Fundamentals of Machine Learning 10/09/2024

Assignment -4

1. Develop a decision tree using Gini Index for the following dataset:

Day	Outlook	Temperature	Humidity	wind	play Tennis
1	Sunny	Hot	High	weak	No
2.	Sunny	Hot	High	Strong	No
3	Overcast	Hot	High	weak	Yes
4	Rain	Mild	High	weak	Yes
5	Rain	Cool	Normal	weak	Yes
6.	Rain	Cool	Normal	Strong	No
7	Overcast	Cool	Normal	Strong	Yes
8	Sunny	Mild	High	weak	No
9	Sunny	Cool	Normal	weak	Yes
10	Rain	Mild	Normal	weak	Yes
11	Sunny	Mild	Normal	Strong	Yes
12	Overcast	Mild	High	Strong	Yes
13	Overcast	Hot	Normal	weak	Yes
14	Rain	Mild	High	Strong	No

$$Gini(overall) = 1 - \sum_{i=1}^{c} (P_i)^2$$

1) Outlook attribute :-

OL values = (Su, Ov, Ra) = (5, 4, 5)

$$Gain(Sunny) = 1 - \left[\left(\frac{2}{5}\right)^2 + \left(\frac{3}{5}\right)^2\right]$$

Gain (Sunny) = 0.48

$$\text{Gain (Overcast)} = 1 - \left[\left(\frac{4}{4}\right)^2 + 0\right]$$

[4+, 0-]

$$\text{Gain (Overcast)} = 0$$

$$\text{Gain (Rain)} = 1 - \left[\left(\frac{3}{5}\right)^2 + \left(\frac{2}{5}\right)^2\right]$$

[3+, 2-]

$$\text{Gain Rain} = 0.48$$

$$\text{Gain Index (Outlook)} = \sum \frac{|S_v|}{|S|} \text{Gini}(S_v)$$

$$= \frac{5}{14} \times 0.48 + \frac{4}{14} \times 0 + \frac{5}{14} \times 0.48$$

$$\text{Gain Index (Outlook)} = 0.3428$$

ii) Temperature attribute

Temp. values = (Hot, Mild, Cool) = (4, 6, 4)

$$\text{Gini (Hot)} = [2+, 2-] = 1 - \left[\left(\frac{2}{4}\right)^2 + \left(\frac{2}{4}\right)^2\right] = 0.5$$

$$\text{Gini (Mild)} = [4+, 2-] = 1 - \left[\left(\frac{4}{6}\right)^2 + \left(\frac{2}{6}\right)^2\right] = 0.444$$

$$\text{Gini (Cool)} = [3+, 1-] = 1 - \left[\left(\frac{3}{4}\right)^2 + \left(\frac{1}{4}\right)^2\right] = 0.375$$

$$\text{Gini Index (Temp.)} = \frac{4}{14} \times 0.5 + \frac{6}{14} \times 0.444 + \frac{4}{14} \times 0.375$$

$$\text{Gini Index (Temp.)} = 0.440 \text{ or } 0.438$$

iii) Humidity attribute

H values = (High, Normal) = (7, 7)

Gini (High) (3+, 4-) $= 1 - \left[\left(\frac{3}{7}\right)^2 + \left(\frac{4}{7}\right)^2\right] = 0.4897$

Gini (Normal) (6+, 1-) $= 1 - \left[\left(\frac{6}{7}\right)^2 + \left(\frac{1}{7}\right)^2\right] = 0.2448$

Gini Index (Humidity) $= \frac{7}{14} \times 0.4897 + \frac{7}{14} \times 0.2448$

Gini Index (Humidity) = 0.367

iv) wind attribute

wind values = (weak, Strong) = (8, 6)

Gini (weak) (6+, 2-) $= 1 - \left[\left(\frac{6}{8}\right)^2 + \left(\frac{2}{8}\right)^2\right] = 0.375$

