

ID3 Algorithm (Decision Tree)

	A	B	C	D	E
1	gender	car	travel cost	income	transport
2	male	0	cheap	low	bus
3	male	1	cheap	medium	bus
4	female	1	cheap	medium	train
5	female	0	cheap	low	bus
6	male	1	cheap	medium	bus
7	male	0	standard	medium	train
8	female	1	standard	medium	train
9	female	1	expensive	high	car
10	male	2	expensive	medium	car
11	female	2	expensive	high	car
12					



YouTube

Search



যদি $\log_2 4$ এর মান জানতে চাই:

$$\log_2 4 = \log_2 2^2 = 2 \log_2 2 \quad [\log_2 2 = 1]$$

$$= 2$$

অথবা, $\log_2 4 = \frac{\log 4}{\log 2}$ যেহেতু $\log a = \frac{\log a}{\log b}$

$$= 2$$

$\log_2 \left(\frac{1}{4}\right) = \frac{\log \left(\frac{1}{4}\right)}{\log 2}$

অথবা: $\log \left(\frac{m}{n}\right) = \log m - \log n$

$$\log_2 \left(\frac{1}{4}\right) = \frac{\log 1 - \log 4}{\log 2}$$

$$= \frac{0 - \log 2^2}{\log 2}$$

$$= -2$$

$$\text{Entropy}(S) = \sum_{i=1}^n -p_i \log_2 p_i$$

$$\text{Gain}(S, A) = \text{Entropy}(S) - \sum_{\text{values of } A} \frac{|S_v|}{|S|} \times \text{Entropy}(S_v)$$

যদি E_A পার্টিশনের আগে এর এন্ট্রপি E_A হয় এবং
পরের এন্ট্রপি E_B হয় তবে আমাদের লক্ষ্য হল E_A
এবং E_B এর মধ্যে পার্থক্য যতটা সম্ভব বড় হওয়া।
 $\therefore E_A \gg E_B$ করতে হবে।

STUDY MART

4:01 / 34:10



#id3 #Decision_Tree_ML



	A	B	C	D	E
1	gender	car	travel cost	income	transport
2	male	0	cheap	low	bus
3	male	1	cheap	medium	bus
4	female	1	cheap	medium	train
5	female	0	cheap	low	bus
6	male	1	cheap	medium	bus
7	male	0	standard	medium	train
8	female	1	standard	medium	train
9	female	1	expensive	high	car
10	male	2	expensive	medium	car
11	female	2	expensive	high	car
12					
13					

Entropy Before partition :

$$E(S) = -\left(\frac{4}{10} \log_2 \frac{4}{10} + \frac{3}{10} \log_2 \frac{3}{10} + \frac{3}{10} \log_2 \frac{3}{10}\right)$$

$$= -(-0.528 - 0.521 - 0.521)$$

$$= 1.57$$

Now calculate entropy for each attribute :

$$E(\text{Gender} - \text{male}) = -\left(\frac{3}{5} \log_2 \frac{3}{5} + \frac{1}{5} \log_2 \frac{1}{5} + \frac{1}{5} \log_2 \frac{1}{5}\right)$$

$$= -(-0.442 - 0.464 - 0.464)$$

$$= 1.37$$

$$E(\text{Gender} - \text{female}) = -\left(\frac{2}{5} \log_2 \frac{2}{5} + \frac{1}{5} \log_2 \frac{1}{5} + \frac{2}{5} \log_2 \frac{2}{5}\right)$$

$$= -(-0.528 - 0.464 - 0.528)$$

$$= 1.52$$

Information Gain \rightarrow "Gender"

$$1.571 - \left(\left(\frac{5}{10} \times 1.37\right) + \left(\frac{5}{10} \times 1.52\right)\right)$$

