SNS COLLEGE OF TECHNOLOGY

Coimbatore-35
An Autonomous Institution
Accredited by NBA – AICTE and Accredited by NAAC – UGC with 'A++'
Grade Approved by AICTE, New Delhi & Affiliated to Anna University,
Chennai

DEPARTMENT OF COMPUTER APPLICATIONS

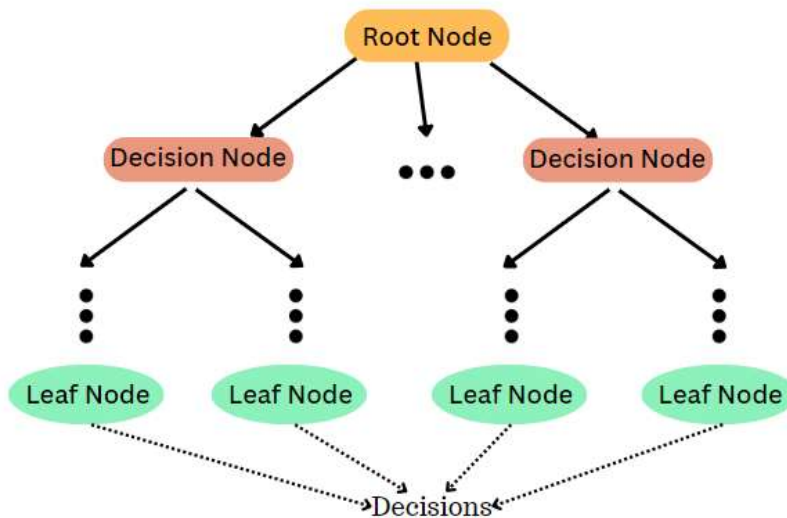## 23CAT702 – MACHINE LEARNING

II YEAR III SEM

UNIT IV – TREE AND RULE MODELS

TOPIC 29 – Trees –regression trees

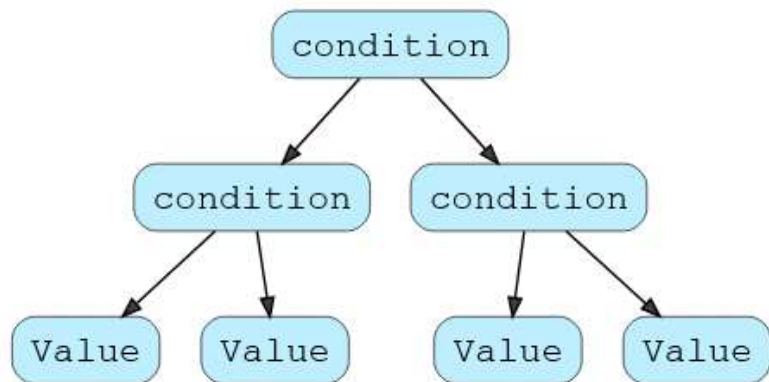**Tree: Non-parametric supervised learning algorithm, which is utilized for both classification and regression tasks**

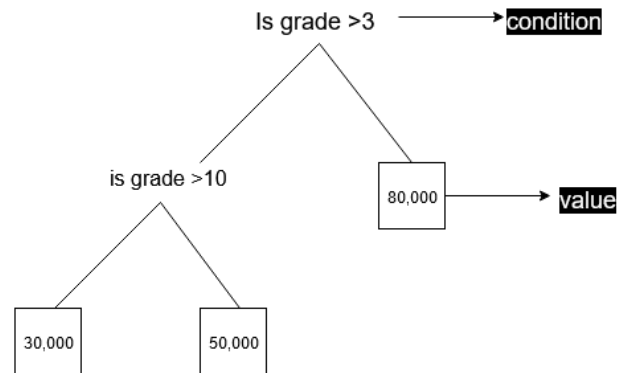Regression Tree/Dr.N.Nandhini/AP/MCA/SNSCT

Regression Tree

Regression trees are decision trees in which the target variables can take continuous values instead of class labels in leaves

How a Regression Tree looks like.

