

Decision Tree Regressor

Decision Tree Regressor

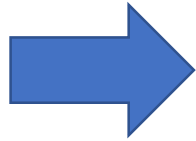
- In regression problems, instead of using Gini, entropy , and information gain, we use measures such MSE, and Standard deviation reduction.
- We use standard deviation to calculate the homogeneity of a numerical sample. If the numerical sample is completely homogenous its standard deviation is zero

Decision Tree Regressor

Outlook	Temp	Humidity	Windy	Hours played
Rainy	Hot	High	False	25
Rainy	Hot	High	True	30
Overcast	Hot	High	False	48
Sunny	Mild	High	False	46
Sunny	Cool	Normal	True	62
Sunny	Cool	Normal	True	23
Overcast	Cool	Normal	False	43
Rainy	Mild	High	False	36
Rainy	Cool	Normal	False	38
Sunny	Mild	Normal	True	48
Rainy	Mild	Normal	True	48
Overcast	Mild	High	High	62
Overcast	Hot	Normal	Normal	44
Sunny	Mild	High	High	30

Decision Tree Regressor

Hours played
25
30
48
46
62
23
43
36
38
48
48
62
44
30



- *Count* = $n = 14$
- *Average* = $\bar{x} = \frac{\sum x}{n}$
- *Standard Deviation* = $S = \sqrt{\frac{\sum (x - \bar{x})^2}{n}} = 9.32$
- *Coefficient of Variation* = $CV = \frac{S}{\bar{x}} \times 100 = 23\%$

Decision Tree Regressor

- *Standard Deviation* = $S(T, X) = \sum_i P(c)S(c)$
- $S(\text{Hours}, \text{Outlook}) = P(\text{Sunny})S(\text{Sunny}) + P(\text{Overcast})S(\text{Overcast}) + P(\text{Rainy})S(\text{Rainy})$
- $S(\text{Hours}, \text{Outlook}) = (4/14)3.49 + (5/14)7.78 + (5/14)10.87 = 7.66$

		Hours played StdDev
Outlook	Overcast	3.49
	Rainy	7.78
	Sunny	10.87
SDR = 1.66		

$$SDR(T, X) = S(T) = S(T, X)$$

$$SDR(\text{Hours}, \text{Outlook}) = 9.32 - 7.66 = 1.66$$

Decision Tree Regressor

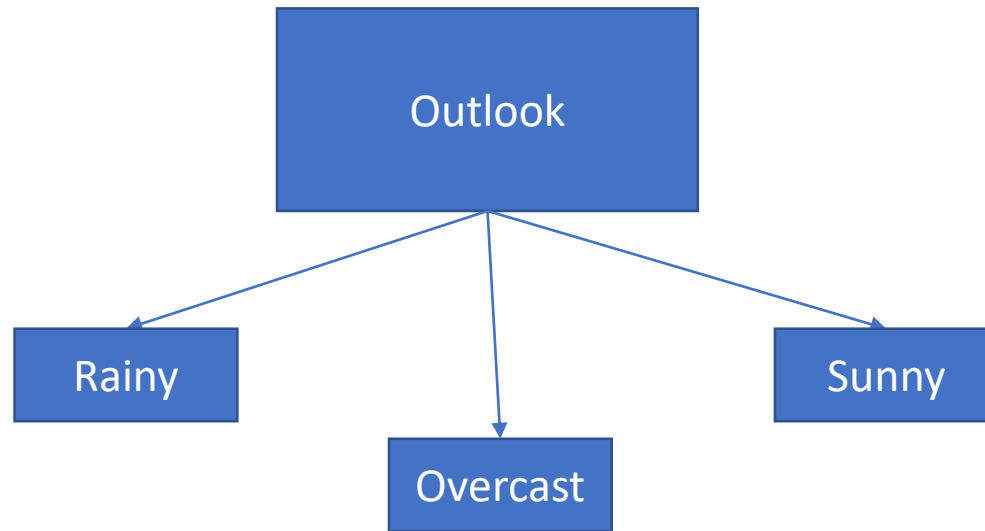
		Hours played StdDev
Outlook	Overcast	3.49
	Rainy	7.78
	Sunny	10.87
SDR = 1.66		

