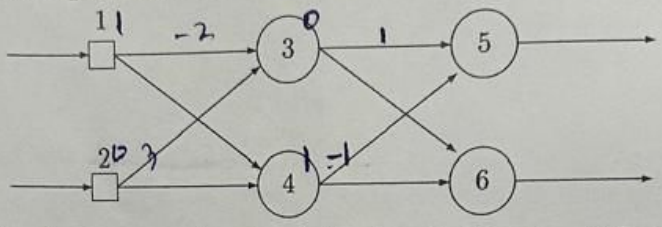
(C) What is problem with Tanh activation function? Why ReLU is better than Tanh activation function?

Q.34. **(A)** Explain K-Nearest Neighbor classifier. What is role of k in KNN? Explain with the help of example.

(B) How non-linearity is introduced in non-linear SVM? What is kernel trick in non-linear SVM?

Q.35. **(A)** The following diagram represents a feed-forward neural network with one hidden layer:



A weight on connection between nodes i and j is denoted by w_{ij}, such as w_{13} is the weight on the connection between nodes 1 and 3. The following table lists all the weights in the network:

$w_{13} = -2$	$w_{35} = 1$
$w_{23} = 3$	$w_{45} = -1$
$w_{14} = 4$	$w_{36} = -1$
$w_{24} = -1$	$w_{46} = 1$

Each of the nodes 3, 4, 5 and 6 uses the following activation function:

$$\varphi(v) = \begin{cases} 1 & \text{if } v \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

where v denotes the weighted sum of a node. Each of the input nodes (1 and 2) can only receive binary values (either 0 or 1). Calculate the output of the network (y_5 and y_6) when the input is 1 for Node 1 and 0 for Node 2. Consider no bias in the network.

(B) What will be the output size and no. of parameters in the following scenario:

a. Input volume: 30x30x256. Use Max pooling layer having 2x2 kernel size with stride 2 and pad 0.
b. Input volume: 128x128x3. Use Batch Normalization layer.
c. Input volume: 16x16x16. Use Conv layer having 8 5x5 filters with stride 2 and pad 0

B. ReLU
C. Sigmoid
D. Tanh

Q.25. What is the primary purpose of the reparameterization trick in a Variational Autoencoder (VAE)?

A. To minimize the reconstruction error
B. To enforce sparsity in the latent space
C. To enable efficient gradient estimation during training
D. To regularize the weights of the neural network

Q.26. You are solving the binary classification task of classifying images as cat vs. non-cat. You design a CNN with a single output neuron. Let the output of this neuron be z. The final output of your network, y' is given by: y' = Sigmoid(ReLU(z)). You classify all inputs with a final value y' ≥ 0.5 as cat images. What problem are you going to encounter?

A. All the inputs will be classified as cat.
B. There will not be any problem.
C. All the inputs will be classified as non-cat.
D. The cat inputs will be classified as non-cat and non-cat inputs will be classified as cat.

Q.27. What is the primary purpose of using Multi-headed Attention in Transformer models?

A. To reduce computational complexity by dividing attention across multiple heads
B. To increase model capacity and capture different types of relationships in the input sequence
C. To improve generalization performance on tasks with large datasets
D. To prevent overfitting during training

Q.28. How does Multi-headed Attention help improve the performance of Transformer models?

A. By reducing the model's memory requirements
B. By enabling the model to focus on different parts of the input sequence simultaneously
C. By increasing the model's ability to generate text
D. By speeding up the training process

Q.29. Which of the following is false for most CNN architectures?

A. Size of input (height and width) decreases, while depth increases
B. Multiple convolutional layers followed by pooling layers
C. Fully connected layers in the first few layers
D. Non-linearity layers are present with conv layers in the network

Q.30. Which of the following propositions are true about a CONV layer?

A. The number of weights depends on the depth of the input volume.
B. The number of biases is equal to the number of input channels.
C. The total number of parameters depends on the stride.
D. The total number of parameters depends on the padding.

PART-B

Q.31. Explain the training procedure of a VAE-GAN, including the joint optimization of the VAE objective and the GAN objective. How do the encoder, generator, and discriminator networks interact during training to learn meaningful representations and generate realistic samples?

Q.32. Define what Large Language Models (LLMs) are and how they are trained. Discuss the architecture and key components of LLMs, including self-attention mechanisms and transformer networks.

Q.33. (A) How are LLMs trained on vast amounts of text data to learn contextual representations and generate coherent text?

(B) Why dropout can be used for regularization? What is inverted dropout?

5. Consider an input volume of size 32×32×3. If you have 5 filters of spatial dimension 7×7 in a conv layer applied with stride 1 and padding 3, then what will the size of output activation map?
   A. 32×32×5
   B. 32×32×7
   C. 29×29×5
   D. None

Q.16. Which statement is false?
   A. Traditional machine learning approaches learn very deep layers of model.
   B. Nearest neighbor works for any number of classes.
   C. In practice, linear regression is very sensitive to outliers.
   D. CNN exploits the local connectivity and weight sharing to extract features from images.