Gini (Strong) (3+, 3-) $= 1 - \left[\left(\frac{3}{6}\right)^2 + \left(\frac{3}{6}\right)^2\right] = 0.5$

Gini Index (wind) $= \frac{8}{14} \times 0.375 + \frac{6}{14} \times 0.5$

Gini Index (wind) = 0.428

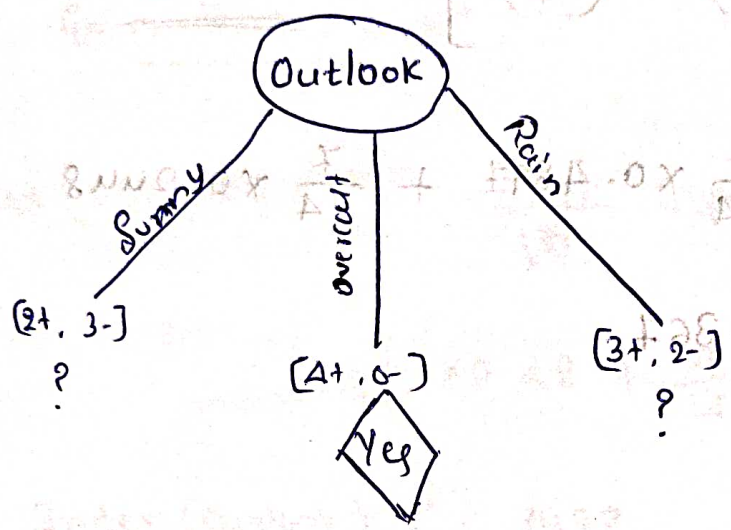
Gini Index (Outlook) = 0.3428 ✓

Gini Index (Temp.) = 0.438

Gini Index (Humidity) = 0.367

Gini Index (wind) = 0.428

Decision Tree



using Sunny updated table

Day	Outlook	Temperature	Humidity	wind	Play Tennis
1	Sunny	Hot	High	weak	NO
2	Sunny	Hot	High	Strong	NO
8	Sunny	Mild	High	weak	NO
9	Sunny	Cool	Normal	weak	yes
11	Sunny	Mild	Normal	Strong	yes

i) Temperature Attribute

Temp. values = (Hot, Cool, Mild) = (2, 1, 2)

$$\text{Gini (Hot)} = 1 - \sum_{i=1}^{c} (P_i)^2$$
(0+, 2-)

$$= 1 - \left[\left(\frac{0}{2}\right)^2 + \left(\frac{2}{2}\right)^2\right]$$

Gini (Hot) = 0

$$\text{Gini (Mild)} = 1 - \left[\left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^2\right] = 0.5$$
(1+, 1-)

$$\text{Gini (Cool)} = 1 - \left[\left(\frac{1}{1}\right)^2 + 0\right] = 0$$
(1+, 0-)

$$\text{Gini Index (Sunny, Temp)} = \sum \frac{|S_v|}{|S|} \text{Gini}(S_v)$$

$$= \frac{2}{5} \times 0 + \frac{2}{5} \times 0.5 + \frac{1}{5} \times 0$$

Gini Index (Sunny, Temp) = 0.2

ii) Humidity attribute

H values = (High, Normal) = (3, 2)

$$\text{Gini (High)} = 1 - \left[\left(\frac{0}{3}\right)^2 + \left(\frac{3}{3}\right)^2\right] = 0$$
(0+, 3-)

$$\text{Gini (Normal)} = 1 - \left[\left(\frac{2}{2}\right)^2 + \left(\frac{0}{2}\right)^2\right] = 0$$
(2+, 0-)

$$\text{Gini Index (Sunny, Humidity)} = \frac{3}{5} \times 0 + \frac{2}{5} \times 0$$

Gini Index (Sunny, Humidity) = 0

iii) wind attribute

wind values = (weak, Strong) = (3, 2)

$$\text{Gini (weak)} = 1 - \left[\left(\frac{1}{3}\right)^2 + \left(\frac{2}{3}\right)^2\right] = 0.444$$

