Lesson 5

Decision Tree (Rule Based Approach)

Example

outlook	temperature	humidity	windy	play
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

Example

Features

outlook	temperature	humidity	windy	play
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

Example

Class

outlook	temperature	humidity	windy	play
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

Example

outlook	temperature	humidity	windy	play
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

Given : **<sunny, cool, high, true>**

Predict, if there will be a match?

Example

outlook	temperature	humidity	windy	play
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

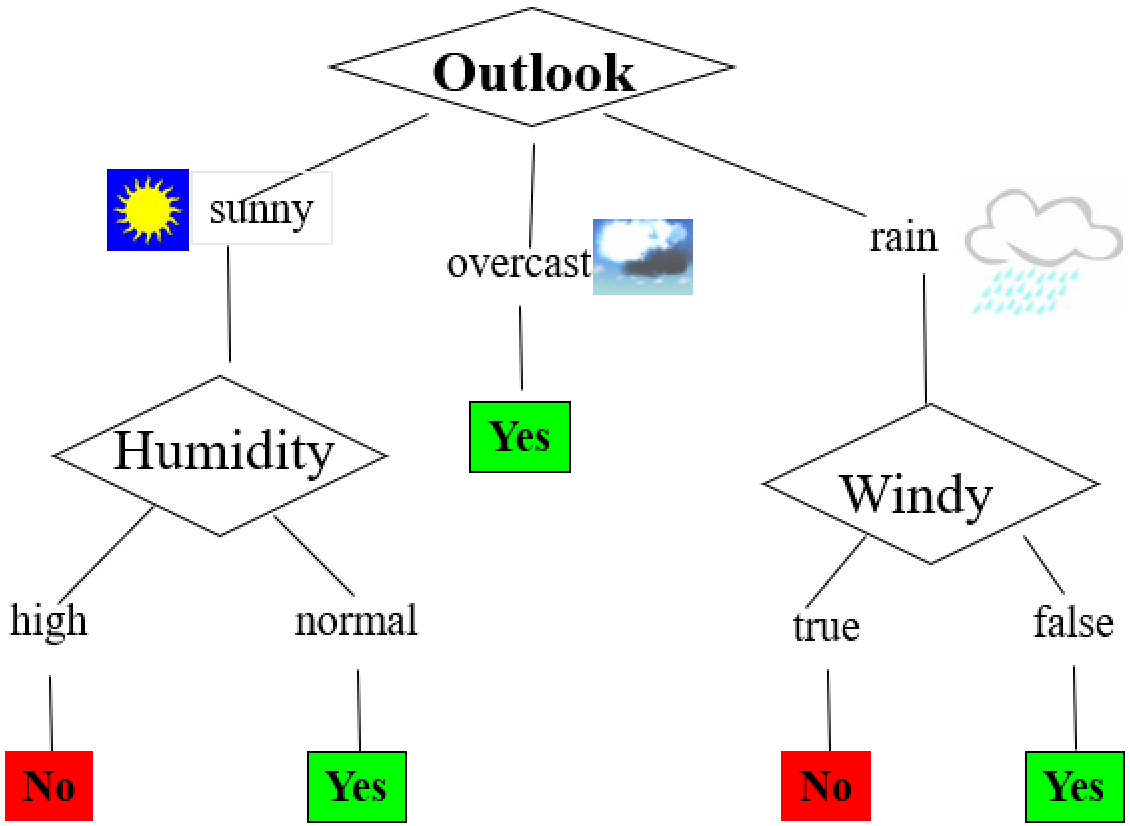
Given : **<sunny, cool, high, true>**

Predict, if there will be a match?

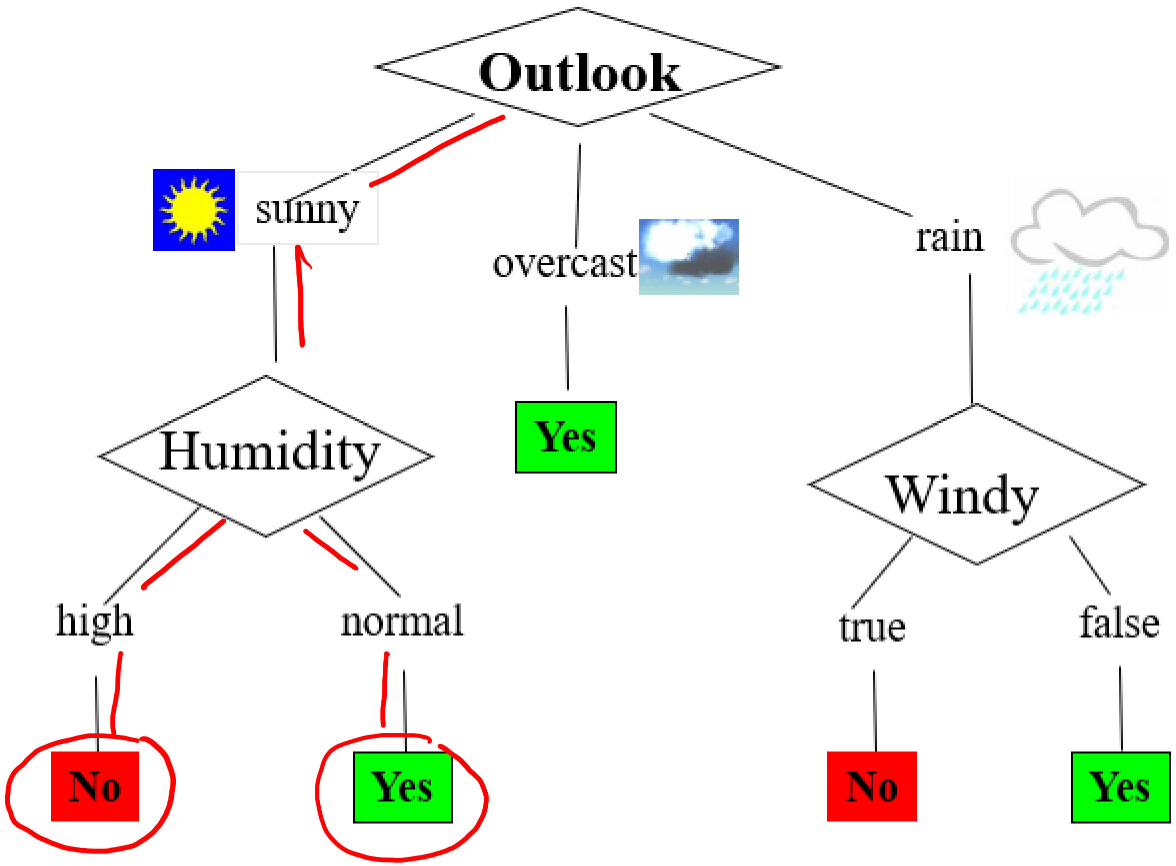
Assume that I have a set of rules:

- If ((*lookout=sunny)* **and** *(humudity=high)* **and** *(windy=false)) then (yes) else (no)*
- If *(lookout=overcast) then (yes)*
- If ((*lookout=sunny)* **and** *(humudity=high)) then (yes) else (no)*
- *so on…..*

Set of rules can be visualized as a tree.

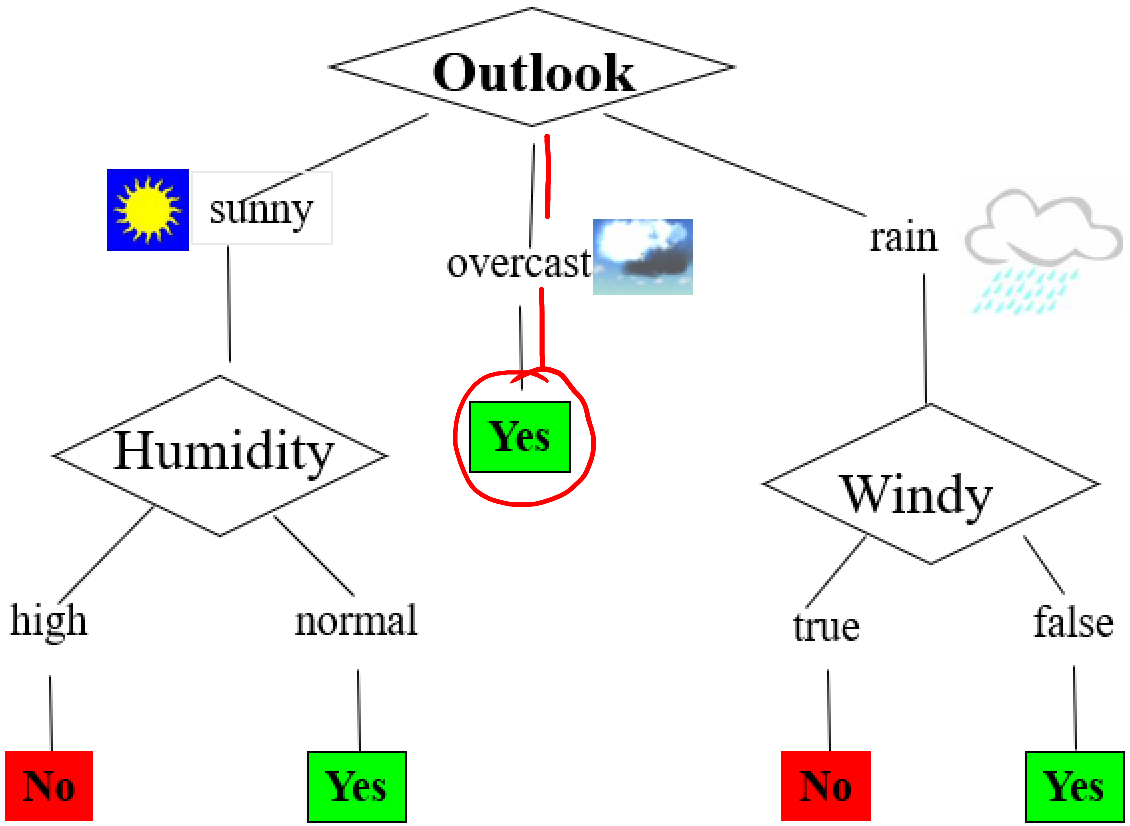


Set of rules can be visualized as a tree.



