

# Fundamentals of Machine Learning

## Assignment-5

1. Apply the decision tree regression to develop the decision tree for the following dataset:

| Outlook  | Temperature | Humidity | Wind  | Hours Played |
|----------|-------------|----------|-------|--------------|
| Rain     | Hot         | High     | False | 25           |
| Rain     | Hot         | High     | True  | 30           |
| Overcast | Hot         | High     | False | 48           |
| Sunny    | Mild        | High     | False | 45           |
| Sunny    | Cool        | Normal   | False | 52           |
| Sunny    | Cool        | Normal   | True  | 23           |
| Overcast | Cool        | Normal   | True  | 43           |
| Rain     | Mild        | High     | False | 35           |
| Rain     | Cool        | Normal   | False | 38           |
| Sunny    | Mild        | Normal   | False | 48           |
| Rain     | Mild        | Normal   | True  | 48           |
| Overcast | Mild        | High     | True  | 52           |
| Overcast | Hot         | Normal   | True  | 44           |
| Sunny    | Mild        | High     | False | 30           |

$$\text{Average} = \frac{25 + 30 + 48 + 45 + 52 + 23 + 43 + 35 + 38 + 48 + 48 + 52 + 44 + 30}{14}$$

14

$$(\bar{x}) = \frac{561}{14} = \underline{\underline{40.07}}$$

$$\text{Standard deviation} = \sqrt{\frac{(x - \bar{x})^2}{n}}$$

$$= \sqrt{(25 - 40.07)^2 + (30 - 40.07)^2 + (48 - 40.07)^2 + \dots + (30 - 40.07)^2}$$

14

$$SD = \sqrt{227.10 + 101.4 + 62.88 + 24.30 + 142.3 + 291.378.58 + 25.7 + 4.28 + 62.88 + 62.88 + 142.3 + 15.44 + 101.4}$$

$$SD = \sqrt{\frac{1272.74}{14}} = \sqrt{90.91} = 9.53$$

$$S.D = 9.53$$

Outlook  $\rightarrow$  (Sunny, Overcast, Rain)

$$\text{Average for Sunny} = \frac{45 + 52 + 23 + 48 + 30}{5} = 39.6$$

$$(S.D)_{\text{Sunny}} = \sqrt{\frac{(45 - 39.6)^2 + (52 - 39.6)^2 + (23 - 39.6)^2 + (48 - 39.6)^2 + (30 - 39.6)^2}{5}}$$

$$(SD)_{\text{Sunny}} = \sqrt{124.24}$$

$$(SD)_{\text{Sunny}} = 11.14$$

$$\textcircled{1} \text{ Average of Overcast} = \frac{48 + 43 + 52 + 44}{4} = 46.75$$

$$(S.D)_{\text{overcast}} = \sqrt{\frac{(48 - 46.75)^2 + (43 - 46.75)^2 + (52 - 46.75)^2 + (44 - 46.75)^2}{4}}$$

$$(S.D)_{\text{overcast}} = \sqrt{12.68}$$

$$(S.D)_{\text{overcast}} = 3.56$$

$$\textcircled{2} \text{ Average of Rain} = \frac{25 + 30 + 35 + 38 + 48}{5} = 35.2$$

$$(S.D)_{\text{Rain}} = \sqrt{\frac{(25 - 35.2)^2 + (30 - 35.2)^2 + (35 - 35.2)^2 + (38 - 35.2)^2 + (48 - 35.2)^2}{5}}$$

$$(S.D)_{\text{Rain}} = \sqrt{60.56}$$

$$(S.D)_{\text{Rain}} = 7.78$$

$$\therefore \text{Weighted Standard deviation for outlook} = \frac{5}{14} \times 11.14 + \frac{4}{14} \times 3.56 + \frac{5}{14} \times 7.78 = 7.76$$

$$\text{Standard deviation Reduction for outlook} = 9.53 - 7.76 = 1.77$$

② Temperature : (Hot, mild, cool)

$$\rightarrow \text{Average of Hot} = \frac{25 + 30 + 48 + 44}{4} = 36.75$$

$$(S.D)_{\text{Hot}} = \sqrt{\frac{138.06 + 45.56 + 126.56 + 56.56}{4}}$$

$$(S.D)_{\text{Hot}} = 9.53$$

$$\rightarrow \text{Average of mild} = \frac{45 + 35 + 48 + 48 + 52 + 30}{6} = 44.43$$

$$(S.D)_{\text{mild}} = \sqrt{\frac{4 + 64 + 25 + 25 + 81 + 169}{6}}$$

$$(S.D)_{\text{mild}} = 7.83$$

$$\rightarrow \text{Average of cool} = \frac{52 + 23 + 43 + 38}{4} = 39$$

$$(S.D)_{\text{cool}} = \sqrt{\frac{169 + 256 + 8 + 1}{4}} = (S.D)_{\text{cool}} = 10.41$$

