P. Rithika Reddy
19K41A05A8

Assignment - 1

Decision Tree. Termination Criteria $CV \leq 10.0$, $n \leq 4$

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	52
Sunny	Cool	Normal	True	23
Overcast	Cool	Normal	True	43
Rainy	Mild	High	False	35
Rainy	Cool	Normal	False	38
Sunny	Mild	Normal	False	46
Rainy	Mild	Normal	True	48
Overcast	Mild	High	True	52
Overcast	Hot	Normal	False	44
Sunny	Mild	High	True	30

SD(Target) = 9.67

		Mean	SD	CV	n	ω(v)
Outlook	Rainy	35.2	8.7	24.7	5	5/14
	Overcast	46.25	4.03	8.72	4	4/14
	Sunny	39.2	12.2	31.0	5	5/14

$SD(Outlook) = \frac{5}{14} \times 8.7 + \frac{4}{14} \times 4.03 + \frac{5}{14} \times 12.2 = 8.59$

SDR(Outlook) = 9.67 − 8.59 = 1.08

		Mean	SD	CV	n	ω(v)
Temperature	Hot	36.25	10.34	30.6	4	4/14
	Cool	39	12.14	31.1	4	4/14
	Mild	42.6	8.38	19.65	6	6/14

$SD(Temperature) = 10.01 \leftarrow \frac{4}{14} \times 10.34 + \frac{4}{14} \times 12.14 + \frac{6}{14} \times 8.38$

SDR(Temperature) = −0.34

		Mean	SD	CV	n	ω(v)
Humidity	High	37.51	10.11	26.92	7	7/14
	Normal	4.2	9.4	22.4	7	7/14

$SD(Humidity) = \frac{7}{14} \times 10.11 + \frac{7}{14} \times 9.4 = 9.77$

SDR(Humidity) = 9.67 − 9.77 = −0.1

		Mean	SD	CV	n	w(v)
Windy	True	37.6	11.6	30.8	6	6/14
	False	41.3	8.41	20.3	8	8/14

$SD = \frac{6}{14} \times 11.6 + \frac{8}{14} \times 8.41 = 9.77$

$SDR(Windy) = 9.67 - 9.77 = -0.1$

→ Outlook becomes the root node. Overcast attr satisfies the hyper parameters so it becomes the leaf node.

Temp	Humidity	Windy	Hours
Hot	High	False	25
Hot	High	True	30
Mild	High	False	35
Cool	Normal	False	38
Mild	Normal	True	48

$SD = 8.7$

		Mean	SD	CV	n	w(v)
Temperature	Hot	27.5	3.5	12.9	2	2/5
	Cool	38	0	0	1	1/5
	Mild	41.5	9.19	22.2	2	2/5

$SD(Temp) = 5.07 \leftarrow \frac{2}{5} \times 3.5 + 0 + \frac{2}{5} \times 9.19$

$SDR(Temp) = 8.7 - 5.07 = \boxed{3.6}$

		M	SD	CV	n	w(v)
Hiamidity	High	30	5	16.67	3	3/5
	Normal	36.5	2.12	5.8	2	2/5

$SD(Hum) = \frac{3}{5} \times 5 + \frac{2}{5} \times 2.12 = 3.84$

$SDR(Hum) = 8.7 - 3.84 = 4.86$

		M	SD	CV	n	w(v)
Windy	True	39	12.7	32.5	2	2/5
	False	32.6	6.8	20.8	3	3/5

$SD(Windy) = \frac{2}{5} \times 12.7 + \frac{3}{5} \times 6.8 = 9.16$

$SDR(Windy) = 8.7 - 9.16 = -0.46$

→ Here Temp has highest SDR. So it becomes decision node.

Sunny.

Temperature	Humidity	Windy	Hours.
Mild	High	False	45
Cool	Normal	False	52
Cool	Normal	True	23
Mild	Normal	False	46
Mild	High	True	30

$SD(Sunny) = 12.15$

		M	SD	CV	n	w(v)
Temp	Mild	40.34	8.96	22.21	3	3/5
	Cool	37.5	20.5	54.67	2	2/5

$SD(Temp) = \frac{3}{5} \times 8.96 + \frac{2}{5} \times 20.5 = 13.57$

$SDR(Temp) = 12.15 - 13.57 = -1.42$

		M	SD	CV	n	w(v)
Humidity	High	37.5	10.6	28.2	2	2/5
	Normal	40.34	15.3	037	3	3/5

$SD(Hum) = \frac{2}{5} \times 10.6 + \frac{3}{5} \times 15.3 = 13.42$

$SDR(Hum) = 12.15 - 13.42 = -1.27$

		M	SD	CV	n	w(v)
Windy	True	26.5	4.94	18.6	2	2/5
	False	47.6	3.78	7.94	3	3/5

$SD(Windy) = \frac{2}{5} \times 4.94 + \frac{3}{5} \times 3.78 = 4.24$

$SDR(Win) = 12.15 - 4.24 = 7.91$

→ Here the highest SDR is in Windy. So, it becomes a decision node.

→ At the node Humidity, it has two branches High, Normal. The values satisfy the hyper parameters so they become the leaf nodes.

→ Similarly at the node Windy, we get leaf nodes.

Decision Tree.