Rule 1: If (*(lookout=sunny)* **and** *(humudity=high))*
then (yes) else (no)

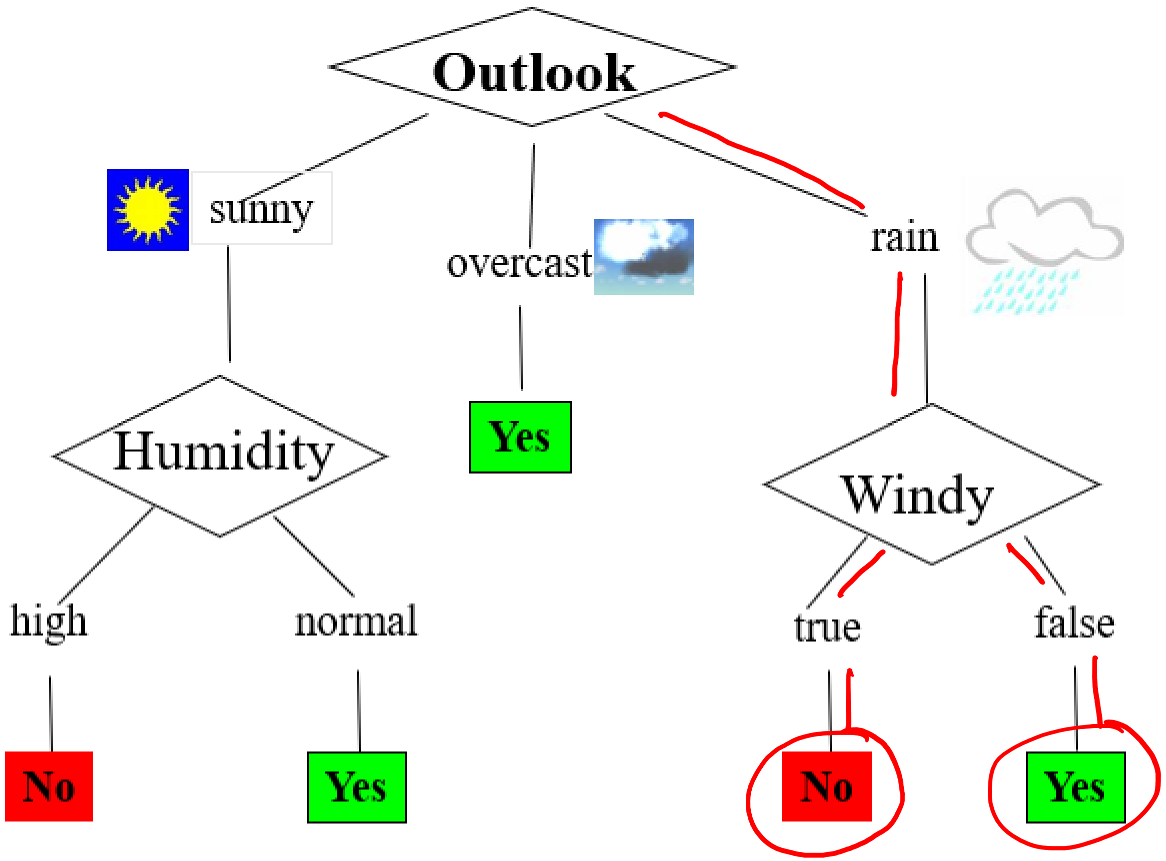
Set of rules can be visualized as a tree.



Rule 1: If ((*lookout=sunny*) **and** (*humudity=high*))
then (yes) else (no)

Rule 2: If *(lookout=overcast) then (yes)*

Set of rules can be visualized as a tree.



Rule 1: If ((*lookout=sunny)* **and** *(humudity=high))*
then (yes) else (no)

Rule 2: If *(lookout=overcast) then (yes)*

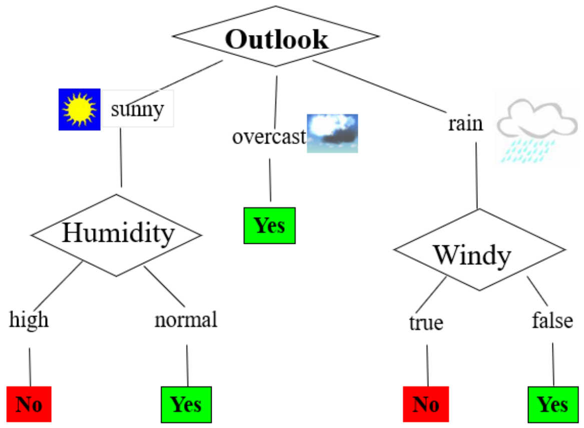
Rule 3: If ((*lookout=rain)* **and** *(windy=true))*
then (no) else (yes)

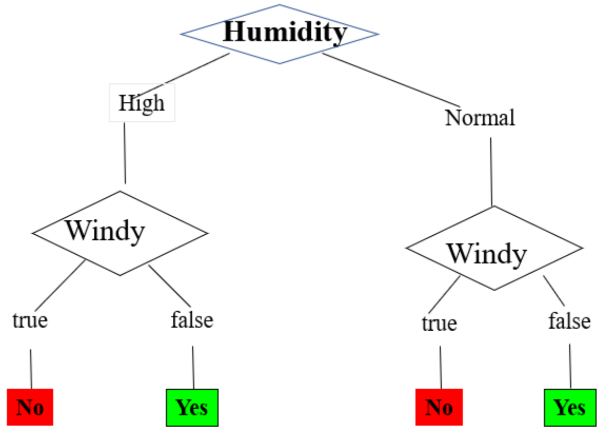
Many possible Trees

outlook	temperature	humidity	windy	play
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

Many possible Trees

outlook	temperature	humidity	windy	play
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





Which Tree is the Best?