Q.17. Which statement is true for a neural network?
   A. Neural network can not handle the non-linearity
   B. More number of neurons increases the capacity of the network
   C. Neural network is same as the SVM
   D. Neural network is made with only one hidden layers

Q.18. How does VAE-GAN address the mode collapse issue observed in traditional GANs?
   A. By minimizing the KL divergence between the encoded distribution and a prior distribution
   B. By using a combination of reconstruction loss and adversarial loss
   C. By introducing additional regularization terms in the loss function
   D. By reducing the capacity of the generator network

Q.19. What is the primary advantage of using both VAE and GAN components in *VAE-GAN*?
   A. Faster training convergence
   B. Enhanced capability to model complex data distributions
   C. Reduced computational resources required
   D. Simplified architecture design

Q.20. What is the purpose of fine-tuning a pre-trained Large Language Model (LLM) on a specific task?
   E. To improve the model's understanding of complex syntax
   A. To adapt the model to the nuances and specifics of the task
   B. To increase the model's vocabulary size
   C. To reduce the risk of overfitting

Q.21. Which of the following techniques is commonly used to prevent catastrophic forgetting during fine-tuning of LLMs?
   A. Gradient clipping
   B. Knowledge distillation
   C. Dropout regularization
   D. Early stopping

Q.22. What is the significance of the attention mechanism in Large Language Models (LLMs)?
   A. It allows the model to focus on relevant parts of the input sequence
   B. It helps in reducing the computational complexity of the model
   C. It improves the model's ability to handle long-range dependencies
   D. It facilitates better interpretation of the model's decisions

Q.23. Consider an input volume of size 32×32×3. If you have 5 filters of spatial dimension 5×5 in a conv layer applied with stride 1 and padding 3, then how many number of parameters are present in this layer excluding bias term?
   A. 25
   B. 125
   C. 375
   D. None

Q.24. Which of the following functions can be used as the class score layer to generate the probabilities for n classes (p1, p2, ..., pn) such that sum of p over all n equals to 1?
   A. Softmax

B. It is continuous and smooth
C. It is discrete and sparse
D. It has no relation to the input data space

Q.7. What is the primary limitation of traditional GANs in terms of training stability?

A. Mode collapse
B. Vanishing gradients
C. Slow convergence
D. Unstable discriminator behavior

Q.8. Consider an input volume of size 32×32×3. If you apply max pooling with kernel size 2×2 at stride 2 without padding, then what will the size of output activation map?

A. None
B. 31×31×3
C. 16×16×2
D. 16×16×3

Q.9. Which of the following options lists only the activation functions?

A. ReLU, Weight Decay, Leaky ReLU
B. Sigmoid, ReLU, Swish
C. Maxout, Tanh, Dropout
D. Xavier Initialization, Batch Normalization, ReLU

Q.10. Which loss function is typically used to train a GAN?

A. Mean Squared Error (MSE)
B. Mini-Max
C. KL Divergence
D. Binary Cross-Entropy

Q.11. In a GAN, what is the role of the generator?

A. To discriminate between real and generated samples
B. To generate data samples that resemble real data
C. To maximize the likelihood of generating real samples
D. To minimize the difference between real and generated samples

Q.12. Consider one layer of weights in a convolutional neural network (CNN) for grayscale images, connecting one layer of nodes to the next layer of nodes. Which type of layer has the highest number of parameters to be learned during training? Choose only one most suitable answer.

E. A convolutional layer with 8 filters of spatial dimension 3×3.
F. A max-pooling layer that reduces a 10×10 image to 5×5.
G. A convolutional layer with 6 filters of spatial dimension 5×5.
H. A fully connected layer from 20 hidden units to 4 output units.

Q.13. Suppose we have a 3-dimensional input $\vec{x} = (x_1, x_2, x_3)$ connected to 4 neurons $\vec{y} = (y_1, y_2, y_3, y_4)$ with the exact same weights $\vec{w} = (w_1, w_2, w_3)$ where: $x_1 = 2, w_1 = 1, x_2 = -1, w_2 = -0.5, x_3 = 1, w_3 = 0$, and the bias $b = 0.5$. We calculate the output of each of the four neurons using the input $\vec{x}$, weights $\vec{w}$ and bias b. If $y_1 = 0.95, y_2 = 3, y_3 = 1, y_4 = 3$, then the valid guesses for the neuron types of y_1, y_2, y_3 and y_4 are:

I. Rectified Linear, Logistic Sigmoid, Binary Threshold, Linear
J. Logistic Sigmoid, Rectified Linear, Binary Threshold, Rectified Linear
K. Logistic Sigmoid, Tanh, Binary Threshold, Rectified Linear
L. Rectified Linear, Linear, Binary Threshold, Logistic Sigmoid

Q.14. What is a common approach to combining VAE and GAN in VAE-GAN architecture?

