Asignmt - 5

Hardik Probh
24105801g

1) Apply the Decision tree regression to develop the Decision tree:

⇒

$$\text{Average} = \frac{25+50+48+45+52+23+43+35+38+48+52+5+30}{14}$$

$$= 40.07$$

$$SD = \sqrt{\frac{(25-40.07)^2 + (30-40.07)^2 + (48-40.07)^2 + \dots}{14}}$$

$$SD = 9.53$$

1) Outlook

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

$$SD_{(Sunny)} = \sqrt{\frac{(45-39.6)^2 + (52-39.6)^2 + (23-39.6)^2 + (30-39.6)^2}{5}}$$

$$= 11.12$$

$$\text{Avg overst} = \frac{48+43+52+46}{4} = 46.75$$

$$SD(\text{overst}) = \sqrt{\frac{(48-46.75)^2 + (43-46.75)^2 + (52-46.75)^2 + (46-46.75)^2}{4}}$$

$$= 3.56$$

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

$$SD(\text{Rain}) = \sqrt{\frac{104.04 + 27.04 + 0.04 + 7.84 + 163.2}{5}}$$

$$= 7.78$$

∴ Weighted SD for outlook $= \frac{4}{14} \times 3.56 + \frac{5}{14} \times 11.12 + \frac{5}{14} \times 7.78$

$$= 7.76$$

SD Reduction for outlook $= 9.53 - 7.76 = 1.77$

2) Temperature Attribute.

$$\text{Average}_{(\text{Hot})} = \frac{25+30+46+44}{4} = 36.75$$

$$SD(\text{Hot}) = \sqrt{\frac{138.06+68.06+126.56+52.56}{4}}$$

$$= 9.52$$

$$\text{Average}_{(\text{mild})} = 43$$

$$SD_{(\text{mild})} = 7.83$$

$$\text{Average}_{(\text{cool})} = 39$$

$$SD_{(\text{cool})} = \sqrt{100.33}$$

$$= 10.57$$

weighted SD for temperature $= \frac{4}{14} \times 9.52 + \frac{6}{14} 7.53 + \frac{6}{14} \times 10.51$

$$= 9.07$$

∴ Standard deviation Reduction for temperature $= 9.53 - 9.07 = 0.46$

3) Wind Attribute

(i) $\text{Average}_{(\text{False})} = 40.12$

$SD = 9.01$

ii) $\text{Average}_{(\text{true})} = 40$

$SD_{(\text{true})} = 10.18$

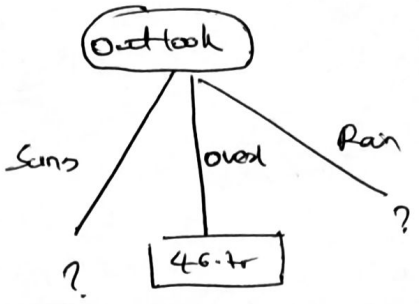
weighted ~~de~~ SD for wind $= \frac{6}{14} \times 10.18 + \frac{8}{14} \times 9.01$

$= 9.51$

$SD_{outlook} = 1.771$

$SD_{temp} = 0.46$

$SD_{wind} = 0.02$



2nd Step. (Sunny)

