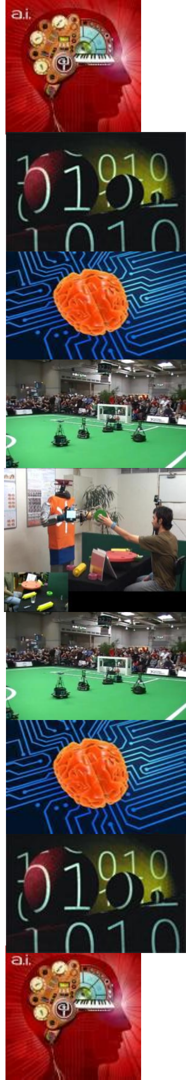
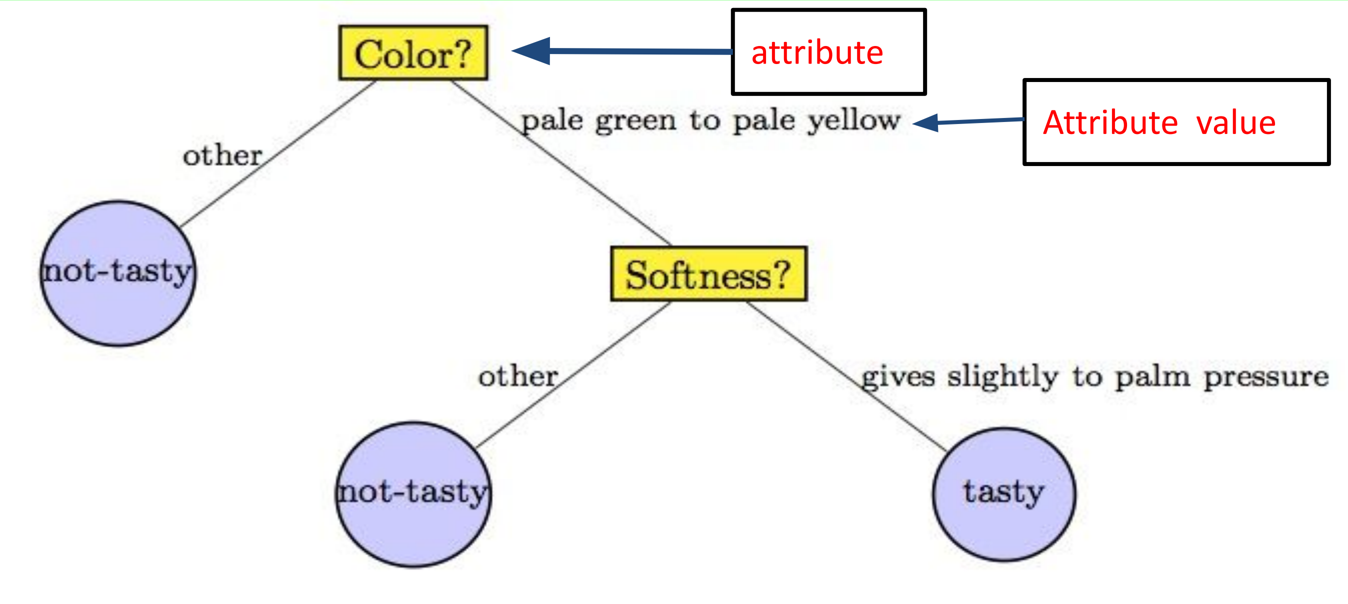
Decision Trees

Introduction

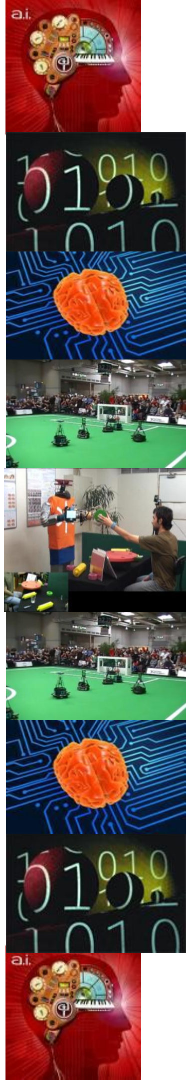
- Perform supervised learning
- Classifiers
  - Training
  - Testing
- Interpretability
- Tree size
- Creating decision trees. (ID3, C4.5 , C5)
- Can be classification, regression trees or hybrid.



Example

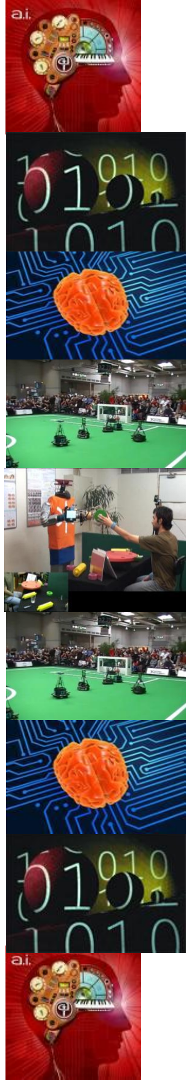


Shalev-Schwartz et al. 2014



Example Dataset

Day	Outlook	Temperature	Humidity	Wind	Play Netball?
1	Sunny	Hot	High	Weak	No
2	Sunny	Hot	High	Strong	No
3	Overcast	Hot	High	Weak	Yes
4	Rain	Mild	High	Weak	Yes
5	Rain	Cool	Normal	Weak	Yes
6	Rain	Cool	Normal	Strong	No
7	Overcast	Cool	Normal	Strong	Yes
8	Sunny	Mild	High	Weak	No
9	Sunny	Cool	Normal	Weak	Yes
10	Rain	Mild	Normal	Weak	Yes
11	Sunny	Mild	Normal	Strong	Yes
12	Overcast	Mild	High	Strong	Yes
13	Overcast	Hot	Normal	Weak	Yes
14	Rain	Mild	High	Strong	No



