AI - Assignment-7

19k4/A05F5

Given data:

outlook	Temperature	Humidity	windy	Hours to play.
Rainy	Hot	High	false	25
Rainy	Hot	High	true	30
Overcast	Hot	High	false	46
Sunny	mild	High	False	45
Sunny	cool	Normal	false	22
overcast	Cool	Normal	True	23
Rainy	mild	Normal	True	43
Rainy	Cool	High	False	35
Sunny	mild	Normal	false	38
Rainy	mild	High	False	46
Overcast	mild	Normal	True	48
Overcast	Hot	high	False	52
Sunny	mild	Normal	True	30

Decission tree for the given data:

SFS (10)

Target table (PGH):

mean = 39.78, SD(T) = 9.32 % CV = SD x 100 = 230.

outlook	no. of samples	mean weight
→ Rainy → 8.7	5	5/14
→ overcast → 4.03	4	4/14
→ sunny → 12.15	5	5/14

$$SD(outlook) = \left[\frac{5}{14} \times 8.7\right] + \left[\frac{4}{14} \times 4.03\right] + \left[\frac{5}{14} \times 12.15\right] = 8.59$$

$$SDR(outlook) = SD(T) - SD(outlook) = 9.32 - 8.59 = 0.73$$

Temperature → Hot - 10.34 | 4 | weight(x)
mild - 8.36 | 6 | 4/14
cool - 12.13 | 4 | 6/14
4/14

SD(Temperature) = $\left[\frac{4}{14} \times 10.34\right] + \left[\frac{6}{14} \times 8.38\right] + \left[\frac{4}{14} \times 12.13\right]$

= 10.01

SDR(temperature) = SD(T) - SD(temperature)

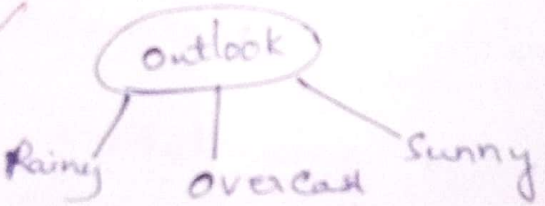
= 9.32 - 10.01 = -0.75

Humidity → High — 9.5169 — 7 | weight(x)
normal — 9.433 — 7 | 7/14
7/14.

SD(Humidity) = $\frac{7}{14} \times 9.51 + \frac{7}{14} \times 9.43 = 9.47$

SDR(Humidity) = SD(T) - SD(Humidity) = 9.32 - 9.47 = -0.15

→ The outlook has the highest SDR so it will became root node



Rainy table:

Temp	Humidity	Windy	Hours to play
Hot	High	False	25
Hot	High	True	30
mild	High.	False	35
Cool	Normal	False	38
mild.	Normal	True.	48.

Target Column (PGH).

mean = 35.2 $\quad$ %Cv = $\frac{50}{\text{mean}} \times 100 = 24.7$.

SD(T) = 8.70

			mean	weight
Temp	Hot	→ 3.53	2	2/5
	mild	→ 9.19	2	2/5
	cool	→ 0	1	1/5

$SD(Temp) = \left[\frac{2}{5} \times 3.53\right] + \left[\frac{2}{5} \times 9.192\right] + \left[\frac{1}{5} \times 0\right] = 5.37.$

SDR (Temp) = SD(Target) - SD(Temp)

= 8.70 - 5.37 = 3.33

Humidity	High	— 7.071	3	3/5
	normal	— 7.07	2	2/5

$SD(Humidity) = \left[\frac{3}{5} 17.071\right] + \left[\frac{2}{5} \times 7.07\right] = 7.07.$

SDR = SD(T) - SD(Hum)

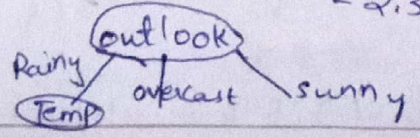
= 8.70 - 7.07 = 1.63

windy	False	— 9.19	3	3/5
	True	— 12.72	2	2/5

$SD(windy) = \left[\frac{3}{5} 19.19\right] + \left[\frac{2}{5} \times 12.72\right]$

SDR = SD(Target) - SD(windy)

= 8.70 - 11.03 = -2.33



OverCast table:

Tempearture	Humidity	windy	Hours to play.
Hot	High	False	46
Cool	normal	True	43
mild	High	True	52
Hot.	normal	False	44

Target Hours to play:

$SD(T) = 46/5^{+} = 4.0311$, mean = 46.25

$\%(T) = \frac{SD}{mean} \times 100 = 8.71$

Here the $\%(T) \leq 10$, number of sample = 4

So here it Contains leaf node.

Sunny table:

Temp	Humidity	windy	Hours to play.
mild	High	False	45
Cool	High	False	52
Cool	normal	True	23
mild	normal	False	46
mild.	High.	False	30

Target Column:

mean = 39.2 SD(Target) = 12.15 %. CV = $\frac{SD(T)}{mean} \times 100$

Temp — mild — 8.96 — 3 — 3/5 = 30.9

Cool — 20.50 — 2 — 2/5

SD = 13.57 SDR = 12.15 − 13.576 = −1.42

Humidity — High — 11.23 — 3 — 3/5
normal — 16.26 — 2 — 2/5

SD(H) = 13.24

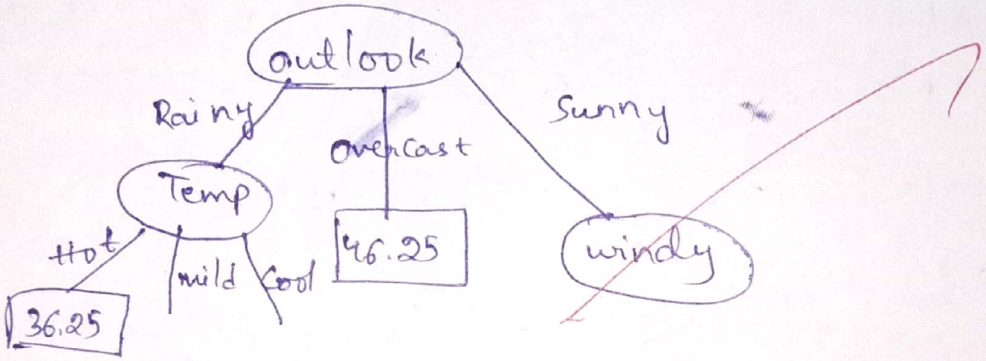
SDR = 12.15 - 13.24 = -1.09

Windy — False — 3.78 — 3 — 3/5
True — 4.94 — 2 — 2/5

SD(TV) = 4.244

SDR = SD(Target) - SD = 12.15 - 4.244

SDR = 7.91



Mild table:

Humidity.	windy	Hours to play.
High	False	45
High	False	35
normal	False	48
normal	True	46
High.	True	52

Target: mean = 45.2 SD = 6.370

%Cv = $\frac{SD}{mean} \times 100$ = 13.938.

Humidity
- High — 8.54 — 3 — 3/5
- normal — 1.41 — 2 — 2/5

SD(Hum) = 5.688

SDR(Hum) = SD(T) − SD(Hum)

= 6.30 − 5.688 = 0.62

Cool table:

Humidity	windy	Hours to play.
Cool → Normal	False	52
Cool → Normal	True	23
Cool → Normal	True	43
Cool → Normal	False.	38

mean(T) = 39 SD(T) = 12.13

%CV = $\frac{SD}{mean} \times 100$ = 31.10

Humidity → Normal → 12.13 5/5

SD(Hum) = 12.13

SD(Hum) = SD(T) − SD(Hum) = 6.30 − 12.13 = − 5.83