Leaf nodes

Branch node

Root node



Root node → decision node

2nd

and

Branch node

Branch node

Leaf node

Leaf node

Leaf node

branch node

Leaf

Leaf

Ex

Salary is bet⁷ 50k to 80k

yes

no

} decision

Office is near to home

Declined offer → leaf nod

yes

No

Cab facility

declined offer

vy

No

declined

Accepted

ex



Age
< h0        > h0
income                    cibiz
  7hok      selin
apphre

---

on

| color | Dia | label |
|-------|-----|-------|
| Green | 3 | Mango |
| yellow | 8 | Mango |
| red | 1 | Grape |
| red | 1 | Grape |
| yellow | 2 | Lemon |

R = 1 = Grape
R = 1 = Grape



no      yes

Dia >= 2
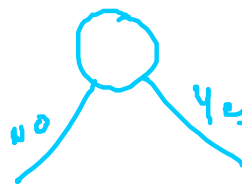
G = 3   = Mango
y = 3   = Mango
y = 2   = Lemon

Gini impurity
↓
Information Gain

for I.G  ↑ > entropy

Gini ⓪      Apple

label

apple

apple
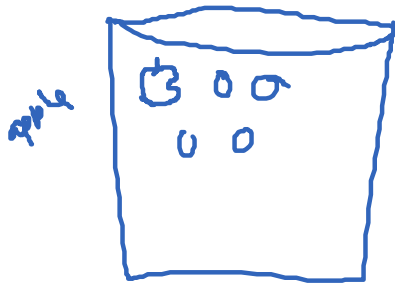


$$P = 1 \quad >> \text{highest}$$

$$\text{impurity} = \text{low} = 0$$

apple

case (i)



$$P = \text{low}$$

$$\neq 1$$

$$\text{impurity} \neq 0$$



$$H(x)$$
$$K(2)$$
$$\text{entropy}(x)$$

$p(x)$  0.5   1
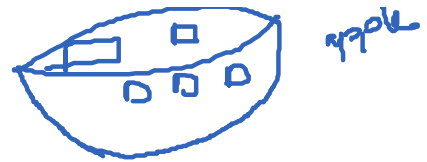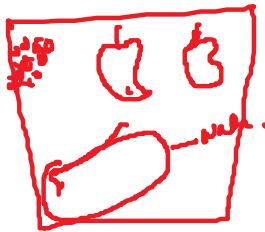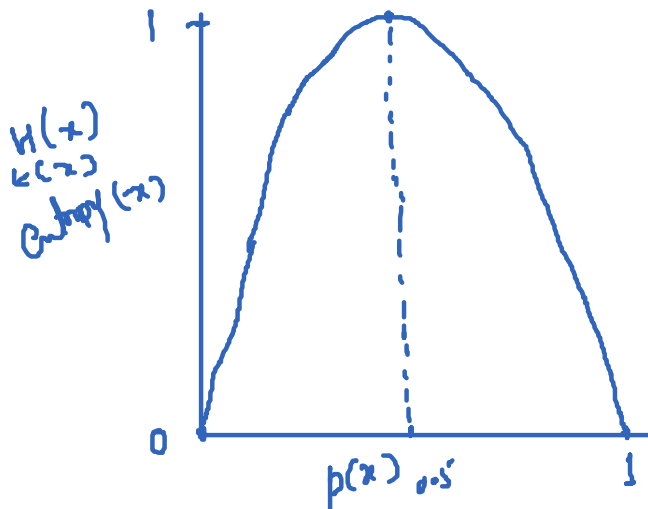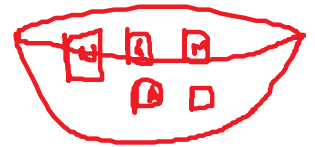
$$\text{Entropy}(s) = -P(y)\log_2 P(y) - P(no)\log_2 P(N)$$