Inducing a Decision Tree -ID3

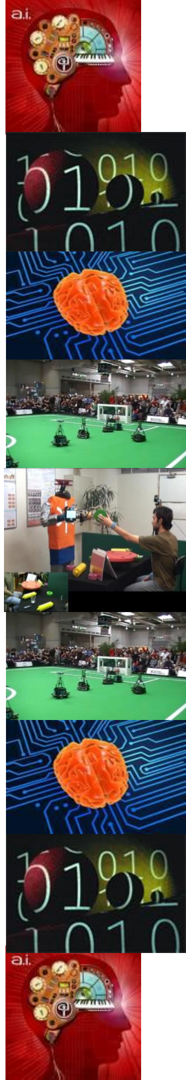
1. Use **Information Gain**

   Attribute with the highest **Gain** is the root node

2. Calculate the **Information Gain**

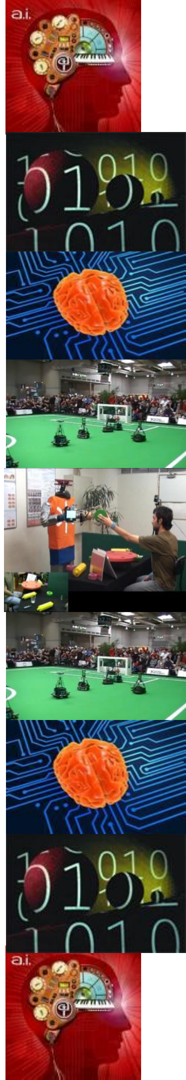
3. First calculate the **Entropy -** measures uncertainty, chaos

$$E(D) = \sum_{i=1}^{n} -p(i) log_2 p(i)$$



Creating The Decision Tree

1. COMPUTE THE **ENTROPY** FOR DATA-SET **ENTROPY(S)**
2. FOR EVERY ATTRIBUTE/FEATURE:
    1. CALCULATE ENTROPY FOR ALL OTHER VALUES **ENTROPY(A)**
    2. TAKE **AVERAGE INFORMATION ENTROPY** FOR THE CURRENT ATTRIBUTE
    3. CALCULATE **GAIN** FOR THE CURRENT ATTRIBUTE
3. PICK THE **HIGHEST GAIN ATTRIBUTE.**
4. **REPEAT** UNTIL WE GET THE TREE WE DESIRED.



How Do We Calculate Gain

- Calculate **Entropy** (Amount of uncertainity in dataset):

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

p = +ve class
n = -ve class

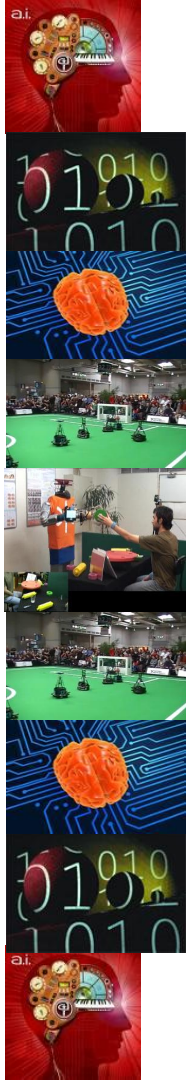
- Calculate **Average Information**:

$$I(Attribute) = \sum \frac{p_i + n_i}{p+n} Entropy(A)$$

- Calculate **Information Gain**: (Difference in Entropy before and after splitting dataset on attribute A)

$$Gain = Entropy(S) - I(Attribute)$$

***S* = dataset**



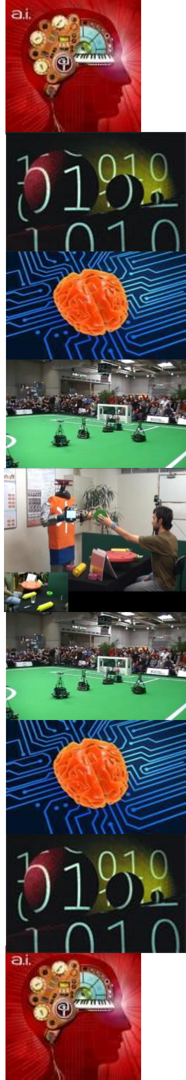
Dataset Entropy

- Calculate **Entropy(S):**

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

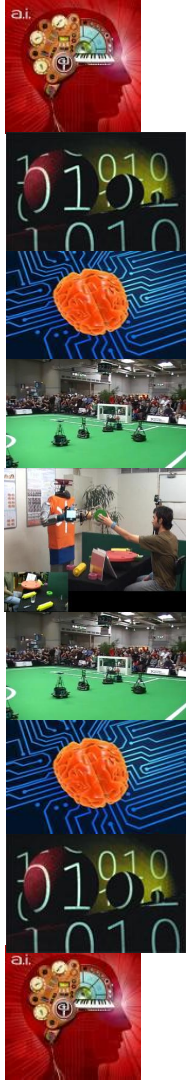
$$Entropy(S) = \frac{-9}{9+5} log_2 (\frac{9}{9+5}) - \frac{5}{9+5} log_2 (\frac{5}{9+5})$$

$$Entropy(S) = \frac{-9}{14} log_2 (\frac{9}{14}) - \frac{5}{14} log_2 (\frac{5}{14}) \quad = 0.940$$



