In [5]:
```
import numpy as np
import pandas as pd
import math
```

In [3]:
```
df = pd.read_csv('play_tennis.csv')
df.head()
```

Out[3]:

	day	outlook	temp	humidity	wind	play
0	D1	Sunny	Hot	High	Weak	No
1	D2	Sunny	Hot	High	Strong	No
2	D3	Overcast	Hot	High	Weak	Yes
3	D4	Rain	Mild	High	Weak	Yes
4	D5	Rain	Cool	Normal	Weak	Yes

In [19]:
```
data = df
del df['day']
```

In [20]:
```
decision_tree = {}
n = len(df)
def calculate_overral_entropy(label):
    entropy = 0
    yes = label.count('Yes')
    no = label.count('No')
    if yes > 0:
        entropy += -(yes/(yes+no)*math.log2(yes/(yes+no)))
    if no > 0:
        entropy += -(no/(yes+no)*math.log2(no/(yes+no)))
    return entropy
```

In [53]:
```
def calculate_entropy(df,attribute,label):
    df_len = len(df)
    entropy = 0
    custom_list = list(set(list(df[attribute])))
    custom_list_len = len(custom_list)
    for i in range(custom_list_len):
        internal_entropy = 0
        yes = len(df[(df[attribute] == custom_list[i]) & (df[label] == 'Ye
        no = len(df[(df[attribute] == custom_list[i]) & (df[label] == 'No'
        if yes > 0:
            internal_entropy += -(yes/(yes+no)*math.log2(yes/(yes+no)))
        if no > 0:
            internal_entropy += -(no/(yes+no)*math.log2(no/(yes+no)))
        if internal_entropy > 0:
            entropy += (yes+no)/df_len*internal_entropy
    return calculate_overral_entropy(list(df[label])) - entropy
```

In [54]:
```
1 calculate_entropy(df,'outlook','play')
```

Out[54]: 0.24674981977443933

In [48]:
```
1 0.9402859586706311 - 0.6935361388961918
```

Out[48]: 0.24674981977443933