∴ weighted standard deviation for outlook

$$= \frac{4}{14}(9.53) + \frac{6}{14}(7.83) + \frac{4}{14}(10.41)$$

$$= 9.05$$

standard deviation reduction for temperature

$$= 0.46$$

③ Humidity = (High, Normal)

$$\rightarrow \text{Average of High} = \frac{25 + 30 + 48 + 45 + 35 + 52 + 30}{7} = 37.85$$

$$(S.D)_{\text{High}} = \sqrt{\frac{165.1 + 61.62 + 103.02 + 51.12 + 8.12 + 200.22 + 61.62}{7}}$$

$$(S.D)_{\text{High}} = 9.64$$

$$\rightarrow \text{Average of Normal} = \frac{52 + 23 + 43 + 38 + 48 + 48 + 44}{7} = 42.29$$

$$(S.D)_{\text{Normal}} = \sqrt{\frac{94.47 + 371.71 + 0.51 + 18.31 + 32.71 + 32.71 + 29.5}{7}}$$

$$(S.D)_{\text{Normal}} = 8.89$$

$$\therefore \text{weighted Standard deviation} = \frac{\pm (9.64)}{14} + \frac{\pm (8.89)}{14}$$

$$\text{for Humidity} = 9.265$$

$$\text{Standard deviation reduction} = \frac{1 + 8 + 228 + 291}{1000} = 9.53 - 9.26$$

for Humidity

$$(A.v) \frac{1}{A_1} + (E.v) \frac{1}{A_1} + (C.P) \frac{1}{A_1} = 0.27$$

④ Wind = (True, False)

$$\text{Average of True} = \frac{30 + 23 + 43 + 48 + 52 + 44}{6} = 40$$

$$(S.D)_{\text{True}} = \sqrt{\frac{100 + 289 + 9 + 64 + 144 + 16}{6}}$$

$$(S.D)_{\text{True}} = 10.18$$

$$\rightarrow \text{Average of } \text{False} = \frac{25 + 30 + 48 + 45 + 52 + 35 + 38 + 48}{8}$$

$$= 40.12$$

$$(S.D)_{\text{False}} = \sqrt{\frac{228.6 + 102.4 + 62.09 + 23.81 + 141.13 + 26.21 + 4.49 + 62.09}{8}}$$

$$(S.D)_{\text{False}} = 9.019$$

$\therefore$  weighted standard deviation

for wind

$$= \frac{6}{14} (10.18) + \frac{8}{14} (9.019)$$

$$= 9.51$$

$$\text{Standard deviation reduction for wind} = 9.53 - 9.51$$

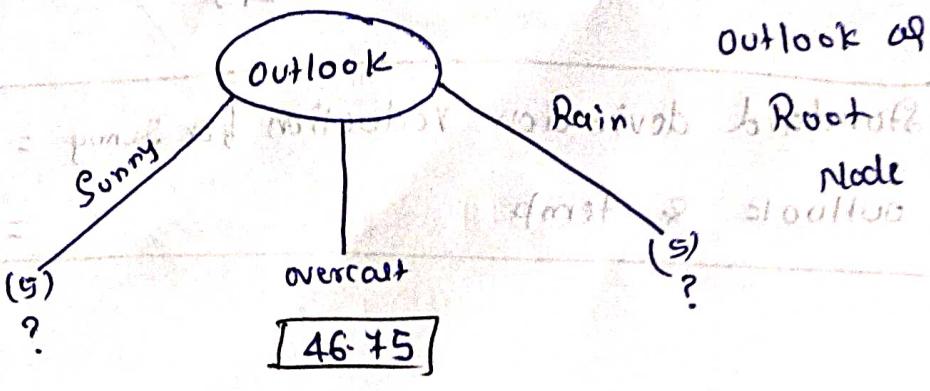
$$= 0.02$$

$$(S.D)_{\text{outlook}} = 1.77 \checkmark$$

$$(S.D)_{\text{temp}} = 0.46$$

$$(S.D)_{\text{Humidity}} = 0.27$$

$$(S.D)_{\text{wind}} = 0.02$$



## Step 2:

| outlook | Temperature | Humidity | Wind  | Hours Played |
|---------|-------------|----------|-------|--------------|
| Sunny   | mild        | High     | False | 45           |
| Sunny   | Cool        | Normal   | False | 52           |
| Sunny   | Cool        | Normal   | True  | 23           |
| Sunny   | mild        | Normal   | False | 48           |
| Sunny   | mild        | High     | False | 30           |

