

Department of Computer Science and Engineering, MNIT Jaipur
MTE, M.Tech. II Semester (spring 2023-24), Software Testing and Validation (21CST838)

Time: 1 and half hour

Max.Marks: 30

- ✓ 1. Why do we really need **testing**? Discuss the significance of **Requirement Specifications** in software development and software testing activity. (1+2=03)
2. Differentiate between:
 - i) Data defects V/S Data flow defects,
 - ii) Positive testing v/s negative testing,
 - iii) Testing v/s Debugging,
 - iv) Black box testing v/s white box testing. (04)
3. Write **short** answers:-
 - ✓ (i) What are **Root Causes** of project failures? Explain. (2)
 - ✓ (ii) Why do we need to **test the tests first**? Explain. (1)
 - (iii) What is **Pesticide Paradox**? *repeated test* (1)
 - (iv) What is "The Convoy and The rags"? (1)
 - (v) Is Complete Testing Possible? Explain and justify. (2)
- ✓ 4. Explain **V model** (**design phase** in detail) with the help of a neat diagram. (6+2=08)
- ✓ 5. Define: Fault model, Phase containment of errors, Test harness, Test case and test oracle. (2*4=08)

Static Driven
□

Test Judge of test
highlighting



MALAVIYA NATIONAL INSTITUTE OF TECHNOLOGY JAIPUR
DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

Iyr M. Tech (CSE/CSIS), II Semester - Mid-Semester Examinations - February 2024
Cyber Security (21CST851)

Date: 28-02-2024

Time: 04:15 PM to 05:45 PM

Max. Marks: 30

Note: Answer all the questions. Attend questions in sequential order; starting from Q1. Calculators and any other electronic gadgets are not allowed.

1. (a) What are the characteristics of Firewall? (2M)
(b) Explain the types (any two) of Firewall with neat diagram. (4M)
Handwritten notes: M/W, D/P, A/rw, TCP
2. Give the importance of Intrusion Detection Systems (IDS). Compare Host-based and Network-based IDS. (2M + 4M = 6M)
3. (a) What are the elements of Access control? (2M)
(b) Explain any two Access control models. (2M + 2M = 4M)
4. (a) What is phishing attack and its different forms? (3M)
(b) Explain the obfuscation techniques (with example) used by Metamorphic malware programs to mutate their code while maintaining the same functionality in each generation. (4M)
5. Explain SQL injection attack with one example. (3M)
6. Write short answers (one or two lines) for the following: (4 x 0.5M = 2M)
(a) Name the type of virus that is interpreted rather than executed directly.
(b) A computer program, which is self-replicating but does not attach itself to an existing program is called as
(c) A user-level rootkit operates in which Ring level of the Operating System.
(d) Write down the name of first polymorphic engine made.



Malaviya National Institute of Technology Jaipur
Social Network Analysis (CST836)
Mid Term Examination
Date: February 29, 2024

Timing: 10:15 to 11:45 AM

Max marks: 30

Attempt all the questions

1. (a) For the given graph H , compute the followings:

1. Clustering Coefficient of nodes 1 and 3 ✓
2. Diameter of the graph H ✓ 3
3. Maximum Matching and size of the maximum matching (Matching Number) ✓
set of edge without common vertex
4. Maximum Clique and size of maximum clique (Clique Number) ✓
 $\{6, 7\}, \{5, 3\} \Rightarrow 2$
5. Maximum Independent Set and size of the maximum independent set (Vertex Independence Number) ✓
 $\{4, 5, 6, 8, 1\}$
(min) kitna vertex dekho an vertex wala
6. Minimum Vertex Cover and size of the minimum vertex cover (Vertex Covering Number) ✓
 $\{4, 5, 6, 8, 2\}$

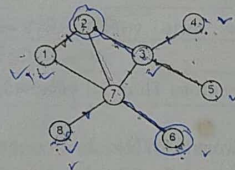
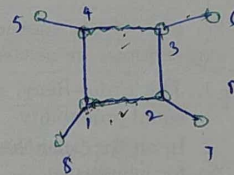


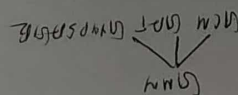
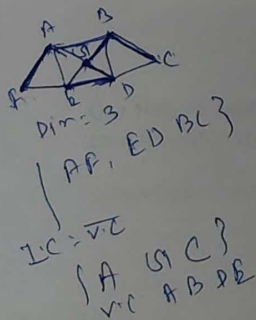
Figure 1: Graph H



matching = $\{1, 2\}, \{3, 4\}$
size = 2

(b) Calculate PageRank scores for the Graph G (shown on Page 2) using power iteration method (show calculations for the first three iterations only).

2. (a) Suggest two different novel ideas to improve the existing Graph Convolutional Networks (GCN) performance.