Example Dataset

Day	Outlook	Temperature	Humidity	Wind	Play Netball?
1	Sunny	Hot	High	Weak	No
2	Sunny	Hot	High	Strong	No
3	Overcast	Hot	High	Weak	Yes
4	Rain	Mild	High	Weak	Yes
5	Rain	Cool	Normal	Weak	Yes
6	Rain	Cool	Normal	Strong	No
7	Overcast	Cool	Normal	Strong	Yes
8	Sunny	Mild	High	Weak	No
9	Sunny	Cool	Normal	Weak	Yes
10	Rain	Mild	Normal	Weak	Yes
11	Sunny	Mild	Normal	Strong	Yes
12	Overcast	Mild	High	Strong	Yes
13	Overcast	Hot	Normal	Weak	Yes
14	Rain	Mild	High	Strong	No



Attribute Entropy

Outlook	Temperature	Humidity	Windy	Play
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

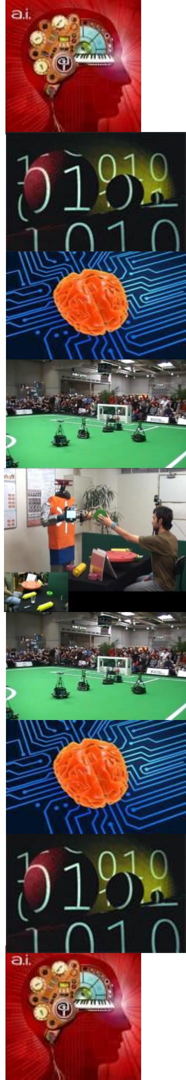
p = 2, n = 3, total = 5

calc entropy of Outlook when it is sunny only

- ENTROPY:

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

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



Attribute Entropy

- For each Attribute: (let say **Outlook**)
  - Calculate Entropy for each Values, i.e for 'Sunny', 'Rainy','Overcast'

Outlook	Play
Sunny	No
Sunny	No
Sunny	No
Sunny	Yes
Sunny	Yes

1

Outlook	Play
Rainy	Yes
Rainy	Yes
Rainy	No
Rainy	Yes
Rainy	No

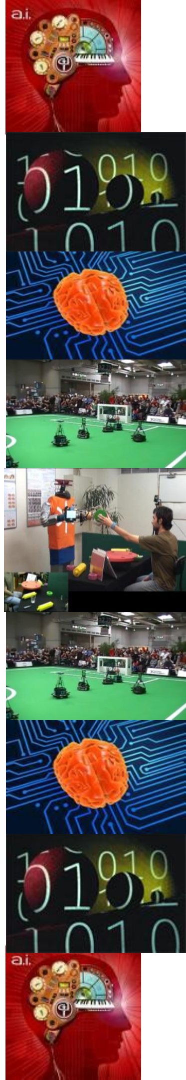
2

Outlook	Play
Overcast	Yes
Overcast	Yes
Overcast	Yes
Overcast	Yes

3

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

$$E(D) = \sum_{i=1}^{n} -p(i) log_2 p(i)$$



How Do We Calculate Gain

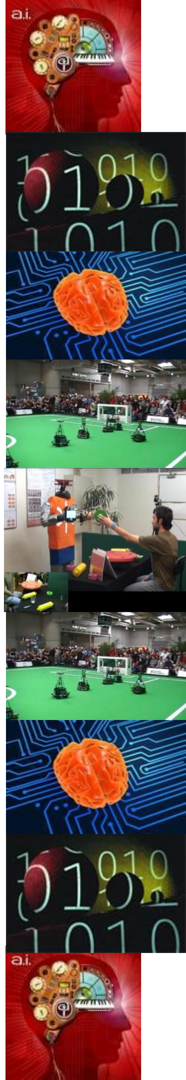
- Calculate **Average Information Entropy**:

$$I(Outlook) = \frac{p_{sunny} + n_{sunny}}{p+n} Entropy(Outlook = Sunny) +$$

$$\frac{p_{rainy} + n_{rainy}}{p+n} Entropy(Outlook = Rainy) +$$

$$\frac{p_{Overcast} + n_{Overcast}}{p+n} Entropy(Outlook = Overcast)$$

$$I(Outlook) = \frac{3+2}{9+5} * 0.971 + \frac{2+3}{9+5} * 0.971 + \frac{4+0}{9+5} * 0 \quad = 0.693$$