- Which feature should be used to break the dataset?
- Types of DT
  - ID3 (Iterative Dichotomiser 3)
  - C4.5 (Successor of ID3)
  - CART (Classification and Regression Tree)
  - Random Forest

ID3

Person	Hair Length	Weight	Age	Class
Homer	0”	250	36	M
Marge	10”	150	34	F
Bart	2”	90	10	M
Lisa	6”	78	8	F
Maggie	4”	20	1	F
Abe	1”	170	70	M
Selma	8”	160	41	F
Otto	10”	180	38	M
Krusty	6”	200	45	M

ID3

Person	Hair Length	Weight	Age	Class
Homer	0″	250	36	M
Marge	10″	150	34	F
Bart	2″	90	10	M
Lisa	6″	78	8	F
Maggie	4″	20	1	F
Abe	1″	170	70	M
Selma	8″	160	41	F
Otto	10″	180	38	M
Krusty	6″	200	45	M

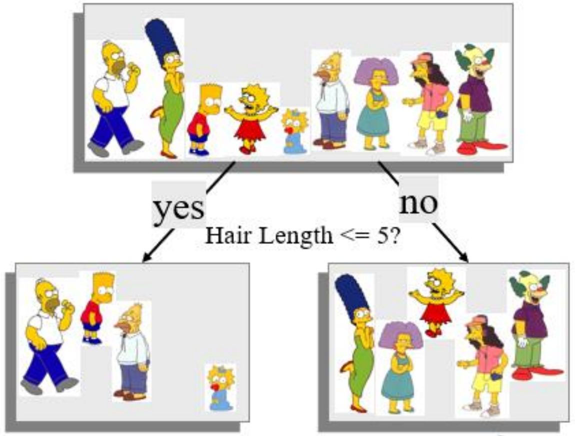
$$Entropy(S) = -\frac{p}{p+n}\log_2\left(\frac{p}{p+n}\right) - \frac{n}{p+n}\log_2\left(\frac{n}{p+n}\right)$$

$$Entropy(4\mathbf{F},5\mathbf{M}) = -(4/9)\log_2(4/9) - (5/9)\log_2(5/9)$$
$$= \mathbf{0.9911}$$

1. Calculate the entropy of the total dataset => H(S)=0.9911

ID3

Person	Hair Length	Weight	Age	Class
Homer	0″	250	36	M
Marge	10″	150	34	F
Bart	2″	90	10	M
Lisa	6″	78	8	F
Maggie	4″	20	1	F
Abe	1″	170	70	M
Selma	8″	160	41	F
Otto	10″	180	38	M
Krusty	6″	200	45	M

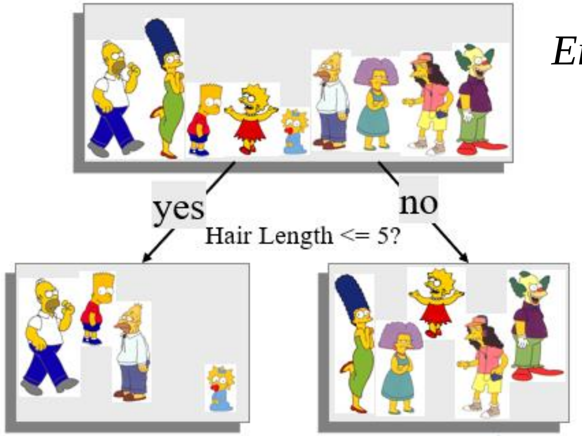


1. Calculate the entropy of the total dataset
2. Choose and attribute and Split the dataset by an attribute

ID3

Person	Hair Length	Weight	Age	Class
Homer	0″	250	36	M
Marge	10″	150	34	F
Bart	2″	90	10	M
Lisa	6″	78	8	F
Maggie	4″	20	1	F
Abe	1″	170	70	M
Selma	8″	160	41	F
Otto	10″	180	38	M
Krusty	6″	200	45	M

1. Calculate the entropy of the total dataset
2. Choose and attribute and Split the dataset by an attribute
3. Calculate the entropy of each branch



$$Entropy(3\mathbf{F},2\mathbf{M}) = -(3/5)\log_2(3/5) - (2/5)\log_2(2/5)$$
$$= \mathbf{0.9710}$$

$$Entropy(1\mathbf{F},3\mathbf{M}) = -(1/4)\log_2(1/4) - (3/4)\log_2(3/4)$$
$$= \mathbf{0.8113}$$

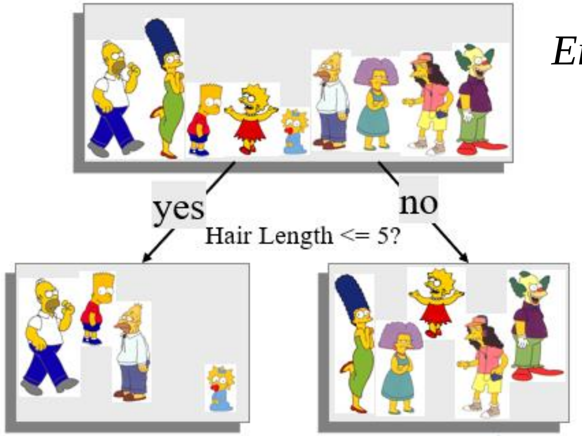
ID3

Person	Hair Length	Weight	Age	Class
Homer	0″	250	36	M
Marge	10″	150	34	F
Bart	2″	90	10	M
Lisa	6″	78	8	F
Maggie	4″	20	1	F
Abe	1″	170	70	M
Selma	8″	160	41	F
Otto	10″	180	38	M
Krusty	6″	200	45	M