A. Training VAE and GAN separately and then combining their outputs
B. Training VAE and GAN simultaneously with a shared latent space
C. Using VAE for feature extraction and GAN for data generation
D. Using GAN for feature extraction and VAE for data generation

Indian Institute of Information Technology Allahabad
End Sem/C3 Question Paper

Course Name: Deep Learning **Course Code:** DEL3001/PC-IT-DEL506/PC-IT-DEL601
Course Instructor: Prof. G. C. Nandi & Dr Satish K. Singh
Program Name(s): B.Tech/M.Tech/PhD **Exam Date:** 9th May 2024 **MM:** 60

Note:

1. *This question paper contains **TWO** parts i.e **Part-A** and **Part-B**.*
2. *Part-A contains **THIRTY** objective type questions of **1 mark** each.*
3. *Part-B contains **FIVE** subjective type questions of **6 marks** each.*
4. *All questions are **compulsory**.*

PART-A

Q.1. Why is it important to place non-linearities between the layers of deep neural networks?

A. It reduces the complexity of the model.
B. It reduces the degrees of freedom to the model.
C. It is not important to use the non-linearity in a deep neural network.
D. A deep neural network without non-linearities is essentially a linear regression.

Q.2. Which probability distribution is typically assumed for the latent variables in a Variational Autoencoder (VAE)?

A. Uniform distribution
B. Gaussian distribution
C. Exponential distribution
D. Poisson distribution

Q.3. In a VAE, what is the purpose of the "reconstruction loss"?

A. To maximize the likelihood of generating data
B. To minimize the difference between the input and the reconstructed output
C. To regularize the latent space
D. To train the discriminator network

Q.4. You are applying batch normalization to a fully connected (dense) layer with an input size of 10 and output size of 20. How many training parameters does this layer have, including batch normalization parameters? Note that the dense layer does not has bias term.

A. 200
B. 220
C. 240
D. None

Q.5. You're asked to build an algorithm estimating the risk of premature birth for pregnant women using ultrasound images. You have 500 examples in total, of which only 175 were examples of preterm births (positive examples, label = 1). To compensate for this class imbalance, you decide to perform the augmentation. Which of the following augmentation technique will be most suitable among the four choices?

A. Duplicate all of the positive examples, and then split the data into train, validation and test sets.
B. Duplicate all of the negative examples, and then split the data into train, validation and test sets.
C. Split the data into training, validation and test sets and then duplicate all of the positive examples of all sets.
D. Split the data into training, validation and test sets and then duplicate all of the positive examples of only training set.

Q.6. Which of the following statements about the latent space of a VAE is true?

A. It is always low-dimensional

Indian Institute of Information Technology Allahabad
End Sem Question Paper

Course Name: Physical Education (Sports) **Course Instructor/ Co-ordinator:** Dr. B S Bhushan

Course Code: HM-XX- PHE **Program Name(s):** BTech, MTech, MBA sem-2 **Exam Date:** 01-05-2024 **MM:** 40

MML2023005

Important Instructions:

- **Present your answers using bullet points and avoid lengthy paragraphs.**
- Keep your answers as short and precise as possible.
- There will be a penalty for very lengthy, mystifying, and unclear handwriting.
- Answer all the 5 questions. Each question carries 8 marks.

1) (a) What is the full form of CPR, and how to perform the same in a medical emergency ?

   (b) Explain the benefits of regular swimming.

2) (a) What are the different types of Diabetes and explain the reasons of diabetes ?.

   (b) Write a brief note on the Diabetic condition of India and world.

   (c) Explain the importance of physical activities in controlling the diabetic epidemic.

3) What are the psychological attributes that make MS Dhoni so special in comparison to other cricketers ?. What can you learn from them ?

4) Explain the basic rules of

   (a) Javelin Throw

   (b) Discus Throw

   (c) Shot put

5) (a) What is Yoga ?.

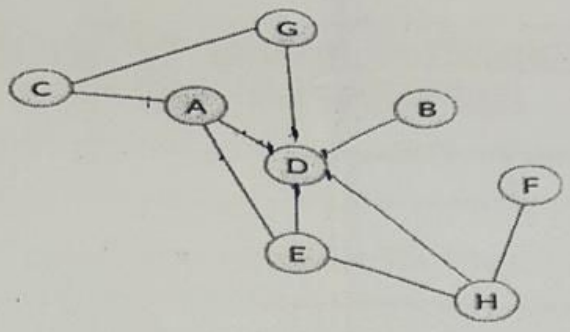
   (b) Explain the procedure of performing the following yoga asanas for concentration and their benefits.

