C4.5 Algorithm

함수활용

```
def E(l):
    entropy = 0
    for n in l:
        entropy += -(n/sum(l)) * (math.log2(n/sum(l)))
    return entropy
```

1-① ooooooooo xxxxx

Outlook

S: ooxxx O: oooo R: ooo xx

$H(Play) = E[9,5] = 0.940$

$H(Play, outlook) = \frac{5}{14}E[2,3] + \frac{4}{14}\times 0 + \frac{5}{14}E[3,2] = 0.694$

$IV = E[5,4,5] = 1.577$

$\therefore IGR = 0.156$

1-② Temperature

hot: ooxx mild: oooo xx cold: ooo x

$H(Play, Temperature) = \frac{4}{14}E[2,2] + \frac{6}{14}E[4,2] + \frac{4}{14}E[3,1]$

$= 0.911$

$IV = E[4,6,4] = 1.557$

$\therefore IGR = 0.019$

1-③ Humidity

high: ooo xxxx normal: oooooo x

$H(Play, Humidity) = \frac{7}{14}E[3,4] + \frac{7}{14}E[6,1] = 0.788$

$IV = E[7,7] = 1$

$\therefore IGR = 0.152$

1-④ Windy

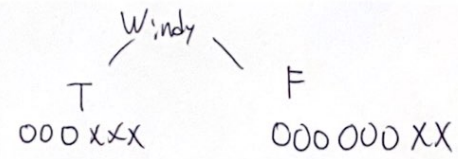
T: ooo xxx F: oooooo xx

$H(Play, Windy) = \frac{6}{14}E[3,3] + \frac{8}{14}E[6,2] = 0.767$

$IV = E[6,8] = 0.985$

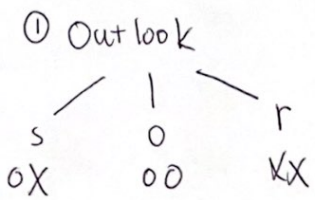
$\therefore IGR = 0.176$

∴ 1. Windy의 IGR이 가장 높다



2.

Windy – True (OOOXXX)

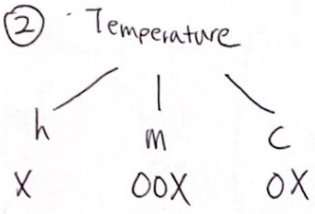


$H(\text{S}) = E([3,3]) = 1$

$H(\text{S, Outlook}) = \frac{1}{3}E([1,1]) + 0 + 0 = \frac{1}{3} = 0.333$

$IV = E([2,2,2]) = 1.585$

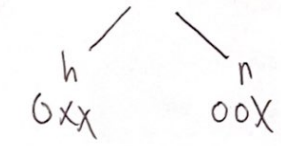
$\therefore IGR = 0.421$



$H(S, \text{Temperature}) = \frac{1}{6}\times 0 + \frac{3}{6}E([2,1]) + \frac{2}{6}E([1,1]) = 0.792$

$IV = E([1,3,2]) = 1.459$

$\therefore IGR = 0.143$



$H(S, \text{Humidity}) = \frac{1}{2}E([1,2]) + \frac{1}{2}E([2,1]) = 0.918$

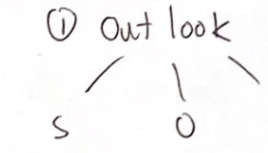
$IV = E([3,3]) = 1$

$\therefore IGR = 0.082$

⇒ Windy – True는 Outlook으로 분류

Windy – False (OOOOOOXX)

