



NATIONAL SCHOOL OF BUSINESS MANAGEMENT

BSc (Hons) in Computer Science – 18.2/19.1

3rd Year 2nd Semester Examination

12 September 2021

CS303.3 – Computational Theory

Instructions to Candidates

- 1) Answer all the questions
- 2) The duration of the question paper is 5 hours. Including half an hour to download the paper and upload your answers in a single file. (Note: No email submissions are accepted under any condition.)
- 3) Weightage of Examination: 60% out of final grade
- 4) Download the paper, provide answers to the selected questions in a word document.
- 5) Please upload the document with answers (Answer Script) to the submission link before the submission link expires
- 6) Answer script should be uploaded in PDF Format
- 7) Under any circumstances E-mail submissions would not be taken into consideration for marking. Incomplete attempt would be counted as a MISSED ATTEMPT.
- 8) The Naming convention of the answer script – Module Code_Subject name_Index No
- 9) You must adhere to the online examination guidelines when submitting the answer script to N-Learn.
- 10) Your answers will be subjected to Turnitin similarity check, hence, direct copying and pasting from internet sources, friend's answers etc. will be penalized.

Question-1 (Monte Carlo Simulations) [20 marks]

1. What is a Monte Carlo Simulation? (03 marks)
2. Describe how Monte Carlo simulation can be used to estimate the value of 'pi' (04 marks)
3. Explain how Monte Carlo Simulation can be used to model a Game of two players shooting at each other sequentially clearly stating your assumptions. Note that probability of person A surviving is 0.4 and the probability of person B surviving is 0.25.
(Code is not required.) (06 marks)
4. Explain how this technique of Monte Carlo Simulation could be used to model the vehicle traffic at a traffic light with the probability a vehicle reaching the light is 0.35 (Code is not required). State any assumptions you have made. (07 marks)

Question-2 (Numerical Methods) [20 marks]

1. Use the trapzoidal rule with 5 ordinates to estimate to 3 decimal places the value of $\int_0^{\pi/3} \cos 2x \, dx$. You need to clearly show your workings. (06 marks)
Hint: Make sure to substitute angles in radians.
2. State whether the estimate in part a) is an overestimate or underestimate of the area under the curve $y = \cos 2x$ between the limits $x = 0$ and $x = \pi/3$. (03 marks)
3. Explain how the above integration can be performed using a Monte Carlo simulation. (07 marks)
4. Find the below integral using the Simpson's 1/3 rd rule to 3 decimal places. (04 marks)
$$\int_0^{\pi/3} \sin 2x \, dx$$

Question -3 (Bayes Theorem and Naïve Bayes Classification) [20 marks]

1. A company produces light bulbs at three factories A, B, C. Factory A produces 45% of the total number of bulbs of which 4% are defective. Factory B produces 30% of the total number of bulbs of which 2% are defective and factory C produces 25% of the total number of bulbs of which 3% are defective. If a defective bulb is found among the output, find the probability that it came from factory A. (04 marks)
2. Assume that you are going to construct a classifier to categorize Salmon from Sea bass. Explain the steps you would use to construct a Naïve Bayes Classifier to separate Salmon from Sea bass. (06 marks)

3. Given $x_1 = 230$ g, and $x_2 = 20$ cm of a newly caught fish, determine the class to which this fish can be categorized? Note that ratio of salmon: sea bass is 3:2. You need to show your calculations. Refer to the table given below. (05 marks)

Weight -x1	Prob.	Weight -x1	Prob.
$P(x_1=200 \text{g} \text{Salmon})$	0.20	$P(x_1=200 \text{g} \text{Sea bass})$	0.24
$P(x_1=225 \text{g} \text{Salmon})$	0.40	$P(x_1=225 \text{g} \text{Sea bass})$	0.34
$P(x_1=230 \text{g} \text{Salmon})$	0.30	$P(x_1=230 \text{g} \text{Sea bass})$	0.32
Height-x2	Prob.	Height-x2	Prob.
$P(x_2=20 \text{ cm} \text{Salmon})$	0.35	$P(x_2=20 \text{ cm} \text{Sea bass})$	0.28
$P(x_2=25 \text{ cm} \text{Salmon})$	0.20	$P(x_2=25 \text{ cm} \text{Sea bass})$	0.25
$P(x_2=30 \text{ cm} \text{Salmon})$	0.35	$P(x_2=30 \text{ cm} \text{Sea bass})$	0.37

4. Consider the problem of binary classification using the Naive Bayes classifier. You are given two dimensional features (X_1, X_2) and the categorical class conditional distributions in the tables below. The entries in the tables correspond to $P(X_1 = x_1 | C_i)$ and $P(X_2 = x_2 | C_i)$ respectively. The two classes are equally likely.