(i) IF no. of $y$ = no. of $NO$ = $P(s) = \underline{\underline{0.5}}$

$$\Rightarrow \text{entropy} = 1$$
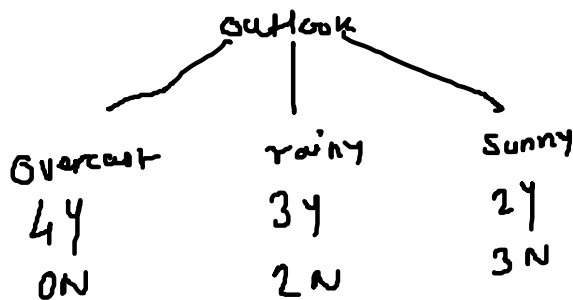
(ii) all yes or all no = $P(s) = 1 \text{ or } 0$

$\Rightarrow$ entropy = 0

$$H(s) : E(s) = -p(y)\log_2 p(y)$$
$$= -p(y)\log_2 p(y) - p(N)\log_2 p(N)$$

$$p(N) = p(y) = 0.5$$
$$E(s) =$$

# # Information Gain



```
              outlook
         /       |       \
   Overcast    rainy    Sunny
     4Y         3Y        2Y
                          3N
     ON         2N
```

$$E(\text{outlook} = \text{overcast}) = -\frac{4}{4}\log_2 \frac{4}{4} - 0\log_2 0$$
$$= 0$$

$$E(ov = rain) = -\frac{3}{5}\log_2 \frac{3}{5} - \frac{2}{5}\log_2 \frac{2}{5}$$
$$= 0.971$$

$$E(ov = Sunny) = -\frac{2}{5}\log_2 \frac{2}{5} - \frac{3}{5}\log_2 \frac{3}{5}$$
$$= 0.971$$

Information :→        $\qquad$ wt vy $= \frac{4}{14} \times 0$

$$I(\text{outlook}) = \frac{4}{14} \times 0 + \frac{5}{14} \times 0.971 + \frac{5}{14}(0.971)$$

$$= 0.693$$

$$I_{feather} = 0.94 - 0.693$$

$$= 0.247$$

## # Temp



Temp
- hot: 2Y, 2N
- cool: 3Y, 1N
- mild: 4Y, 2N
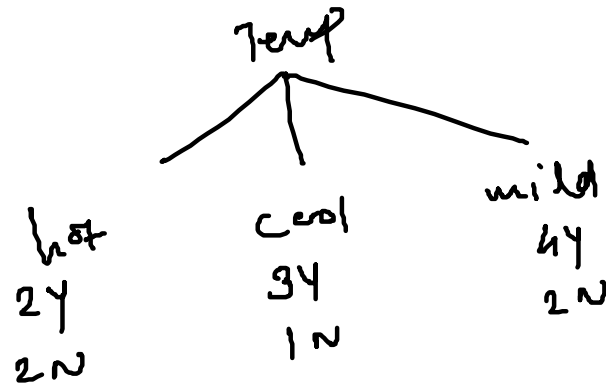
$$E(Hot) = -\frac{2}{4}\log_2 \frac{2}{4} - \frac{2}{4}\log_2 \frac{2}{4}$$

