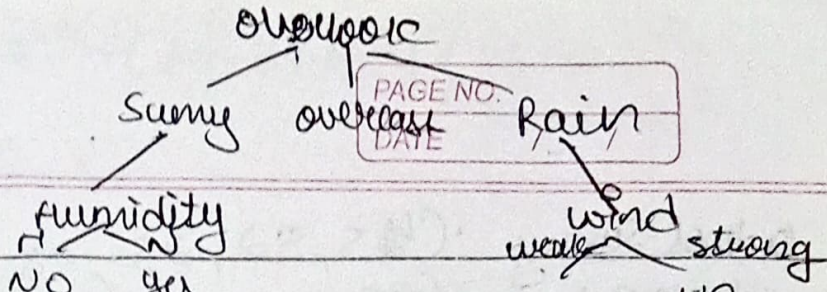
sunny → humidity →
overcast → yes
Rain → wind →



Assignment-5

Apply: The division tree regression to develop the division tree for the following dataset.

outlook	Temperature	humidity	wind	hourse 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

39.78

Variance of the Dataset:

$$\text{Variance} = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2 = 40.14.$$

$$\text{variance} = (25-40.14)^2 + (30-40.14)^2 + (48-40.14)^2 + (45-40.14)^2 + (52-40.14)^2 + (23-40.14)^2 + (43-40.14)^2 + (35-40.14)^2 + (38-40.14)^2 + (48-40.14)^2 + (48-40.14)^2 + (52-40.14)^2 + (44-40.14)^2 + (30-40.14)^2$$

$$14. = 91$$

outlook Sunny [45, 52, 23, 48, 30] = 39.6 ≈ 124.24

outlook Rain [25, 30, 35, 38, 48] = 35.2 ≈ 75.7

outlook overcast [46, 43, 52, 44] = 46.25 ≈ 135.68

$\frac{5}{14} \times 124.24 + \frac{5}{14} \times 75.7 + \frac{4}{14} \times 135.68 = 109.64$

Average of golf player: all no of all.

$\frac{25+30+46+45+52+23+43+35+38+46+48+52+44+30}{14} = 39.78$

Variance / standard deviation of golf.

$\sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}$

$\sqrt{\frac{(25-39.78)^2 + (30-39.78)^2 + (46-39.78)^2 + (45-39.78)^2 + (52-39.78)^2 + (23-39.78)^2 + (43-39.78)^2 + (35-39.78)^2 + (38-39.78)^2 + (46-39.78)^2 + (48-39.78)^2 + (58-39.78)^2 + (44-39.78)^2 + (30-39.78)^2}{14}} = \sqrt{99.284} = 9.32$

844.7472

outlook sunny (25, 30, 35, 38, 48) = 35.2

outlook Rain (45, 52, 23, 46, 30) = 39.2

outlook overcast (46, 43, 52, 44) = 46.25.

Sunny SD $= \sqrt{\frac{(25-35.2)^2 + (30-35.2)^2 + (35-35.2)^2 + (38-35.2)^2 + (48-35.2)^2}{5}} = 7.78$

SD Rain $\sqrt{\frac{(45-39.2)^2 + (52-39.2)^2 + (23-39.2)^2 + (46-39.2)^2 + (30-39.2)^2}{5}} = 12.87.$

overcast.

SD $\sqrt{\frac{(46-46.25)^2 + (43-46.25)^2 + (52-46.25)^2 + (44-46.25)^2}{4}} = 3.491$

for outlook (+0 other features.

$\frac{5}{14} \times 7.78 + \frac{5}{14} \times 10.87 + \frac{4}{14} \times 3.491 = 7.65.$

total SD − outlook SD = 9.32 − 7.66 = 1.66

Temp Hot (25 30 46 44) = 36.25

Temp mild (45 35 46 48 52 30) = 42.66

Temp cool (52, 23 43, 38) = 39.

Hot SD 8.93 $\sqrt{\frac{(25-36.25)^2 + (30-36.25)^2 + (46-36.25)^2 + (44-36.25)^2}{4}}$ 319.6275/4 79.906875

Mild SD 7.65 $\sqrt{\frac{(45-42.66)^2 + (35-42.66)^2 + (46-42.66)^2 + (48-42.66)^2 + (52-42.66)^2 + (30-42.66)^2}{6}}$ 351.3336 = 58.5556

cool 10.51 $\sqrt{\frac{(52-39)^2 + (23-39)^2 + (43-39)^2 + (38-39)^2}{4}}$ 442/4 = 110.5

for Temp.

$4/14 \times 8.95 + 6/14 \times 7.65 + 4/14 \times 10.51 = 8.81$

Total SD − Temp SD = 9.32 − 8.84 = 0.47.