Basic example of Regression Tree

Regression Tree/Dr.N.Nandhini/AP/MCA/SNSCT

Features of regression trees

1. **Root**: Beginning of the decision tree. The first node represents the first condition based on the criteria of the data provided.
2. **Leaf**: Last node in the tree is represented by the value in the decision tree above. Terminal node that does not point to any condition or value.
3. **Decision Node**: Nodes after the root where any decision or condition is further divided into different categories.
4. **Child Node**: The node that is further divided into different categories is called a parent node. The nodes that result from this division are called child nodes.
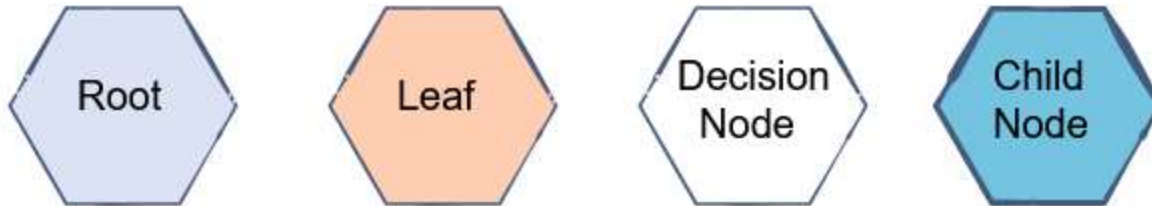
Advantages of regression trees

1. **Visualization of data becomes easier** as users can identify and process each and every step.

2. A specific decision node could be set to have a **priority** against other decision nodes.

3. As the regression tree progresses, **undesired data will be filtered at each step**. As a result, only important data is left to process, which increases the efficiency and accuracy of our design.

4. It is easy to prepare regression trees – they can be used to present data during meetings, presentations, etc.

Example

Day	Outlook	Temp	Humidity	Wind	GolfPlayers
D1	Sunny	Hot	High	Weak	25
D2	Sunny	Hot	High	Strong	30
D3	Overcast	Hot	High	Weak	46
D4	Rain	Mild	High	Weak	45
D5	Rain	Cool	Normal	Weak	52
D6	Rain	Cool	Normal	Strong	23
D7	Overcast	Cool	Normal	Strong	43
D8	Sunny	Mild	High	Weak	35
D9	Sunny	Cool	Normal	Weak	38
D10	Rain	Mild	Normal	Weak	46
D11	Sunny	Mild	Normal	Strong	48
D12	Overcast	Mild	High	Strong	52
D13	Overcast	Hot	Normal	Weak	44
D14	Rain	Mild	High	Strong	30

Standard deviation

$$Average\ of\ golf\ players\ =$$

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

$$Average\ of\ golf\ players\ =\ 39.78$$

$$Standard\ deviation\ of\ golf\ players$$

$$sd = \sqrt{\frac{(25-39.78)^2+(30-39.78)^2+\cdots+(30-39.78)^2}{14}}$$

SD=9.32

Example

Day	Outlook	Temp	Humidity	Wind	Golf Players
D1	Sunny	Hot	High	Weak	25 ✓
D2	Sunny	Hot	High	Strong	30
D8	Sunny	Mild	High	Weak	35
D9	Sunny	Cool	Normal	Weak	38
D11	Sunny	Mild	Normal	Strong	48

- **Outlook**

- Outlook →{sunny, overcast, rain}

- Calculate standard deviation of golf players for all of these outlook candidates.

- **Sunny outlook**

- *Average of golf players for sunny outlook* $= \dfrac{25+30+35+38+48}{5} = 35.2$

- *SD of golf players for sunny outlook* $= \sqrt{\dfrac{(25-35.2)^2 + (30-35.2)^2 + (35-35.2)^2 + (38-35.2)^2 + (48-35.2)^2}{5}}$

- ***SD of golf players for sunny outlook*** $= 7.78$

Regression Tree/Dr.N.Nandhini/AP/MCA/SNSCT

Example

Day	Outlook	Temp	Humidity	Wind	Golf Players
D3	Overcast	Hot	High	Weak	46
D7	Overcast	Cool	Normal	Strong	43
D12	Overcast	Mild	High	Strong	52
D13	Overcast	Hot	Normal	Weak	44

- **Outlook**

- Outlook →{sunny, overcast, rain}

- Calculate standard deviation of golf players for all of these outlook candidates.