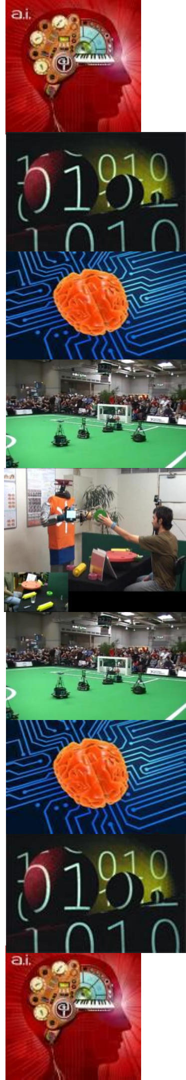
How Do We Calculate Gain

- Calculate **Gain**: attribute is Outlook

$$Gain = Entropy(S) - I(Attribute)$$

$$Entropy(S) = 0.940$$

$$Gain(Outlook) = 0.940 - 0.693 = 0.247$$



Temperature Entropy

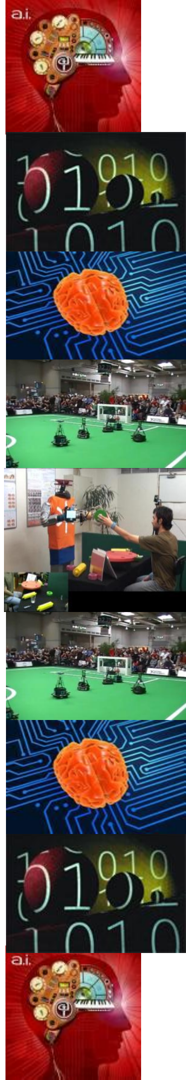
- Calculate **Average Information Entropy**:

$$I(Temperature) = \frac{p_{hot} + n_{hot}}{p+n} Entropy(Temperature = Hot) +$$

$$\frac{p_{mild} + n_{mild}}{p+n} Entropy(Temperature = Mild) +$$

$$\frac{p_{Cool} + n_{Cool}}{p+n} Entropy(Temperature = Cool)$$

$$I(Temperature) = \frac{2+2}{9+5} * 1 + \frac{4+2}{9+5} * 0.918 + \frac{3+1}{9+5} * 0.811 => 0.911$$



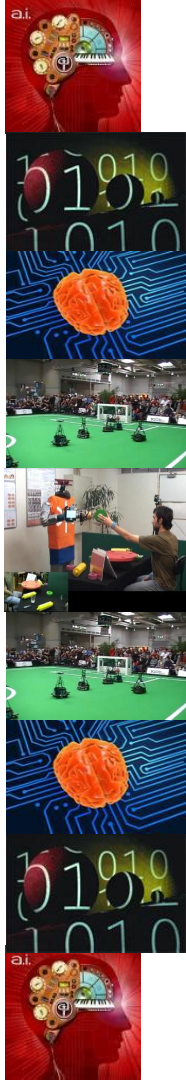
Temperature Gain

- Calculate **Gain**: attribute is Temperature

$$Gain = Entropy(S) - I(Attribute)$$

$$Entropy(S) = 0.940$$

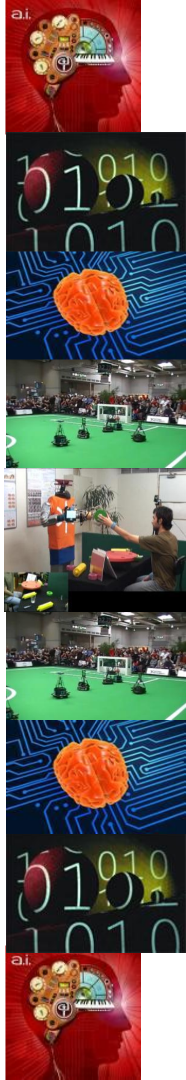
$$Gain(Temperature) = 0.940 - 0.911 = 0.029$$