1) Temperature attribute = (mild, cool)

$$\text{Average of mild} = \frac{45 + 48 + 30}{3} = 41$$

$$(S.D)_{\text{mild}} = \sqrt{\frac{16 + 49 + 121}{3}} = (S.D)_{\text{mild}} = 7.87$$

$$\rightarrow \text{Average of (Sunny, cool)} = \frac{52 + 23}{2} = 37.5$$

$$(S.D)_{(\text{Sunny, cool})} = \sqrt{\frac{210.25 + 210.25}{2}} = S.D_{(\text{Sunny, cool})} = 14.5$$

$$\therefore \text{Weighted S.D of temp} = \frac{3}{45} (7.87) + \frac{2}{5} (14.5) = 10.52$$

$$\begin{aligned} \text{Standard deviation reduction for sunny} &= 11.14 - 10.52 \\ \text{outlook \& temp} &= 0.62 \end{aligned}$$

2) Humidity attribute = (High, Normal)

$$\rightarrow \text{Average}(\text{Sunny}, \text{High}) = \frac{45 + 30}{2} = 37.5$$

$$S.D(\text{Sunny}, \text{High}) = \sqrt{\frac{56.25 + 56.25}{2}}$$

$$S.D(\text{Sunny}, \text{High}) = 7.5$$

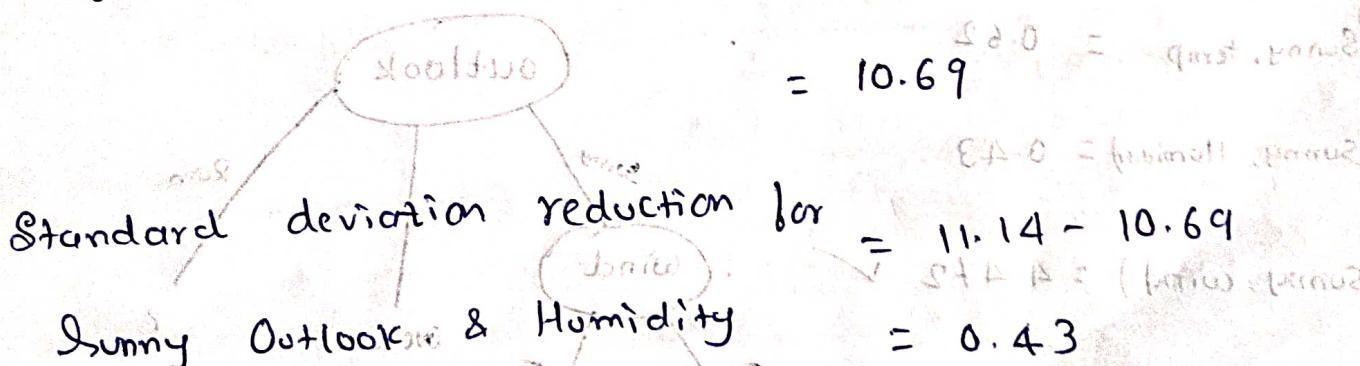
$$\rightarrow \text{Average}(\text{Sunny}, \text{Normal}) = \frac{52 + 23 + 48}{3} = 41$$

$$S.D(\text{Sunny}, \text{Normal}) = \sqrt{\frac{121 + 324 + 44}{3}}$$

$$S.D(\text{Sunny}, \text{Normal}) = 12.83$$

$$\therefore \text{weighted standard deviation} = \frac{2}{5}(7.5) + \frac{3}{5}(12.83)$$

$$= 10.69$$



3) wind attribute = (False, True)

$$\rightarrow \text{Average}(\text{Sunny}, \text{False}) = \frac{45 + 52 + 48 + 30}{4} = 43.75$$

$$S.D = \sqrt{\frac{1.56 + 68.06 + 18.06 + 189.06}{4}}$$

$$S.D = \boxed{8.31}$$

$$\rightarrow \text{Average} = \frac{23}{4}$$

$$S.D = \boxed{0.18}$$

$$\therefore \text{weighted standard deviation} = \frac{4}{5} (8.31) + 0 = 6.648$$

