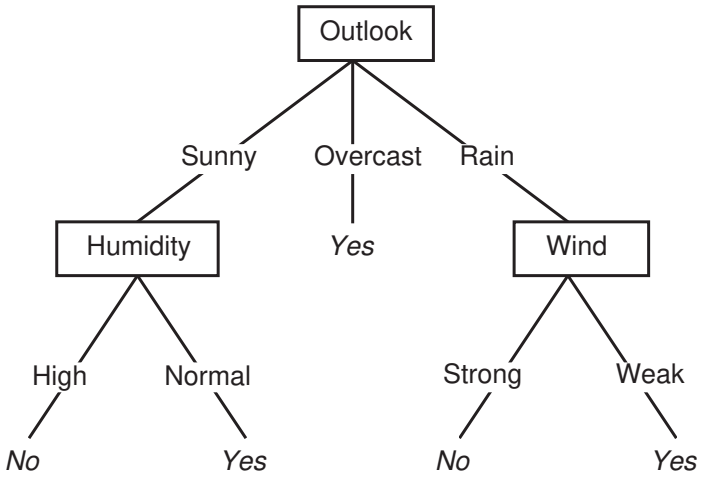
# Typical Decision Tree

---

## Decision Tree for Play-Tennis?




**Copyright Tom Mitchell, 15 September 1994**

# Decision Trees

---

How would we represent:

- AND, OR, XOR
- (A $\wedge$ B) #$\vee$# (C $\wedge$ D $\wedge$ E)
- X of N

Decision tree representation:

- Attempt to test most discriminating property first
- Drastic misclassification possible if early attribute error
- Each tree path corresponds to a rule

**Copyright Tom Mitchell, 15 September 1994**

# Appropriate Problems for Decision Tree Learning

- Instances represented by attribute-value pairs
- Target function has discrete output values
- Disjunctive descriptions might be required
- Training data may contain errors

Copyright Tom Mitchell, 15 September 1994

# Top Down Induction of Decision Trees

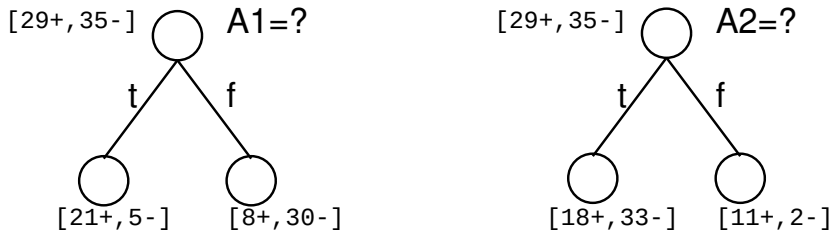
---

after ID3 [Quinlan, 1983]

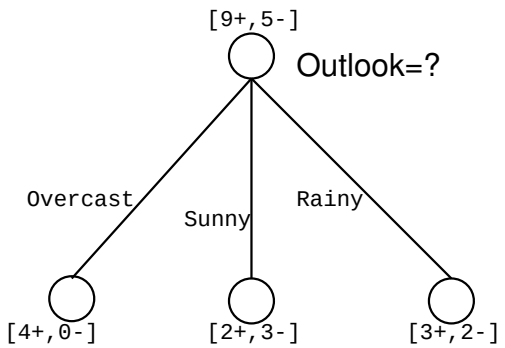
Basic Decision Tree Learning Method:

1. Choose "best" attribute
2. Extend tree by adding new branch for each attribute value
3. Sort training examples to leaf nodes
4. If training examples unambiguously classified Then Stop Else Repeat steps 1-4 for leaf nodes

Which attribute is best??



