

Decision Trees

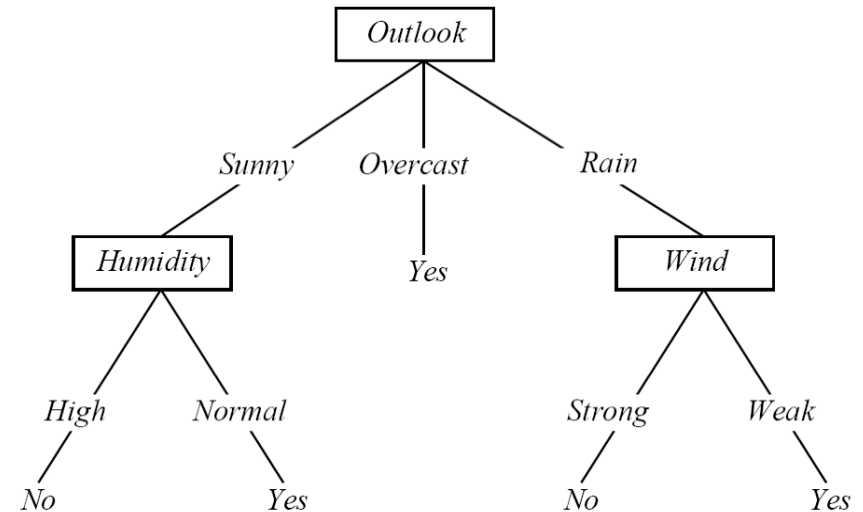
- Learn from labeled observations - supervised learning
- Represent the knowledge learned in form of a tree

Example: learning when to play tennis.

- Examples/observations are days with their observed characteristics and whether we played tennis or not