		Hours played StdDev
Temp	Cool	10.51
	Hot	8.95
	Mild	7.65
SDR = 0.17		

		Hours played StdDev
Humidity	High	9.36
	Normal	8.37
SDR = 0.28		

		Hours played StdDev
Windy	False	7.87
	True	10.59
SDR = 0.29		

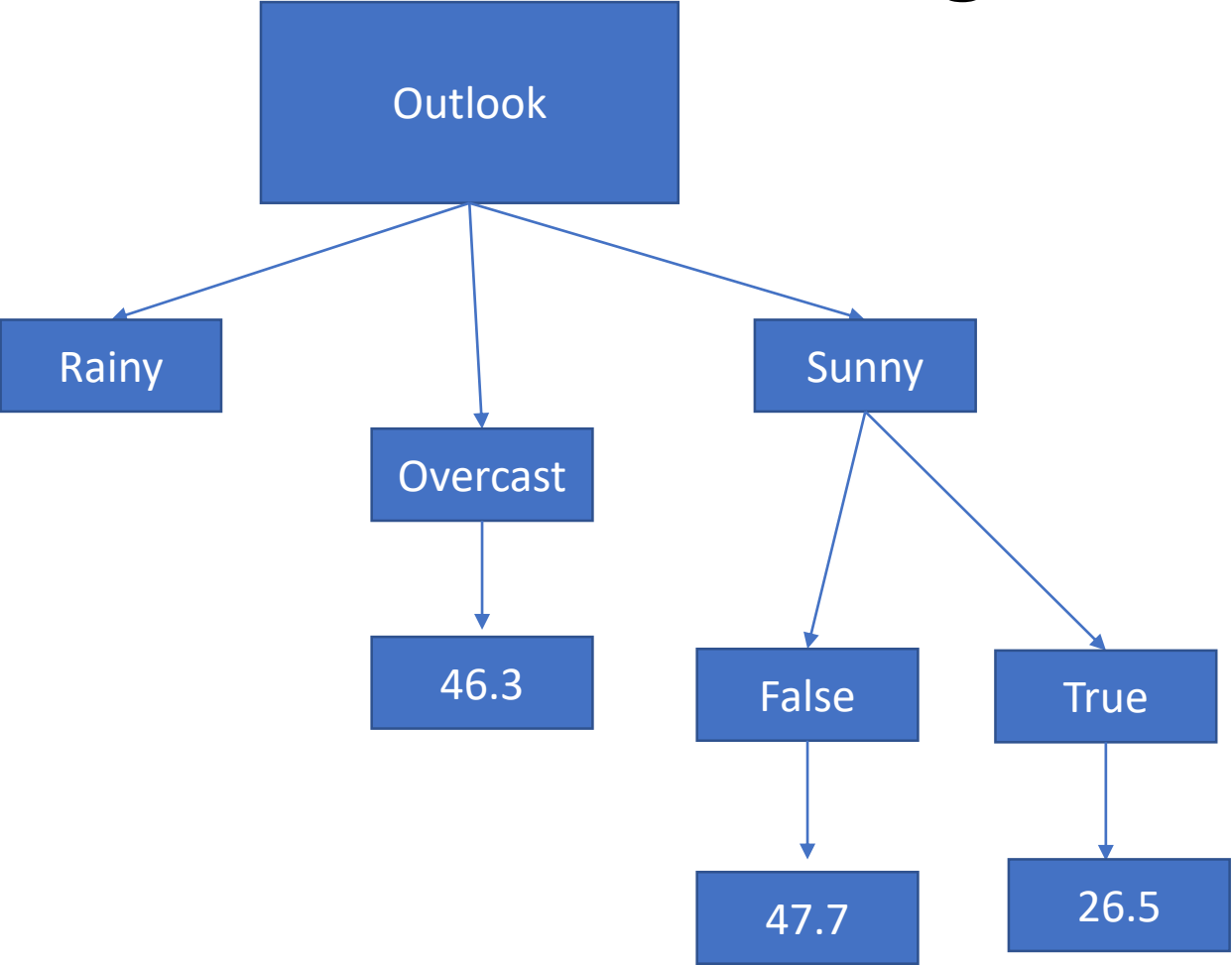
Decision Tree Regressor



- Termination criteria $CV \leq 10\%$, $n \leq 3$

		Hours played StdDev	Hours played Avg	Hours played CV (%)	Count n
Outlook	Overcast	3.49	46.3	8	4
	Rainy	7.78	35.2	22	5
	Sunny	10.87	39.2	28	5