1. Calculate the entropy of the total dataset
2. Choose and attribute and Split the dataset by an attribute
3. Calculate the entropy of each branch
4. Calculate Information Gain of the split

$$IG(A_1) = H(S) - [p(S_1)H(S_1) + p(S_2)H(S_2)]$$

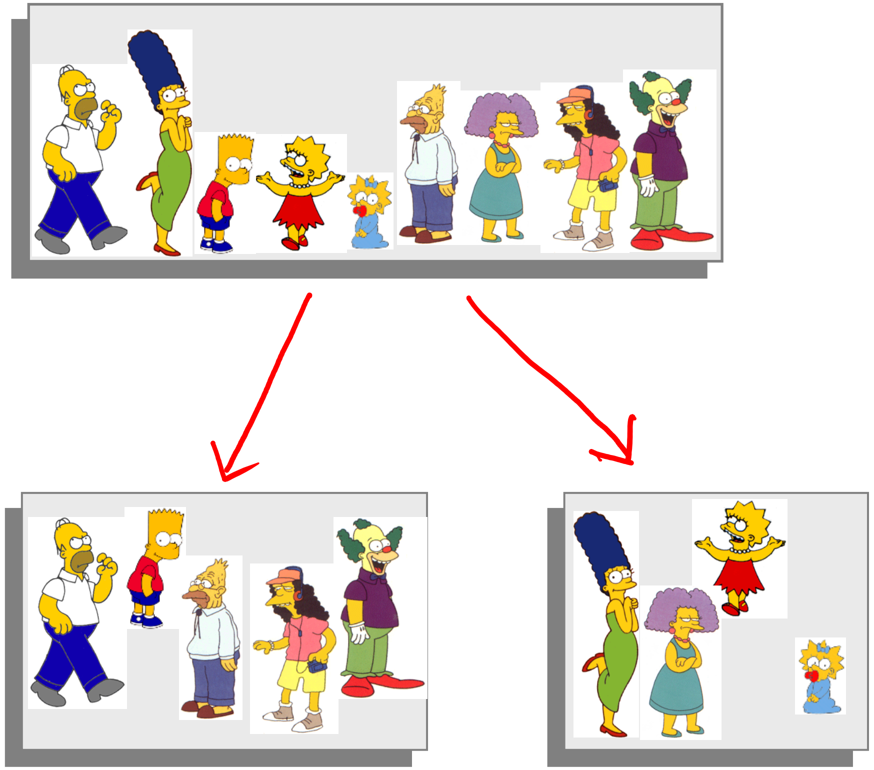
Gain(Hair Length <= 5) = **0.9911** – (4/9 * **0.8113** + 5/9 * **0.9710**) = **0.0911**



Entropy(3**F**,2**M**) = -(3/5)$\log_2$(3/5) - (2/5)$\log_2$(2/5)
= **0.9710**

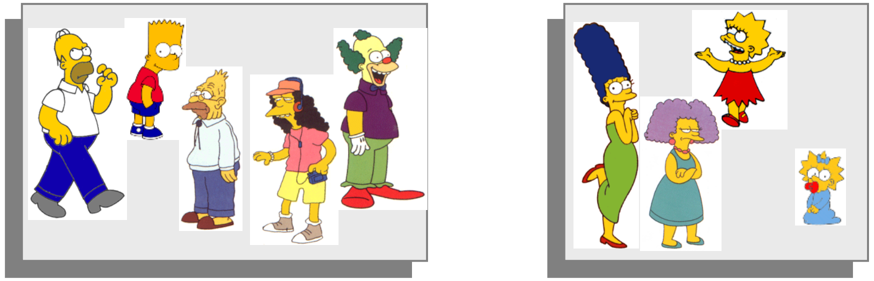
Entropy(1**F**,3**M**) = -(1/4)$\log_2$(1/4) - (3/4)$\log_2$(3/4)
= **0.8113**

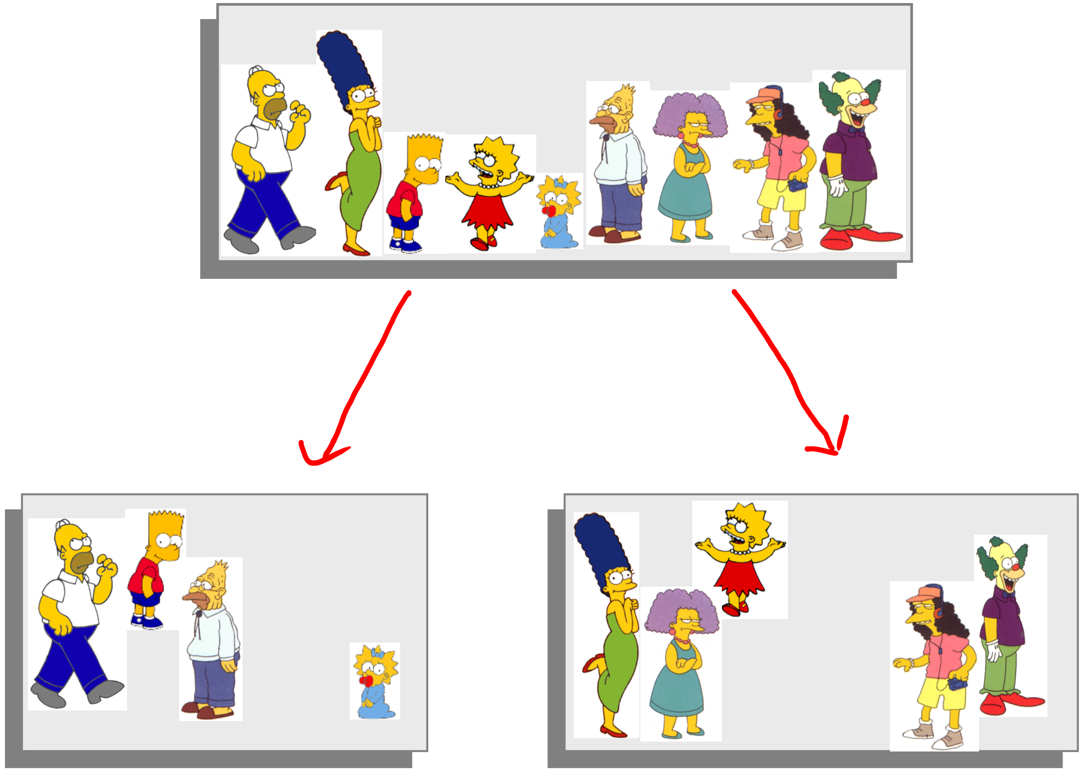
What is Information Gain?