(1+, 2-)

$$\text{Gini (Strong)} = 1 - \left[\left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^2\right] = 0.5$$

(1+, 1-)

$$\text{Gini Index (Sunny, Wind)} = \frac{3}{5} \times 0.444 + \frac{2}{5} \times 0.5$$

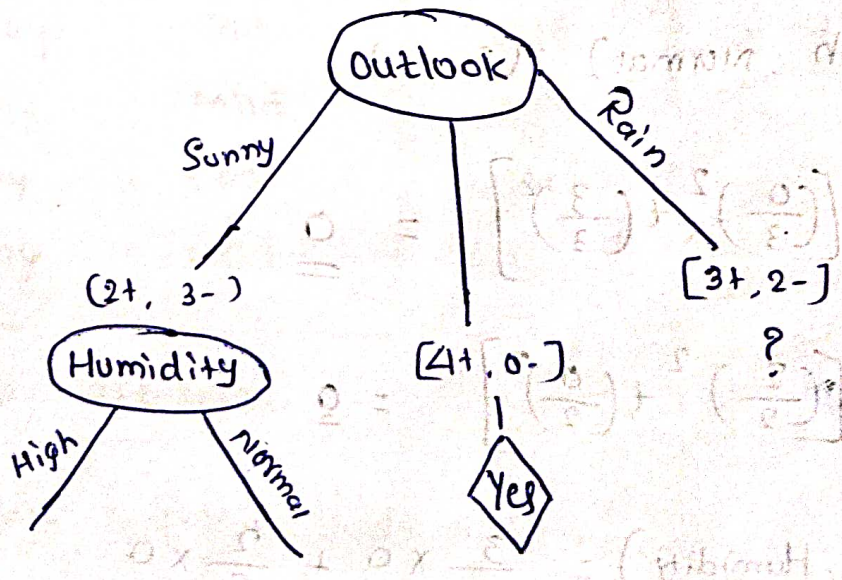
Gini Index (Sunny, Wind) = 0.464

Gini Index (Sunny, Temp) = 0.2

Gini Index (Sunny, Humidity) = 0 ✓

Gini Index (Sunny, wind) = 0.464

Decision tree



using Rain update table

Day	Outlook	Temperature	Humidity	Wind	Play Tennis
4	Rain	Mild	High	weak	Yes
5	Rain	Cool	Normal	weak	Yes
6	Rain	Cool	Normal	Strong	NO
10	Rain	Mild	Normal	weak	Yes
14	Rain	Mild	High	Strong	NO

i) Temperature attribute

Temp values = (Mild, Cool) = (3, 2)

$$\text{Gini}(\text{mild}) = 1 - \left[\left(\frac{2}{3}\right)^2 + \left(\frac{1}{3}\right)^2\right] = 0.44$$

(2+, 1-)

$$\text{Gini}(\text{cool}) = 1 - \left[\left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^2\right] = 0.5$$

(1+, 1-)

$$\text{Gini Index}(\text{Rain}, \text{Temp}) = 0 + \frac{3}{5} \times 0.44 + \frac{2}{5} \times 0.5$$

$$\text{Gini Index}(\text{Rain}, \text{Temp}) = 0.464$$

ii) Humidity attribute

H values = (High, Normal) = (2, 3)

$$\text{Gini}(\text{High}) = 1 - \left[\left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^2\right] = 0.5$$

(1+, 1-)

$$\text{Gini}(\text{Normal}) = 1 - \left[\left(\frac{2}{3}\right)^2 + \left(\frac{1}{3}\right)^2\right] = 0.44$$

(2+, 1-)

Gini Index (Rain, Humidity) $= \frac{2}{5} \times 0.5 + \frac{3}{5} \times 0.44$

Gini Index (Rain, Humidity) $= 0.464$

iii) wind attribute

wind values = (weak, Strong) = (3, 2)