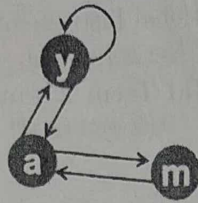
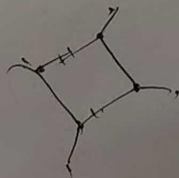


Figure 2: Graph G

- (b) Explain how dropout layers function in preventing overfitting in neural networks. What is the principle behind their operation, and how do they impact the training process? (2)
- (c) How can you apply GNN for tabular data? (assume n data points and p features for each point) (2)
- (d) What is the purpose of using a softmax function in the output layer of a neural network for classification tasks? (1)
- (e) Why citation networks are acyclic? (1)
- (f) Why Email networks have self loops? (1)
3. (a) For each of the questions below, write your answer either True or False. No explanations are needed.
1. The Erdős-Rényi model assumes that all connections in a network have equal probability. (1)
 2. In an Artificial Neural Network (ANN), the activation function is optional for neurons. (1)
 3. In deep learning, a higher number of layers always leads to better model performance. (1)
 4. Bipartite graphs do not contain even length cycles. (1)
- (b) Explain node embedding generation process (with equations) in Graph Convolutional Networks (GCN). (4)
- (c) What does the "strength of weak ties" theory propose in social network analysis? (3)

Best wishes



Department of Computer Science and Engineering, MNIT, Jaipur
Natural Language Processing 21CST822 MTE, February 2024

Max. Marks: 30

Attempt all questions

Time: 90 Minutes

1. Given some Weekdays, the task is to check if they are valid or not using regular expressions. Rules for the valid Weekdays; It should contain specific only words as a string. They are mentioned as: Monday, Tuesday, Wednesday, Thursday, Friday, Saturday, and Sunday, Mon, Tues, Wed, Thurs, Fri, Sat, Sun, Mon., Tues., Wed., Thurs., Fri., Sat., Sun, mon, tues, wed, thurs, fri, sat, sun (Don't use simple disjunction of all words) [2]
2. We are given the following corpus:
 a. <s> You are Rio Jam </s> b. <s> Rio You are Jio </s> c. <s> You are Rio </s> d. <s> You like Jio orange Rio </s> Using a bigram model with add-one smoothing, what is $P(\text{Rio}|\text{are})$ and $P(\text{<s> You like Jio Rio </s>})$? Include <s> and </s> in your counts just like any other token. [4]
3. Compute minimum edit distance in between "zwiliamcohen" and "wwoiamcohon". (consider all operations costs as same i.e. 1 for each operation) [show complete table][5]
4. Computer the class of test document using Naïve Bays Classifier (consider add one smoothing). [3]

Training	Doc	Words	Class
	1	Rio sport great India Proud Dish	N
	2	India sport bad player	N
	3	Sam injustice poverty India	P
	4	Rio games feel good sport	N
Test	5	India injustice Sam Rio	?

5. Computer the cross entropy loss for both classes for the following observation: [3]
 $X=[2, 1, 0, 2, 3.12, 1.2]$ $W=[3.5, -6, -1.2, 0.4, 3.5, 0.3]$ $b=0.20$
6. Use gradient descent learning used in logistic regression to updates the weights and bias after one iteration for the sample $x_1=6$ (count of positive lexicon) $x_2=5$ (count of negative lexicon), initial values of $w_1=1.2$ $w_2=-3.2$, $b=1.5$ and learning rate is 0.1. [3]
7. Consider the following Tagged sentences:

Rahul	can	play	Chess
N	M	V	N
Karan	should	clean	Table
N	M	V	N
will	Shyam	play	Chess
M	N	V	N
Karan	should	beat	Rahul
N	M	V	N

Use Hidden Markov Model (HMM) to predict the probability of tagging the sentence "Rahul should beat Karan" with tag sequence "N M V N". [5]

8. Differentiate: (a) Stemming and Lemmatization (b) Winner take all and Outstar Learning Rule [2.5+2.5]

Malaviya National Institute of Technology Jaipur

Department of Computer Science and Engineering

Mid Term Examination, February, 2024

Research Methodology (21CST507)

Duration: 1.5 hrs

Max. Marks: 30

Part I: Basic Mathematics for CSE

Q1. (*Vector and Matrices*) Suppose, you are given a set of investment portfolios represented by vectors in \mathbb{R}^n , where each component corresponds to the proportion of the total investment allocated to a specific asset class. For example, given three investment portfolios:

$$P_1 = \begin{pmatrix} 0.2 \\ 0.5 \\ 0.3 \end{pmatrix}, \quad P_2 = \begin{pmatrix} 0.3 \\ 0.4 \\ 0.3 \end{pmatrix}, \quad P_3 = \begin{pmatrix} 0.1 \\ 0.6 \\ 0.3 \end{pmatrix}$$

Analyze whether these portfolios are linearly independent and discuss the span of the set. If a new portfolio with allocation $\begin{pmatrix} 0.25 \\ 0.45 \\ 0.3 \end{pmatrix}$ is introduced, how does it affect the diversification potential? **4 marks**

Q2. Given a covariance matrix Σ representing a dataset with three features:

$$\Sigma = \begin{pmatrix} 4 & 2 & 1 \\ 2 & 3 & 0 \\ 1 & 0 & 2 \end{pmatrix}$$

Find the eigenvalues of Σ and discuss their significance in PCA. How many principal components are needed to retain 90% of the variance in the data? **4 marks**

Q3. (*Descriptive statistics*) In a class of 50 students, the teacher recently conducted a midterm exam in Research Methodology. Upon grading the exams, the teacher notices a wide range of scores, ranging from 30 to 95 out of a possible 100 points. While some students performed exceptionally well, scoring in the 90s, others struggled and scored below 50.