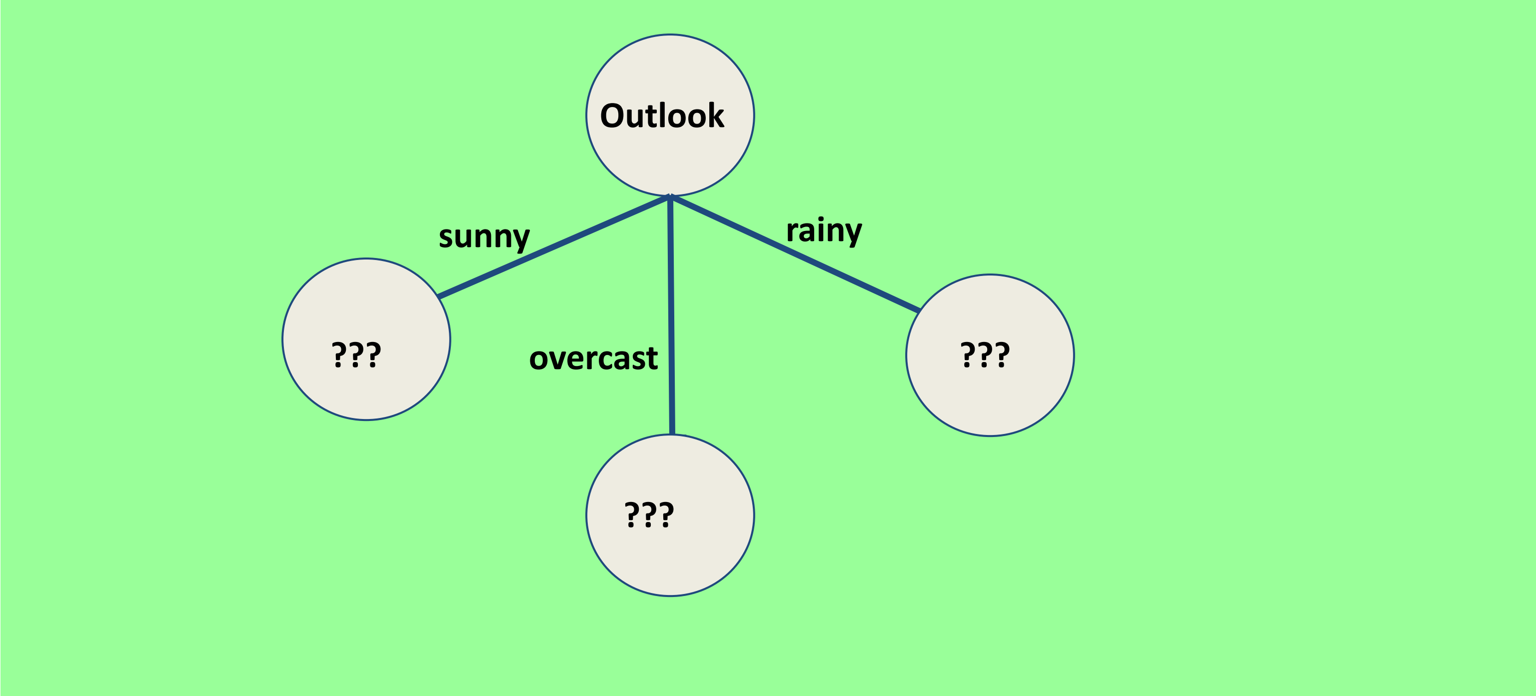
Information Gain

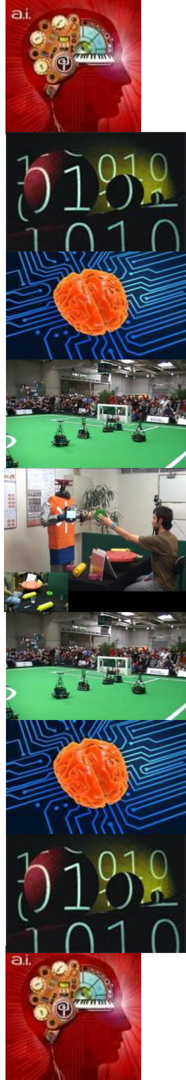
Attribute	IG Value
Outlook	**0.247**
Temperature	0.029
Humidity	0.152
Wind	0.048

Root Node = Outlook



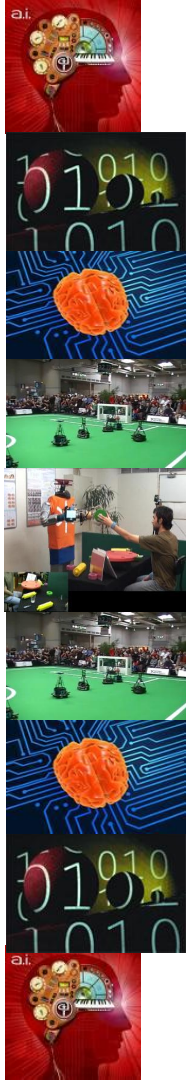
How Do We Building the DT





Example Dataset

Day	Outlook	Temperature	Humidity	Wind	Play Netball?
1	Sunny	Hot	High	Weak	No
2	Sunny	Hot	High	Strong	No
3	Overcast	Hot	High	Weak	Yes
4	Rain	Mild	High	Weak	Yes
5	Rain	Cool	Normal	Weak	Yes
6	Rain	Cool	Normal	Strong	No
7	Overcast	Cool	Normal	Strong	Yes
8	Sunny	Mild	High	Weak	No
9	Sunny	Cool	Normal	Weak	Yes
10	Rain	Mild	Normal	Weak	Yes
11	Sunny	Mild	Normal	Strong	Yes
12	Overcast	Mild	High	Strong	Yes
13	Overcast	Hot	Normal	Weak	Yes
14	Rain	Mild	High	Strong	No



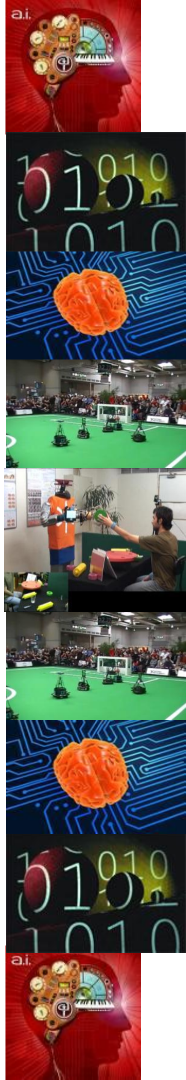
IG wrt Sunny

- For each Attribute: (let say **Humidity**):
  - Calculate Entropy for each Humidity, i.e for 'High' and 'Normal'

Outlook	Humidity	Play
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

  - Calculate **Average Information Entropy**: I(Humidity) = 0
  - Calculate **Gain:** = Sunny - Humidity = 0.971 -0 Gain = 0.971



