Q3 Decision Tree

In [73]:

```
import math
```

In [97]:

```
def exp_inf(p):
    s = sum(p)
    inf = 0
    for i in p:
        if i == 0:
            inf += 0
        else:
            inf += (float(i)/s)*math.log(float(i)/s,2)
    return abs(inf)
```

In [99]:

```
exp_inf([3,2])
```

Out[99]:

```
0.9709505944546686
```

In [76]:

```
a = [['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']]
```

In [77]:

```
sunny = [0,0] overcast = [0,0] rain = [0,0] hot = [0,0] mild = [0,0]
cool = [0,0] high = [0,0] normal = [0,0] strong = [0,0] weak = [0,0]
length = len(a)
```

In [78]:

```
for i in range(len(a)):
    if a[i][0] == 'Sunny':
        if a[i][4] == 'Yes':
            sunny[0] += 1
        else:
            sunny[1] += 1
    elif a[i][0] == 'Overcast':
        if a[i][4] == 'Yes':
            overcast[0] += 1
        else:
            overcast[1] += 1
    elif a[i][0] == 'Rain':
        if a[i][4] == 'Yes':
            rain[0] += 1
        else:
            rain[1] += 1
```

In [79]:

```
for i in range(len(a)):
    if a[i][1] == 'Hot':
        if a[i][4] == 'Yes':
            hot[0] += 1
        else:
            hot[1] += 1
    elif a[i][1] == 'Mild':
        if a[i][4] == 'Yes':
            mild[0] += 1
        else:
            mild[1] += 1
    elif a[i][1] == 'Cool':
        if a[i][4] == 'Yes':
            cool[0] += 1
        else:
            cool[1] += 1
```

In [80]:

```
for i in range(len(a)):
    if a[i][2] == 'High':
        if a[i][4] == 'Yes':
            high[0] += 1
        else:
            high[1] += 1
    elif a[i][2] == 'Normal':
        if a[i][4] == 'Yes':
            normal[0] += 1
        else:
            normal[1] += 1
```

In [81]:

```
for i in range(len(a)):
    if a[i][3] == 'Strong':
        if a[i][4] == 'Yes':
            strong[0] += 1
        else:
            strong[1] += 1
    elif a[i][3] == 'Weak':
        if a[i][4] == 'Yes':
            weak[0] += 1
        else:
            weak[1] += 1
```

In [82]:

```
Total_yes = strong[0] + weak[0]
Total_no = strong[1] + weak[1]
Total_inf = exp_inf([Total_yes,Total_no])
```

In [83]:

```
Total_inf
```

Out[83]:

```
0.9402859586706309
```

In [85]:

```
def one_inf(one):
    return float(sum(one))/length*exp_inf(one)
```

In [103]:

```
Gain_weather = Total_inf - one_inf(sunny) - one_inf(overcast)- one_inf(rain)
Gain_temp = Total_inf - one_inf(hot) - one_inf(mild)- one_inf(cool)
Gain_hum = Total_inf - one_inf(high) - one_inf(normal)
Gain_wind = Total_inf - one_inf(weak) - one_inf(strong)
```

In [104]:

```
Gain_weather
```

Out[104]:

```
0.2467498197744391
```

In [105]:

```
Gain_temp
```

Out[105]:

```
0.029222565658954647
```

In [106]:

```
Gain_hum
```

Out[106]:

```
0.15183550136234136
```

In [107]:

```
Gain_wind
```

Out[107]:

```
0.04812703040826932
```

Since Gain_weather is highest, it is the root

In [108]:

```
print exp_inf(sunny),exp_inf(overcast),exp_inf(rain)
```

```
0.970950594455 0.0 0.970950594455
```

In [117]:

```
float(overcast[0])/sum(overcast)
```

Out[117]:

```
1.0
```

We can find overcast is 0, there is no need to extend

In [109]:

```
sunny_hot = [0,0]
sunny_mild = [0,0]
sunny_cool = [0,0]
rain_hot = [0,0]
rain_mild = [0,0]
rain_cool = [0,0]
sunny_high = [0,0]
sunny_normal = [0,0]
rain_high = [0,0]
rain_normal = [0,0]
sunny_strong = [0,0]
sunny_weak = [0,0]
rain_strong = [0,0]
rain_weak = [0,0]
```

In [110]:

```
for i in range(len(a)):
    if a[i][0] == 'Sunny':
        if a[i][1] == 'Hot':
            if a[i][4] == 'Yes':
                sunny_hot[0] += 1
            else:
                sunny_hot[1] += 1
        elif a[i][1] == 'Mild':
            if a[i][4] == 'Yes':
                sunny_mild[0] += 1
            else:
                sunny_mild[1] += 1
        elif a[i][1] == 'Cool':
            if a[i][4] == 'Yes':
                sunny_cool[0] += 1
            else:
                sunny_cool[1] += 1
    elif a[i][0] == 'Rain':
        if a[i][1] == 'Hot':
            if a[i][4] == 'Yes':
                rain_hot[0] += 1
            else:
                rain_hot[1] += 1
        elif a[i][1] == 'Mild':
            if a[i][4] == 'Yes':
                rain_mild[0] += 1
            else:
                rain_mild[1] += 1
        elif a[i][1] == 'Cool':
            if a[i][4] == 'Yes':
                rain_cool[0] += 1
            else:
                rain_cool[1] += 1
```