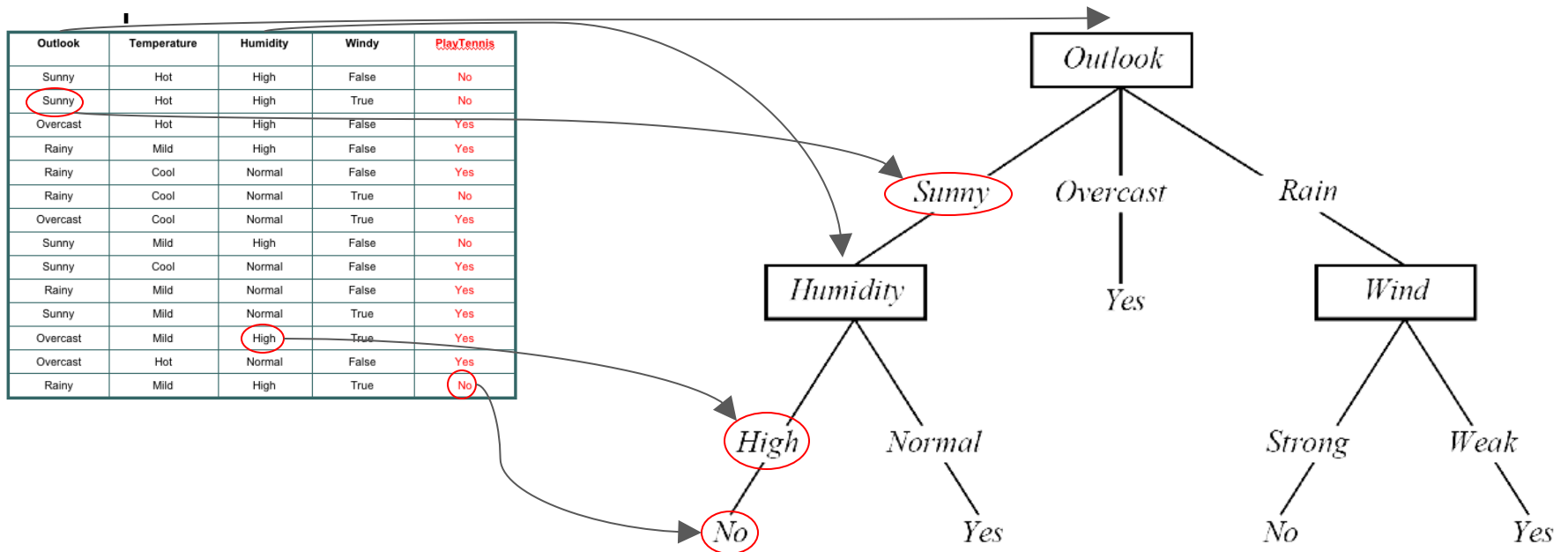
Play Tennis Example

Outlook	Temperature	Humidity	Windy	PlayTennis
Sunny	Hot	High	False	No
Sunny	Hot	High	True	No
Overcast	Hot	High	False	Yes
Rainy	Mild	High	False	Yes
Rainy	Cool	Normal	False	Yes
Rainy	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Sunny	Mild	High	False	No
Sunny	Cool	Normal	False	Yes
Rainy	Mild	Normal	False	Yes
Sunny	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Rainy	Mild	High	True	No



Interpreting a DT

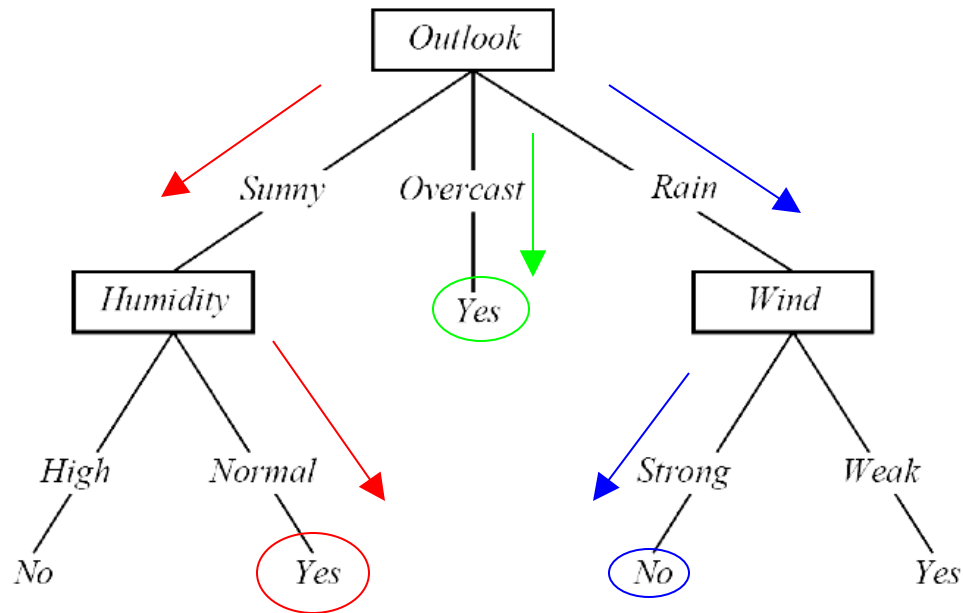
DT \equiv Decision
Tree



- A DT uses the features of an observation table as nodes and the feature values as links.
- All feature values of a particular feature need to be represented as links.
- The target feature is special - its values show up as leaf nodes in the DT.

Interpreting a DT

Each path from the root of the DT to a leaf can be interpreted as a decision rule.