$H(S_1) = 0$ $H(S_2) = 0$

What is Information Gain?

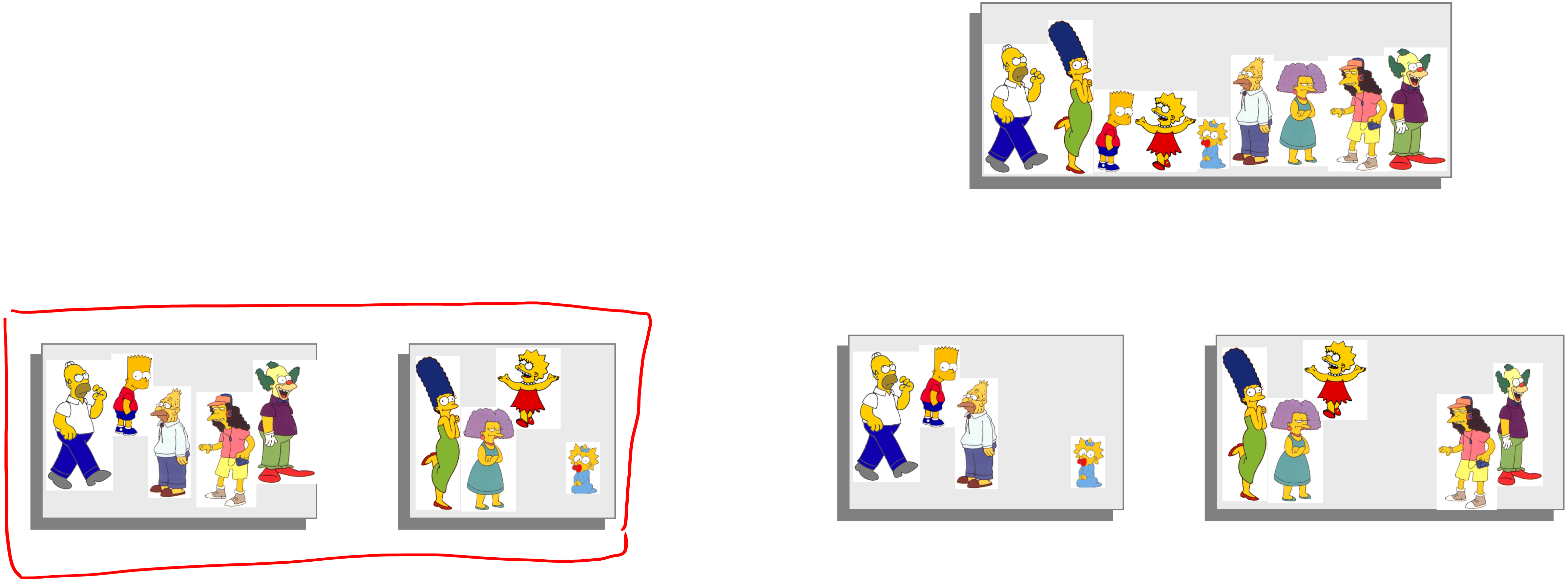




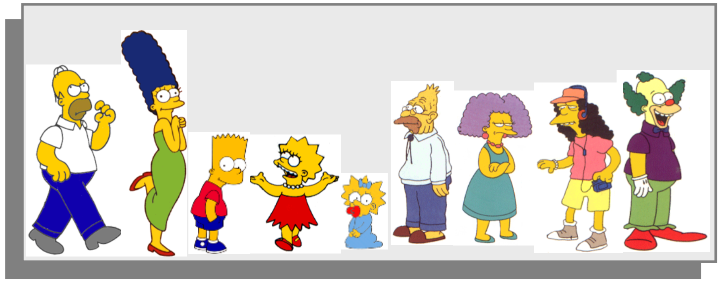
$H(S_1) = 0.8113$

$H(S_2) = 0.971$

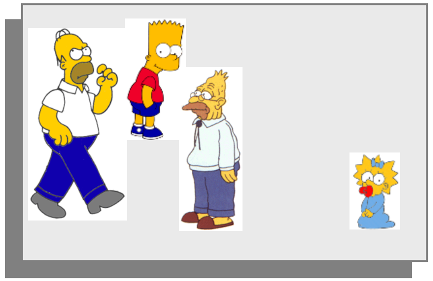
Which split is better?