In [111]:

```
for i in range(len(a)):
    if a[i][0] == 'Sunny':
        if a[i][2] == 'High':
            if a[i][4] == 'Yes':
                sunny_high[0] += 1
            else:
                sunny_high[1] += 1
        elif a[i][2] == 'Normal':
            if a[i][4] == 'Yes':
                sunny_normal[0] += 1
            else:
                sunny_normal[1] += 1
    elif a[i][0] == 'Rain':
        if a[i][2] == 'High':
            if a[i][4] == 'Yes':
                rain_high[0] += 1
            else:
                rain_high[1] += 1
        elif a[i][2] == 'Normal':
            if a[i][4] == 'Yes':
                rain_normal[0] += 1
            else:
                rain_normal[1] += 1
```

In [112]:

```
for i in range(len(a)):
    if a[i][0] == 'Sunny':
        if a[i][3] == 'Strong':
            if a[i][4] == 'Yes':
                sunny_strong[0] += 1
            else:
                sunny_strong[1] += 1
        elif a[i][3] == 'Weak':
            if a[i][4] == 'Yes':
                sunny_weak[0] += 1
            else:
                sunny_weak[1] += 1


    elif a[i][0] == 'Rain':
        if a[i][3] == 'Strong':
            if a[i][4] == 'Yes':
                rain_strong[0] += 1
            else:
                rain_strong[1] += 1
        elif a[i][3] == 'Weak':
            if a[i][4] == 'Yes':
                rain_weak[0] += 1
            else:
                rain_weak[1] += 1
```

In [113]:

```
Gain_sunny_temp =one_inf(sunny) - one_inf(sunny_hot) - one_inf(sunny_mild)- one_inf(sunny_c
Gain_sunny_hum = one_inf(sunny)- one_inf(sunny_high) - one_inf(sunny_normal)
Gain_sunny_wind = one_inf(sunny) - one_inf(sunny_weak) - one_inf(sunny_strong)
Gain_rain_temp =one_inf(rain)- one_inf(rain_hot) - one_inf(rain_mild)- one_inf(rain_cool)
Gain_rain_hum =one_inf(rain)- one_inf(rain_high) - one_inf(rain_normal)
Gain_rain_wind =one_inf(rain) - one_inf(rain_weak) - one_inf(rain_strong)
```

In [114]:

```
print Gain_sunny_temp,Gain_sunny_hum,Gain_sunny_wind
```

```
0.203910926591 0.346768069448 0.00713324786499
```

In [115]:

```
print Gain_rain_temp, Gain_rain_hum, Gain_rain_wind
```

```
0.00713324786499 0.00713324786499 0.346768069448
```

In [119]:

```
float(rain_strong[0])/sum(rain_strong)
```

Out[119]:

```
0.0
```

In [120]:

```
float(rain_weak[0])/sum(rain_weak)
```

Out[120]:

```
1.0
```

if we set a prune parameter that all value under 0.01 can be ignored

we can find in sunny condition, humidity is higher than temp

In [122]:

```
sunny_high_hot =[0,0]
sunny_high_mild =[0,0]
sunny_high_cool =[0,0]
sunny_normal_hot =[0,0]
sunny_normal_mild =[0,0]
sunny_normal_cool =[0,0]
```

In [123]:

```
for i in range(len(a)):
    if a[i][0] == 'Sunny':
        if a[i][2] == 'High':
            if a[i][1] == 'Hot':
                if a[i][4] == 'Yes':
                    sunny_high_hot[0] += 1
                else:
                    sunny_high_hot[1] += 1
            elif a[i][1] == 'Mild':
                if a[i][4] == 'Yes':
                    sunny_high_mild[0] += 1
                else:
                    sunny_high_mild[1] += 1
            elif a[i][1] == 'Cool':
                if a[i][4] == 'Yes':
                    sunny_high_cool[0] += 1
                else:
                    sunny_high_cool[1] += 1
        elif a[i][2] == 'Normal':
            if a[i][1] == 'Hot':
                if a[i][4] == 'Yes':
                    sunny_normal_hot[0] += 1
                else:
                    sunny_normal_hot[1] += 1
            elif a[i][1] == 'Mild':
                if a[i][4] == 'Yes':
                    sunny_normal_mild[0] += 1
                else:
                    sunny_normal_mild[1] += 1
            elif a[i][1] == 'Cool':
                if a[i][4] == 'Yes':
                    sunny_normal_cool[0] += 1
                else:
                    sunny_normal_cool[1] += 1
```

In [124]:

```
Gain_sunny_high_temp = one_inf(sunny_high)-one_inf(sunny_high_hot) - one_inf(sunny_high_mil
Gain_sunny_normal_temp = one_inf(sunny_normal)-one_inf(sunny_normal_hot) - one_inf(sunny_no
```

In [125]:

```
print Gain_sunny_high_temp,Gain_sunny_normal_temp
```

```
0.0 0.0
```

we need to exclude all 0, so there is no need to extend from wind to temperature

In [130]:

```
float(sunny_normal[0])/sum(sunny_normal)
```

Out[130]:

```
1.0
```

In [131]:

```
float(sunny_high[0])/sum(sunny_high)
```

Out[131]:

0.0

3.a

the decesion tree should be:

```
    Outlook-sunny    -- Humidity ---high   ----No
                                 ---Normal ----Yes
           -overcast -- Yes
           -rain     -- Wind     ---strong ---- No
                                 ---weak   ---- Yes
```

3.b

for the 1st one, it predict to be Yes but actually No;

for the 2nd one, it predict to be Yes and actually Yes;

for the 3rd one, it predict to be No and actually No;

for the 4th one, it predict to be Yes and actually Yes.