   (i) Budhapadmasana

   (ii) Ardhamatsyendrasana

   (iii) Shalabhasana

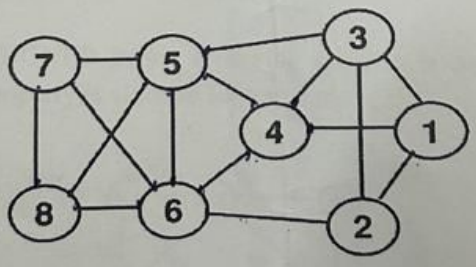
   (iv) Naukasana



9. Assume there are two communities in this network: {A, B, C, D, G} and {E, F, H}. Calculate
   a) Common Neighbor Soundarajan-Hopcroft score of node A and node G **[2.5]**
   b) Resource Allocation Soundarajan-Hopcroft score of node A and node G **[2.5]**

10. A political party approaches you with the network of an online social network. Links in the network are undirected, and and edge (u, v) indicates that user u and v have tweeted about the same topic and also follow each other. The party wants to know what are the topical clusters in this network so that they know how to frame the manifesto to attract all types of voters. State an algorithm that would work here and why? What are this algorithms pros and cons. **[3]**

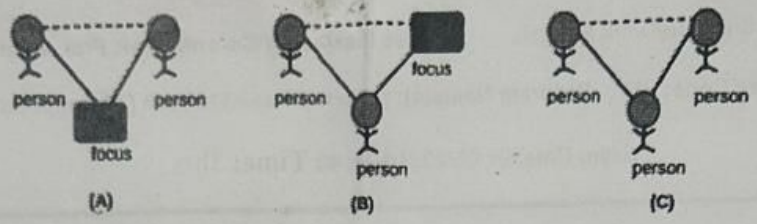
11. Find the number of best possible community in given network and calculate the modularity score of the communities using **[6:3 ea]**
    (a) modularity maximization and
    (b) complete clustering linkage.



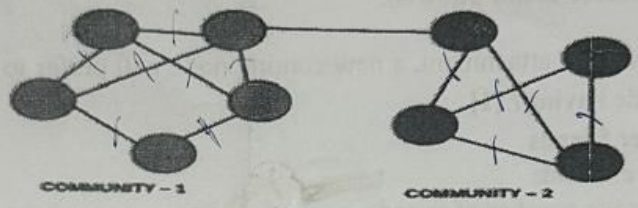
12. Explain any two from the following with examples: **[6:3ea]**
    (a) Minimum Connected Dominating Set (MCDS)
    (b) Information cascades
    (c) Link partition

***

5. Consider figures A, B and C in Fig and write and explain the right kind of closure they represent (Please note that the solid line represents the existing friendship and, the dotted line represents the new friendship) [2]

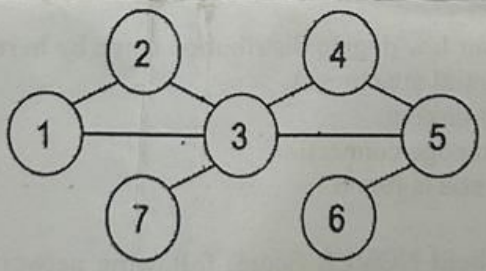


6. What are the densities of communities 1 and 2 in the network shown in Figure? **[2]**



7. For the network shown in the Figure, calculate the following similarity measures of nodes 3 and 5. **[6: 2ea]**

   a) Jaccard similarity
   b) Preferential attachment
   c) Adamic Adar



Use the network below for both Questions 8 and 9

8. **Calculate the following:**

   a) Suppose we want to apply the common neighbours' measures to add an edge from node H. Which is the most probable node to connect to H and why? **[2]**

   b) if we apply Resource Allocation method to predict the new edges, what is the value of the Allocation index of node C and D? **[3]**

and the discriminator? Discuss the adversarial training process between these networks. 3+3+4=10

b) Describe the training algorithm of a GAN, including the alternating optimization process. How do the generator and discriminator update their parameters during training to improve their respective objectives? Discuss the challenges and potential solutions in stabilizing GAN training. 3+2+2=7

c) Discuss the process of generating new samples using a trained GAN. How does the generator network produce realistic data samples, and how can the quality of generated samples be evaluated? Compare and contrast the testing phase of a GAN with that of autoencoders. 3+2+3=8

d) Explore two practical applications of GAN. Explain in detail how you used pix2pix GAN for data generation. 1+9=10

****

d) It speeds up the training process.

13. What is the primary objective of an Autoencoder?
a) Classify input data into different categories
b) Maximize the likelihood of generating data
c) Minimize the reconstruction error between input and output
d) Minimize the KL divergence between the encoded distribution and a prior distribution

14. Which of the following is a characteristic feature of the latent space in an Autoencoder?
a) It is always high-dimensional
b) It is continuous and smooth
c) It is fixed and cannot be modified during training
d) It has no relation to the input data space

15. What is the purpose of the decoder component in an Autoencoder?
a) To encode the input data into a low-dimensional representation
b) To generate new data samples from random noise
c) To reconstruct the input data from the encoded representation
d) To discriminate between real and generated samples

Group-B
(Full Marks in this section=70)