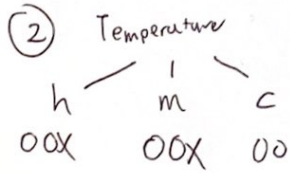
$H(S) = E([6,2]) = 0.811$



$H(S, \text{Outlook}) = \frac{3}{8}E([1,2]) = 0.344$

$IV = E([3,2,3]) = 1.561$

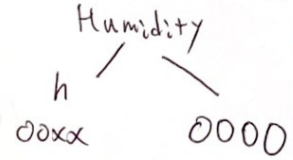
$\therefore IGR = 0.299$



$H(S, \text{Temp}) = \frac{3}{8}E([1,2]) + \frac{3}{8}E([1,2]) + 0 = 0.689$

$IV = E([3,3,2]) = 1.561$

$\therefore IGR = 0.098$

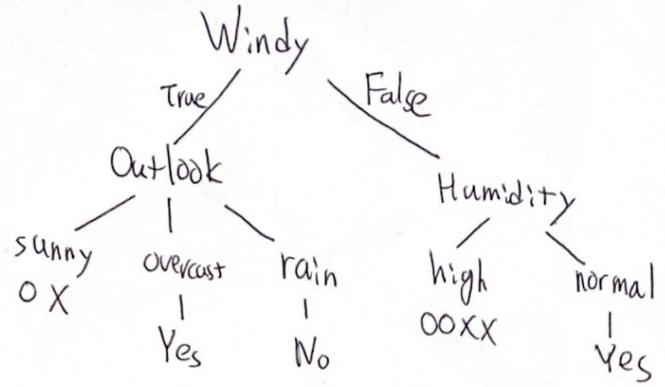


$H(S, \text{Humidity}) = \frac{4}{8}E([2,2]) = 0.5$

$IV = E([4,4]) = 1$

$\therefore IGR = 0.311$

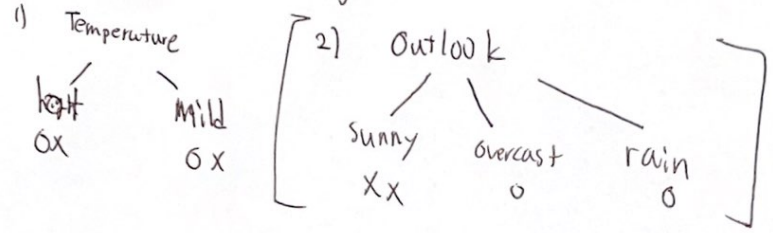
⇒ Windy-False는 Humidity로 분류



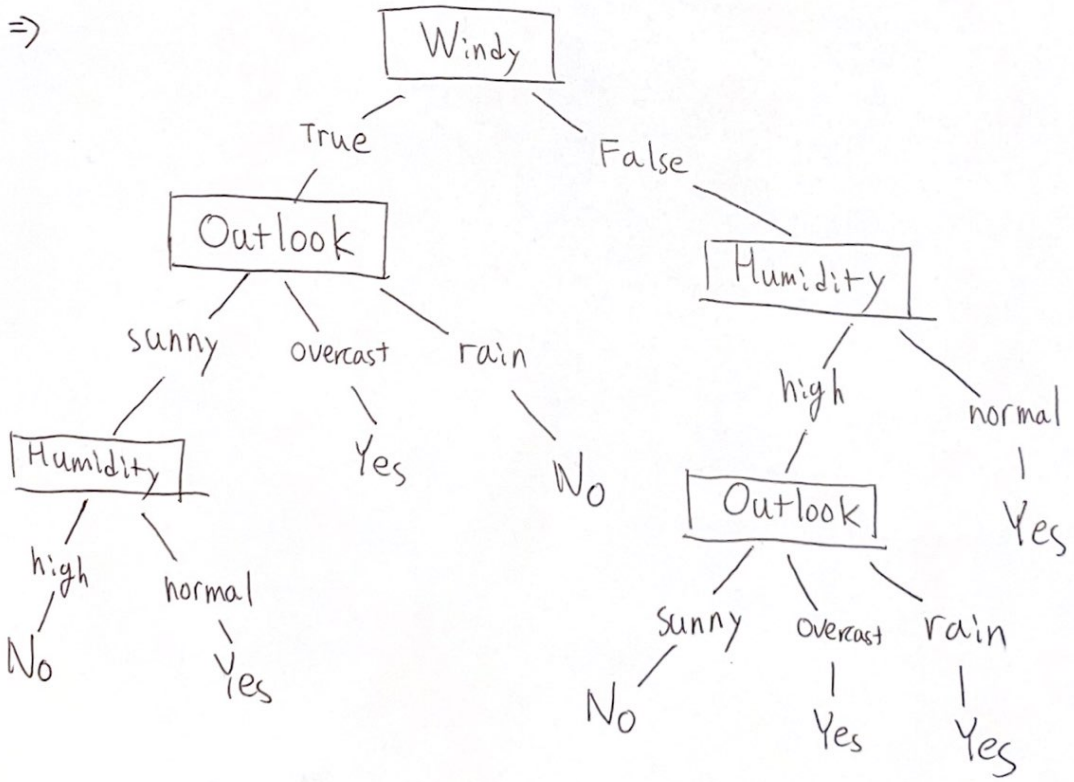
3.