$$= 1$$

$$E(cool) = -\frac{3}{4}\log_2 \frac{3}{4} - \frac{1}{4}\log_2 \frac{1}{4}$$

$$= 0.81$$

$$E(mild) = 0.91$$

$$I(temp) = 0.91$$

$$I_h(temp) = 0.03$$

## # Humidity

Humidity
- high
- normal: 6Y
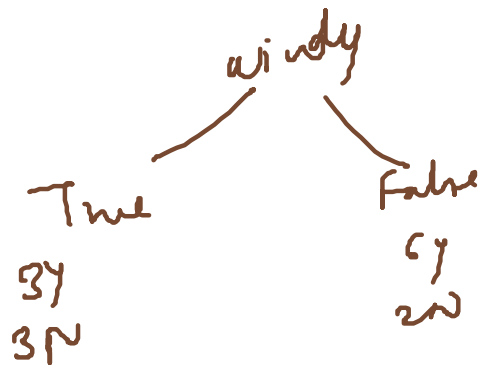
3Y
4N

normal
6Y
1N

$E(high) = 0.98$

$E(nom) = 0.59$

$I(num) = 0.78$

$I_G(hum) = \underline{0.16}$

## # Windy

windy

True
3Y
3N

False
6Y
2N

$E(T) = 1$

$E(F) = 0.81$

$I(windy) = 0.89$

$I_G(windy) = 0.05$

Outlook

sunny → Humidity
overcast → yes
rainy → windy → Pure

Humidity:
high → NO
normal → yes

windy:
True → NO
False → yes

Outlook
|
humid

humidity

Temp high | norm

mild hot

$IG\_Out : 0.24) \rightarrow$ Root node

$IG\_Hum = 0.15 \rightarrow$

$IG\_windu = 0.05$

$IG\_Temp = 0.03$

Out

sun rain

humid windy Temp

Temp

# Gini index

$$G.I = 1 - \Sigma(P)^2$$
$$= 1 - \left[ P(4)^2 + P(N)^2 \right]$$

$P(T)$ | $P(W)$
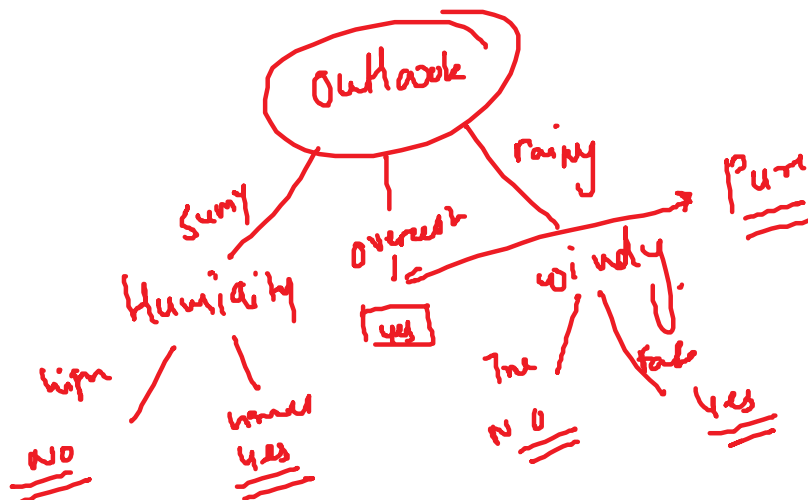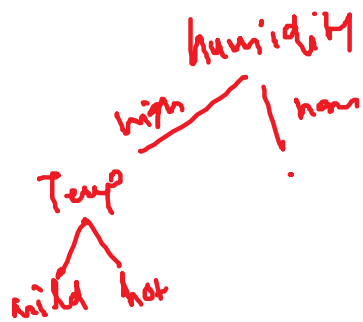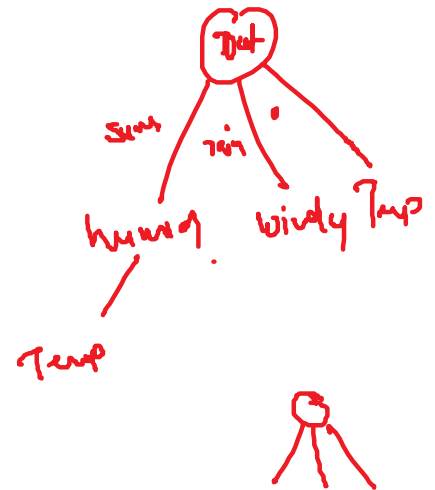
case ① All yes $\stackrel{==}{\text{or}}$ All No $\quad = P = 1$

$$G.I = 1 - (1)^2$$
$$= 0$$

case ② equal yes or equal No $\quad = P = 0.5 \ \& \ 0.5$

$$G.I = 1 - \left[ (0.5)^2 + (0.5)^2 \right]$$

$$= 1 - [0.25 + 0.25]$$

$$= 1 - 0.5$$

$$\boxed{G.I = 0.5}$$

highest value of G.I

hi



0.5

p(0-1)

