

11.11.2020

## ML ASSIGNMENT

11T2018178

MANAV AGRAWAL

Decision Tree Problems

①

Age	Heart Disease		Age	Heart Disease
44	0	Sort the list ⇒	40	1
52	0		43	0
57	0		44	0
48	1		48	1
54	0		48	0
48	0		48	0
54	0		49	0
48	0		50	0
49	0		52	0
64	0		54	0
58	0		54	0
58	1		57	0
58	1		58	0
60	1		58	0
50	0		58	1
58	0		58	1
66	0		60	1
43	0		64	0
40	1		64	1
64	1		65	0

Age	Heart Disease
41.5	
43.5	
46	
48	
48.5	
49.5	
51	
53	
54	
55	
57	
58	
59	
62	
64, 65	

←  
getting all  
average distinct  
values for  
comparison



Age	Greater than Equal to		Less than	
	(+ve) labels (l)	(-ve) labels (o)	(l) (+ve) labels	(o) (-ve) labels
41.5	5	14	1	0
43.5	5	13	1	1
46	5	12	1	2
48	<del>5</del>	<del>10</del> 12	1	2
49.5	4	9	2	<del>5</del>
51	4	8	2	6
53	4	7	2	<del>4</del> 7
54	4	7	2	7
55.5	4	5	2	<del>7</del> 9
57.5	4	4	2	<del>8</del> 10
58	4	4	2	10
59	2	2	<del>2</del> 4	<del>10</del> 12
62	1	2	<del>5</del>	12
64	1	2	5	12
65	0	1	6	13
48.5	4	10	2	4

Next we calculate Weight for each of our Ages (Nodes)

To calculate Weight of for any branch of node,

$$\text{Weight} = 1 - \left( \frac{(\text{+ve}) \text{ labels}}{(\text{+ve}) + (\text{-ve}) \text{ labels}} \right)^2 - \left( \frac{(\text{-ve}) \text{ labels}}{(\text{+ve}) \text{ labels} + (\text{-ve}) \text{ labels}} \right)^2$$

We calculate Weight for values less than the age and more than or equal to age for all ages.



Age	Weights	
	<del>Greater</del> Greater than Equal to	less than
41.5	0.38	0
43.5	0.401	0.5
46	<del>0.415</del> 0.415	0.44
48	<del>0.415</del> 0.415	0.44
48.5	0.408	0.44
49.5	0.426	0.408
51	0.444	0.375
53	0.462	0.345
54	0.462	0.345
55.5	0.493	0.297
57.5	<del>0.493</del> 0.5	0.27
58	0.5	0.27
59	0.5	0.375
62	0.44	0.415
64	0.44	0.415
65	0	0.432

Lastly, we calculate Gini ~~weights~~ Value for each Age.

Gini

$$Gini = \left( \frac{\text{Total Values Greater Than Equal to Age}}{\text{Total Nos of Ages}} \right) \times \text{Pos Gini Weight} + \left( \frac{\text{Total Nos of Values less than Age}}{\text{Total nos of Ages}} \right) \times \text{Neg Gini Weight}$$

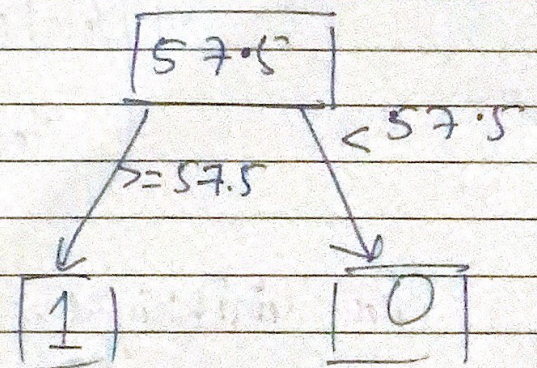
Pos Gini Weight  $\Rightarrow$  Corresponds to weight calculated for ages greater than equal to the particular Age (node)

$\rightarrow$  Neg Gini Weight  $\Rightarrow$  Corresponds to weight calculated for ages less than ~~the~~ the Node (Age)



Age	Gini Impurity
41.5	0.368
43.5	0.411
46	0.4196
48	0.4196
48.5	0.4190
49.5	0.4197
51	0.416
53	0.410
54	0.410
55.5	0.385
57.5	0.366
58	0.366
59	0.4
62	0.419
64	0.419
65	0.410

Take select Age with least Gini Impurity Value as our Node to ~~split~~ classify heart disease.





2)

Slope	Heart Disease
3	0
2	1
2	1
3	0
1	0
1	0
3	1
1	0
2	1
3	1
2	0
2	0
2	1
1	0
1	0
1	0
3	1
1	0
1	0
1	0

We calculate the total  
nos of values of slope 1, 2 and  
3, and calculate nos of 1s and  
0's of heart disease for  
each of them.

Slope	Heart Disease
1	
2	
3	

Slope	Heart Disease		Total Values
	(+ve)'s	(-ve)'s	
1	0	9	9
2	4	2	6
3	3	2	5



Now, we calculate the Entropy for each of these slopes using Formula.

$$\text{Entropy} = - \left( \frac{\text{total Positive Values}}{\text{total Vals}} \right) \times \log_2 \left( \frac{\text{total Pos}}{\text{total Vals}} \right) - \left( \frac{\text{total Negative values}}{\text{total Vals}} \right) \times \log_2 \left( \frac{\text{Negative Vals}}{\text{total Vals}} \right)$$

Slope	Entropy
1	0.45
2	0.3
2	0.25

Now, we calculate Global Entropy where

Total Vals = 20, Negative Vals = 13

Positive Vals = 7.

$$\therefore \text{Global Entropy} = 0.9340$$

Cumulative

Next, We calculate Cumulative Entropy.

$$\text{Cumulative Entropy} = \left( \frac{9}{20} \right) \times 0.45 + \left( \frac{6}{20} \right) \times 0.3 + \left( \frac{5}{20} \right) \times 0.25$$

$$(E = 0.355)$$

$$\text{Gain} = \text{Global Entropy} - \text{Cumulative Entropy}$$

$$| \text{Gain} = 0.934 - 0.355 = 0.579 |$$