① Windy True – Outlook, Sunny 다음은 Temperature, Humidity 중 아무거나

② Windy False – Humidity high



⇒



CART Algorithm

$G(S) = G([5,9]) = 0.459$

~~$G(\text{Outlook=Sunny}) = G([2,3])$~~

함수활용
```
def G(l):
    gini = 1
    for n in l:
        gini -= (n/sum(l))**2
    return gini
```

1. ① Outlook (sunny / overcast / rain)

sunny: OOXXX overcast: OOOO rain: OOOXX

$G(\text{outlook=sunny}) = \frac{5}{14} G([2,3]) + \frac{9}{14} G([7,2])$

$= 0.394$

$G(\text{outlook=overcast}) = \frac{4}{14} G([4,0]) + \frac{10}{14} G([5,5])$ ✓

$= 0.357$

$G(\text{outlook=rain}) = \frac{5}{14} G([3,2]) + \frac{9}{14} G([6,3])$

$= 0.457$

② Temperature (hot / mild / cool)

hot: OOXX mild: OOOXX cool: OOOX

$G(\text{Temp=hot}) = \frac{4}{14} G([2,2]) + \frac{10}{14} G([7,3])$ ✓

$= 0.443$

$G(\text{Temp=mild}) = \frac{6}{14} G([4,2]) + \frac{8}{14} G([5,3])$

$= 0.458$

$G(\text{Temp=cool}) = \frac{4}{14} G([3,1]) + \frac{10}{14} G([6,4])$

$= 0.45$

③ Humidity (high / normal)

high: OOOXXXX normal: OOOOOOX

$G(\text{Humidity=high/normal}) = \frac{7}{14} G([3,4]) + \frac{7}{14} G([6,1])$ ✓

$= 0.367$

④ Windy (T / F)

T: OOOXXX F: OOOOOOXX

$G(\text{Windy=T/F}) = \frac{6}{14} G([3,3]) + \frac{8}{14} G([6,2])$ ✓

$= 0.429$

⇒ ∴ Outlook = overcast

Outlook (overcast / S/R)

overcast: Yes S/R: OOOOO XXXXX

2. Outlook = sunny/rain

① Outlook

sunny: ooxxx rain: oooxx

$G(\text{Outlook} = \text{sunny/rain}) = \frac{5}{10} G([2,3]) + \frac{5}{10} G([3,2])$

$= 0.48$

② Temperature

hot: xx mild: oooxx cool: oox

$G(\text{Temp} = \text{hot}) = \frac{2}{10} G([2,0]) + \frac{8}{10} G([5,3])$ ✓

$= 0.375$

$G(\text{Temp} = \text{mild}) = \frac{5}{10} G([3,2]) + \frac{5}{10} G([2,3])$

$= 0.48$

$G(\text{Temp} = \text{cool}) = \frac{3}{10} G([2,0]) + \frac{7}{10} G([3,4])$

$= 0.476$

③ Humidity

high: oxxxx normal: oooox

$G(\text{Humidity} = \text{high/normal}) = \frac{5}{10} G([1,4]) + \frac{5}{10} G([4,1])$ ✓

