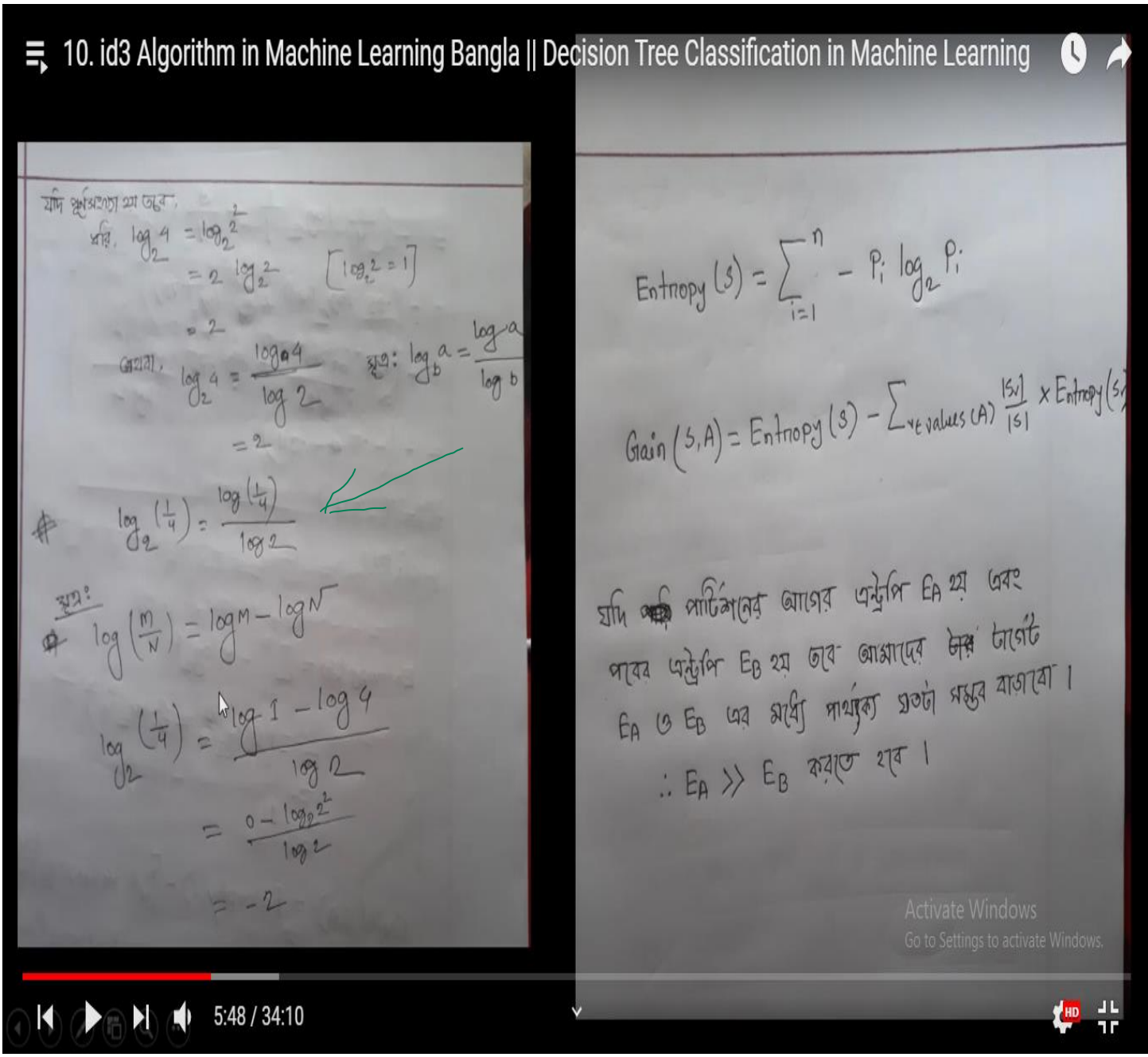

10. id3 Algorithm in Machine Learning Bangla || Decision Tree Classification in Machine Learning
যদি পূর্ণসংখ্যা না হবে:
ধরি, $\log_2 4 = \log_2 2^2$
$= 2 \log_2 2$ $[\log_2 2 = 1]$
$= 2$
অথবা, $\log_2 4 = \frac{\log_a 4}{\log 2}$ সূত্র: $\log_b a = \frac{\log a}{\log b}$
$= 2$
$\log_2 \left(\frac{1}{4}\right) = \frac{\log\left(\frac{1}{4}\right)}{\log 2}$
সূত্র:
$\log\left(\frac{m}{N}\right) = \log m - \log N$
$\log_2 \left(\frac{1}{4}\right) = \frac{\log 1 - \log 4}{\log 2}$
$= \frac{0 - \log_2 2^2}{\log 2}$
$= -2$
$Entropy(S) = \sum_{i=1}^{n} - P_i \log_2 P_i$
$Gain(S, A) = Entropy(S) - \sum_{v \in values(A)} \frac{|S_v|}{|S|} \times Entropy(S_v)$
যদি কোন পার্টিশনের আগের এন্ট্রপি E_A হয় এবং
পরের এন্ট্রপি E_B হয় তবে আমাদের লক্ষ্য টার্গেট
E_A ও E_B এর মধ্যে পার্থক্য যতটা সম্ভব বাড়াবো।
$\therefore E_A \gg E_B$ করতে হবে।
Activate Windows
Go to Settings to activate Windows.
5:48 / 34:10