Humidity high [25, 30, 46, 45, 35, 52, 30] = 37.85

Humidity normal [52, 23, 43, 38, 48, 46, 44] = 42.28

high SD $\sqrt{\frac{(25-37.85)^2 + (30-37.85)^2 + (48-37.85)^2 + (45-37.85)^2 + (35-37.85)^2 + (52-37.85)^2 + (30-37.85)^2}{7}} = 650.8575 = 9.64$

Normal SD $\sqrt{\frac{(52-37.85)^2 + (23-37.85)^2 + (43-37.85)^2 + (38-37.85)^2 + (48-37.85)^2 + (48-37.85)^2 + (44-37.85)^2}{7}}$ 691.1575 = 98.7 = 9.93

9.32 = 0.27

wind strong (30, 23, 45, 46, 52, 30) = 37.66
, wind weak (25, 46, 45, 52, 35, 38, 46, 44) = 41.37

S SD $\sqrt{(30-37.66)^2+(23-37.66)^2+(45-37.66)^2+(46-37.66)^2+(52-37.66)^2+(30-37.66)^2/6}$ = 673.3336

W SD $\sqrt{(25-41.37)^2+(46-41.37)^2+(45-41.37)^2+(52-41.37)^2+(35-41.37)^2+(38-41.37)^2+(46-41.37)^2+(44-41.37)^2/8}$

= 495.8750

= 7.87

112.2222667 = 10.59

weight SD = $\left(\frac{6}{14}\right) \times 10.59 + \left(\frac{8}{14}\right) \times 7.87 = 9.035$

SD reduction wind = 9.32 − 9.03 = 0.29.

SD reduction

outlook	1.66 ✓
Temp.	0.47
Humidity	0.27
wind	0.29

Root node → outlook

Sunny	Overcast	Rain
	3.491	10.87

Sunny = 7.78.

If there are less than 5 data points then it may lead to overfitting so it is better to take the whole age.

outlook	Temp	Humidity	wind	golf players
sunny	hot	high	weak	25
sunny	hot	high	strong	30
sunny	mild	high	weak	35
sunny	cool	normal	weak	38
sunny	mild	normal	strong	48

PAGE NO
DATE

SD Temp hot = (25, 30) = 27.5

mild = (35, 48) = 41.5

cool. = 0 because only one.

hot $\sqrt{(25-27.5)^2+(30-27.5)^2/2}$. mild = $\sqrt{(35-41.5)^2+(48-41.5)^2}$

12.5/2 = 6.252.5 84.5/2 = 42.25

weighted standard deviation $\left(\frac{3.6}{10}\right)$ ⇒ $\frac{2}{5}\times 2.5 + \frac{2}{5}\times 6.5 + \frac{1}{5}\times 0$

standard deviation reduction = 7.78 − 3.6 = 4.18.

Sunny High = (25+30+35) = 30 $\sqrt{(25-30)^2+(30-30)^2+(35-30)^2/3}$.

Normal = (38+48) = 43 $\sqrt{(38-43)^2+(48-45)^2/2}$.

SD 25 SD 25 S 5 $\frac{3}{5}\times 4.08 + \frac{2}{5}\times 5 = 5$

16.66 = 4.08 = 4.45 weight SD.

7.78 − 4.45 = 3.33 SD reduct

S wind weak (25, 35, 38) = 32.66 = $\sqrt{(25-32.66)^2+(35-32.66)^2+(38-32.66)^2}$

S wind strong (30, 48) = 39 ⇒ $\sqrt{(30-39)^2+(48-39)^2/2}$ = 9

5.55. $\frac{3}{5}\times 5.55 + \frac{2}{5}\times 9 = 6.93$

7.78 − 6.93 = 0.85.

Sunny ⇒ Temp Hum Wind

4.18 3.33 0.85

Outlook	Temp	Humidity	wind	play
Rain	mild	High	weak	45
Rain	cool	Normal	weak	52
Rain	cool	Normal	strong	23
Rain	mild	Normal	weak	46
Rain	mild	High	strong	30

Rain Temp mild $(45+46+30) = 40.33$ $\sqrt{(45-40.33)^2+(46-40.33)^2+(30-40.33)^2/3}$

Rain Temp cool $(52+23) = 37.5$ $\sqrt{(52-37.5)^2+(23-37.5)^2/2}$

$\sqrt{160.667/3}$ $216.25 = 14.5$, 7.31

$(3/5 \times 7.31 + 2/5 \times 14.5) = 10.186$

$10.87 - 10.186 = 0.67$

Rain Hum High $(45+30) = 37.5$ $\sqrt{(45-37.5)^2+(30-37.5)^2/2} = 112.5/3$

Normal $(52+23+46) = 40.33$ $\sqrt{(52-40.33)^2+(23-40.33)^2+(46-40.33)^2/3}$

$468.6667 = 156.2 = 12.5$ 7.5

$2/5 \times 7.5 + 3/5 \times 12.5 = 10.5$ $10.87 - 10.5 = 0.37$

Rain wind weak $(45, 52, 46) = 47.6$ $\sqrt{(45-47.6)^2+(52-47.6)^2+(46-47.6)^2/3}$

strong $(23, 30) = 26.5 = \sqrt{(23-26.5)^2+(30-26.5)^2/2}$

$28.68/3$ $24.5/2$ $12.25 = 3.5 + 3.09$

$(3/5)(3.09) + (2/5)(3.5) = 3.254$

$10.87 - 3.25 = 7.62$

Rain so Reduct 0.67 0.37, 7.62 ✓

Temp Humidity wind

now further Rain to wind we have 3 weak 2 strong

now further sunny to Temp we have 2 hot 2 mild 1 cool.

make the averages of these all values in the sub branches