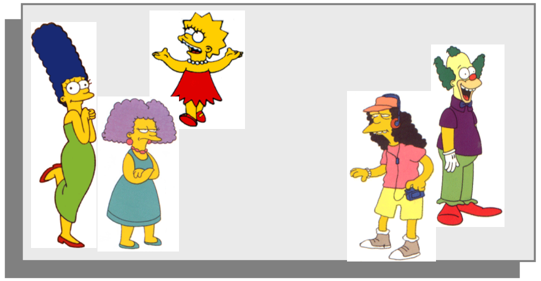
What is information gain?



$H(S_1) = 0$ $H(S_2) = 0$



$H(S_1) = 0.8113$

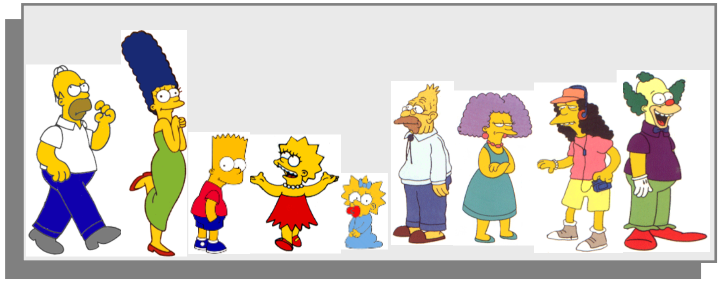


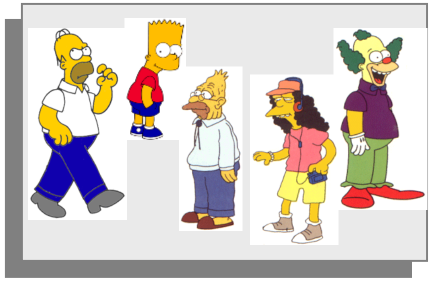
$H(S_2) = 0.971$

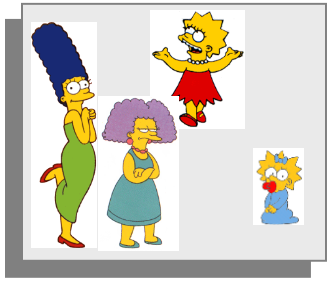
What is information gain?

Reduction in uncertainty of the parent dataset after the split.

$$IG(A_1) = H(S) - [p(S_1)H(S_1) + p(S_2)H(S_2)]$$

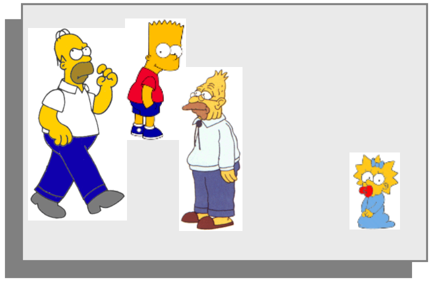


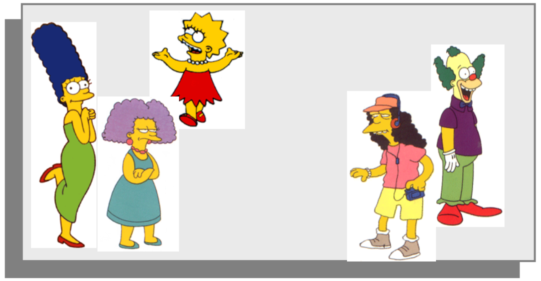




$H(S_1) = 0$

$H(S_2) = 0$





$H(S_1) = 0.8113$

$H(S_2) = 0.971$

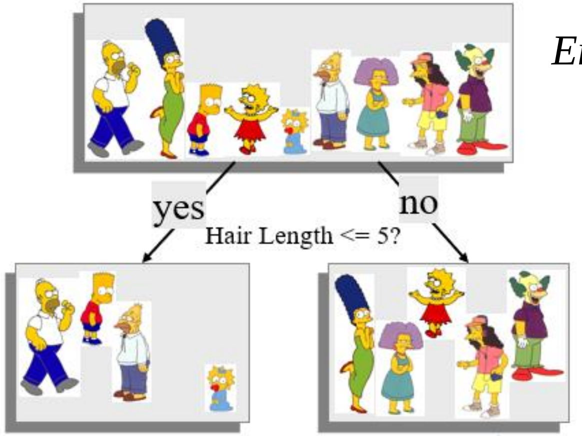
ID3

Person	Hair Length	Weight	Age	Class
Homer	0″	250	36	M
Marge	10″	150	34	F
Bart	2″	90	10	M
Lisa	6″	78	8	F
Maggie	4″	20	1	F
Abe	1″	170	70	M
Selma	8″	160	41	F
Otto	10″	180	38	M
Krusty	6″	200	45	M

1. Calculate the entropy of the total dataset
2. Choose and attribute and Split the dataset by an attribute
3. Calculate the entropy of each branch
4. Calculate Information Gain of the split

$$IG(A_1) = H(S) - [p(S_1)H(S_1) + p(S_2)H(S_2)]$$

Gain(Hair Length <= 5) = **0.9911** – (4/9 * **0.8113** + 5/9 * **0.9710**) = **0.0911**