- **Sunny Overcast**

- *Average of golf players for Overcast outlook* $= \frac{46+43+52+44}{4} = 46.25$

- *SD of golf players for Overcast outlook* $= \sqrt{\frac{(46-46.25)^2+(43-46.25)^2+(52-46.25)^2+(44-46.25)^2}{4}}$

- ***SD of golf players for sunny Overcast = 3.49***

Regression Tree/Dr.N.Nandhini/AP/MCA/SNSCT

Example

Day	Outlook	Temp	Humidity	Wind	Golf Players
D4	Rain	Mild	High	Weak	45
D5	Rain	Cool	Normal	Weak	52
D6	Rain	Cool	Normal	Strong	23
D10	Rain	Mild	Normal	Weak	46
D14	Rain	Mild	High	Strong	30

- **Outlook**

- Outlook → {sunny, overcast, rain}

- Calculate standard deviation of golf players for all of these outlook candidates.

- **Sunny Rain**

- $Average\ of\ golf\ players\ for\ Rain\ outlook = \dfrac{45+52+23+46+30}{5} = 39.2$

- $SD\ of\ golf\ players\ for\ Rain\ outlook = \sqrt{\dfrac{(45-39.2)^2+(52-39.2)^2+(23-39.2)^2+(46-39.2)^2+(30-39.2)^2}{5}}$

- $SD\ of\ golf\ players\ for\ sunny\ Rain = 10.87$

Example

Outlook	Stdev of Golf Players	Instances
Overcast	3.49	4
Rain	10.87	5
Sunny	7.78	5

- $Weighted\ standard\ deviation\ for\ outlook = \left(\frac{4}{14}\right) * 3.49 + \left(\frac{5}{14}\right) * 10.87 +$

$\left(\frac{5}{14}\right) * 7.78 = 7.66$

- Global standard deviation of golf players 9.32

- $Standard\ deviation\ reduction\ for\ outlook\ = 9.32 - 7.66 = 1.66$

Regression Tree/Dr.N.Nandhini/AP/MCA/SNSCT

Example

Temperature

Temperature can be hot, cool or mild. We will calculate standard deviations for those candidates.

Hot temperature

Day	Outlook	Temp.	Humidity	Wind	Golf Players
1	Sunny	Hot	High	Weak	25
2	Sunny	Hot	High	Strong	30
3	Overcast	Hot	High	Weak	46
13	Overcast	Hot	Normal	Weak	44

Golf players for hot temperature = {25, 30, 46, 44}

Standard deviation of golf players for hot temperature = 8.95

Regression Tree/Dr.N.Nandhini/AP/MCA/SNSCT

Example

Mild temperature

Day	Outlook	Temp.	Humidity	Wind	Golf Players
4	Rain	Mild	High	Weak	45
8	Sunny	Mild	High	Weak	35
10	Rain	Mild	Normal	Weak	46
11	Sunny	Mild	Normal	Strong	48
12	Overcast	Mild	High	Strong	52
14	Rain	Mild	High	Strong	30

Golf players for mild temperature = {45, 35, 46, 48, 52, 30}

Standard deviation of golf players for mild temperature = 7.65

Regression Tree/Dr.N.Nandhini/AP/MCA/SNSCT

Example

Temp	Stdev of Golf Players	Instances
Hot	8.95	4
Mild	10.51	4
Cool	7.65	6

- $Weighted\ standard\ deviation\ for\ Temp = \left(\frac{4}{14}\right) * 8.95 + \left(\frac{4}{14}\right) * 10.51 + \left(\frac{6}{14}\right) *$

$7.65 = 8.84$

- Global standard deviation of golf players 9.32

- $Standard\ deviation\ reduction\ for\ Temp = 9.32 - 8.84 = 0.47$

Regression Tree/Dr.N.Nandhini/AP/MCA/SNSCT

Humidity

Humidity is a binary class. It can either be normal or high.

