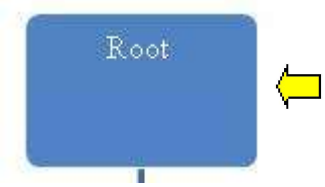
Note: You may find it helpful to get the tree structure completely working before coding up the entropy calculations

Calculating entropy for data set: tennis.arff



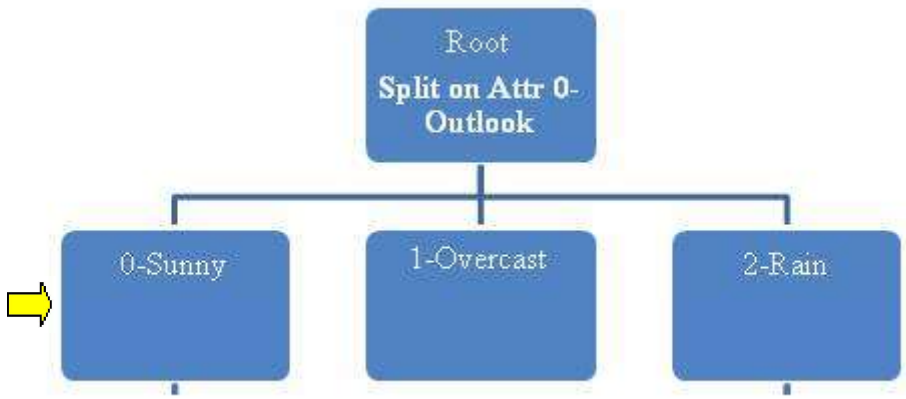
Start at the root node with all the instances:

outlook	temperature	humidity	wind	playTennis
sunny	hot	high	weak	no
sunny	hot	high	strong	no
overcast	hot	high	weak	yes
rain	mild	high	weak	yes
rain	cool	normal	weak	yes
rain	cool	normal	strong	no
overcast	cool	normal	strong	yes
sunny	mild	high	weak	no
sunny	cool	normal	weak	yes
rain	mild	normal	weak	yes
sunny	mild	normal	strong	yes
overcast	mild	high	strong	yes
overcast	hot	normal	weak	yes
rain	mild	high	strong	no

(Note that this example uses $\log_2$. The decisions will be identical if you use $\log_e$, or any other base, but the entropy and gain values will be scaled differently.)

```
Node: ( 5/14 9/14 ) Entropy=0.9402859586706309
      Attribute 0-Outlook:
            Value 0-Sunny: ( 3/5 2/5 ) Entropy=0.9709505944546686
            Value 1-Overcast: ( 0/4 4/4 ) Entropy=0.0
            Value 2-Rain: ( 2/5 3/5 ) Entropy=0.9709505944546686
            InfoGain=0.2467498197744391
      Attribute 1-Temperature:
            Value 0-Hot: ( 2/4 2/4 ) Entropy=1.0
            Value 1-Mild: ( 2/6 4/6 ) Entropy=0.9182958340544896
            Value 2-Cool: ( 1/4 3/4 ) Entropy=0.8112781244591328
            InfoGain=0.029222565658954647
      Attribute 2-Humidity:
            Value 0-High: ( 4/7 3/7 ) Entropy=0.9852281360342516
            Value 1-Normal: ( 1/7 6/7 ) Entropy=0.5916727785823275
            InfoGain=0.15183550136234136
      Attribute 3-Wind:
            Value 0-Weak: ( 2/8 6/8 ) Entropy=0.8112781244591328
            Value 1-Strong: ( 3/6 3/6 ) Entropy=1.0
            InfoGain=0.04812703040826932
```

Maximum InfoGain=0.247
Split on Attribute 0-Outlook

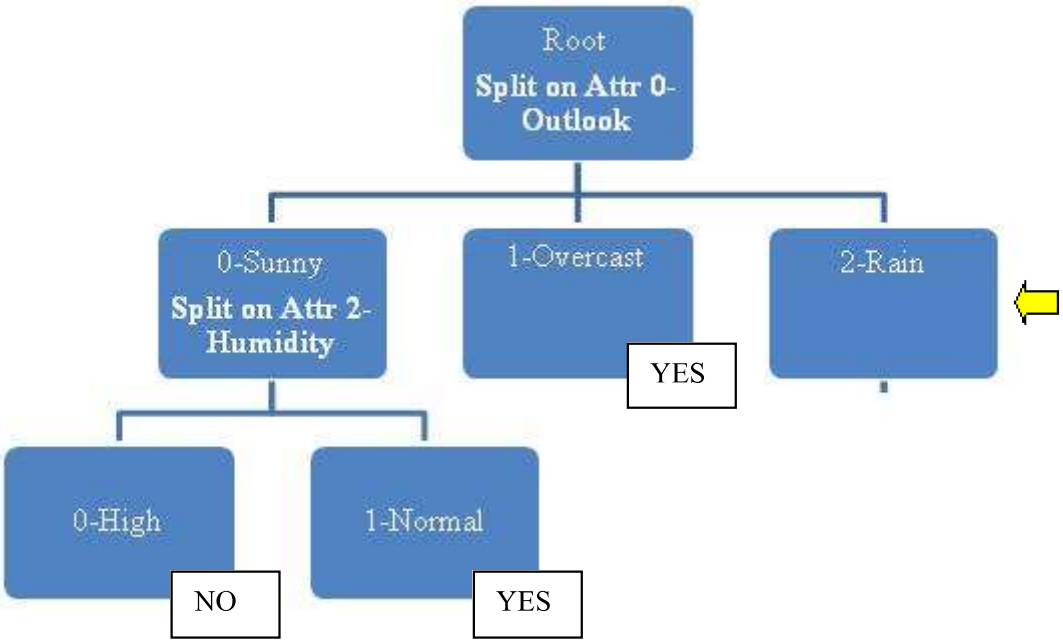


Move on to the next node in the tree that can be expanded:

outlook	temperature	humidity	wind	playTennis
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

```
Node: ( 3/5 2/5 ) Entropy=0.9709505944546686
      Attribute 1-Temperature:
            Value 0-Hot: ( 2/2 0/2 ) Entropy=0.0
            Value 1-Mild: ( 1/2 1/2 ) Entropy=1.0
            Value 2-Cool: ( 0/1 1/1 ) Entropy=0.0
            InfoGain=0.5709505944546686
      Attribute 2-Humidity
            Value 0-High: ( 3/3 0/3 ) Entropy=0.0
            Value 1-Normal: ( 0/2 2/2 ) Entropy=0.0
            InfoGain=0.9709505944546686
      Attribute 3-Wind:
            Value 0-Weak: ( 2/3 1/3 ) Entropy=0.9182958340544896
            Value 1-Strong: ( 1/2 1/2 ) Entropy=1.0
            InfoGain=0.01997309402197489
```

Maximum InfoGain=0.971
Split on Attribute 2-Humidity

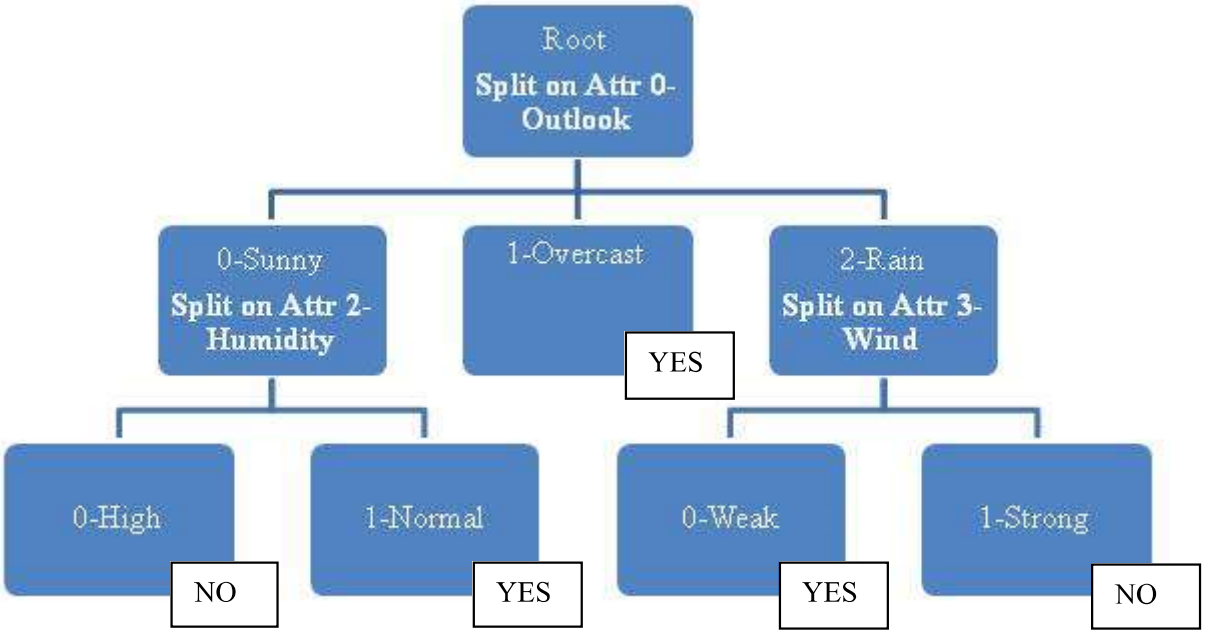


Move on to the next node in the tree that can be expanded:

outlook	temperature	humidity	wind	playTennis
rain	mild	high	weak	yes
rain	cool	normal	weak	yes
rain	cool	normal	strong	no
rain	mild	normal	weak	yes
rain	mild	high	strong	no

```
Node: ( 2/5 3/5 ) Entropy=0.9709505944546686
      Attribute 1-Temperature:
            Value 0-Hot: ( 0/0 0/0 ) Entropy=0.0
            Value 1-Mild: ( 1/3 2/3 ) Entropy=0.9182958340544896
            Value 2-Cool: ( 1/2 1/2 ) Entropy=1.0
            InfoGain=0.01997309402197489
      Attribute 2-Humidity:
            Value 0-High: ( 1/2 1/2 ) Entropy=1.0
            Value 1-Normal: ( 1/3 2/3 ) Entropy=0.9182958340544896
            InfoGain=0.01997309402197489
      Attribute 3-Wind:
            Value 0-Weak ( 0/3 3/3 ) Entropy=0.0
            Value 1-Strong: ( 2/2 0/2 ) Entropy=0.0
            InfoGain=0.9709505944546686
```

Maximum InfoGain=0.971
Split on Attribute 3-Wind


Root
Split on Attr 0-Outlook
0-Sunny
Split on Attr 2-Humidity
1-Overcast
YES
2-Rain
Split on Attr 3-Wind
0-High
NO
1-Normal
YES
0-Weak
YES
1-Strong
NO