$$= 0.125$$

	A	B	C	D	E
1	gender	car	travel cost	income	transport
2	male	0	cheap	low	bus
3	male	1	cheap	medium	bus
4	female	1	cheap	medium	train
5	female	0	cheap	low	bus
6	male	1	cheap	medium	bus
7	male	0	standard	medium	train
8	female	1	standard	medium	train
9	female	1	expensive	high	car
10	male	2	expensive	medium	car
11	female	2	expensive	high	car
12					
13					

$$E(\text{car}_0) = -\left(\frac{2}{3} \log_2 \frac{2}{3} + \frac{1}{3} \log_2 \frac{1}{3}\right)$$

$$= -(-.387 - .528)$$

$$= .915$$

$$E(\text{car}_1) = -\left(\frac{2}{5} \log_2 \frac{2}{5} + \frac{2}{5} \log_2 \frac{2}{5} + \frac{1}{5} \log_2 \frac{1}{5}\right)$$

$$= -(-.528 - .528 + .464)$$

$$= 1.52$$

$$E(\text{car}_2) = -\left(\frac{2}{2} \log_2 \frac{2}{2}\right)$$

$$= 0$$

Information Gain: "car"

$$= 1.571 - \left(\left(\frac{0}{10} \times .915\right) + \left(\frac{5}{10} \times 1.52\right) + \left(\frac{2}{10} \times 0\right)\right)$$

$$= 1.571 - (.274 + .76 + 0)$$

$$= .537$$

	A	B	C	D	E
1	gender	car	travel cost	income	transport
2	male	0	cheap	low	bus
3	male	1	cheap	medium	bus
4	female	1	cheap	medium	train
5	female	0	cheap	low	bus
6	male	1	cheap	medium	bus
7	male	0	standard	medium	train
8	female	1	standard	medium	train
9	female	1	expensive	high	car
10	male	2	expensive	medium	car
11	female	2	expensive	high	car
12					
13					

$$E(\text{cost-cheap}) = -\left(\frac{4}{5} \log_2 \frac{4}{5} + \frac{1}{5} \log_2 \frac{1}{5}\right)$$

$$= -(-.257 + .464)$$

$$= .721$$

$$E(\text{cost-standard}) = -\left(\frac{2}{2} \log_2 \frac{2}{2}\right)$$

$$= 0$$

$$E(\text{cost-Expensive}) = -\left(\frac{3}{3} \log_2 \frac{3}{3}\right)$$

$$= 0$$

Information Gain: "cost"

$$1.57 - \left(\left(\frac{5}{10} \times .721\right) + 0 + 0\right)$$

$$= 1.57 - .36$$

$$= 1.21$$

	A	B	C	D	E
1	gender	car	travel cost	income	transport
2	male	0	cheap	low	bus
3	male	1	cheap	medium	bus
4	female	1	cheap	medium	train
5	female	0	cheap	low	bus
6	male	1	cheap	medium	bus
7	male	0	standard	medium	train
8	female	1	standard	medium	train
9	female	1	expensive	high	car
10	male	2	expensive	medium	car
11	female	2	expensive	high	car
12					
13					

$$E(\text{Income-low}) = -\left(\frac{2}{2} \log_2 \frac{2}{2}\right) = 0$$

$$E(\text{Income-medium}) = -\left(\frac{2}{6} \log_2 \frac{2}{6} + \frac{3}{6} \log_2 \frac{3}{6} + \frac{1}{6} \log_2 \frac{1}{6}\right) = -(-.528 - .5 - .430) = 1.459$$

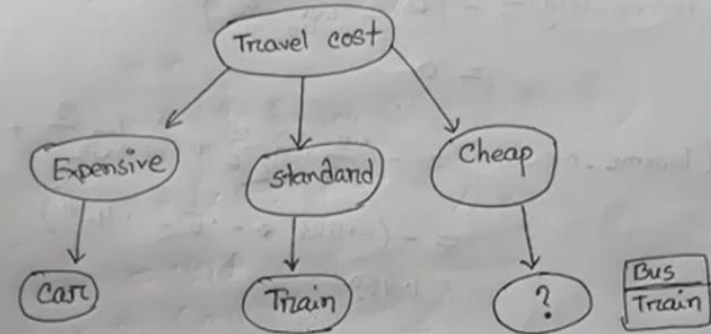
$$E(\text{Income-High}) = -\left(\frac{2}{2} \log_2 \frac{2}{2}\right) = 0$$

