

①

### 03 Regression Trees

→ Standard Deviation

$$\text{Golf players} = \{25, 30, 46, 45, 52, 23, 43, 35, 38, 46, 48, 52, 44, 30\}$$

$$\begin{aligned}\text{Average of golf players} &= \frac{(25+30+46+45+52+23+43+35+38+46+48+52+44+30)}{14} \\ &= 39.78\end{aligned}$$

Standard deviation of golf players =

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

$$= \sqrt{\frac{(218.45 + 95.65 + 33.68 + 27.25 + 149.33 + 281.57 + 10.3684 + 22.8484 + 3.14 + 38.68 + 67.57 + 149.33 + 17.81 + 95.65)}{14}}^{1/2}$$

$$\begin{aligned}&= \sqrt{\frac{1211.3568}{14}} = \sqrt{86.52} \\ &\approx 9.30\end{aligned}$$

(2)

## Finding Decision Features for

- 1) Outlook
- 2) Temperature
- 3) Humidity
- 4) Wind

### Outlook

- 1) Sunny

Golf players for sunny outlook = { 25 + 30 +

$$35 + 38 + 48 \}$$

Average of golf players =  $\frac{(25+30+35+38+48)}{5}$

$$= 35.2$$

Standard deviation of golf players for sunny outlook.

$$= \sqrt{\left[ (25-35.2)^2 + (30-35.2)^2 + (35-35.2)^2 + (38-35.2)^2 + (48-35.2)^2 \right] / 5}^{1/2}$$

$$= \sqrt{\left[ 104.04 + 27.04 + 0.04 + 7.84 + 163.84 \right] / 5}^{1/2}$$

$$= 7.78$$

### 2) Overcast

Golf players for overcast outlook = {

$$46 + 43 + 52 + 44 \}$$

(3)

Average of golf players for overcast outlook =

$$(46 + 43 + 52 + 44) / 4 \\ = 46.25$$

Standard Deviation of golf players  
for overcast outlook

$$= \sqrt{\left[ \frac{((46 - 46.25)^2 + (43 - 46.25)^2 + (52 - 46.25)^2 + (44 - 46.25)^2)}{4} \right] / 4}$$

$$= \sqrt{(0.0625 + 10.5625 + 33.0625 + 5.0625) / 4}$$

$$= \sqrt{12.18}$$

$$= 3.49$$

3) Rainy.

Golf players for rainy outlook

$$= \{45, 52, 23, 46, 30\}$$

Average of golf players for rainy outlook

$$= \frac{(45 + 52 + 23 + 46 + 30)}{5}$$

$$= 39.2$$

Standard Deviation of golf players for  
rainy outlook =

(4)

$$= \left( \left[ (45 - 39.2)^2 + (52 - 39.2)^2 + (23 - 39.2)^2 + (46 - 39.2)^2 + (30 - 39.2)^2 \right] / 5 \right)^{1/2}$$

$$= \left[ (33.64 + 163.84 + 262.44 + 46.24 + 84.64) / 5 \right]^{1/2}$$

$$\approx \sqrt{118.16} = 10.87$$

### Summarized Standard Deviation

Outlook	StdDev of golf players	Instances
Overscast	3.49	4
Rain	10.87	5
Sunny	7.78	5

Weight standard deviation for outlook

$$= \left( \frac{4}{14} \times 3.49 \right) + \left( \frac{5}{14} \times 10.87 \right) + \left( \frac{5}{14} \times 7.78 \right)$$

$$= 7.66$$

reduction

$$\text{Standard Deviation for outlook} = 9.30 - 7.66 = 1.66$$

### Temperature

1) Hot

Golf players for hot temperature =

$$\{25, 30, 46, 44\}$$

(5)

1) Average = ?

2) Standard Deviation = ?

Standard Deviation =  $8.95$ 

2) Cool

Golf players for cool temperature =  $\{52, 23, 43, 38\}$ 

→ Average

→ Std Deviation =  $10.51$ 

3) Mild

Golf players for mild Temp =  $\{45, 35, 48, 52, 30\}$ 

→ Average

→ Std Deviation =  $7.65$ Summarized Std Deviation  
Temperature Feature

Temp	Std Dev	Instances
Hot	$8.95$	4
Mild	$10.51$	6
Cool	$7.65$	4

Weighted standard deviation

$$= \left( \frac{4}{14} \times 8.95 \right) + \left( \frac{6}{14} \times 10.51 \right) + \left( \frac{4}{14} \times 7.65 \right)$$

$$= 8.84$$

(6)

$$\text{Stand Deviation Reduction} = 9.32 - 8.84 \\ = 0.47$$

### Humidity

- 1) High  $\rightarrow$  Std Dev = ?
- 2) Normal  $\rightarrow$  Std Dev = ?
- 3) Weighted Standard Dev for Humidity
- 4) Std Dev Reduction = ~~Old Std Dev~~ - Weighted Std Dev

### Wind

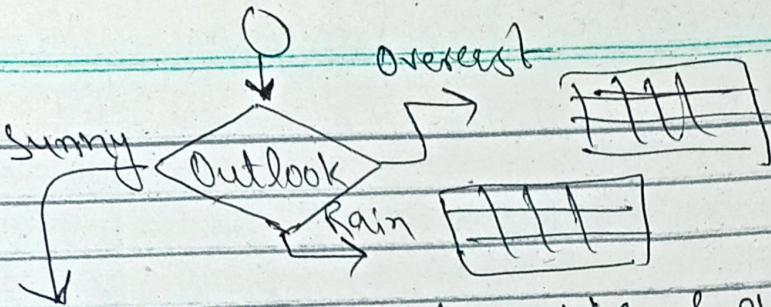
- 1) Weak  $\rightarrow$  Std Dev = ?
- 2) Strong  $\rightarrow$  Std Dev = ?
- 3) Weighted Std Dev for Wind
- 4) Std Dev Reduction = Std Dev of y - Weighted Std Dev

### Comparison of Std Dev

Feature	Std Dev Reduction
Outlook	1.66
Temperature	0.47
Humidity	0.27
Wind	0.29

The winner is outlook because it has highest reduction value

7



Day	Outlook	Temp	Humidity	Wind	Golf Players
1	Sunny	Hot	High	Weak	25
2	Sunny	Hot	High	Strong	30
8	Sunny	Mild	High	Weak	35
9	Sunny	Cool	Normal	Weak	38
11	Sunny	Mild	Normal	Strong	48

Each Branch as separate Dataset

1) Std Dev of Golf players = ?

### Temp

- A) Hot  $\rightarrow$  Std Dev = ?
- B) Mild  $\rightarrow$  Std Dev = ?
- C) Cool  $\rightarrow$  Std Dev = ?
- D) Weight Std Dev = ?
- E) Std Dev Reduction = ?

### Humidity

- A) High  $\rightarrow$  Std Dev
- B) Normal  $\rightarrow$  Std Dev
- C) Weight Std Dev = ?
- D) Std Dev Reduction = ?

Wind

- A) Weak  $\rightarrow$  Std Dev = ?
- B) Strong  $\rightarrow$  Std Dev = ?
- C) Windy Std Dev = ?
- D) Std Dev Reduction = ?

Comparison of Std Dev

Feature	Std Dev	Reduction
Temp	$x_1$	
Humidity	$x_2$	
Wind	$x_3$	

→ The winner will be the feature who has highest std dev reduction value.