Gini (weak) (3+, 0-) $= 1 - \left[\left(\frac{3}{3}\right)^2 + \left(\frac{0}{3}\right)^2\right] = 0$

Gini (Strong) (0+, 2-) $= 1 - \left[\left(\frac{0}{2}\right)^2 + \left(\frac{2}{2}\right)^2\right] = 0$

Gini Index (Rain, wind) $= \frac{3}{5} \times 0 + \frac{2}{5} \times 0$

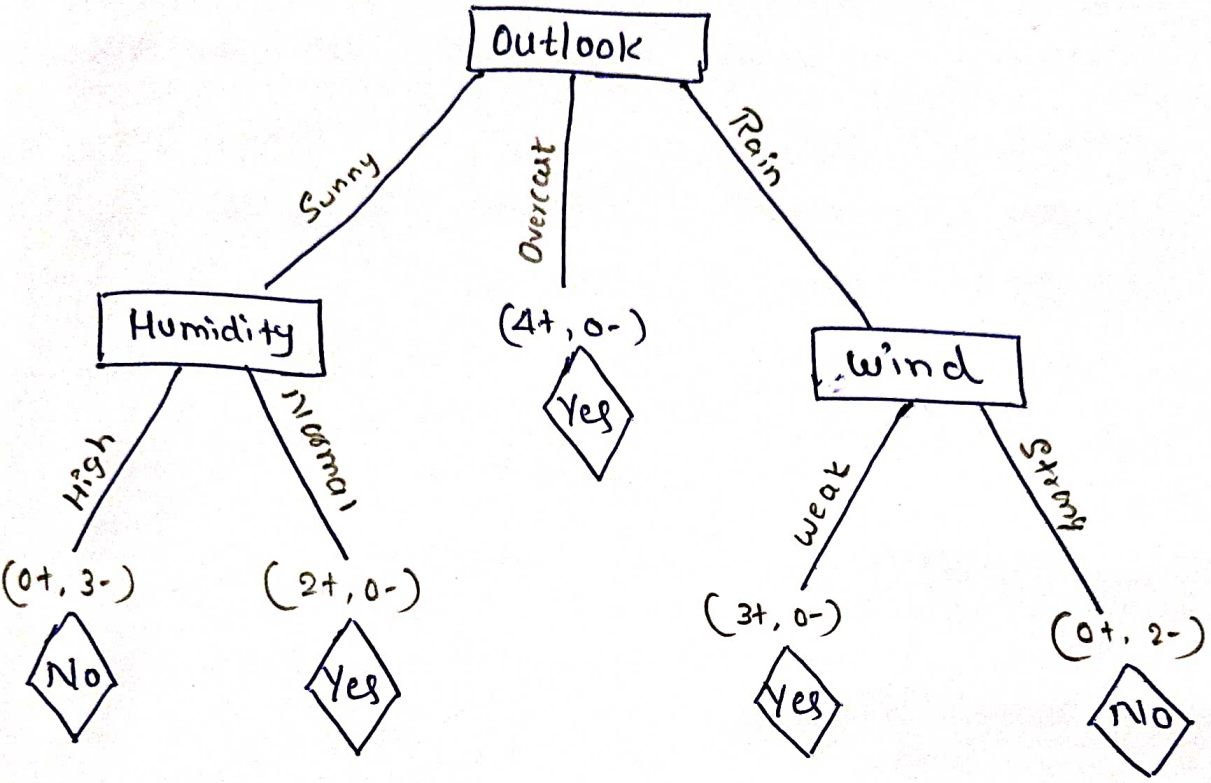
Gini Index (Rain, wind) $= 0$

Gini Index (Rain, Temp) $= 0.464$

Gini Index (Rain, Humidity) $= 0.464$

Gini Index (Rain, wind) $= 0$ ✓

Decision tree



FML

Assignment 4:- Decision tree using Gini index

Submitted by :- Karthik. C

Reg. No :- 240158016