Homework 4

Tuesday, April 1, 2025 8:14 PM

1. Consider the following data set as training data and do the following tasks (In your calculation, you can discard the attribute **Day**):

Day	Outlook	Temp.	Humidity	Wind	Decision
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

A. to select the attribute in each step of building a decision tree. using **Gini Index/Gini gain**. (***Please show your solutions/calculations in detail.***)
B. Draw the final decision tree.
C. Obtain the **class label** (based on the constructed tree) for the following **test data** (fill-out the table):

Answer for 1-C

Outlook	Temp	Humidity	Wind	Decision/Class label
Sunny	Mild	High	Strong	?
Rain	Hot	Normal	Strong	?
Overcast	Mild	Normal	Weak	?

Gini(class variable) Yes: 8/12 No: 4/12

$= 1 - \left[\left(\frac{8}{12}\right)^2 + \left(\frac{4}{12}\right)^2\right]$

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

$= 0.444$

			Decision: Yes	Decision: No
Outlook:	Sunny	5/12	3/5	3/5
	Overcast	3/12	3/3	0/3
	Rain	4/12	3/4	1/4

Gini(Outlook)

$= (5/12) * gini(2, 3) + (3/12) * gini(3, 0) + (4/12) * gini(3, 1)$

$= (5/12) * \left(1 - \left[\left(\frac{2}{5}\right)^2 + \left(\frac{3}{5}\right)^2\right]\right) + (3/12) * (1 - [(1)^2 + (0)^2]) + (4/12) * \left(1 - \left[\left(\frac{3}{4}\right)^2 + \left(\frac{1}{4}\right)^2\right]\right)$

$= (5/12) * (0.48) + (1/4) * (0) + (1/3) * (0.375)$

$= 0.325$

		Decision: Yes	Decision: No

= 0.320

			Decision: Yes	Decision: No
Temp:	Hot	3/12	1/3	2/3
	Mild	5/12	4/5	1/5
	Cool	4/12	3/4	1/4

Gini(Temp)
= (3/12) * gini(1,2) + (5/12) * gini(4,1) + (4/12) * gini(3,1)
= (1/4) * (1 - [(1/3)² + (2/3)²]) + (5/12) * (1 - [(4/5)² + (1/5)²]) + (1/3) * (1 - [(3/4)² + (1/4)²])
= (1/4) * (0.444) + (5/12) * (0.32) + (1/3) * (0.375)
= 0.3694

			Decision: Yes	Decision: No
Humidity:	High	6/12	3/6	3/6
	Normal	6/12	5/6	1/6

Gini(Humidity)
= (6/12) * gini(3,3) + (6/12) * gini(5,1)
= (1/2) * (1 - [(3/6)² + (3/6)²]) + (1/2) * (1 - [(5/6)² + (1/6)²])
= (1/2) * (0.5) + (1/2) * (0.2778)
= 0.3889

			Decision: Yes	Decision: No
Wind:	Strong	5/12	3/5	2/5
	Weak	7/12	5/7	2/7

Gini(Wind)
= (5/12) * gini(3,2) + (7/12) * gini(5,2)
= (5/12) * (1 - [(3/5)² + (2/5)²]) + (7/12) * (1 - [(5/7)² + (2/7)²])
= (5/12) * (0.48) + (7/12) * (0.408)
= 0.438

Gini Gain:
= gini(class data) - weighted gain(variable)

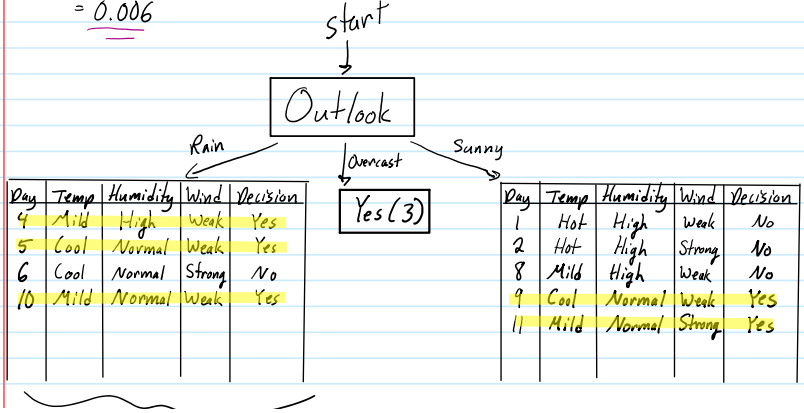
Outlook Gini Gain
= gini(class data) - weighted gain(outlook)
= 0.444 - 0.325
= 0.119

Temp Gini Gain
= 0.444 - 0.3694
= 0.0746

Humidity Gini Gain
= 0.444 - 0.3889
= 0.0551

Wind Gini Gain
= 0.444 - 0.438
= 0.006

Want maximum gain, which is Outlook
So that becomes root of the tree.