## What Information is Gained By Testing Attribute A?




Entropy, E(I), of a set of instances I, containing
p positive and n negative examples

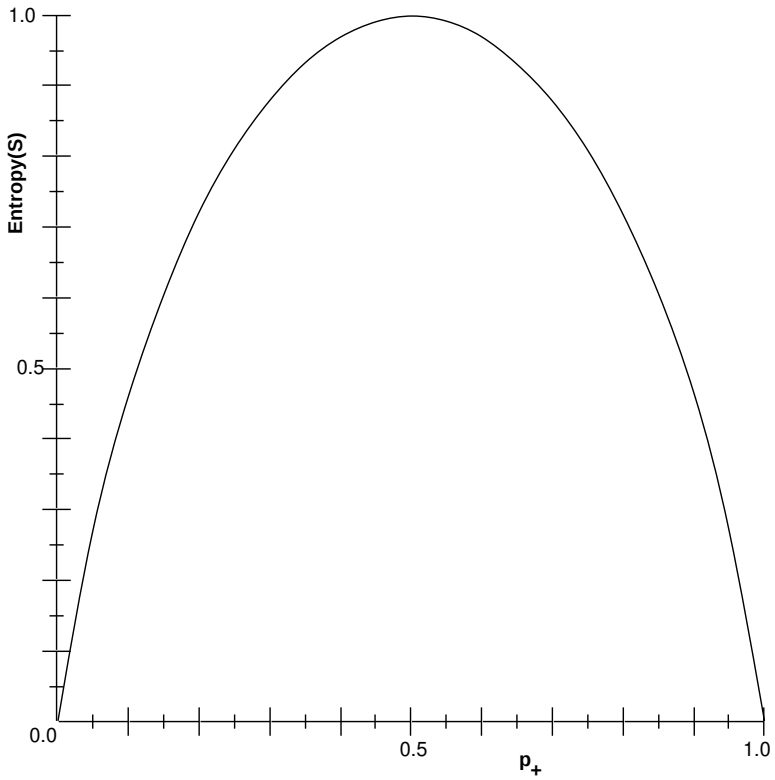
$$E(I) = -\frac{p}{p+n}\log_2\frac{p}{p+n} - \frac{n}{p+n}\log_2\frac{n}{p+n}$$

Gain(A,I) is the expected reduction in Entropy due to attribute A:

$$Gain(A,I) = E(I) - \sum_{\text{descendant } j} \frac{p_j + n_j}{p+n} E(I_j)$$

the jth descendant of I is the set of instances with value Vj for A

# Entropy Relative to a Boolean Classification



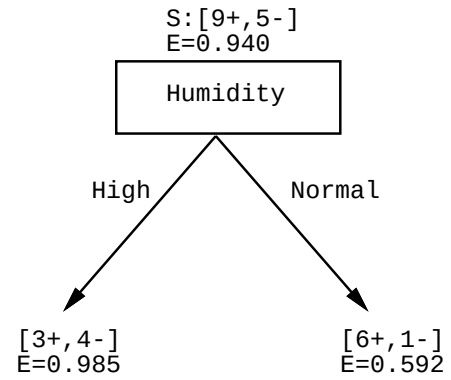

Copyright Tom Mitchell, 15 September 1994

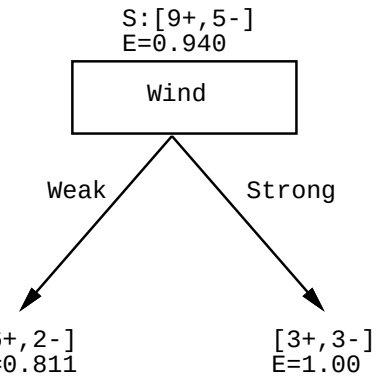
# A Concept Learning Task

| Day | Outlook | Temp | Humidity | Wind | Play-Tenni |
|---|---|---|---|---|---|
| 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 |

