United International University (UIU)

Dept. of Computer Science & Engineering (CSE)

Final Exam: Fall 2023

Course Code: CSE 3811, Course Title: Artificial Intelligence

Total Marks: 40 Duration: 2 hours

Answer all questions. Marks are indicated in the right side of each question.

---

1. A person uses his car 30% of the time, walks 30% of the time and rides the bus 40% of the time as he goes to work. He is late 10% of the time when he walks; he is late 3% of the time when he drives; and he is late 7% of the time he takes the bus. What is the probability he walked if he arrives on time? **[4]**

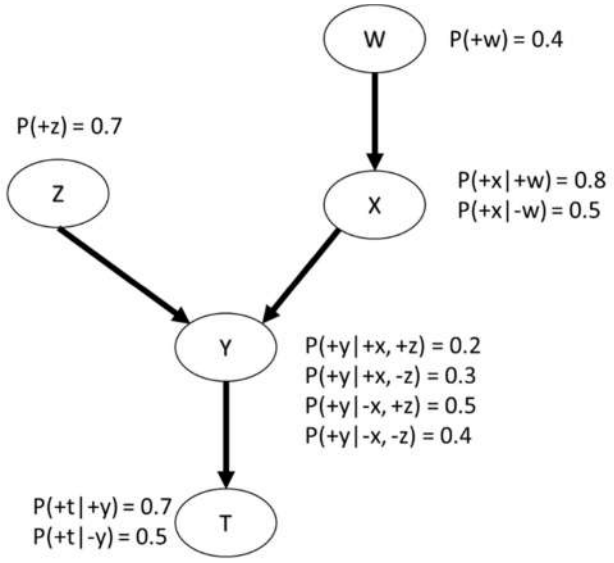
2. You are analyzing the trend of people's migration across continents using Markov Models. The table below displays the transition probability that you calculated for annual population migration.

Next Year → / Current Year ↓	Asia	Europe	Australia
Asia	0.6	0.3	0.1
Europe	0.05	0.9	0.05
Australia	0.1	0.2	0.7

   a. Given this year (Year **2023**) 80% people live in **Asia**, 15% people in **Europe** and the rest in **Australia**.
      i. In **2024**, what percentage of people will reside in Europe? **[1.5]**
      ii. In **2027**, how do you anticipate the distribution of people throughout the continents? **[1.5]**
   b. Determine the distribution of people across continents in the long run (stationary distribution) **[3]**

3. Suppose you are doing a survey on CSE department students to assess their interest for department tour. The survey is done on 200 students, half of them 3rd year and half 4th year. Among the 3rd year students, 60% are male students. Among the male 3rd year students, 30% are interested to go to Sundarbans, 40% Cox's bazar and the rest of them Sylhet. Among the female 3rd year students, 30% want to go to Sundarbans, 60% to Cox's bazar and the rest to Sylhet. 50% of the 4th year students are male and the ratio of liking Sundarbans, Cox's bazar and Sylhet is 3:2:5. For the female 4th year students, the ratio is 2:4:4.
   a. Based on this data, construct a full joint distribution among the three random variables Student category(C), Gender(G) and Destination(D). **[3]**
   b. Calculate the following probabilities from your table:
      i. Probability of a student being interested in visiting Sundarbans. **[1]**
      ii. $P(D|C=4^{th}$ year) **[2]**

4. Consider the following Bayes Net:



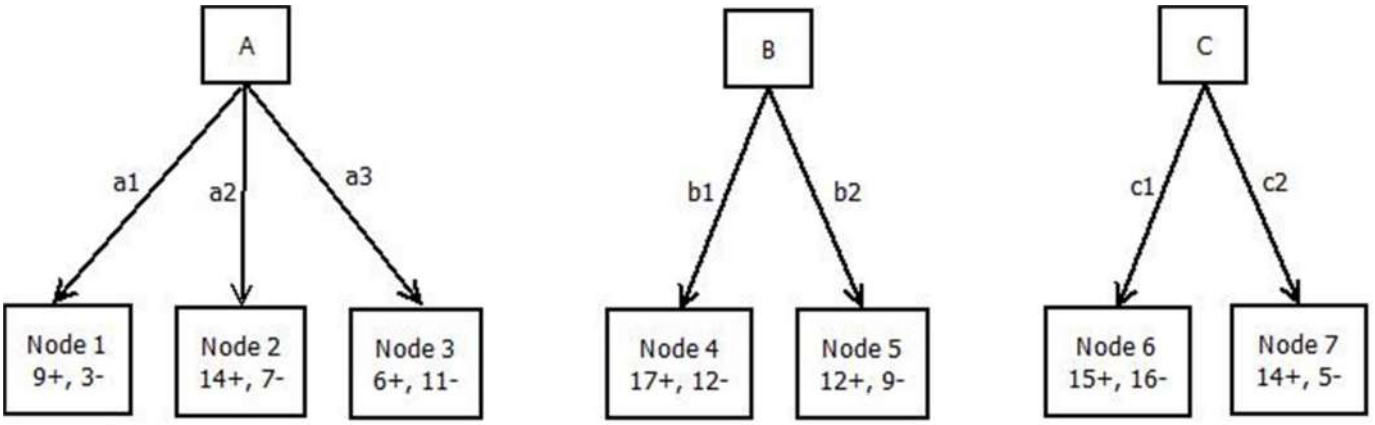
   a. Now how many entries are in total required to represent the above network? **[1]**
   b. Determine the following probability information from the given bayes net. At each query, +p denotes P = true, and -p denotes P = false. Show the steps.
      i. P(+w, +x, -z, +t) **[2]**
      ii. P(+w | -z, -y, +t) using Inference by Enumeration **[3]**
      iii. P(+w | -y, +t) using Variable Elimination **[3]**

5. The following heart disease prediction dataset contains four features and a label predicting whether the person is likely to have heart disease or not.

Smoker?	Blood Pressure	Cholesterol	Diabetes?	Heart Disease?
Yes	High	Medium	Yes	Yes
No	Medium	High	Yes	Yes
No	Low	Medium	No	No
No	High	Low	No	No
Yes	High	High	No	Yes
No	High	High	No	No
Yes	High	High	Yes	Yes
No	Low	Low	No	No
No	High	High	Yes	Yes
Yes	Medium	Medium	Yes	Yes

Use Naive Bayes classifier to determine whether a person with the features <Yes, Low, Medium, Yes> is likely to have heart disease or not. Use Laplacian smoothing with k=1. Show detailed calculation for both training step and testing step. **[7]**

6. Consider a dataset containing 50 samples with 29 positive and 21 negative samples and three attributes: **A, B** and **C**.
   Attribute **A** has three unique values: (a1, a2, a3)
   Attribute **B** has two unique values: (b1, b2)
   Attribute **C** has two unique values: (c1, c2)
   Target/Label Y has two class: yes represented by **+** and no represented by **-**
   The figure below displays three splits of the dataset that resemble the ID3 decision tree algorithm at the top level. The splits are based on the values of attribute A, B and C respectively.



Now answer the following questions: **[4+3+1]**

a. Calculate Entropy values for each of the seven nodes shown and the parent node
b. Calculate Information Gain values for both all three attributes.
c. Which attribute will ID3 decision tree select first to partition the dataset at this level?