	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 :

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

$$= -(-.528 - 0.521 - 0.521)$$

$$= 1.57$$

$$\frac{\log\left(\frac{4}{10}\right)}{\log 2} \times \frac{4}{10}$$

Now calculate entropy for each attribute :

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

$$= -(-.442 - .464 - .464)$$

$$= 1.37$$

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

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

$$= 1.52$$

Information Gain (→ "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					

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

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

$$= .915$$

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

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

$$= 1.52$$

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

$$= 0$$

Information Gain: "car"

$$= 1.571 - \left(\left(\frac{3}{10} \times .915\right) + \left(\frac{5}{10} \times 1.52\right) + \left(\frac{2}{10} \times 0\right)\right)$$

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

$$= .537$$

	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"

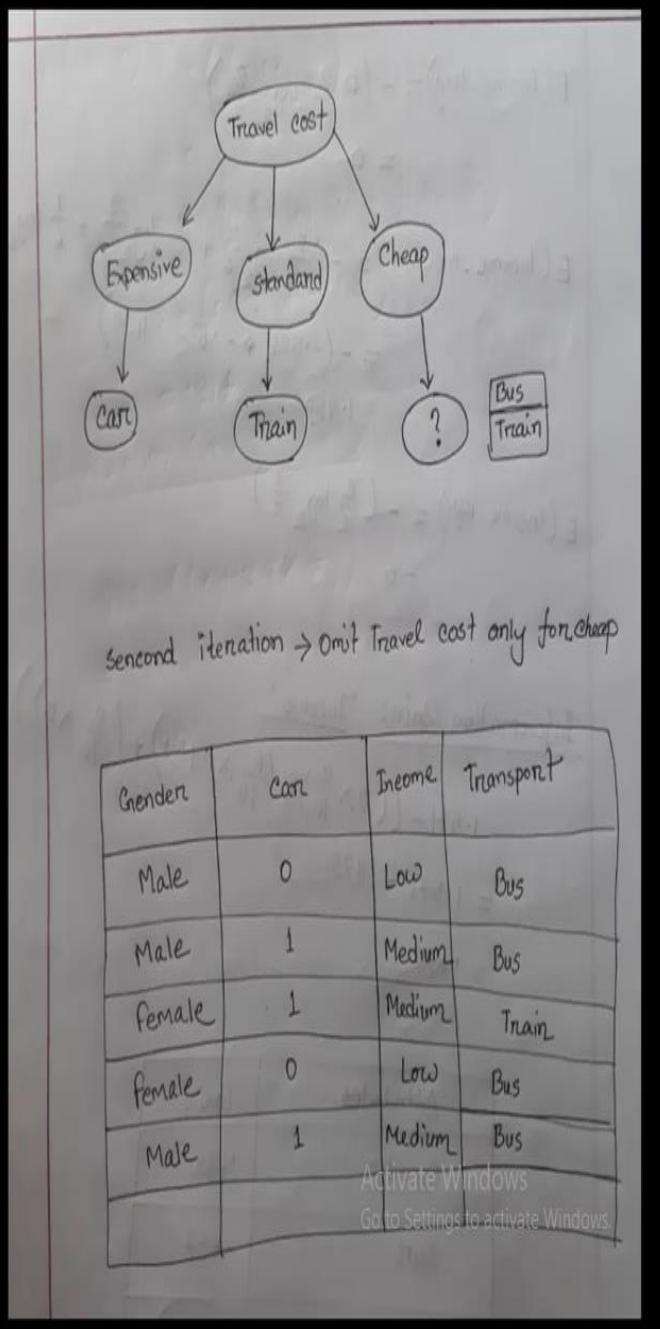
$1.571 - \left(\left(0 \times \frac{2}{10}\right) + \left(\frac{6}{10} \times 1.459\right) + \left(\frac{2}{10} \times 0\right)\right)$

$= 1.571 - .875$

$= .695$

Attributes	Gain
Gender	.125
car	.537
Travel cost	1.021
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					



sencond 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				

Entropy Before partition :

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

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

$$= .721$$

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"

$$.721 - \left(\left(\frac{2}{5} \times 1\right) + 0\right)$$

$$= .721 - .40$$

$$= .321$$

	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				

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

$$= 0$$

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

$$= -(-.389 - .528)$$

$$= .917$$

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				

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

$= 0$

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

$= -(.389 - .528)$

$= .917$

Information Gain: 'income'

$.721 - ((0) + (.917 \times 3/5))$

$= .170$

Attributes	Info Gain
Gender	.322
Car	.170
Income	

	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