High humidity

Day	Outlook	Temp.	Humidity	Wind	Golf Players
1	Sunny	Hot	High	Weak	25
2	Sunny	Hot	High	Strong	30
3	Overcast	Hot	High	Weak	46
4	Rain	Mild	High	Weak	45
8	Sunny	Mild	High	Weak	35
12	Overcast	Mild	High	Strong	52
14	Rain	Mild	High	Strong	30

Golf players for high humidity = {25, 30, 46, 45, 35, 52, 30}

Standard deviation for golf players for high humidity = 9.36

Example

Humidity	Stdev of Golf Players	Instances
High	9.36	7
Normal	8.73	7

- *Weighted standard deviation for Humidity* $= \left(\frac{7}{14}\right) * 9.36 + \left(\frac{7}{14}\right) * 8.73 = 9.04$

- Global standard deviation of golf players 9.32

- *Standard deviation reduction for Humidity* $= 9.32 - 9.04 = 0.27$

Regression Tree/Dr.N.Nandhini/AP/MCA/SNSCT

Example

Strong Wind

Day	Outlook	Temp.	Humidity	Wind	Golf Players
2	Sunny	Hot	High	Strong	30
6	Rain	Cool	Normal	Strong	23
7	Overcast	Cool	Normal	Strong	43
11	Sunny	Mild	Normal	Strong	48
12	Overcast	Mild	High	Strong	52
14	Rain	Mild	High	Strong	30

Golf players for strong wind= {30, 23, 43, 48, 52, 30}

Standard deviation for golf players for strong wind = 10.59

Regression Tree/Dr.N.Nandhini/AP/MCA/SNSCT

Weak Wind

1	Sunny	Hot	High	Weak ✓	25
3	Overcast	Hot	High	Weak	46
4	Rain	Mild	High	Weak	45
5	Rain	Cool	Normal	Weak	52
8	Sunny	Mild	High	Weak	35
9	Sunny	Cool	Normal	Weak	38
10	Rain	Mild	Normal	Weak	46
13	Overcast	Hot	Normal	Weak	44

Golf players for weakk wind= (25, 46, 45, 52, 35, 38, 46, 44)

Standard deviation for golf players for weak wind = 7.87

Wind	Stdev of Golf Players	Instances
Strong	10.59 ✓	6
Weak	7.87	8

- $Weighted\ standard\ deviation\ for\ Wind = \left(\frac{6}{14}\right) * 10.59 + \left(\frac{8}{14}\right) * 7.87 = 9.03$

- Global standard deviation of golf players 9.32

- $Standard\ deviation\ reduction\ Wind = 9.32 - 9.03 = 0.29$

Regression Tree/Dr.N.Nandhini/AP/MCA/SNSCT

Example

Feature	Standard Deviation Reduction
Outlook	1.66
Temperature	0.47
Humidity	0.27
Wind	0.29