$= 0.320$

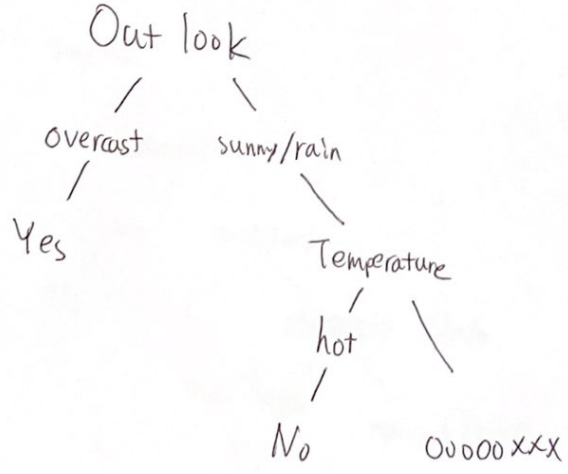
④ Windy

T: oxxx F: ooooxx

$G(\text{Windy} = T/F) = \frac{4}{10} G([1,3]) + \frac{6}{10} G([4,2])$ ✓

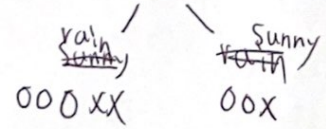
$= 0.417$

⇒ Temperature = hot



3. Temperature = mild/cool

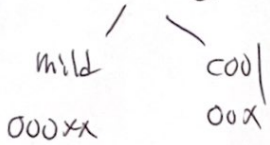
① Outlook



$G(\text{outlook} = \text{rain/sunny}) = \frac{5}{8} G([3,2]) + \frac{3}{8} G([2,1])$

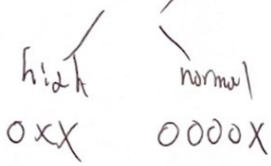
$= 0.467$

② Temperature



$G(\text{Temp} = \text{mild/cool}) = 0.467$

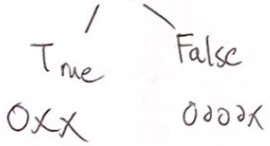
③ Humidity



$G(\text{Humidity} = \text{high/normal}) = \frac{3}{8} G([1,2]) + \frac{5}{8} G([4,1])$

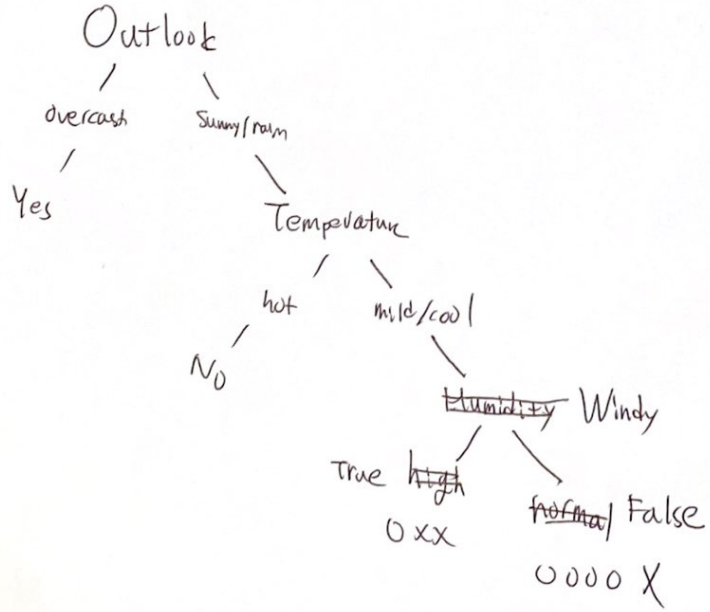
$= 0.367$ ✓

④ Windy



$G(\text{Windy} = \text{True/False}) = 0.367$ ✓

⇒ ~~Humidity = high/normal~~ Windy = True/False



4. Windy = True/False

4-1 Windy = True

OXX

Outlook

sunny / rain

O / XX

4-2 Windy = False

① Outlook $G(\text{Outlook} = \text{sunny/rain}) = \frac{2}{5} G([1,1]) = 0.2$ ∨

sunny / rain

OX / ~~OOO~~

② Temperature. $G(\text{Temp} = \text{mild/cool}) = \frac{3}{5} G([2,1]) = 0.267$

mild / cool

OOX / OO

③ Humidity $G(\text{Humidity} = \text{high/normal}) = 0.2$ ∨

high / normal

OX / OOO

=> 둘다 Outlook = sunny / rain