i) temperature

$Avg_{(sunny, mild)} = \frac{45+43+30}{3}$

$= 41$

$SD_{(Sunny, mild)} = \sqrt{\frac{16+49+121}{3}} = 7.87$

$Avg_{(sunny, cool)} = \frac{52+23}{2} = 37.5$

$SD = \sqrt{\frac{210.25+210.25}{2}} = 14.5$

2) Humidity

$Avg_{(Sunny, Normal)} = 41$

$SD_{(Sunny, Normal)} = \sqrt{\frac{121+324+49}{3}} = 12.83$

$Avg_{(Sunny, high)} = \frac{45+30}{2} = 37.5$

$SD_{(Sunny, high)} = \sqrt{\frac{56.25+56.25}{2}} = 7.5$

Weighted SD $= \frac{3}{5} \times 12.83 + \frac{2}{5} \times 14.5 = 10.52$

SD reduction for Sunny outlook & temp

$= 11.12 - 10.52$

$= 0.6$

∴ SD reduction for Sunny, outlook & humidity

$= 11.12 - 10.69 = 0.43$

(2) Wind Attribute

$Avg_{(Sunny, F)} = \frac{45+52+48+30}{4}$

$= 43.75$

$SD_{(Sunny, F)} = 8.31$

$Avg_{(Sunny, true)} = 23$

$SD_{(Sunny, true)} = 0$

Weighted SD $= \frac{4}{5} \times 9.31 + 0$

$= 6.648$

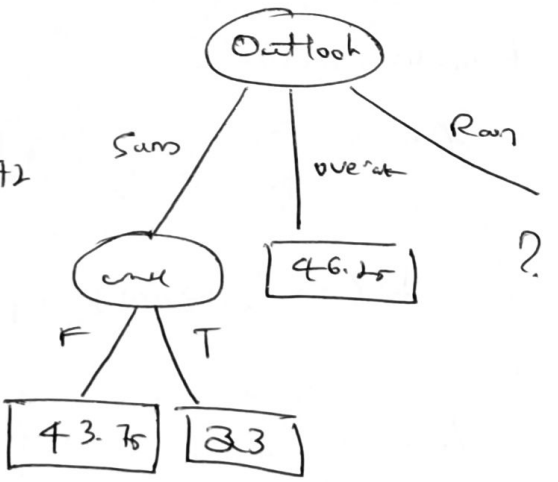
∴ SD reduction $= 11.12 - 6.64$

$= 4.472$

$SD_{(Sunny, temp)} = 0.6$

$SD_{(Sunny, Humidity)} = 0.6$

$SD_{(Sunny, wind)} = 4.472$



Step 3.

i) Temperature

$Avg_{(Rainy), (Hot)} = \frac{25+30}{2} = 27.5$

$SD_{(Rainy, hot)} = 2.5$

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

$$SD_{(\text{Rain, mild})} = \sqrt{\frac{42.25+42.25}{2}}$$

$$= 6.5$$

$$\text{Avg}_{(\text{Rain, cool}=0)} \quad SD_{(\text{Rain, cool})=0}$$

$$\therefore \text{weighted } SD = \frac{2}{5} \times 2.5 + \frac{2}{5} \times 2.5 + 0$$

$$= 3.6$$

Standard Deviation Reduction $= 7.78 - 3.6$

$$= 4.18$$

(2) Humidity Attributed

(High, Normal)

$$\text{Avg}_{(\text{Rain, high})} = 30$$

$$\text{Avg } SD_{(\text{Rain, high})} = 4.08$$

$$\text{Avg}_{(\text{Rain Normal})} = 43$$

$$SD_{(\text{Rain Normal})} = 5$$

$$\therefore \text{weighted } SD = \frac{3}{5} \times 4.08 + \frac{2}{5} \times 5 = 4.65$$

$$\therefore SD \text{ Reduction} = 7.78 - 4.48$$

$$= 3.32$$

(2) Wind Attribute

$$\text{Avg}_{(\text{Rain, False})} = 36.5$$

$$SD_{(\text{Rain False})} = 9.25$$

$$\text{Avg}_{(\text{Rain, true})} = 34.33$$

$$SD_{(\text{Rain, true})} = 9.87$$

$\therefore$ weighted SD $= \frac{2}{5} \times 1.05 + \frac{3}{5} \times 9.87$

$= 6.522$

$\therefore$ SD Reduction $= 7.78 - 6.522 = 1.258$

SD (Rain, temp) = (4.18)

SD (Rain, humidity) = 3.32

SD (Rain, wind) = 1.258

Find Decision tree.