Regression Tree/Dr.N.Nandhini/AP/MCA/SNSCT
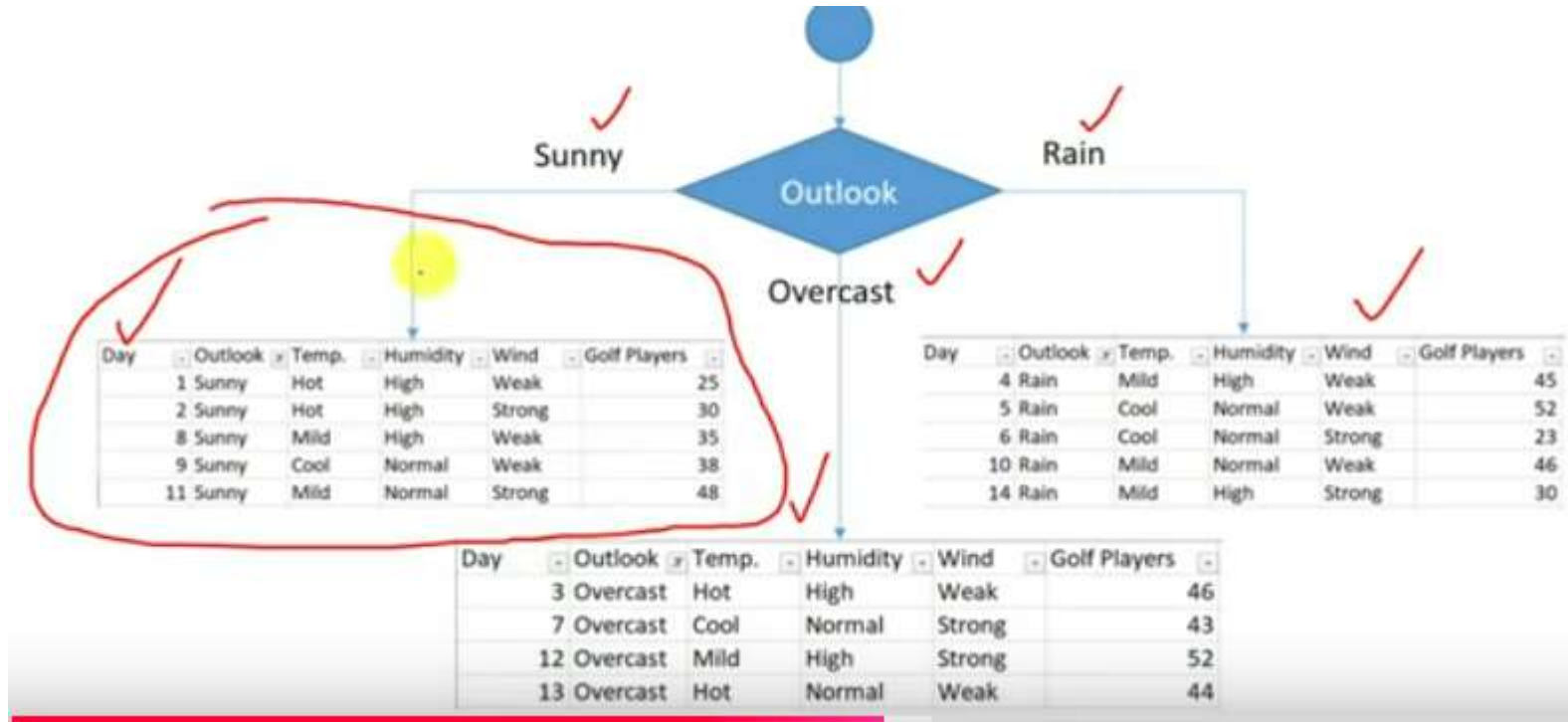
Example

Example

Sunny Outlook

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

Golf players for sunny outlook = {25, 30, 35, 38, 48}

Standard deviation for sunny outlook = 7.78

Regression Tree/Dr.N.Nandhini/AP/MCA/SNSCT

Sunny outlook and Hot Temperature

Day	Outlook	Temp.	Humidity	Wind	Golf Players
1	Sunny	Hot	High	Weak	25
2	Sunny	Hot	High	Strong	30

Standard deviation for sunny outlook and hot temperature = 2.5

Sunny outlook and Cool Temperature

Day	Outlook	Temp.	Humidity	Wind	Golf Players
9	Sunny	Cool	Normal	Weak	38

Standard deviation for sunny outlook and cool temperature = 0

Sunny outlook and Mild Temperature

Day	Outlook	Temp.	Humidity	Wind	Golf Players
8	Sunny	Mild	High	Weak	35
11	Sunny	Mild	Normal	Strong	48