- Standard deviation reduction for

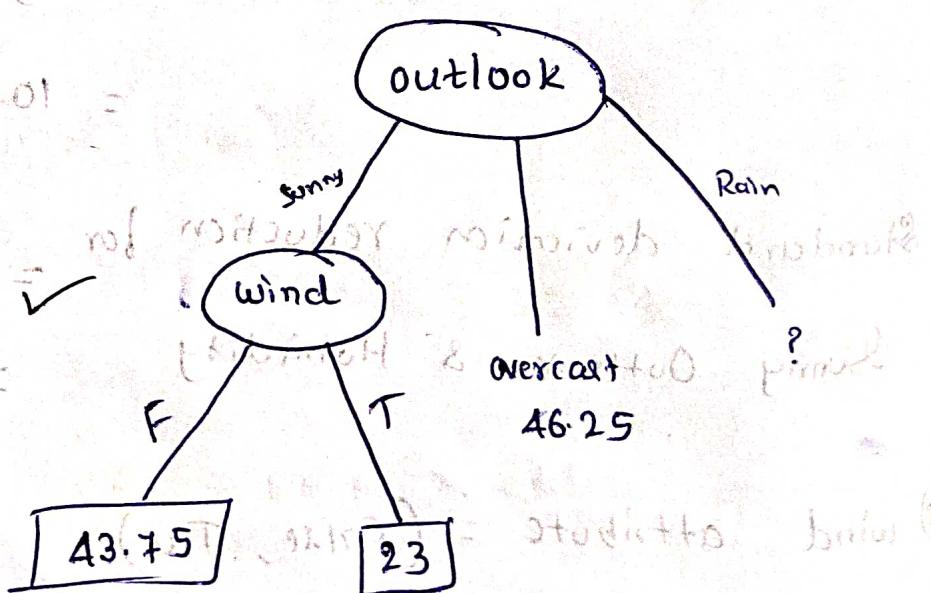
Sunny outlook & wind

$$= 4.472$$

$$(S.D)_{\text{Sunny, temp}} = 0.62$$

$$S.D_{(\text{Sunny, Humidity})} = 0.43$$

$$S.D_{(\text{Sunny, wind})} = 4.472$$



Step 3:-

| outlook | Temperature | Humidity | wind  | Hours Played |
|---------|-------------|----------|-------|--------------|
| Rain    | Hot         | High     | False | 25           |
| Rain    | Hot         | High     | True  | 30           |
| Rain    | mild        | High     | False | 35           |
| Rain    | cool        | Normal   | False | 39           |
| Rain    | mild        | Normal   | True  | 48           |

① Temperature attribute = (Hot, mild, cool)

$$\rightarrow \text{Average}_{(\text{Rain, Hot})} = \frac{25 + 30}{2} = 27.5$$

$$S.D_{(\text{Rain, Hot})} = \sqrt{\frac{6.25 + 6.25}{2}} = 2.5$$

$$\rightarrow \text{Average}_{(\text{Rain, mild})} = \frac{35 + 48}{2} = 41.5$$

$$S.D_{(\text{Rain, mild})} = \sqrt{\frac{42.25 + 42.25}{2}} = 6.5$$

$$\rightarrow \text{Average}_{(\text{Rain, cool})} = 38$$

$$S.D_{(\text{Rain, cool})} = 0$$

$$\therefore \text{weighted standard deviation} = \frac{2}{5}(2.5) + \frac{2}{5}(6.5) + 0$$

$$(2 \cdot 2.5 + 2 \cdot 6.5 + 0) = 3.6$$

$$S.D \text{ for Rain outlook & temp} = 7.48 - 3.6$$

$$= 4.18$$

② Humidity attribute : (High, Normal)