Copyright Tom Mitchell, 15 September 1994

## Which Attribute is the Best Classifier?



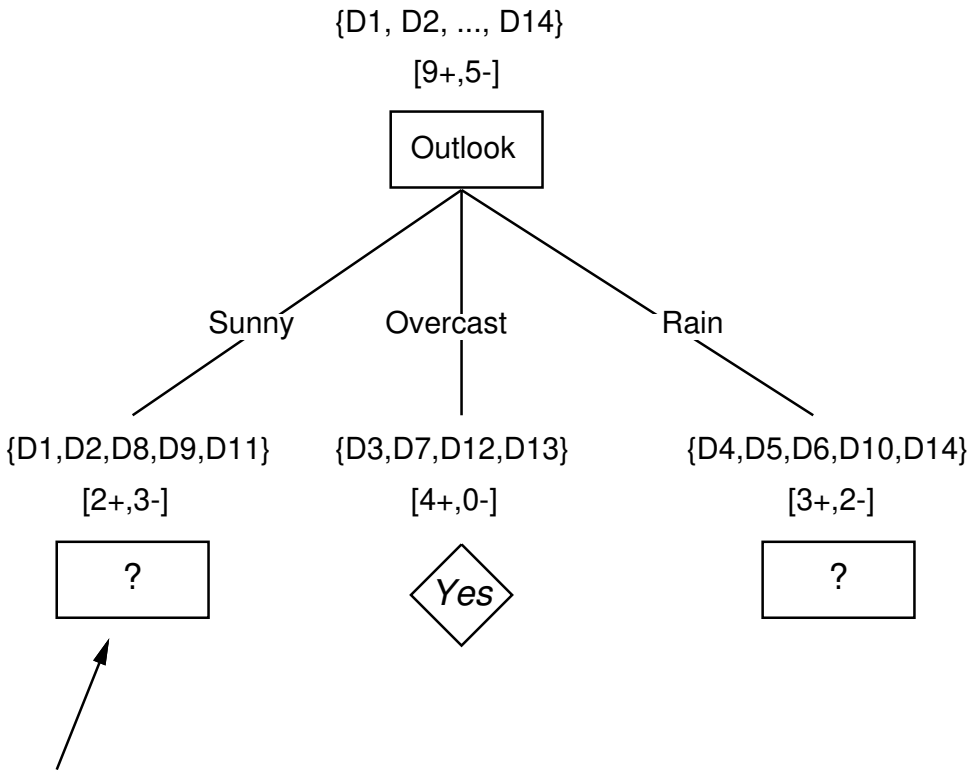

Gain(S,Humidity)
= .940 - (7/14).985 - (7/14).592
= .151

Gain(S,Wind)
= .940 - (8/14).811 - (6/14)1.0
= .048

Classifying examples by Humidity provides more information gain than by Wind

**Copyright Tom Mitchell, 15 September 1994**

## Partially Learned Decision Tree




Which attribute should be tested here?

$S_{sunny}$ = {D1,D2,D8,D9,D11}

Gain($S_{sunny}$,Humidity)= .970 - (3/5)0.0 - (2/5)0.0 = .970

Gain($S_{sunny}$,Temperature)= .970 - (2/5)0.0 - (2/5)1.0 - (1/5)0.0 = .570

Gain($S_{sunny}$,Wind)= .970 - (2/5)1.0 - (3/5).918 = .019

**Copyright Tom Mitchell, 15 September 1994**

# Hypothesis Space Search in ID3

- Hypothesis space is complete!
    - Target function surely in there...
- Maintain a single current hypothesis
    - Can’t play 20 questions...
- No back tracking
    - Local minima...
- Statisically-based search choices
    - Robust to noisy data...

Copyright Tom Mitchell, 15 September 1994

# Restriction Bias versus Preference Bias

Restriction bias limits the set of available hypotheses

- Also called language bias

Preference bias orders the available hypotheses

- Also called search bias

Copyright Tom Mitchell, 15 September 1994

# ID3 Bias: Prefer Short Hypotheses

---

but why?

**Occam's Razor:** Prefer the simplest hypothesis

Argument in favor:

- There are fewer short hypotheses than long
- A short hypothesis that fits training data is unlikely to be coincidence
- A long hypothesis that fits data might be

Argument opposed:

- There are many small sets of hypotheses
- E.g., All trees that use attributes beginning with "M"
- What's special about small sets that are short?

Copyright Tom Mitchell, 15 September 1994