Standard deviation for sunny outlook and mild temperature = 6.5

Regression Tree/Dr.N.Nandhini/AP/MCA/SNSCT

Temperature	Stdev for Golf Players	Instances
Hot	2.5	2
Cool	0	1
Mild	6.5	2

Weighted standard deviation for sunny outlook and temperature = (2/5)x2.5 + (1/5)x0 + (2/5)x6.5 = 3.6

Standard deviation reduction for sunny outlook and temperature = 7.78 - 3.6 = 4.18

Sunny outlook and high humidity

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

Standard deviation for sunny outlook and high humidity = 4.08

Sunny outlook and normal humidity

Day	Outlook	Temp.	Humidity	Wind	Golf Players
9	Sunny	Cool	Normal	Weak	38
11	Sunny	Mild	Normal	Strong	48

Standard deviation for sunny outlook and normal humidity = 5

Regression Tree/Dr.N.Nandhini/AP/MCA/SNSCT

Humidity	Stdev for Golf Players	Instances
High	4.08	3
Normal	5.00	2

Weighted standard deviations for sunny outlook and humidity = (3/5)x4.08 + (2/5)x5 = 4.45

Standard deviation reduction for sunny outlook and humidity = 7.78 − 4.45 = 3.33

Apply the same procedure to wind to calculate the SD…..

Standard deviation for sunny outlook and weak wind = 5.56

Wind	Stdev for Golf Players	Instances
Strong	9	2
Weak	5.56	3

Weighted standard deviations for sunny outlook and wind = (2/5)x9 + (3/5)x5.56 = 6.93

Standard deviation reduction for sunny outlook and wind = 7.78 – 6.93 = 0.85

Feature	Standard Deviation Reduction
Temperature	4.18
Humidity	3.33
Wind	0.85

Cool=38

Mild=(35+48)/2=41.5

Hot=(25+30)/2=27.5
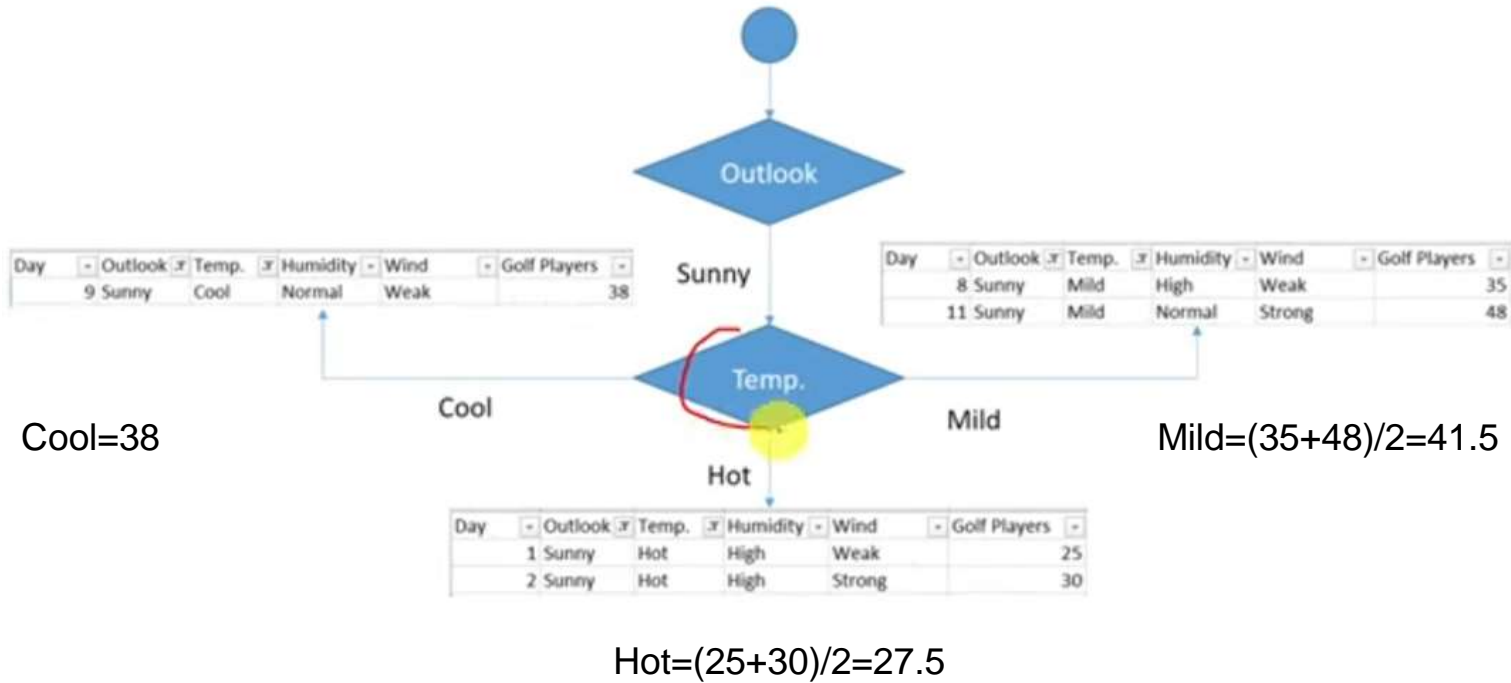
If number of instance less than 5 then prune (ie: find the Average-leaf node)

