CSE 422 (Section 16 & 17)
Assignment on post midterm syllabus
Submission deadline in hardcopy (9.1.2025)

1. A tourism company offers special discount card to its customers. Last year, they called many customers and a fraction of the customers accepted the offer. Here is the data that was collected by the Manager:

Serial No.	Job Type	Income Level	Likes to Hangout	Tours per year	Offer Taken
1	Engineer	High	Yes	2	Yes
2	Doctor	High	Yes	1	No
3	Engineer	Medium	No	3	Yes
4	Teacher	Medium	No	2	Yes
5	Doctor	High	Yes	3	Yes
6	Engineer	Medium	No	2	Yes
7	Teacher	High	Yes	1	No
8	Doctor	High	No	1	No
9	Teacher	High	No	2	Yes
10	Teacher	Medium	Yes	3	Yes
11	Engineer	High	No	1	No
12	Engineer	High	No	2	No

Your task is to learn a Decision tree based on this data to predict whether a particular customer will take the offer or not. Among Income Level and Likes to Hangout which is the better attribute for the root of the decision tree?

2. There are two boxes containing coins. The first box contains 60 gold coins and 40 silver coins. The second box contains 25 gold coins and 75 silver coins. One of the two boxes is randomly chosen (both boxes have probability 0.5 of being chosen) and then a coin is picked up at random from the chosen box. What is the probability that a silver coin is picked up? What is the probability that it comes from the first box?

3. A survey has been done on students to assess their interest in hostel accommodation. The data obtained is as follows:
   200 of them were male students. Among the male students, 80 were juniors (first and second year) with 50% interested in hostel accommodation and the rest were seniors with 70% interested in hostels. Among the 100 females, 70 were juniors with 60% interested in hostels and the rest were seniors with 80% interested in hostels.

a. Based on this data, construct a full joint distribution among the three random variables Gender(G), Category(C) and Interest in hostel accommodation(H)
b. Calculate the following probabilities from your table:
    i. Probability of a student being a junior
    ii. Probability of a female student not being interested in hostels

X	Y
2	11
4	23
6	28

4. Suppose, given a set of data samples in the table above, you are to fit the data points with a straight-line hypothesis Y = mX + c, where m is the slope and c is the y-intercept. The values of X and Y come from the data samples.
If the initial value of m is 9 and c is 2, **calculate** the error produced by the proposed hypothesis mean square error (MSE) or sum of square residual (SSR) function.
Update the m and c values with one iteration of gradient descent. If applicable, you can assume a learning rate of 0.001

5. What will be the prediction of PlayTennis in the following Table when Outlook = Sunny, Temperature = 20, Humidity = Normal and Wind = Strong using Naïve Bayesian classifier? For continuous attribute assume Gaussian distribution.

	Outlook	Temperature	Humidity	Wind	PlayTennis
D1	Sunny	40	High	Strong	No
D2	Sunny	45	High	Strong	No
D3	Overcast	42	High	Weak	Yes
D4	Rain	28	High	Weak	Yes
D5	Rain	18	Normal	Strong	No
D6	Rain	15	Normal	Strong	Yes
D7	Overcast	12	Normal	Strong	Yes
D8	Sunny	26	High	Strong	No
D9	Sunny	16	Normal	Weak	Yes
D10	Rain	27	Normal	Weak	Yes
D11	Sunny	25	Normal	Strong	Yes
D12	Overcast	29	High	Strong	Yes

D13 D14	Overcast Rain	38 24	Normal High	Weak Strong	Yes No

6.

a. Suppose X is a discrete random variable whose domain is exhaustive and mutually exclusive. Now the domain of X = {A, B, C}. Assume P(A) = 0.5 and P(B) = 0.3, then **determine** (i) P(C) and (ii) P(A U B).

b. Suppose two coins are tossed simultaneously. Assume Event A = the 1st coin coming up heads and Event B = the 2nd coin coming up tails. Now **determine** the value of P(A∩B).

c.

	A		A'	
	B	B'	B	B'
C	0.1	0.2	0.2	Y
C'	X	0.1	0.1	0.1

Using the given table answer the following questions:

(i) Assume the events A and B showcase absolute independence and P(B|A) = 0.5. Now **determine** the value of X

(ii) Using the ans obtained from (I), **determine** the value of Y

(iii) Using the ans obtained from (I) and (II), **determine** the value of P(A|B∩C)

7.

Answer the following questions based on the given Joint probability matrix.

	Male	**Female**	**Total**
Football	0.24	0.15	0.39
Rugby	0.2	0.05	0.25
Other	0.1	0.26	0.36
	0.54	0.46	1

a. **Estimate** the probability of someone playing Rugby if they are male.

b. **Estimate** the probability of being female if someone plays Football.

c. If you pick a person randomly, **estimate** the probability of that person playing one of Football or other games.

d. **Infer** whether playing rugby depends on females.