

**DECISION TREES - Example**  
**Prof. Carolina Ruiz**  
**Department of Computer Science**  
**Worcester Polytechnic Institute**

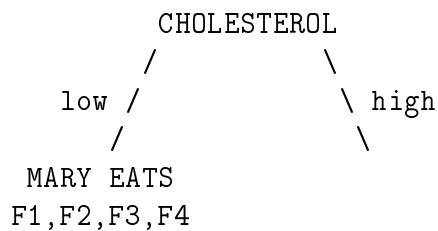
The following examples illustrate Mary's dietary preferences.

FOOD	CHOLESTEROL	SODIUM	TASTE	MARY EATS
F1	low	low	good	yes
F2	low	low	bad	yes
F3	low	high	good	yes
F4	low	high	bad	yes
F5	high	low	good	yes
F6	high	low	bad	no
F7	high	high	good	no
F8	high	high	bad	no

Using information theory we can construct a minimal decision tree that predicts whether or not Mary would eat a particular food.

MARY EATS	YES	NO
CHOLESTEROL		
Low	$\frac{4}{8}[-\frac{4}{4}\lg_2\frac{4}{4} - 0]$	
High	$\frac{4}{8}[-\frac{1}{4}\lg_2\frac{1}{4} - \frac{3}{4}\lg_2\frac{3}{4}]$	
	$= 0.41$	
SODIUM		
Low	$\frac{4}{8}[-\frac{3}{4}\lg_2\frac{3}{4} - \frac{1}{4}\lg_2\frac{1}{4}]$	
High	$\frac{4}{8}[-\frac{2}{4}\lg_2\frac{2}{4} - \frac{2}{4}\lg_2\frac{2}{4}]$	
	$= 0.91$	
TASTE		
Good	$\frac{4}{8}[-\frac{3}{4}\lg_2\frac{3}{4} - \frac{1}{4}\lg_2\frac{1}{4}]$	
Bad	$\frac{4}{8}[-\frac{2}{4}\lg_2\frac{2}{4} - \frac{2}{4}\lg_2\frac{2}{4}]$	
	$= 0.91$	

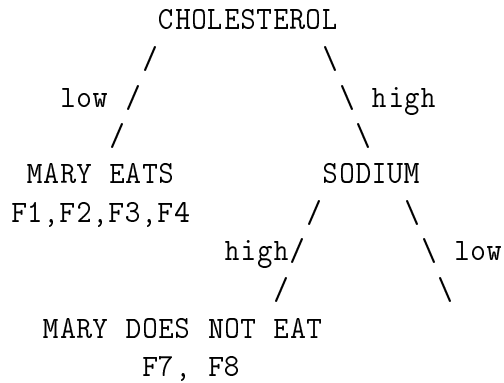
Since the attribute CHOLESTEROL minimizes the entropy, it is selected as the root of the tree:



Now we need to choose between SODIUM and TASTE to split the food with high cholesterol, that is, examples F5, F6, F7, and F8.

MARY EATS	YES	NO
SODIUM		
Low	$\frac{2}{4}[-\frac{1}{2} \lg_2 \frac{1}{2}]$	$-\frac{1}{2} \lg_2 \frac{1}{2}]$
High	$\frac{2}{4}[-\frac{0}{2} \lg_2 \frac{0}{2}]$	$-\frac{2}{2} \lg_2 \frac{2}{2}]$
	= 0.5	
TASTE		
Good	$\frac{2}{4}[-\frac{1}{2} \lg_2 \frac{1}{2}]$	$-\frac{1}{2} \lg_2 \frac{1}{2}]$
Bad	$\frac{2}{4}[-\frac{0}{2} \lg_2 \frac{0}{2}]$	$-\frac{2}{2} \lg_2 \frac{2}{2}]$
	= 0.5	

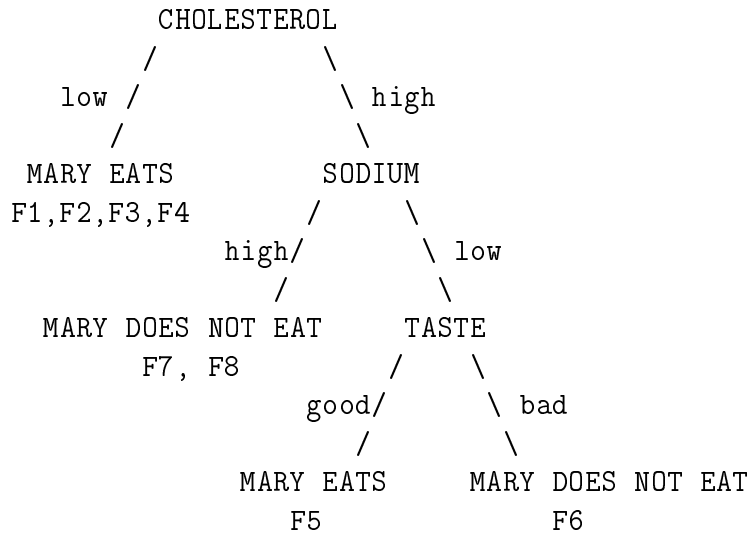
Since both attributes SODIUM and CHOLESTEROL have the same entropy value, either one can be selected for the next level of the tree. I'll select SODIUM:



The only attribute left is TASTE and so it is selected for the last level of the tree:

**DECISION TREES - Example**  
**Prof. Carolina Ruiz**  
**Department of Computer Science**  
**Worcester Polytechnic Institute**  
 Mary's dietary preferences - continuation

**Decision Tree**



**Printing out this decision tree**

```

CHOLESTEROL=low : MARY EATS=yes
CHOLESTEROL=high
  | SODIUM=high : MARY EATS=no
  | SODIUM=low
  |   | TASTE=good : MARY EATS=yes
  |   | TASTE=bad  : MARY EATS=no
  
```

**Rule set obtained from from the decision tree**

```

if CHOLESTEROL=low then MARY EATS=yes
if CHOLESTEROL=high and SODIUM=high then MARY EATS=no
if CHOLESTEROL=high and SODIUM=low and TASTE=good then MARY EATS=yes
if CHOLESTEROL=high and SODIUM=low and TASTE=bad then MARY EATS=no
  
```