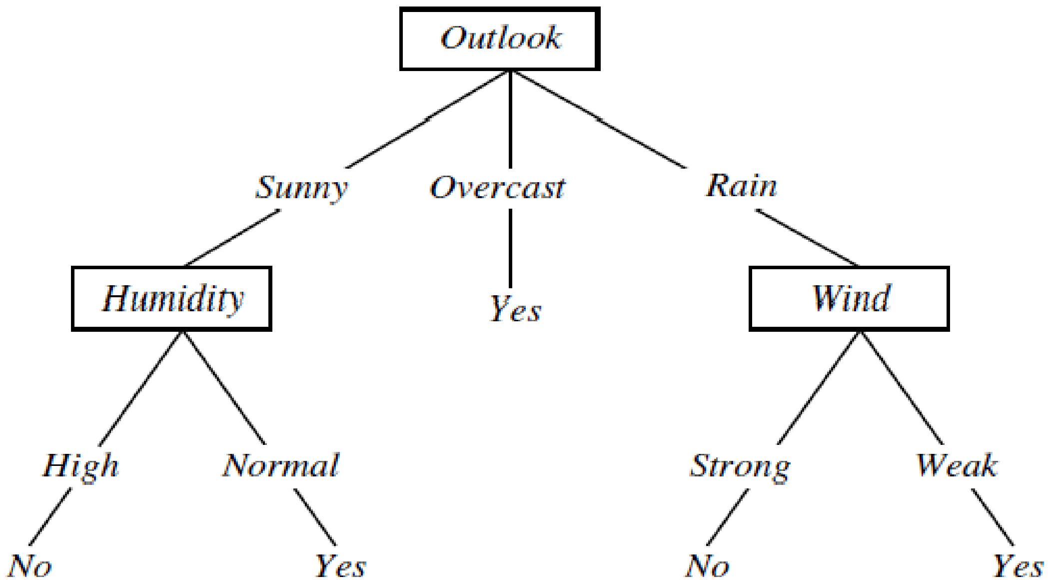
Effect of reduced-error pruning



Rule post-pruning

1. Convert tree to equivalent set of rules
2. Prune each rule independently of others
3. Sort final rules into desired sequence for use

Rule post-pruning (Cont.)



$$\begin{array}{ll} \text{IF} & (Outlook = Sunny) \wedge (Humidity = High) \\ \text{THEN} & PlayTennis = No \end{array}$$

$$\begin{array}{ll} \text{IF} & (Outlook = Sunny) \wedge (Humidity = Normal) \\ \text{THEN} & PlayTennis = Yes \end{array}$$

$\ldots$

Why convert the decision tree to rules before pruning? There are three main advantages.

1. Allow to prune only some paths leaving the attribute rather to pruning all subtree under the attribute.

2. Converting to rules removes the distinction between attribute tests that occur near the root of the tree and those that occur near the leaves.

1. Converting to rules improves readability. Rules are often easier for to understand.

References

- Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani, An Introduction to Statistical Learning with Applications in R, second edition, Springer, New York, 2021.
- Tom M. Mitchell. Machine Learning, McGraw Hill Inter. Editions, 1997.