Sunny

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

Rain

Day	Outlook	Temp.	Humidity	Wind	Golf Players
4	Rain	Mild	High	Weak	45
5	Rain	Cool	Normal	Weak	52
6	Rain	Cool	Normal	Strong	23
10	Rain	Mild	Normal	Weak	46
14	Rain	Mild	High	Strong	30

Overcast

Day	Outlook	Temp.	Humidity	Wind	Golf Players
3	Overcast	Hot	High	Weak	46
7	Overcast	Cool	Normal	Strong	43
12	Overcast	Mild	High	Strong	52
13	Overcast	Hot	Normal	Weak	44

Overcast=(46+43+52+44)/4=46.25

Rainy Outlook

Day	Outlook	Temp.	Humidity	Wind	Golf Players
4	Rain	Mild	High	Weak	45
5	Rain	Cool	Normal	Weak	52
6	Rain	Cool	Normal	Strong	23
10	Rain	Mild	Normal	Weak	46
14	Rain	Mild	High	Strong	30

We need to find standard deviation reduction values for the rest of the features in same way for the sub data set above.

Standard deviation for rainy outlook = 10.87

Sunny outlook and high humidity

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

Standard deviation for sunny outlook and high humidity = 4.08

Sunny outlook and normal humidity

Day	Outlook	Temp.	Humidity	Wind	Golf Players
9	Sunny	Cool	Normal	Weak	38
11	Sunny	Mild	Normal	Strong	48

Standard deviation for sunny outlook and normal humidity = 5

Rainy outlook and humidity

Humidity ✓	Standard deviation for golf players	instances
High ✓	7.50 ✓	2
Normal ✓	12.50	3

Weighted standard deviation for rainy outlook and humidity = (2/5)x7.50 + (3/5)x12.50 = 10.50

Standard deviation reduction for rainy outlook and humidity = 10.87 – 10.50 = 0.37

Rainy outlook and temperature

Temperature	Standard deviation for golf players	instances
Cool	14.50	2
Mild	7.32	3

Weighted standard deviation for rainy outlook and temperature = (2/5)x14.50 + (3/5)x7.32 = 10.19

Standard deviation reduction for rainy outlook and temperature = 10.87 – 10.19 = 0.67

Humidity	Stdev for Golf Players	Instances
High	4.08	3
Normal	5.00	2

Weighted standard deviations for sunny outlook and humidity = (3/5)×4.08 + (2/5)×5 = 4.45