Given a data point (-1, 1), Determine $P(C_1 | X_1 = -1, X_2 = 1) = ?$ (05 marks)

$X_1 =$	Class	
	C_1	C_2
-1	0.2	0.3
0	0.4	0.6
1	0.4	0.1

$X_2 =$	Class	
	C_1	C_2
-1	0.4	0.1
0	0.5	0.3
1	0.1	0.6

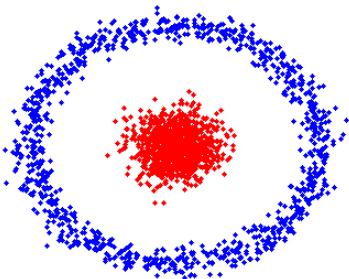
Question-4 (Linear Regression and Logistic Regression) [20 marks]

1. Suppose Linear Regression with multiple variables is to be applied on the below data set.

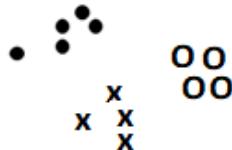
Size (x_1)	# of Bedrooms (x_2)	Price (y)
xx	xx	xx

- I. Develop a mathematical model to predict the Price in terms of x_1 & x_2 using multiple linear regression. (02 marks)
- II. Write down the pseudo code of an algorithm that can be used to estimate the parameters of your model. (04 marks)

2. Compare and contrast the techniques of linear regression and logistic regression, discussing when it is appropriate to use each. (05 marks)
3. Suppose you are required to train a classifier using the 2D training data given below. Explain how you can use logistic regression to build a classifier that can separately identify the two classes and then classify a given new point to its corresponding class. (05 marks)

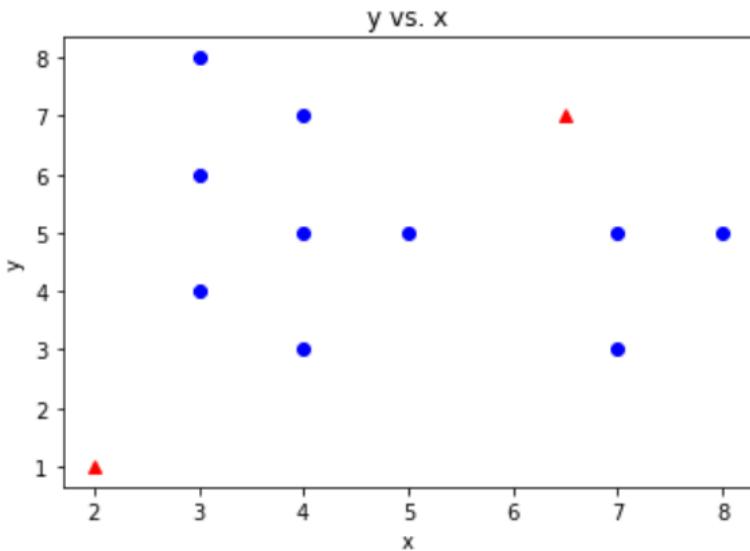


4. Explain the strategy you would use to classify a new data point to its corresponding class if logistic regression is applied on the below 2D training data comprising three classes. (04 marks)



Question-5 (k-means Clustering) [20 marks]

1. Describe the k-means clustering algorithm in your own words. (05 marks)
2. Following diagram provides a visualization of 10 pairs of 2D data points. Triangles indicate the initial points of two cluster centers $C_1 (2,1)$ & $C_2 (6.5,7)$. Read the data points from the diagram and answer the questions given below by applying the k-means clustering algorithm on the given data set.



- I. Show the class label assignment of each data point in the first iteration. You need to show your calculations. (05 marks)
 - II. Find the coordinates of the updated cluster centers. (03 marks)
 - III. Explain a suitable stopping criterion that you can use. (02 marks)
3. Given a new point (6,5) find its cluster assignment. You may assume that the data is clustered into two clusters in the final iteration of k-means algorithm by clustering the last 3 points (for all $x > 6$) in the diagram above into one cluster and the rest of the points into the other cluster. (03 marks)
4. State a limitation of k-means clustering algorithm and suggest a way to overcome it. (02 marks)

-----End-----