

Artificial Intelligence

G. Shiva

20K45A0414

Assignment - 7

Decision Tree Regression Algorithm

Draw a decision tree diagram to Predict number of hours to play based on weather conditions like outlook, temperature, humidity, windy.

outlook	temperature	Humidity	Windy	Hours played
Rainy	Hot	High	False	25
Rainy	Hot	High	true	30
overcast	Mild	High	true	59
Sunny	Mild	High	False	45
Sunny	cool	Normal	False true	53
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	Hot	High	False	46
overcast	Hot	Normal	False	44
Sunny	Mild	High	true	30

→ termination criteria $CV \leq 10\%$

$$n = 4$$

⇒ Decision tree to predict the number of hours played

⇒ standard deviation of hours played

$$SD(\text{Hours played}) = 9.39$$

⇒ calculation of standard deviation of feature columns:-

outlook	Mean	standard deviation	count(n)
Sunny	39.2	10.87	5
Rainy	35.2	7.78	5
overcast	46.25	3.49	4

$$SD(\text{outlook}) = \frac{5}{14} \times 10.87 + \frac{5}{14} \times 7.78 + \frac{4}{14} \times 3.49 = 7.66$$

$$SDR = 1.66$$

temperature	Mean	standard deviation	count(n)
Hot	36.25	8.95	4
cool	39	10.51	4
Mild	42.66	7.65	6

$$SD(\text{temperature}) = \frac{4}{14} \times 8.95 + \frac{4}{14} \times 10.51 + \frac{6}{14} \times 7.65 = 8.84$$

$$SDR = 0.48$$

Humidity	Mean	Standard deviation	count(n)
High	37.57	9.36	7
Normal	42	8.73	7

$$SD(\text{Humidity}) = \frac{7}{14} \times 9.36 + \frac{7}{14} \times 8.73 = 9.05$$

$$SDR = 0.27$$

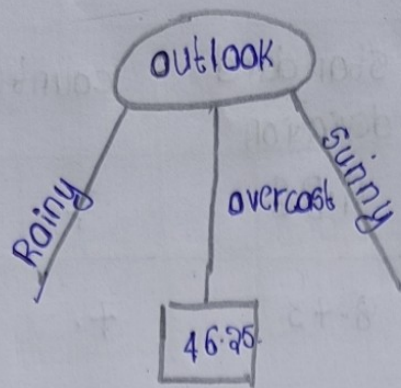
Windy	Mean	Standard deviation	count(n)
True	37.66	10.59	6
False	41.37	7.87	8

$$SD(\text{Windy}) = \frac{6}{14} \times 10.59 + \frac{8}{14} \times 7.87 = 9.04$$

$$SDR = 0.28$$

⇒ Since outlook has the highest SDR, outlook becomes the root node

⇒ For outlook overcast, $n_1 = 4$ and $CV < 10$.
Therefore it will have a leaf node with output as
mean of overcast values i.e. 46.25



⇒ For outlook rainy:

Temperature	Humidity	Windy	Hours played
Hot	High	False	25.
Hot	High	True	30
Mild	High	False	35
Cool	Normal	False	38
Mild	Normal	True	48

SD (Hours played) = 7.78

⇒ calculation of standard deviation to find the next node on outlook rainy.

temperature	Mean	standard deviation	count(n)
Hot	27.5	2.5	2
Cool	37	0	1
Mild	41.5	6.5	2

$$SD (Temperature) = \frac{2}{5} \times 2.5 + \frac{1}{5} \times 0 + \frac{2}{5} \times 6.5 = 3.6$$

$$SDR = 4.18$$

Humidity	Mean	standard deviation	Count(n)
High	30	4.08	3
Normal	43	5	2

$$SD(\text{Humidity}) = \frac{3}{5} \times 4.08 + \frac{2}{5} \times 5 = 4.45$$

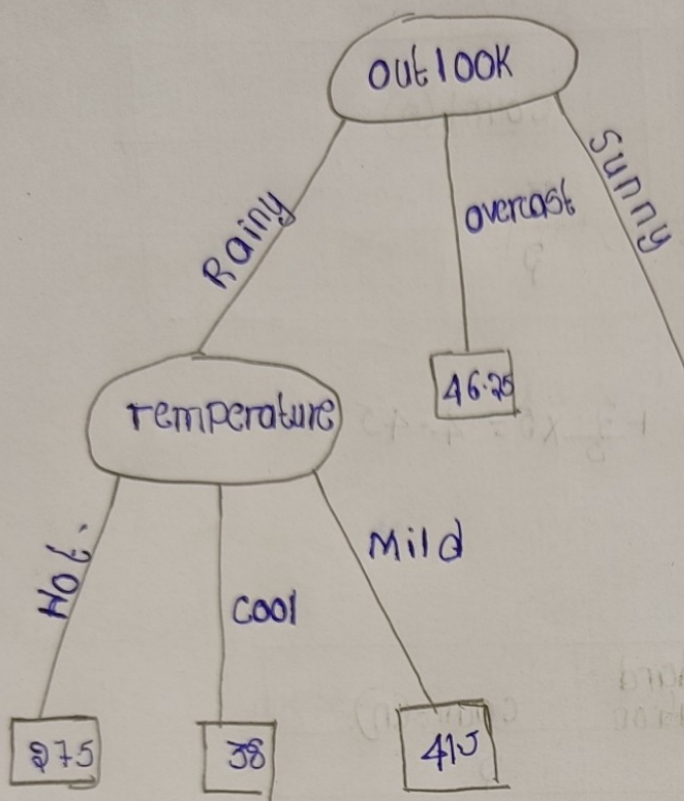
$$SDR = 3.33.$$

Windy	Mean	standard deviation	count(n)
True	39	9	2
False	32.66	5.56	3

$$SD(\text{Windy}) = \frac{2}{5} \times 9 + \frac{3}{5} \times 5.56 = 6.93$$

$$SDR = 0.85.$$

⇒ since temperature has the highest SDR, it becomes the next node on the rainy branch of outlook and the leaf nodes are added with mean value of each attribute.



For outlook sunny :-

temperature	Humidity	Windy	Hours played
Mild	High	False	45
cool	Normal	False	59
cool	Normal	True	23
Mild	Normal	False	46
Mild	High	True	30

$$SD(\text{Hours played}) = 10.87$$

⇒ calculation of standard deviation to find the mode on
but look sunny ÷

temperature	Mean	standard deviation	count(n)
Mild	40.33	7.32	3
cool	37.5	14.5	2

$$SD(\text{temperature}) = \frac{3}{5} \times 7.32 + \frac{2}{5} \times 14.5 = 10.19$$

$$SDR = 0.68$$

Humidity	Mean	standard deviation	count(n)
High	37.5	7.5	2
normal	40.33	12.50	3

$$SD(\text{Humidity}) = \frac{2}{5} \times 7.5 + \frac{3}{5} \times 12.5 = 10.50$$

$$SD = 0.37$$

Windy	Mean	Standard deviation	count (n)
True	26.5	3.5	2
False	47.67	3.09	3

$$SD(\text{Windy}) = \frac{2}{5} \times 3.5 + \frac{3}{5} \times 3.09 = 3.25$$

$$SDR = 7.61$$

⇒ Since Windy has the highest SDR, it becomes next node on sunny branch and all the attributes satisfy the termination criteria therefore leaf nodes are added with mean values as output