IG wrt Sunny

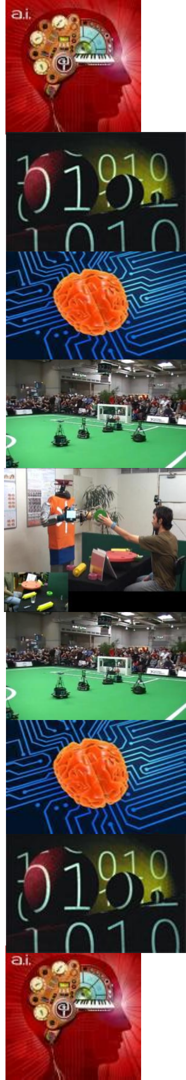
- For each Attribute: (let say **Windy**):
  - Calculate Entropy for each Windy, i.e for 'Strong' and 'Weak'

Outlook	Windy	Play
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

Avg = $\frac{2}{5}$ *1 + $\frac{3}{5}$*0.918 = 0.951

- Calculate **Average Information Entropy**: I(Windy) = 0.951
- Calculate **Gain:** **= 0.971 - 0.951 =** Gain = 0.020



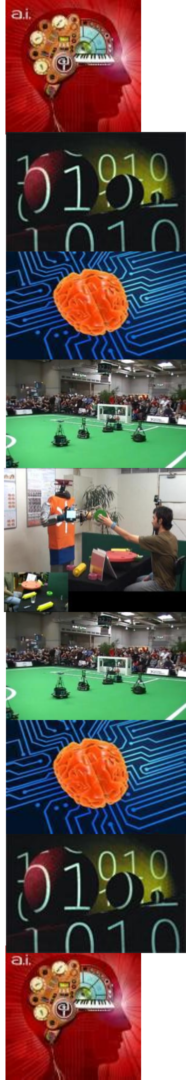
How Do We Calculate Gain

- For each Attribute: (let say **Temperature**):
  - Calculate Entropy for each Windy, i.e for 'Cool', 'Hot' and 'Mild'

Outlook	Temperature	Play
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

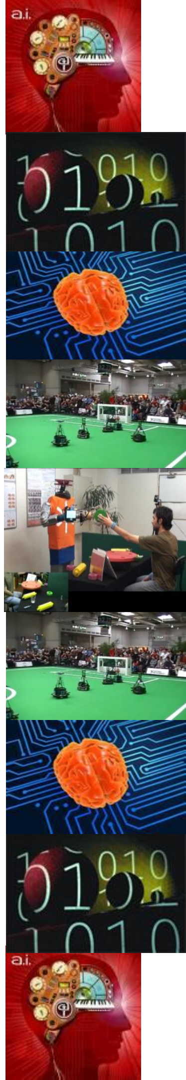
- Calculate **Average Information Entropy**: I(Temp) = 0.4
- Calculate **Gain:** = 0.971 - 0.571 = Gain = 0.571