Entropy(3**F**,2**M**) = $-(3/5)\log_2(3/5) - (2/5)\log_2(2/5)$

= **0.9710**

Entropy(1**F**,3**M**) = $-(1/4)\log_2(1/4) - (3/4)\log_2(3/4)$

= **0.8113**

ID3

Person	Hair Length	Weight	Age	Class
Homer	0″	250	36	M
Marge	10″	150	34	F
Bart	2″	90	10	M
Lisa	6″	78	8	F
Maggie	4″	20	1	F
Abe	1″	170	70	M
Selma	8″	160	41	F
Otto	10″	180	38	M
Krusty	6″	200	45	M

$$\mathit{Gain}(\text{Hair Length} <= 5) = \mathbf{0.0911}$$

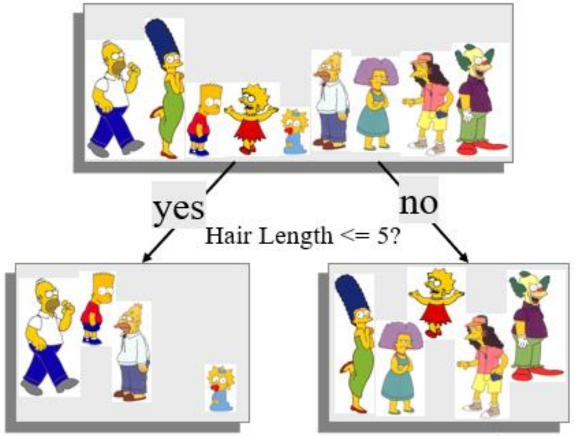
$$\mathit{Gain}(\text{Weight} <= 160) = \mathbf{0.5900}$$

$$\mathit{Gain}(\text{Age} <= 40) = \mathbf{0.0183}$$

1. Calculate the entropy of the total dataset
2. Choose and attribute and Split the dataset by an attribute
3. Calculate the entropy of each branch
4. Calculate Information Gain of the split
5. Repeat 2, 3, 4 for all Attributes
6. The attribute that yields the largest IG is chosen for the decision node.

ID3

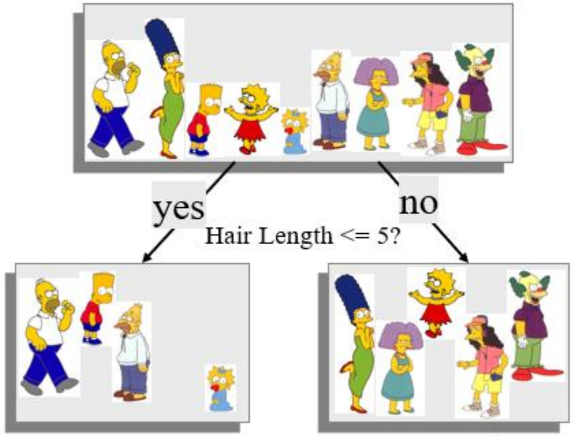
Person	Hair Length	Weight	Age	Class
Homer	0″	250	36	M
Marge	10″	150	34	F
Bart	2″	90	10	M
Lisa	6″	78	8	F
Maggie	4″	20	1	F
Abe	1″	170	70	M
Selma	8″	160	41	F
Otto	10″	180	38	M
Krusty	6″	200	45	M



1. Calculate the entropy of the total dataset
2. Choose and attribute and Split the dataset by an attribute
3. Calculate the entropy of each branch
4. Calculate Information Gain of the split
5. Repeat 2, 3, 4 for all Attributes
6. The attribute that yields the largest IG is chosen for the decision node.

ID3

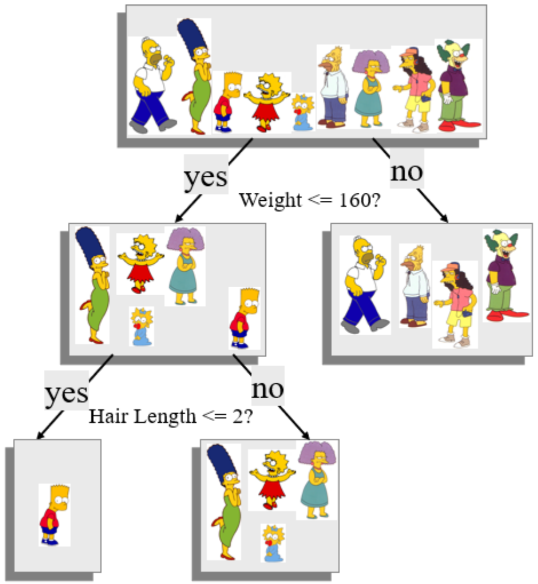
Person	Hair Length	Weight	Age	Class
Homer	0”	250	36	M
Marge	10”	150	34	F
Bart	2”	90	10	M
Lisa	6”	78	8	F
Maggie	4”	20	1	F
Abe	1”	170	70	M
Selma	8”	160	41	F
Otto	10”	180	38	M
Krusty	6”	200	45	M



1. Calculate the entropy of the total dataset
2. Choose and attribute and Split the dataset by an attribute
3. Calculate the entropy of each branch
4. Calculate Information Gain of the split
5. Repeat 2, 3, 4 for all Attributes
6. The attribute that yields the largest IG is chosen for the decision node.
7. Repeat 1 to 6 for all sub-databases till we get sub-databases with single class

ID3

Person	Hair Length	Weight	Age	Class
Homer	0″	250	36	M
Marge	10″	150	34	F
Bart	2″	90	10	M
Lisa	6″	78	8	F
Maggie	4″	20	1	F
Abe	1″	170	70	M
Selma	8″	160	41	F
Otto	10″	180	38	M
Krusty	6″	200	45	M



1. Calculate the entropy of the total dataset
2. Choose and attribute and Split the dataset by an attribute
3. Calculate the entropy of each branch
4. Calculate Information Gain of the split
5. Repeat 2, 3, 4 for all Attributes
6. The attribute that yields the largest IG is chosen for the decision node.
7. Repeat 1 to 6 for all sub-databases till we get sub-databases with single class