16. a) Draw and explain the fundamental architecture and components of an autoencoder. How does it transform input data into a latent space representation? Describe the role of the encoder and decoder in the process. 2+3+2+3=10

b) Outline the training process of an autoencoder using backpropagation and gradient descent. Discuss the objective function used for training and how it encourages the model to learn meaningful representations. 3+3+2+2=10

c) Describe the process of reconstructing input data using a trained autoencoder. How does the decoder transform the latent space representation back into the original input space? Discuss any reconstruction evaluation metrics that can be used to assess the quality of reconstructions. 2+2+2+2=8

d) Explore two distinct applications of autoencoders in real-world scenarios. How can autoencoders be used for dimensionality reduction and feature learning in unsupervised learning tasks? Provide examples of domains where autoencoders have shown Significant utility. 2+3+2=7

17 a) Draw and explain the core principles of a Generative Adversarial Network (GAN) architecture. How does a GAN consist of two neural networks, namely the generator

6. In a GAN, what is the training objective of the discriminator?
   a) Minimize the difference between real and generated samples
   b) Maximize the difference between real and generated samples
   c) Maximize the likelihood of generating real samples
   d) Minimize the likelihood of generating real samples

7. What is the primary advantage of combining VAE with GAN?
   a) Improved reconstruction accuracy
   b) Faster training convergence
   c) Better control over generated samples' diversity
   d) Enhanced ability to model complex data distributions

8 Which of the following is a characteristic feature of VAE-GAN compared to VAE or GAN alone?
   a) It has only an encoder and generator, but no decoder.
   b) It uses a combination of reconstruction loss and adversarial loss.
   c) It does not have a latent space representation.
   d) It requires less computational resources.

9. In VAE-GAN, what is the role of the adversarial loss?
   a) It encourages the encoder to produce more diverse latent representations.
   b) It helps the decoder to better reconstruct the input data.
   c) It forces the generator to produce samples that are more similar to the true data distribution.
   d) It reduces the mode collapse problem in generating diverse samples.

10.What is the primary task of Large Language Models (LLMs)?
   a) Image classification
   b) Text generation and understanding
   c) Speech recognition
   d) Reinforcement learning

11 .Which of the following is a commonly used architecture for Large Language Models?
   a) LSTM (Long Short-Term Memory)
   b) CNN (Convolutional Neural Network)
   c) Transformer
   d) ResNet (Residual Neural Network)

12. What is the main advantage of pre training Large Language Models on large text corpora?
a) It reduces computational resources required for fine-tuning.
b) It enables the model to capture rich semantic representations.
c) It eliminates the need for fine-tuning on specific tasks.

MML 2023005

Indian Institute of Information technology, Allahabad
QUIZ test
Deep Learning
April, 2024

Full Marks-100 **Time -2hrs**

Group-A

(Answer the following objective type questions with justifications. Each question carries 2 marks -1 for correctly selecting the option and another 1 for giving justification)

Full marks in this section-30

1. What is the primary objective of a Variational Autoencoder (VAE)?
   a) Minimize reconstruction error
   b) Maximize likelihood of generating data
   c) Minimize the KL divergence between the encoded distribution and a prior distribution
   d) Maximize feature extraction

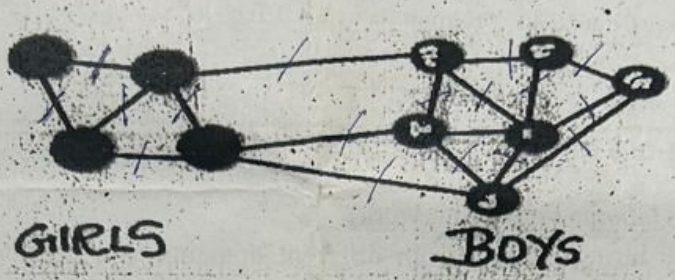
2. Which of the following is NOT a component of a VAE?
   a) Encoder
   b) Decoder
   c) Discriminator
   d) Latent space

3. What is the role of the reparameterization trick in training a VAE?
   a) It helps in generating diverse samples
   b) It enables efficient gradient estimation
   c) It reduces the computational complexity of the model
   d) It increases the model's capacity

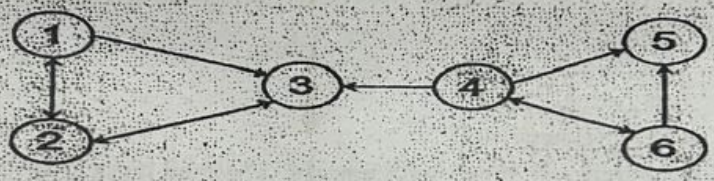
4. What is the main idea behind Generative Adversarial Networks (GAN)?
   a) Minimize the reconstruction error between input and output
   b) Minimize the KL divergence between the generated distribution and the target distribution
   c) Train a generator to generate data that cannot be distinguished from real data by a discriminator
   d) Maximize the likelihood of generating data samples

5. Which component of a GAN is responsible for generating new data samples?
   a) Generator
   b) Discriminator
   c) Encoder
   d) Decoder

7. Suppose a network of 10 students of classroom in which 4 are girls and 6 are boys. The edge represents the friendship among them as shown in Figure. Does the network exhibit homophily? If yes then to what extent does it exhibit? **[4 Marks]**