Entropy = yellow
Gini = Green

$$GI = 1 - [py^2 + pNO^2]$$

$$= 1 - [1^2 + 0^2]$$      All yes

$$= 1 - 1$$

$$= 0$$

$$GI = 1 - [0^2 + 1^2]$$      All No
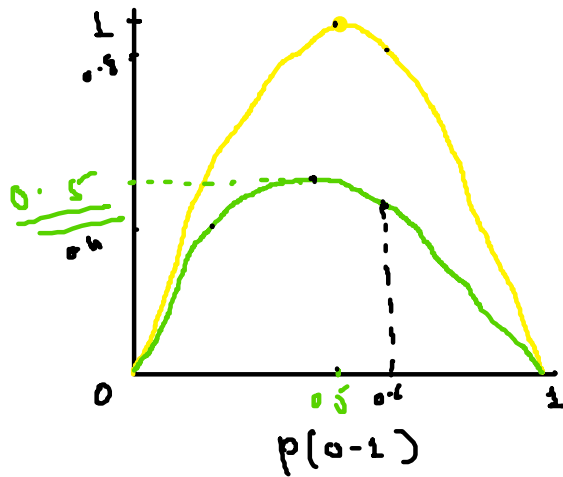
$$= 1 - 1$$

$$= 0$$

1) = Gini < Entropy

2) = Gini doesnot use log
it is faster in terms of entropy

# G.I.

Outlook

- Sunny: 2Y, 3N
- Ocast: 4Y, 0N
- Rainy: 3Y, 2N0

$$G(s) = 1 - \left[\left(\frac{2}{5}\right)^2 + \left(\frac{3}{5}\right)^2\right]$$

$$= 0.48$$

$$G(0) = 1 - \left[1^2 + 0^2\right]$$

$$= 0$$

$$G(r) = 1 - \left[\frac{3}{5}^2 + \left(\frac{2}{3}\right)^2\right]$$

$$= 0.48$$
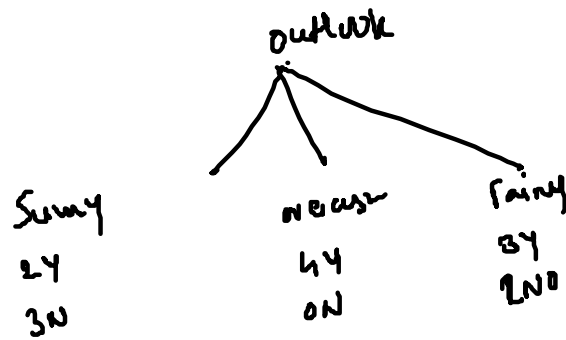
$$Gini\ index = \varepsilon\ \frac{Di}{D} \times G.I\ (Di)$$

$$= \frac{5}{14} \times 0.48 + \frac{4}{14} \times 0 + \frac{5}{14} \times 0.48$$

$$=$$

g lacs          75 h          All glac are Fixed | no variable

Basics   = 4.5      50%.

HRA     = 2.5      20%.  25%.

TA
new.
1mt
special
PF → employee      gov.
P → employer    _____
                  700000        750000
                                 800
             ESIC    -           1800
             PF      =           1800
                              _____
                                 712000

EL Variables = Joining =
              Relocation =
                                        18k
              Special =

              Quar ⟶    3      4      5      6   = 18
              Annual =

regime

new    ⟶    comp in hr  less

old  - →  old tax → in hand  salary  more

                                   ‾
        3  to 5  →   10       Salary slip
      _____            __            Tent
      _____                          ___
                      30       29

      1                              Emp . UAN
      =

pan --

hi

pan --