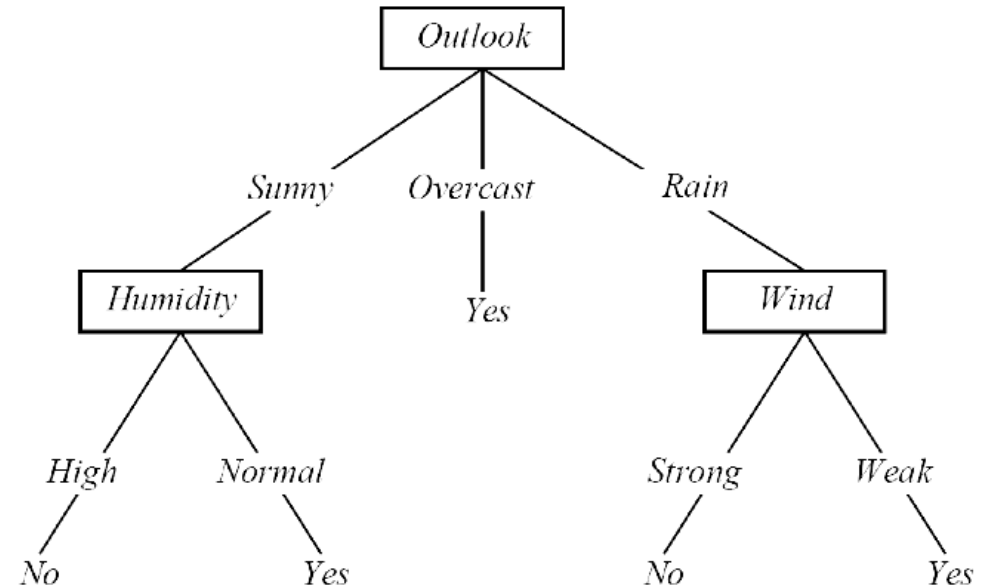
IF *Outlook* = *Sunny* AND *Humidity* = *Normal* THEN *Playtennis* = *Yes*

IF *Outlook* = *Overcast* THEN *Playtennis* = *Yes*

IF *Outlook* = *Rain* AND *Wind* = *Strong* THEN *Playtennis* = *No*

DT: Explanation & Prediction

Outlook	Temperature	Humidity	Windy	PlayTennis
Sunny	Hot	High	False	No
Sunny	Hot	High	True	No
Overcast	Hot	High	False	Yes
Rainy	Mild	High	False	Yes
Rainy	Cool	Normal	False	Yes
Rainy	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Sunny	Mild	High	False	No
Sunny	Cool	Normal	False	Yes
Rainy	Mild	Normal	False	Yes
Sunny	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Rainy	Mild	High	True	No



Explanation: the DT summarizes (explains) all the observations in the table perfectly \Rightarrow 100% Accuracy

Prediction: once we have a DT (or model) we can use it to make predictions on observations that are not in the original training table, consider:

Outlook = Sunny, Temperature = Mild, Humidity = Normal, Windy = False, Playtennis = ?

Constructing DTs

- How do we choose the attributes and the order in which they appear in a DT?
 - Recursive partitioning of the original data table
 - Heuristic - each generated partition has to be “less random” (entropy reduction) than previously generated partitions

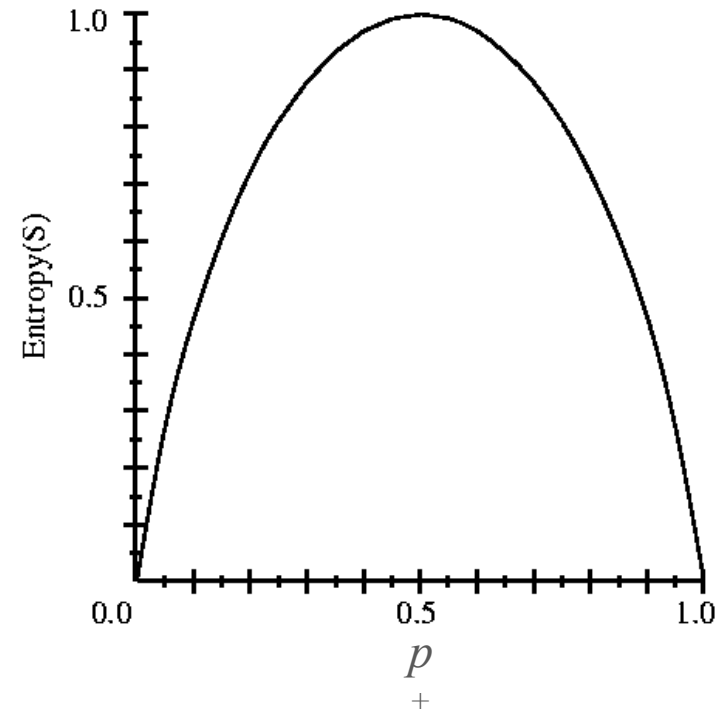
Entropy

- S is a sample of training examples
- p^+ is the proportion of positive examples in S
- p^- is the proportion of negative examples in S
- Entropy measures the impurity (randomness) of S