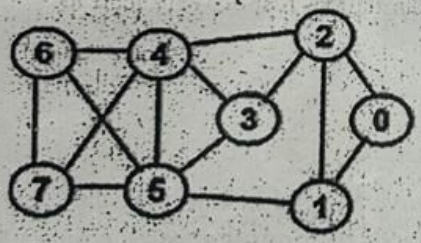


8. Calculate the page rank of the given network as shown in the figure. Let Assume that the initial rank of node 1, 2, 3, 4, 5, and 6 are 4, 3, 6, 1, 5, 2 respectively i.e.
P0= [4, 3, 6, 1, 5, 2]. **[5 Marks]**

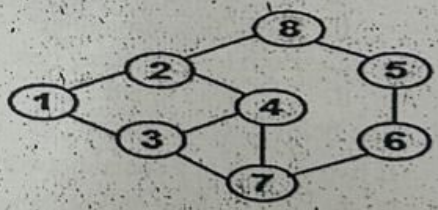


Markov chain (matrix)

9. **Calculate the Eigenvector centrality of the given networks (a) and (b).** **[8 Marks]**

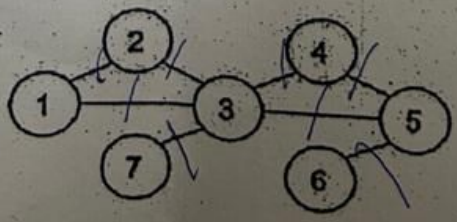


(a)



(b)

10. Calculate the following of the given networks : **[ea: 1Mark]**
    (a) Diameter of the network
    (b) Density of the network
    (c) Degree distribution of the network
    (d) Closeness Centrality for nodes [2, 3, 6]
    (e) Betweenness Centrality for node [3, 4]



Indian Institute of Information Technology Allahabad

Mid Sem Question Paper

Course Name: Social Network Analysis **Course Instructor/ Co-ordinator:** Prof. Vrijendra Singh

Course Code: SNA **Program Name(s):** B.Tech (6th sem)/M.Tech (1st sem)/Phd

Exam Date: 28/02/2024 **MM:** 25 **Time:** 2hrs

Note: Question 1 to 6 carry 0.5 marks each

1. How does Google Page Rank work?
   a. By hiring experts from different domains who maintain a database of the rankings of all web pages.
   b. Using machine learning and natural language processing.
   c. Using web graph and random walk algorithm.
   d. Using web graph and breadth first traversal.
2. The expected number of edges between nodes i and j in a network is given as:
   a. (deg(i)deg(j))/|E|
   b: (deg(i)deg(j))/(2|E|)
   c. (deg(i)deg(j))/|E|2
   d. (2*deg(i)deg(j))/|E|
3. Let G be a complete graph with n nodes. The total number of triangles possible in G is:
   a. n^3 b. (n (n-1) (n-2))/6 c. 3n d. none of the above
4. In social networks, friends and acquaintances respectively lead to:
   a. Strong ties, weak ties
   b. Weak ties, strong ties
   c. Both lead to strong ties
   d. Both lead to weak ties
5. The below mentioned principle is referred as:
   If two people in a social network have a friend in common, then there is an increased likelihood that they will become friends themselves at some point in the future.
   a. Structural holes b. Social capital c. Triadic closure d. None of the above
6. Homophily refers to the friendship between people
   a. who have a common friend
   b. who have same ethnicity
   c. who are similar to each other
   d. who are dissimilar to each other

Indian Institute of Information Technology Allahabad

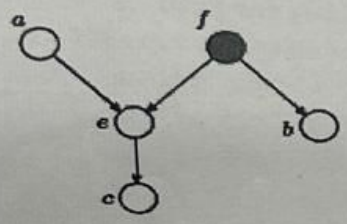
Mid Sem Question Paper

Course Name: Probabilistic Machine Learning and Graphical Model **Co-ordinators: SHV/KPS** **Exam Date:** 28/02/2024

Course Code: PC-IT-PGM507 **Program Name(s):** MTech 2nd Sem &B.Tech 6th Sem (IT, IT-BIN)

[Maximum Marks is 25] **[Duration 2 Hours]**

[Q1] Prove that $a \perp b|f.$, **ie.,** a is independent of b given f. Also prove that $a \perp c|e,.$ ie, a is independent of c given a. **[4+3]**



[Q2] The data observed during an experiment is noisy. The noise is additive gaussian noise with zero mean and a variance, sigma (d). Linear regression is performed to determine the distribution of the input output relation. Determine the parameters of the distribution. **[7]**

[Q3] Find the discriminant function that separates two classes of data with mean m1 and m2. The data points are not independent. Also determine the expression for Fisher discriminant. **[6]**

[Q4] Suppose that X is a discrete random variable with the following probability mass function: where $0 \leq \theta \leq 1$ is a parameter. The following 10 independent observations

X	0	1	2	3
P(X)	2θ/3	θ/3	θ/3	(1 − θ)/3

were taken from such a distribution: (3,0,2,1,3,2,1,0,2,1). What is the maximum likelihood estimate of θ? **[5]**

Q3. Short Answers **[5 Marks]**

(a) **[1 Mark]** Compare Softmax classifier with SVM classifier.

