BRAC UNIVERSITY
Department of Computer Science and Engineering
CSE 422: Artificial Intelligence
Assignment -04

Question-01

SL no.	unique_ID	Age	Income	Student ?	Credit Rating	Buys_Computer ?
1	#5	30	High	NO	Fair	NO
2	#19	30	High	NO	Excellent	NO
3	#24	32	High	NO	Fair	YES
4	#2	41	Medium	NO	Fair	YES
5	#10	42	Low	YES	Fair	YES
6	#32	44	Low	YES	Excellent	NO
7	#4	37	Low	YES	Excellent	YES
8	#8	28	Medium	NO	Fair	NO
9	#6	31	Low	YES	Fair	YES
10	#7	36	Medium	YES	Fair	YES
11	#9	22	Medium	YES	Excellent	YES
12	#14	35	Medium	NO	Excellent	YES
13	#13	37	High	YES	Fair	YES
14	#21	42	Medium	NO	Excellent	NO

(a) Among the features "**Income**" and "**Student**" which one do you think would be more suited for classifying if a person would buy a computer or not? Provide answer using ID3 **(4)**

(b) Do you think unique_ID would be a good feature for classification here? Provide a brief explanation to support your claim **(2)**

(c) Can we use Age as a building node for the decision tree? If not then what must we do beforehand? **(2)**

(d) For row number 4,10 and 11 if the value of Buys_Computer becomes "**Soon**", then what would be the Information Gain for feature "**Credit Rating**" **(2)**

Question-02

Dog_ID	Fur Color	Size	Tail_Length	Disease?
1	Black	Large	5.6	NO
28	White	Large	2.2	YES
3	Black	Small	3.8	YES
34	Black	Small	4.2	YES
26	Black	Large	1.2	NO
11	White	Small	2.3	NO
32	Black	Small	3.5	YES
13	White	Large	1.4	NO

a) Apply ID-3 and figure out among the 3 features which is best suited for being the root node of a decision tree.(**Fur Color**, **Size**, **Tail_Length**)
b) If an event X has **n** possible outcomes each with probabilities **p1, p2,...., pn**, then the entropy of X is defined as H(X) = -p1log2(p1) -p2log2(p2) ..... -pnlog2(pn) . If **n = 8**, what is the **maximum** possible value for H(X)? When does it happen?

Question-03

Vehicle ID	Engine Size (liters)	Fuel Efficiency (mpg)
1	1.6	30
2	2.0	27
3	2.5	24
4	3.0	22
5	3.5	20

(a) Given the following dataset, **find** what could be the Fuel Efficiency when the Engine Size is 2.9 liters.

Use Gradient Descent upto 3 iterations to estimate the unknown parameters and solve the problem. You can assume the two unknown parameters to be 0.1 and 0.2 and the learning rate to be 0.01

Let the loss function here sum of squared residuals, and linear function to fit is $y = ax+b$

(b) Explain why it is important for the derivative of the loss function with respect to each weight to be (near) zero for achieving better predictions in a model.

(c) Discuss the impact on the gradient descent algorithm when the learning rate increases over time.

(d) For a large data set **discuss** the effect of a decaying learning rate "α" (a learning rate that reduces over time), on the efficiency of the learning process.

(e) Discuss the impact on the gradient descent algorithm when the learning rate is too small.