S {

Outlook	Temperature	Humidity	Windy	PlayTennis
Sunny	Hot	High	False	No
Sunny	Hot	High	True	No
Overcast	Hot	High	False	Yes
Rainy	Mild	High	False	Yes
Rainy	Cool	Normal	False	Yes
Rainy	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Sunny	Mild	High	False	No
Sunny	Cool	Normal	False	Yes
Rainy	Mild	Normal	False	Yes
Sunny	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Rainy	Mild	High	True	No

$$Entropy(S) = Entropy([9+, 5-]) = .94$$



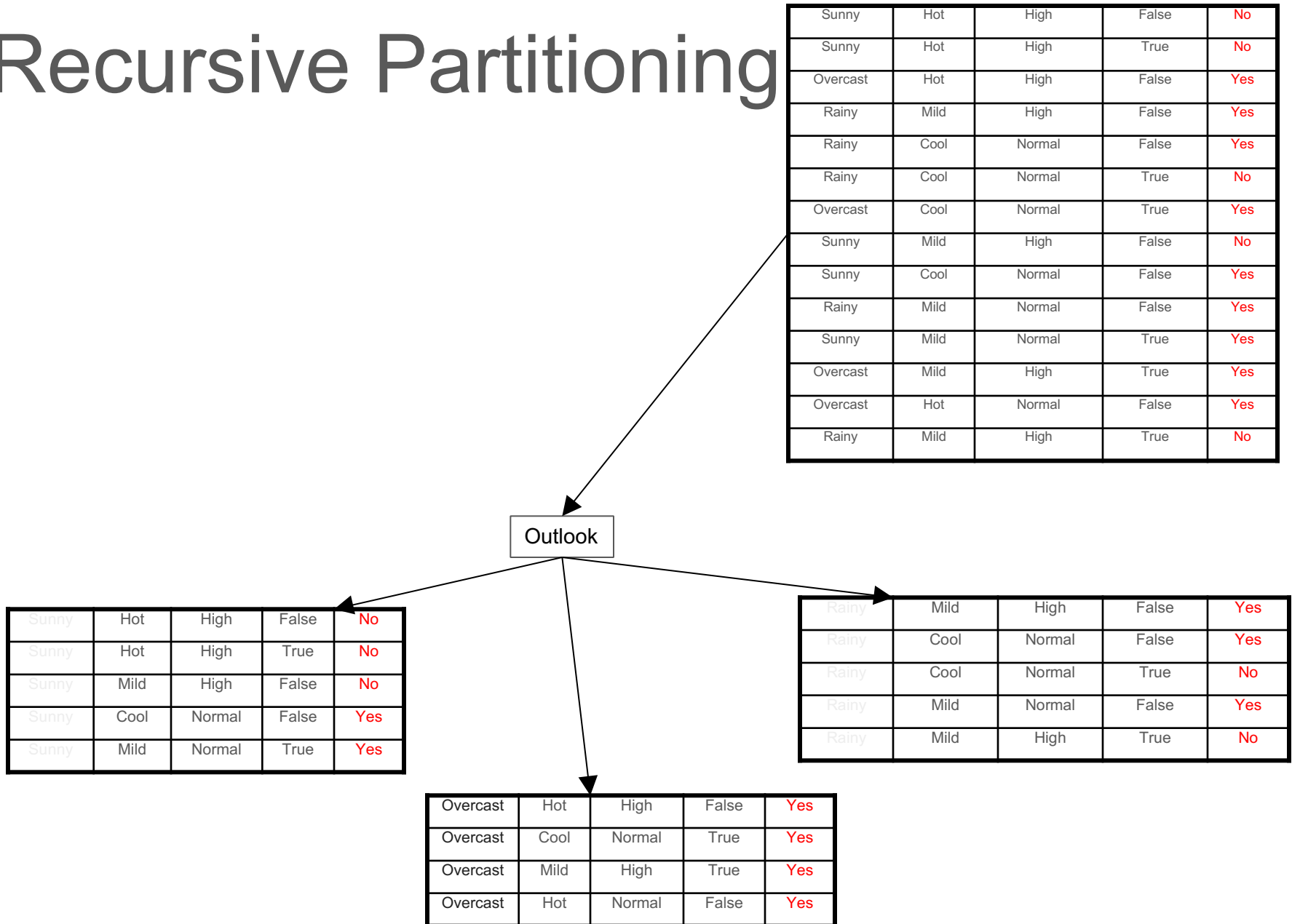
$$\square Entropy(S) \equiv -p^+ \log_2 p^+ - p^- \log_2 p^-$$