(b) **[2 Marks]** Explain kernel-trick in SVM classifier with diagrams and examples.

(c) **[2 Marks]** Consider the convolutional neural network defined by the layers in the left column below. Fill in the shape of the output volume and the number of parameters at each layer. You can write the shapes in the numpy format (e.g. (128,128,3)). Notation:

- CONV5-N denotes a convolutional layer with N filters with height and width equal to 5. Padding is 2, and stride is 1.
- POOL2 denotes a 2x2 max-pooling layer with stride of 2 and 0 padding.
- FC-N denotes a fully-connected layer with N neurons

Layer Activation	Volume Dimensions	Number of parameters
Input	(32, 32, 1)	0
CONV5-10		
POOL2		
CONV5-10		
FC-10		

Q4. Movie Posters **[10 Marks]**

You have been tasked by a company to build a deep learning model to help them decide what movie to watch. Specifically you're asked to build a classifier that takes in an image of a movie poster and classifies it into one of four genres: comedy, horror, action, and *romance*. You have been provided with a large dataset of movie posters where each movie poster corresponds to a move with exactly one of these genres. You now estimate that the human level performance on this task is 95%. You decide to use cross entropy loss $L_{CE}(\hat{y}, y)$ to train your network, where $\hat{y} = (\hat{y}_1, \hat{y}_2, \hat{y}_3, \hat{y}_4)^T$ represents the predicted probability distribution over the *comedy*, *horror*, *action*, and *romance* classes, respectively and $y = (y_1, y_2, y_3, y_4)^T$ represents the ground truth vector, which is zero everywhere except for the correct class (e.g. $y = (1, 0, 0, 0)^T$ for *comedy*, and $y = (0, 0, 1, 0)^T$ for *action*).

(a) **[2 Marks]** Suppose you're given an example poster of a *horror* movie. If the model correctly predicts the resulting probability distribution as $\hat{y} = (0.1, 0.4, 0.3, 0.2)^T$, what is the value of the cross-entropy loss? You can give an answer in terms of logarithms.

(b) **[2 Marks]** After some training, the model now incorrectly predicts *romance* with distribution $\hat{y} = (0, 0.4, 0, 0.6)^T$ for the same poster. What is the new value of the cross-entropy loss for this example?

(c) **[2 Marks]** You train an initial model that achieves only a 75% accuracy on the training dataset. What kind of problems is your model experiencing, and suggest a possible solution.

(d) **[2 Marks]** After tuning the model architecture, you find that the softmax classifier works well. Specifically, the last layer of your network computes logits $z = (z_1, z_2, z_3, z_4)$, which are then fed into the softmax activation. The model achieves 100% accuracy on the training data. However, you observe that the training loss doesn't quite reach zero. Show why the cross-entropy loss can never be zero if you are using a softmax activation.

(e) **[2 Marks]** While the model does well on the training set, it only achieves an accuracy of 85% on the validation set. You conclude that the model is overfitting, and plan to use L1 or L2 regularization to fix the issue. However, before you can do so, you learn from your class mate that some of the examples in the data may be incorrectly labeled. Which form of regularisation would you prefer to use and why?

Indian Institute of Information Technology Allahabad

Mid Sem Question Paper

Course Name: Deep Learning **Course Instructor/ Co-ordinator:** SKS

Course Code: PC-IT-DEL506 **Program Name:** BTech/MTech/PhD **Exam Date:** 26.02.24 **MM:** 25

Q1. MCQ: Write only one most suitable answer in each question **[5 Marks]**

a) **[1 Mark]** Your model for classifying different dog species is getting a high training set error. Which of the followings are promising things to try to improve your classifier?

(i) Use a bigger neural network (ii) Get more training data
(iii) Use data augmentation (iv) Increase the regularization parameter lambda

b) **[1 Mark]** Which of the followings are false about Batch Normalization?

(i) Batch Norm layers are skipped at test time because a single test example cannot be normalized.
(ii) Its learnable parameters can only be learned using gradient descent or mini-batch gradient descent.
(iii) It helps speed up learning in the network.
(iv) It introduces noise to a hidden layer's activation, because the mean and the standard deviation are estimated with a mini-batch of data.

c) **[1 Mark]** Which one of the following statements on initialization is false?

(i) The variances of layer outputs remain unchanged during training when Xavier initialization is used.
(ii) Initializing all weights to a positive constant value isn't sufficient to break learning symmetry during training.
(iii) Different activation functions may benefit from different types of initializations.
(iv) It's possible to break symmetry by initializing weights to be sampled uniformly from the set $\{-1, 1\}$.

d) **[1 Mark]** If your input image is 64x64x16, how many parameters are there in a single 1x1 convolution filter, including bias?