$$\rightarrow \text{Average (Rain, High)} = \frac{25+30+35}{3} = 30$$

$$S.D (Rain, High) = \sqrt{\frac{25+0+25}{3}} = 4.08$$

$$\rightarrow \text{Average (Rain, Normal)} = \frac{23.8+48}{2} = 43$$

$$S.D (Rain, Normal) = \sqrt{\frac{25+25}{2}} = 5$$

$$S.D (Rain, Normal) = 5$$

$$\text{weight } S.D = \frac{3}{5} (4.08) + \frac{2}{5} (5) \\ = 4.48$$

$$S.D \text{ for Rain outlook \& Humidity} = \frac{7.78 - 4.48}{2} = 3.32$$

③ wind attribute {False, True}

$$\rightarrow \text{Average (Rain, False)} = \frac{25+35+38}{3} = 32.66$$

$$S.D (Rain, False) = \sqrt{\frac{58.67 + 5.47 + 28.51}{3}}$$

$$S.D (Rain, False) = 5.55$$

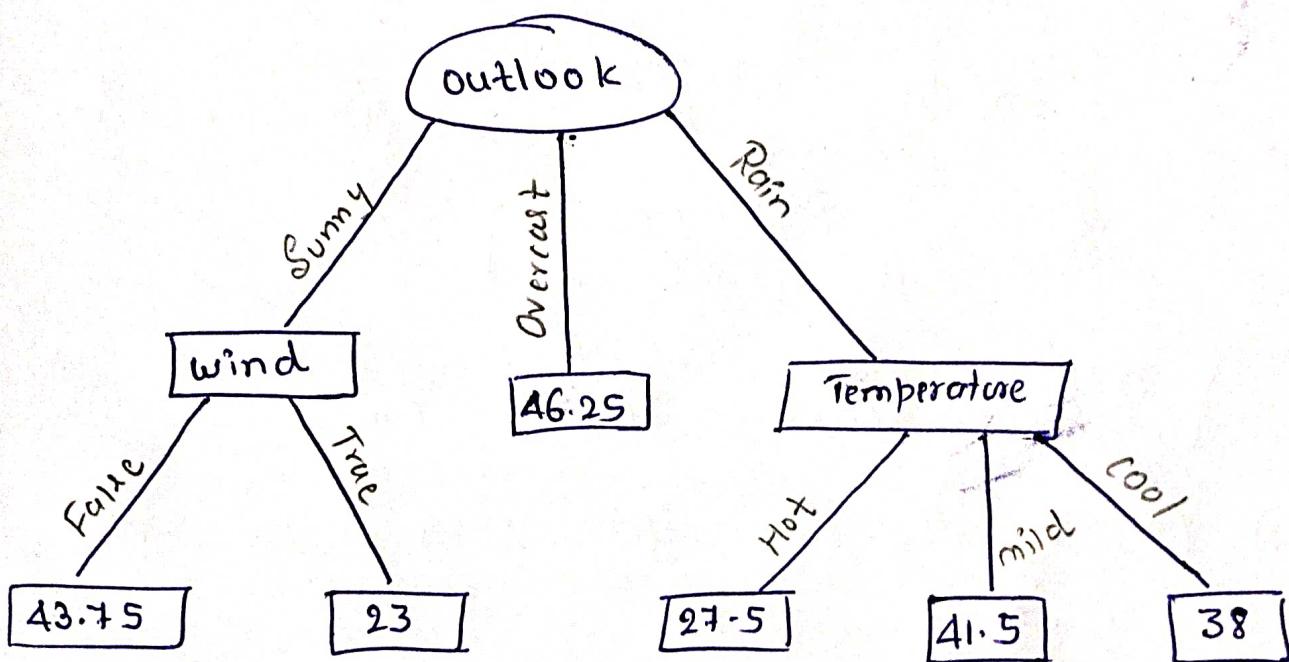
$$\rightarrow \text{Average}_{(\text{Rain}, \text{True})} = \frac{30 + 48}{2} = 39$$

$$S.D_{(\text{Rain}, \text{True})} = \sqrt{\frac{81 + 81}{2}} = 9 //$$

$$\text{Weighted Standard deviation} = \frac{3}{5}(5.55) + \frac{2}{5}(9) \\ = 6.93$$

$$S.D \text{ for Rain outlook \& wind} = \underline{7.48 - 6.93} \\ = \underline{0.85}$$

$$S.D_{(\text{Rain}, \text{temp})} = 4.18 \quad | \quad S.D_{(\text{Rain}, \text{Humidity})} = 3.32 \quad | \quad S.D_{(\text{Rain}, \text{wind})} = 0.85$$



FML

Assignment 5:- Decision tree Regression

Submitted by:- Karthik - C

Reg No :- 240158016