Upon further analysis, the teacher observes that the majority of students scored between 60 and 80, with a peak around 75. However, there are notable variations, with a few students achieving scores significantly higher or lower than the class average. Additionally, the distribution of scores appears to be slightly skewed towards the lower end, suggesting that a greater proportion of students performed below average compared to those who performed above average.

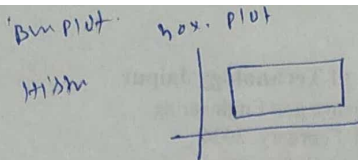
Discuss the overall performance of the class in terms of the center of the score distribution. How might the teacher use this information to plan future lessons or provide additional support to students? Consider the implications of both the average score and the presence of outliers on teaching strategies and student learning outcomes. **5 marks**

Q4. (*Random variables and Probability*) Considered two discrete random variables, A and B, with the following joint probability distribution:

	B=1	B=2	B=3
A=1	0.1	0.2	0.1
A=2	0.2	0.1	0.1
A=3	0.1	0.1	0.1

- Verify that this is a valid joint probability distribution.
- Calculate $P(A = 2 \cap B = 3)$.
- Determine the marginal probability distributions $P(A)$ and $P(B)$.
- Are A and B independent? Justify your answer.

4 marks



Q5. Description for the next three questions: a 2×2 matrix $M = \begin{bmatrix} p & q \\ r & s \end{bmatrix}$ defines a function from \mathbb{R}^2 to \mathbb{R}^2 given by-

$$\begin{bmatrix} p & q \\ r & s \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} pu + qv \\ ru + sv \end{bmatrix},$$

for each point $\begin{bmatrix} u \\ v \end{bmatrix} \in \mathbb{R}^2$.

- Determine the image of the x-axis under the function defined by the matrix $M_1 = \begin{bmatrix} 1 & 3 \\ 0 & 1 \end{bmatrix}$
- Determine the image of the y-axis under the function defined by the matrix $M_2 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$
- Determine the image of the line $y=x$ under the function defined by the matrix $M_3 = \begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix}$

3 marks

Q6. (EDA) Suppose you have the following dataset representing the final exam marks (out of 100) of a class of students: {75, 80, 85, 90, 95, 85, 90, 92, 88, 82, 78, 85, 88, 91, 83}. Create a box plot based on this data and then answer the following question:

- Identify the median (50th percentile) of the exam marks from the box plot. What does this value indicate about the central tendency of the class's performance?
- Determine the upper and lower quartiles (Q3 and Q1) from the box plot. What do these values represent in terms of the distribution of exam marks?
- Calculate the interquartile range (IQR) of the exam marks. What does the IQR signify about the spread or dispersion of the data?
- Estimate the approximate skewness of the distribution of exam marks based on the box plot. Is the distribution skewed to the left, right, or approximately symmetrical?

4 marks

Part II MCQs

6 marks

1. Which of the following research methodologies is primarily concerned with understanding human behavior and interactions within a specific context?

- A) Experimental research B) Qualitative research C) Quantitative research D) Action research

2. What is the primary aim of exploratory research in the context of computer science?

- A) To test hypotheses B) To establish causality C) To gain insights and understanding D) To generalize findings to a population

3. Which sampling technique involves selecting participants who are easily accessible and readily available for study?

- A) Stratified sampling B) Systematic sampling C) Convenience sampling D) Snowball sampling

4. A researcher is conducting a survey on internet usage habits among college students. Out of a total population of 5000 students, the researcher randomly selects 500 students to participate in the survey. What is the sampling fraction for this survey?

- A) 0.1 B) 0.01 C) 0.001 D) 0.0001
- Handwritten calculation: $\frac{500}{5000} = \frac{1}{10}$*

5. A computer program is tested on three different hardware configurations, and the execution times are recorded as follows: Configuration A: 10 seconds, Configuration B: 12 seconds, Configuration C: 15 seconds. What is the median execution time for these configurations?

- A) 10 seconds B) 12 seconds C) 13 seconds D) 15 seconds

6. In a study investigating the relationship between programming experience (in years) and software development productivity (measured in lines of code written per hour), the correlation coefficient is calculated to be -0.75. What does this value indicate about the relationship between programming experience and productivity?

- A) Strong positive correlation B) Weak positive correlation C) Strong negative correlation D) Weak negative correlation