Day	Temp	Humidity	Wind	Decision
4	Mild	High	Weak	Yes
5	Cool	Normal	Weak	Yes
6	Cool	Normal	Strong	No
10	Mild	Normal	Weak	Yes

Day	Temp	Humidity	Wind	Decision
1	Hot	High	Weak	No
2	Hot	High	Strong	No
8	Mild	High	Weak	No
9	Cool	Normal	Weak	Yes
11	Mild	Normal	Strong	Yes

$\frac{4}{5} + \frac{9}{25} = \frac{13}{25} \rightarrow \frac{12}{25}$

11	Mild	Normal	Strong	Yes

$\frac{4}{25} + \frac{9}{25} = \frac{13}{25} \rightarrow \frac{12}{25}$

Rain

Gini(class variable) Yes : 3/4 No : 1/4
= 1 − [(3/4)² + (1/4)²]
= 0.375

			Decision: Yes	Decision: No
Temp :	Hot	0/4	–	–
	Mild	2/4	2/2	0/2
	Cool	2/4	1/2	1/2

Gini(Temp)
= (2/4) * gini(2,0) + (2/4) * gini(1,1)
= (2/4) * [1 − [(2/2)² + (0)²]) + (2/4) * (1 − [(1/2)² + (1/2)²])
= (1/2) * (0) + (1/2) * (1/2)
= 0.25

			Decision: Yes	Decision: No
Humidity :	High	1/4	1/1	0/1
	Normal	3/4	2/3	1/3

Gini(Humidity)
= (1/4) * gini(1,0) + (3/4) * gini(2,1)
= (1/4) * [1 − [(1/1)² + (0/1)²]) + (3/4) * (1 − [(2/3)² + (1/3)²])
= (1/4) * (0) + (3/4) * (0.444)
= 0.333

1 − 5/9

			Decision: Yes	Decision: No
Wind :	Strong	1/4	0/1	1/1
	Weak	3/4	3/3	0/3

Gini(Wind)
= (1/4) * gini(0,1) + (3/4) * gini(3,0)
= (1/4) * [1 − [(0/1)² + (1/1)²]) + (3/4) * (1 − [(3/3)² + (0/3)²])
= (1/4) * (0) + (3/4) * (0)
= 0.00

Gini Gain :

Temp Gini Gain
= Gini(class variable) − Weighted Gain(Temp)
= 0.375 − 0.25
= 0.125

Humidity Gini Gain
= 0.375 − 0.333
= 0.042

Want maximum gain, which is wind

Wind Gini Gain
= 0.375 − 0
= 0.375

Sunny

Gini(class variable) Yes : 2/5 No : 3/5
= 1 − [(2/5)² + (3/5)²]
= 0.48

			Decision: Yes	Decision: No
Temp :	Hot	2/5	0/2	2/2
	Mild	2/5	1/2	1/2
	Cool	1/5	1/1	0/1

Gini(Temp)
= (2/5) * gini(0,2) + (2/5) * gini(1,1) + (1/5) * gini(1,0)
= (2/5) * [1 − [(0/2)² + (2/2)²]) + (2/5) * (1 − [(1/2)² + (1/2)²]) + (1/5) * (1 − [(1/1)² + (0/1)²])
= (2/5) * ( ) + (2/5) * ( )
= 0.

			Decision: Yes	Decision: No
Humidity :	High	1/4	1/1	0/1
	Normal	3/4	2/3	1/3

Gini(Humidity)
= (1/4) * gini(1,0) + (3/4) * gini(2,1)
= (1/4) * [1 − [(1/1)² + (0/1)²]) + (3/4) * (1 − [(2/3)² + (1/3)²])
= (1/4) * (0) + (3/4) * (0.444)
= 0.333

1 − 5/9

			Decision: Yes	Decision: No
Wind :	Strong	1/4	0/1	1/1
	Weak	3/4	3/3	0/3

Gini(Wind)
= (1/4) * gini(0,1) + (3/4) * gini(3,0)
= (1/4) * [1 − [(0/1)² + (1/1)²]) + (3/4) * (1 − [(3/3)² + (0/3)²])
= (1/4) * (0) + (3/4) * (0)
= 0.00

Gini Gain :

Temp Gini Gain
= Gini(class variable) − Weighted Gain(Temp)
= 0.375 − 0.25
= 0.125

Humidity Gini Gain
= 0.375 − 0.333
= 0.042

Want maximum gain, which is wind

Wind Gini Gain
= 0.375 − 0
= 0.375