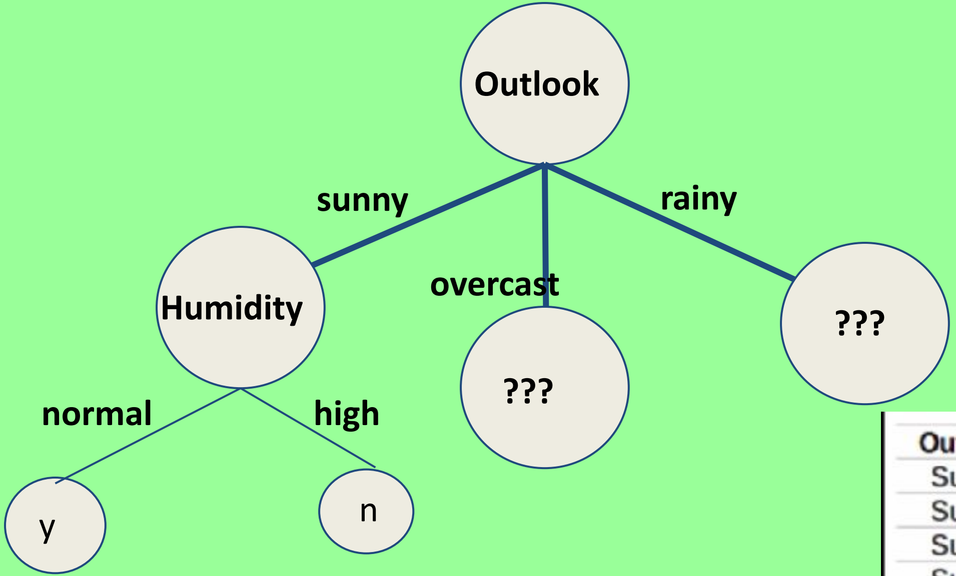
Information Gain wrt Sunny

Attribute	IG Value
Temperature	0.571
Humidity	**0.971**
Wind	0.02

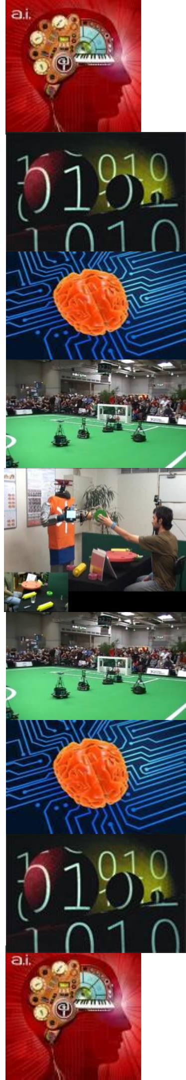
Next Node = Humidity



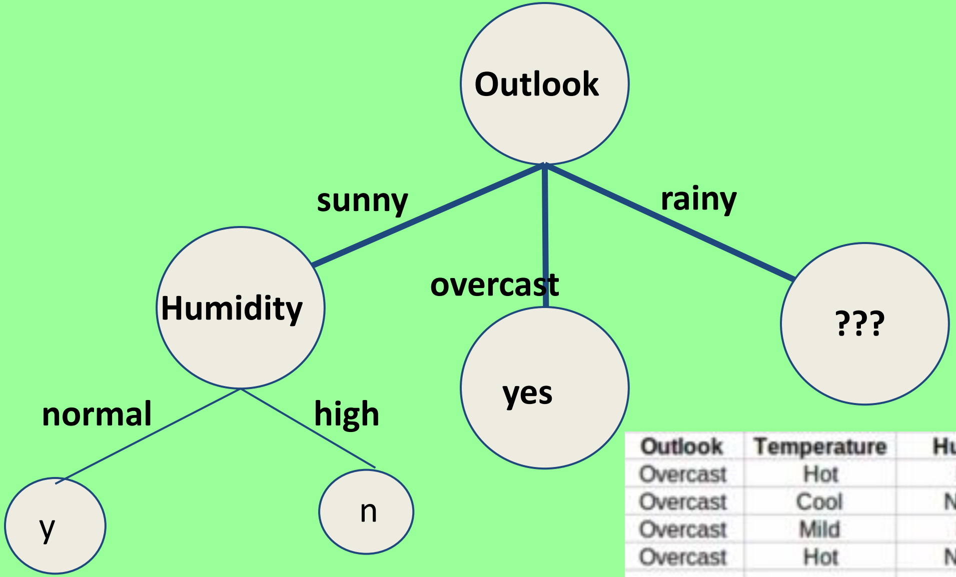
How Do We Calculate Gain



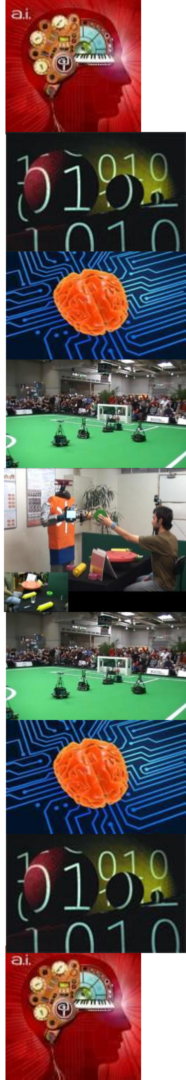
Outlook	Humidity	Play
Sunny	High	No
Sunny	High	No
Sunny	High	No
Sunny	Normal	Yes
Sunny	Normal	Yes



How Do We Calculate Gain



Outlook	Temperature	Humidity	Windy	Play
Overcast	Hot	High	Weak	[illegible]
Overcast	Cool	Normal	Strong	Yes
Overcast	Mild	High	Strong	Yes
Overcast	Hot	Normal	Weak	Yes