Decision Tree Regressor



- Sunny

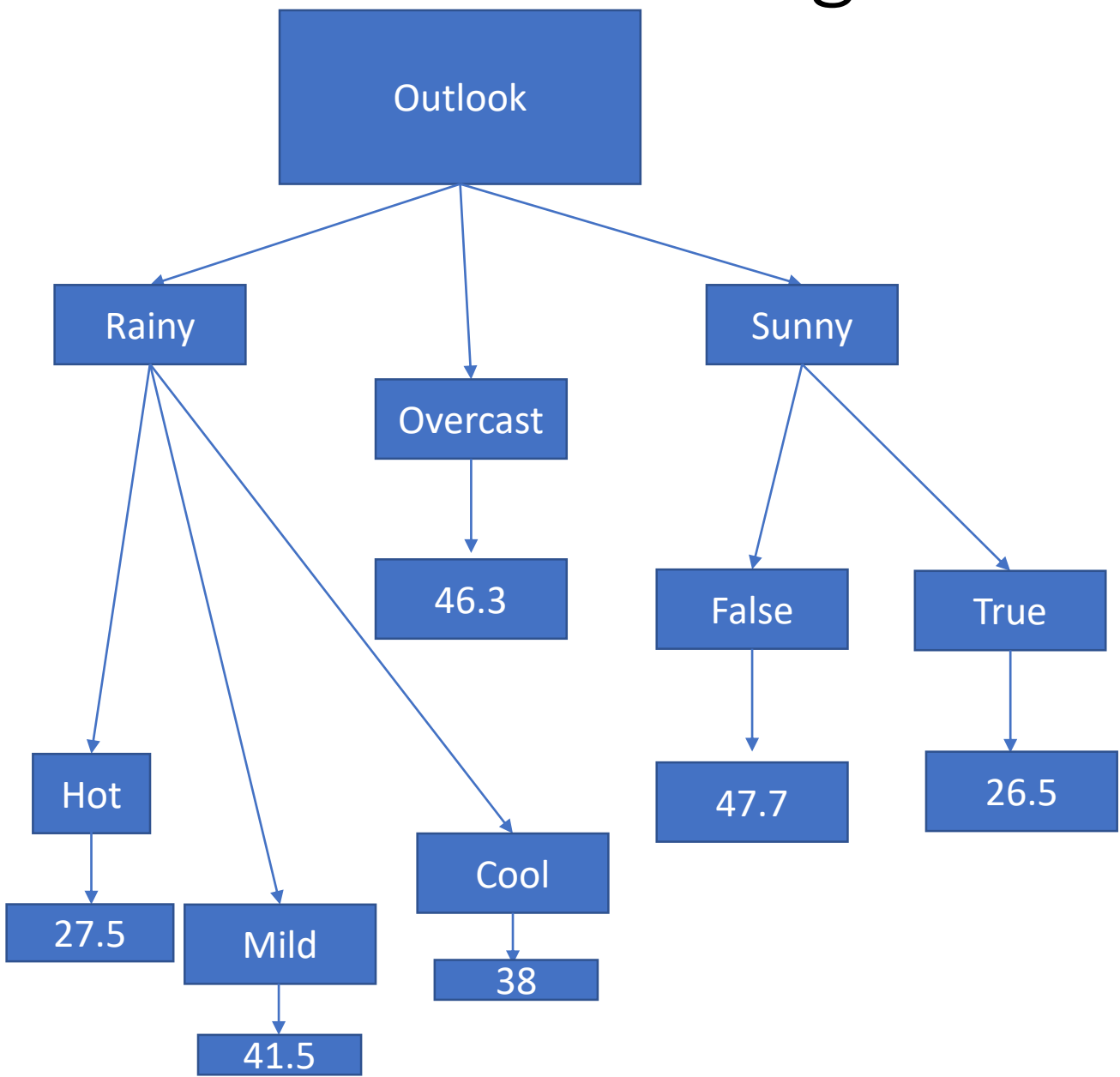
		Hours played StdDev	Count
Temp	Cool	14.5	2
	Mild	7.32	3
SDR = 0.678			

		Hours played StdDev	Count
humidity	High	7.5	2
	Normal	12.5	3
SDR = 0.370			

		Hours played StdDev	Count
Windy	False	3.09	2
	True	3.5	3
SDR = 7.62			

- Termination criteria $CV \leq 10\%$, $n \leq 3$

Decision Tree Regressor



- Rainy

		Hours played StdDev	Count
Temp	Cool	0	1
	Hot	2.5	2
	Mild	6.5	2
SDR = 4.18			

		Hours played StdDev	Count
Humidity	High	4.1	3
	Low	5.0	2
SDR = 3.32			

		Hours played StdDev	Count
Windy	False	5.6	3
	True	9.0	2
SDR = 0.82			

- Termination criteria $CV \leq 10\%$, $n \leq 3$