Decision Tree:
---
x1,x2,...xn
y

Node:features
leaf:y

ALGO'S:

ID3:
(iterative dichotomiser3)

Entropy & Information Gain functions

CART:
(classification and regression tree)

Gini function

steps in ID3:
---
1. calculate entropy of dataset(target)
2. calculate entropy of each feature
3. calculate information gain of each feature
4. highest gain fetaure becomes root node.
5. repeat same steps to buid complete tree.

1.

findout classes in dataset.
prob of class yes:
frequency of yes/no of samples
9/14
.642

prob of class no:
frequency of yes/no of samples
5/14
.358
p(yes)*log2(p(yes))=.642*(-0.63935479754)=-0.409152
p(no)*log2(p(no))=.358*(-1.4819685074)=-0.530544

entrop(t)=-[p(yes)*log2(p(yes))+p(no)*log2(p(no))]
=0.94

H(T)=-Sum[p(ci)*log2(p(ci))]

2.

2.1 Outlook:
find out unique values:
rainy
overcast

```
                sunny

        entropy of rainy:
                -[3/5*log2(3/5)+2/5*log2(2/5)]
                =.971
        entropy of overcast:
                -[4/4*log2(4/4)]
                =0
        entropy of sunny:
                -[3/5*log2(3/5)+2/5*log2(2/5)]
                =.971

        average entropy of outlook:
        e(rainy)*p(r)+e(o)*p(o)+e(s)*p(s)
        (.971*5/14)+0+.(971*5/14)
        =.69

2.2 Temperature:
        find out unique values:
                hot
                cool
                mild

        entropy of hot:
                -[2/4*log2(2/4)+2/4*log2(2/4)]
                =1
        entropy of cool:
                -[3/4*log2(3/4)+1/4*log2(1/4)]
                =0.811
        entropy of mild:
                -[4/6*log2(4/6)+2/6*log2(2/6)]
                =0.918

        average entropy of temperature:
        e(hot)*p(hot)+e(cool)*p(cool)+e(mild)*p(mild)
        (1*4/14)+(.811*4/14)+.(918*6/14)
        =.92

2.3 Humidity:
        find out unique values:
                high
                normal

        entropy of high:
                -[3/7*log2(3/7)+4/7*log2(4/7)]
                =.985
        entropy of normal:
                -[6/7*log2(6/7)+1/7*log2(1/7)]
                =.591

        average entropy of :
        e(high)*p(high)+e(normal)*p(normal)
        (.98*7/14)+(.591*7/14)
        =.79

2.3 Windy:
        find out unique values:
```

true
false

entropy of true:
-[3/6*log2(3/6)+3/6*log2(3/6)]
=1
entropy of false:
-[6/8*log2(6/8)+2/8*log2(2/8)]
=.811

average entropy of :
e(true)*p(true)+e(false)*p(false)
(1*6/14)+(.811*8/14)
=.892

3.

Information Gain of Outlook:
H(t)-H(Outlook)
=.94-.69
=.25

Information Gain of Temperature:
H(t)-H(temperature)
=.94-.92
=.02

Information Gain of Humidity:
H(t)-H(Humidity)
=.94-.79
=.15

Information Gain of Windy:
H(t)-H(Windy)
=.94-.892
=.048

Now outlook is root node:
overcast is always yes
rainy,sunny need further splitting.

Rainy:

Outlook	Temp	Humidity	Windy	Play
Rainy	Hot	High	FALSE	No
Rainy	Hot	High	TRUE	No
Rainy	Mild	High	FALSE	No
Rainy	Cool	Normal	FALSE	Yes
Rainy	Mild	Normal	TRUE	Yes

repeat untill we get leaf....