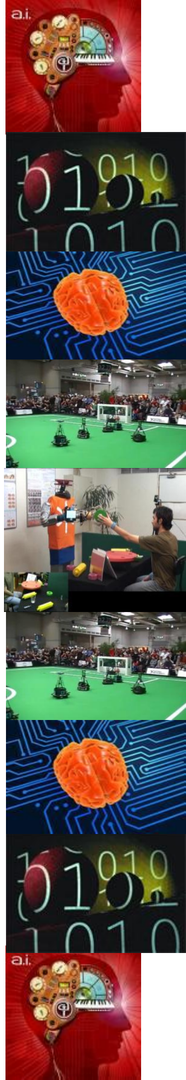


Algorithm

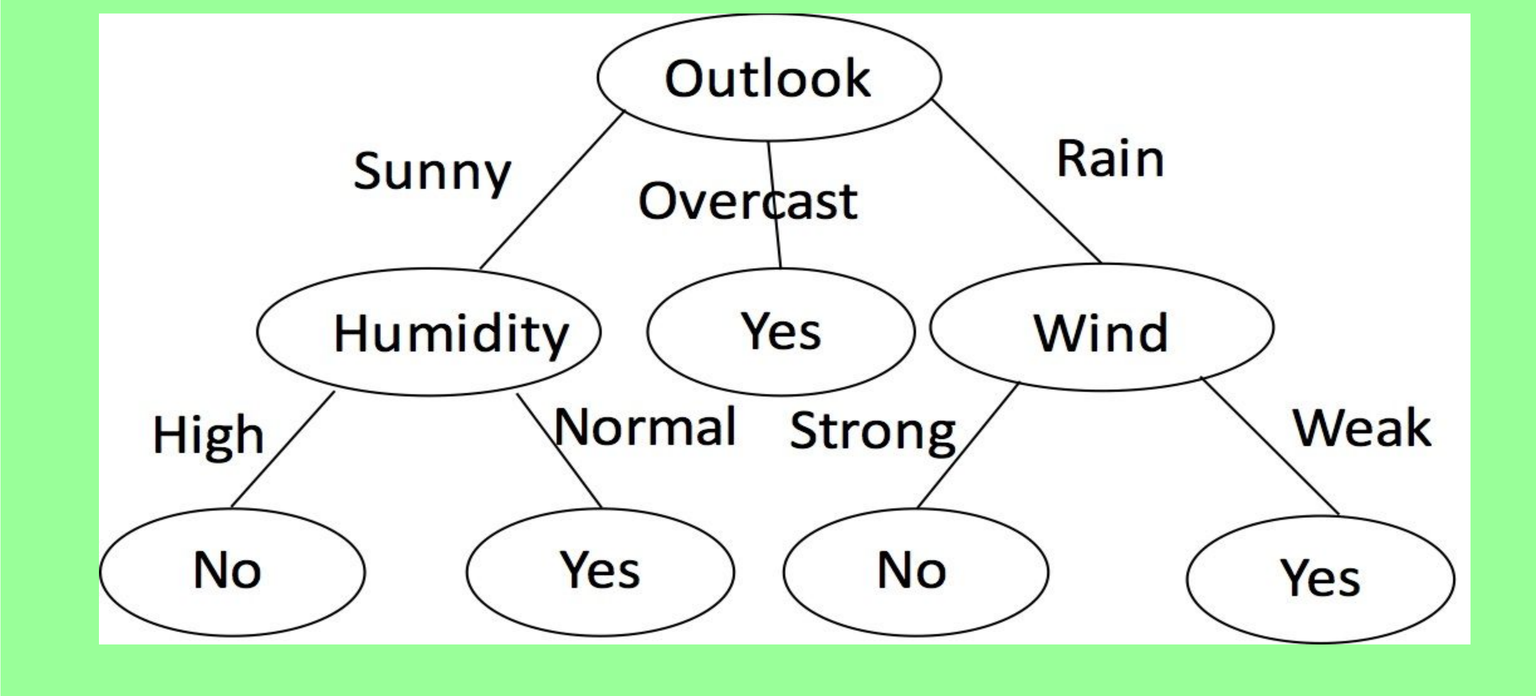
Algorithm 1 ID3 Algorithm

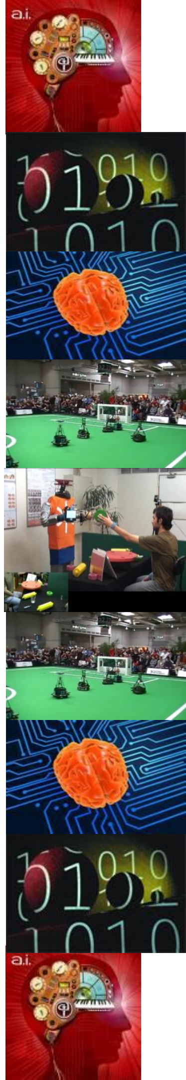
```
procedure ID3(S, A)
    if all labels in S are 1 then
        return a leaf node of 1
    end if
    if all labels in S are 0 then
        return a leaf node of 0
    end if
    if A is empty then
        return a node with the most frequently occurring label in S
    else
        maxAttr = be the attribute with the maximum gain (equation 4)
        if all the instances in S have the same label then
            return a leaf node with the majority label in S
        else
            Make the root of the decision tree maxAttr
            A' = A/maxAttr
            for v ← 1, number of values for j do

                S'= instances in S containing value c_v of A
                if S' is empty then
                    return a leaf node with the most frequently occurring label S
                else
                    Add ID3(S',A') as a subtree
                end if
            end for
        end if
    end if
end procedure
```

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA



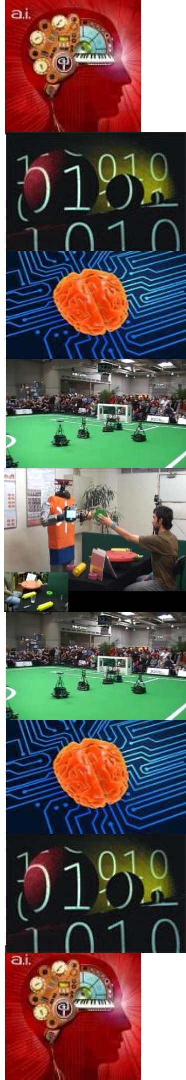
Example Induced DTree





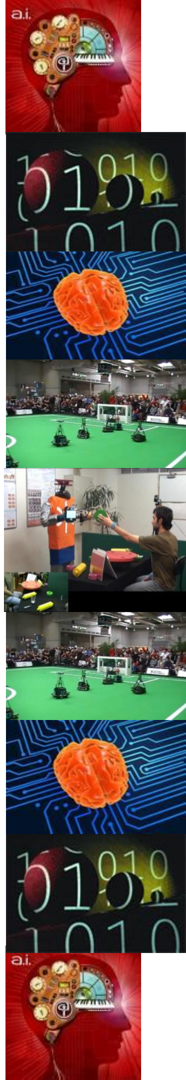
Disadvantages of DTree

- **Overfitting** - particularly for high dimensional datasets.
- Tree becomes overly complex and captures noise or specific patterns in the training data that might not generalize well to unseen examples.
    - *Pruning.*
    - *Tree depth limit.*
- **Sensitive to changes -**ensemble methods like Random Forests
- **Bias -** towards training data, data balancing (oversampling or undersampling).
- **Difficulty continuous features:** more suitable for discrete.



Random Forests

- Is an ensemble of decision trees.
- Performs classification
- Process
  - Choosing a subset
  - Creating a decision tree
  - Majority voting
- Overcomes overfitting



Questions