Standard deviation reduction for sunny outlook and humidity = 7.78 − 4.45 = 3.33

Rainy outlook and wind

Wind	Standard deviation for golf players	instances
Weak	3.09	3
Strong	3.5	2

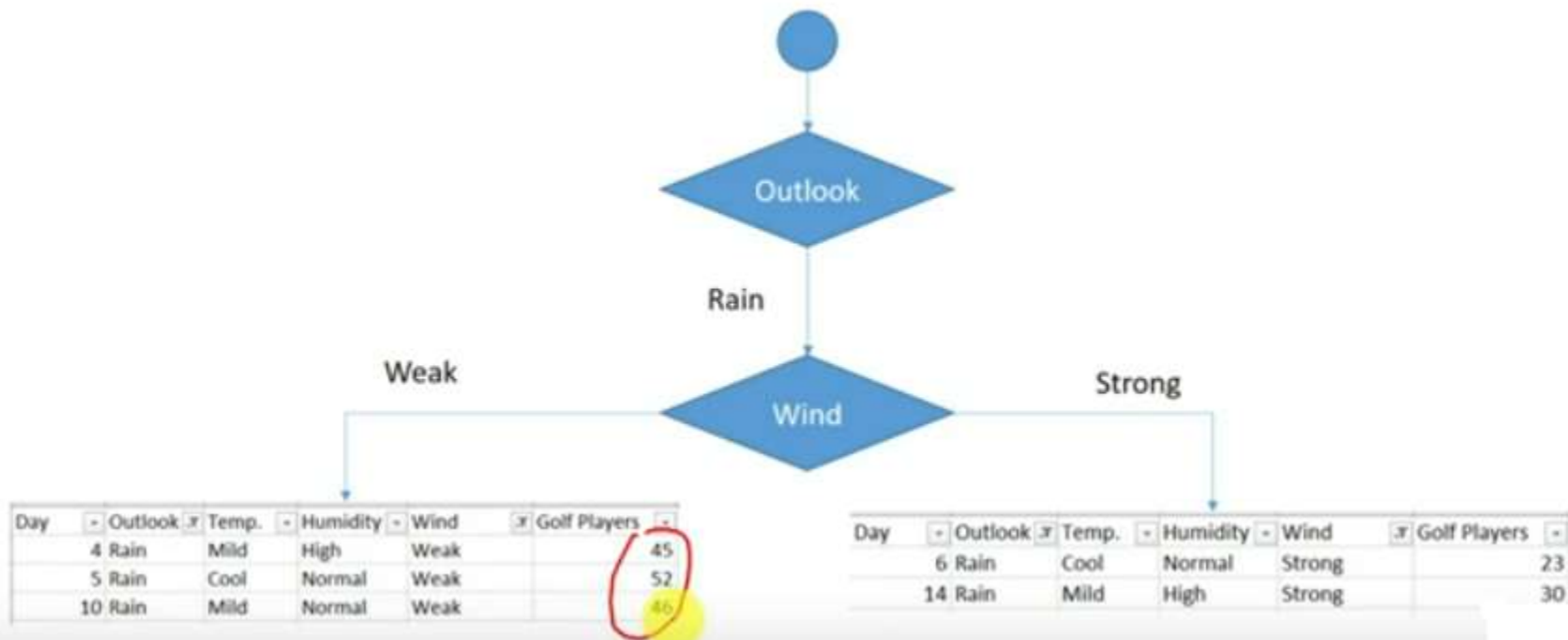Weighted standard deviation for rainy outlook and wind = (3/5)x3.09 + (2/5)x3.5 = 3.25

Standard deviation reduction for rainy outlook and wind = 10.87 – 3.25 = 7.62

Feature	Standard deviation reduction
Temperature	0.67
Humidity	0.37
Wind	7.62

Day	Outlook	Temp.	Humidity	Wind	Golf Players
4 Rain	Mild	High	Weak		45
5 Rain	Cool	Normal	Weak		52
10 Rain	Mild	Normal	Weak		46

Day	Outlook	Temp.	Humidity	Wind	Golf Players
6 Rain	Cool	Normal	Strong		23
14 Rain	Mild	High	Strong		30

weak=(45+52+46)/3=47.6

Strong=(23+30)/2=26.5

Regression Tree/Dr.N.Nandhini/AP/MCA/SNSCT

Final Tree

Regression Tree/Dr.N.Nandhini/AP/MCA/SNSCT