Recursive Partitioning

Our data set:

Outlook	Temperature	Humidity	Windy	PlayTennis
Sunny	Hot	High	False	No
Sunny	Hot	High	True	No
Overcast	Hot	High	False	Yes
Rainy	Mild	High	False	Yes
Rainy	Cool	Normal	False	Yes
Rainy	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Sunny	Mild	High	False	No
Sunny	Cool	Normal	False	Yes
Rainy	Mild	Normal	False	Yes
Sunny	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Rainy	Mild	High	True	No

Recursive Partitioning



Recursive Partitioning

Outlook

Sunny	Hot	High	False	No
Sunny	Hot	High	True	No
Sunny	Mild	High	False	No
Sunny	Cool	Normal	False	Yes
Sunny	Mild	Normal	True	Yes

Rainy	Mild	High	False	Yes
Rainy	Cool	Normal	False	Yes
Rainy	Cool	Normal	True	No
Rainy	Mild	Normal	False	Yes
Rainy	Mild	High	True	No

Overcast	Hot	High	False	Yes
Overcast	Cool	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes

Recursive Partitioning

Outlook

Sunny	Hot	High	False	No
Sunny	Hot	High	True	No
Sunny	Mild	High	False	No
Sunny	Cool	Normal	False	Yes
Sunny	Mild	Normal	True	Yes

Humidity

Overcast	Hot	High	False	Yes
Overcast	Cool	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes

Rainy	Mild	High	False	Yes
Rainy	Cool	Normal	False	Yes
Rainy	Cool	Normal	True	No
Rainy	Mild	Normal	False	Yes
Rainy	Mild	High	True	No

Sunny	Cool	Normal	False	Yes
Sunny	Mild	Normal	True	Yes

Sunny	Hot	High	False	No
Sunny	Hot	High	True	No
Sunny	Mild	High	False	No

Recursive Partitioning

Outlook

Sunny	Hot	High	False	No
Sunny	Hot	High	True	No
Sunny	Mild	High	False	No
Sunny	Cool	Normal	False	Yes
Sunny	Mild	Normal	True	Yes

Humidity

Overcast	Hot	High	False	Yes
Overcast	Cool	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes

Sunny	Cool	Normal	False	Yes
Sunny	Mild	Normal	True	Yes

Rainy	Mild	High	False	Yes
Rainy	Cool	Normal	False	Yes
Rainy	Cool	Normal	True	No
Rainy	Mild	Normal	False	Yes
Rainy	Mild	High	True	No

Windy

Rainy	Mild	High	False	Yes
Rainy	Cool	Normal	False	Yes
Rainy	Mild	Normal	False	Yes

Rainy	Cool	Normal	True	No
Rainy	Mild	High	True	No

Sunny	Hot	High	False	No
Sunny	Hot	High	True	No
Sunny	Mild	High	False	No