Information gain: "Income"

$$1.571 - \left(0 \times \frac{2}{10} + \left(\frac{6}{10} \times 1.459\right) + \left(\frac{2}{10} \times 0\right)\right) = 1.571 - .875 = .695$$

Attributes	Gain
Gender	.125
car	.537
Travel cost	1.21
Income	.695

	A	B	C	D	E
1	gender	car	travel cost	income	transport
2	male	0	cheap	low	bus
3	male	1	cheap	medium	bus
4	female	1	cheap	medium	train
5	female	0	cheap	low	bus
6	male	1	cheap	medium	bus
7	male	0	standard	medium	train
8	female	1	standard	medium	train
9	female	1	expensive	high	car
10	male	2	expensive	medium	car
11	female	2	expensive	high	car
12					
13					



second iteration → omit Travel cost only for cheap

Gender	car	Income	Transport
Male	0	Low	Bus
Male	1	Medium	Bus
Female	1	Medium	Train
Female	0	Low	Bus
Male	1	Medium	Bus

	A	B	C	D
1	gender	car	income	transport
2	male	0	low	bus
3	male	1	medium	bus
4	female	1	medium	train
5	female	0	low	bus
6	male	1	medium	bus
7				
8				

Entropy Before partition:

$$E(s) = - \left(\frac{4}{5} \log_2 \frac{4}{5} + \frac{1}{5} \log_2 \frac{1}{5} \right)$$

$$= -(-0.257 - 0.464)$$

$$= 0.721$$

Now Calculate entropy for each attribute:

$$E(\text{gender_male}) = - \left(\frac{3}{3} \log_2 \frac{3}{3} \right)$$

$$= 0$$

$$E(\text{gender_female}) = - \left(\frac{2}{2} \log_2 \frac{1}{2} + \frac{1}{2} \log_2 \frac{1}{2} \right)$$

$$= 1$$

Information Gain: "Gender"

$$0.721 - \left(\left(\frac{2}{5} \times 1 \right) + 0 \right)$$

$$= 0.721 - 0.40$$

$$= 0.321$$

	A	B	C	D
1	gender	car	income	transport
2	male	0	low	bus
3	male	1	medium	bus
4	female	1	medium	train
5	female	0	low	bus
6	male	1	medium	bus
7				
8				

$$E(\text{car}_0) = -\left(\frac{2}{2} \log_2 \frac{2}{2}\right)$$

$$= 0$$

$$E(\text{car}_1) = -\left(\frac{2}{3} \log_2 \frac{2}{3} + \frac{1}{3} \log_2 \frac{1}{3}\right)$$

$$= -\left(-.389 - .528\right)$$

$$= .917$$

Information Gain : "car"

$$.721 - \left((0) + (.917 \times \frac{3}{5})\right)$$

$$= .721 - .550$$

$$= .170$$

	A	B	C	D
1	gender	car	income	transport
2	male	0	low	bus
3	male	1	medium	bus
4	female	1	medium	train
5	female	0	low	bus
6	male	1	medium	bus
7				
8				

$$E(\text{income_low}) = -\left(\frac{2}{2} \log_2 \frac{2}{2}\right)$$

$$= 0$$

$$E(\text{income_Medium}) = -\left(\frac{2}{3} \log_2 \frac{2}{3} + \frac{1}{3} \log_2 \frac{1}{3}\right)$$

$$= -(\cdot389 - \cdot528)$$

$$= \cdot917$$

Information Gain: 'income'

$$\cdot721 - ((0) + (\cdot917 \times 3/5))$$

$$= \cdot170$$

Attributes	Info Gain
✓ Gender	<u>·322</u>
Car	·170
Income	·170