(i) 2 (ii) 17 (iii) 4097 (iv) 1

e) **[1 Mark]** Which of the below can you implement to solve the exploding gradient problem?

(i) Use SGD optimization (ii) Oversample minority classes
(iii) Increase the batch size (iv) Impose gradient clipping

Q2. Short Answers **[5 Marks]**

(a) **[1 Mark]** Explain using example, what is overfitting in deep learning?
(b) **[1 Mark]** What are different ways to tackle overfitting and how it tackles?
(c) **[1 Mark]** What is problem with Tanh activation function? Why ReLU is better than Tanh activation function? Justify your answer.
(d) **[1 Mark]** What is bias correction in Adam optimizer?
(e) **[1 Mark]** Explain Xavier Weight Initialization.

Indian Institute of Information Technology Allahabad

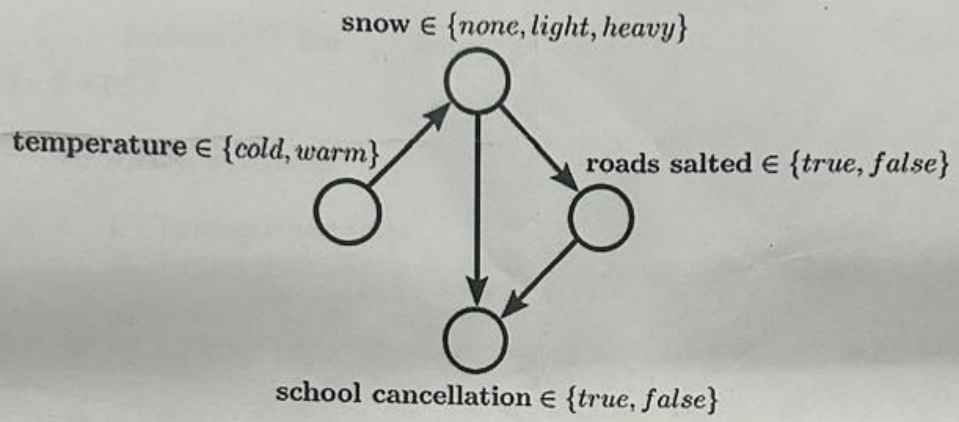
End Sem Question Paper

Course Name: Probabilistic Machine Learning and Graphical Model **Co-ordinators:** SHV/KPS **Exam Date:** 8.5.24

Course Code: PC-IT-PGM507 **Program Name(s):** MTech 2nd Sem &B.Tech 6th Sem (IT, IT-BIN)

[Maximum Marks is 40] [Duration 3 Hours]

[Q1]



Above is depicted a graphical model with four discrete random variables that can be used to predict whether school will be closed due to inclement weather.

a. Answer the following questions about the conditional independence structure in the model: [2+2+2+2]
   i. Which variables are independent of temperature given that snow is observed?
   ii. Which variables are independent of snow given that no variables are observed?
   iii. Which variables are independent of snow given that temperature is observed?
   iv. Which variables are independent of school cancellation given that snow and roads salted are observed?

b. The joint probability is given by p(temperature; snow; roads salted; school cancellation). Write the factorized form of the joint probability (as a product of simpler probabilities) for the model above. [2]

[Q2] Consider a Gaussian Mixture Model. Develop a learning model using latent variables. Write likelihood estimation of the parameters and EM algorithm to solve it. How we can use this model for outlier detection. [8]

[Q3] A latent variable deep generative model is (usually) just a model that turns random numbers into valid samples. instead of training an encoder, can we just train the whole model to generate images that look similar to real images at the population level? What kind of model will be able to achieve this? Can we just employ a model to learn to distinguish between a real sample from a

3. Answer all the questions $[3 \times 5 = 15]$

(a) Draw the ReLU function. Prove that it is a nonlinear function. Find out a point where the ReLU function is discontinuous. Argue in support of your claim.

(b) Construct the adjacency matrix for the below graph. Quantify the degree centrality and closeness centrality for the node B.

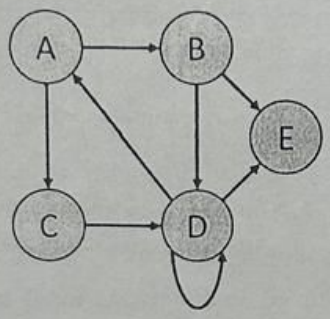


Figure 1: Directed graph.

(c) Plot the following functions:

i. Sinc function
ii. Heaviside step function
iii. Signum function
iv. Sigmoid function
v. Dirac delta function

Indian Institute of Information Technology Allahabad

Mid Sem Question Paper

Course Name: Introduction to Machine Learning **Course Instructor/ Co-ordinator:** Dr. Anjali Gautam

Course Code: PC-IT-IML503 **Program Name(s):** M.Tech (IT) MRH, MSD, DD M.Tech,-PhD(IT)MRH, Ph.D(IT) including WP

Exam Date: 19/10/2023.. **MM:** 25

Note: All questions are compulsory.

1. Suppose that the two variables x and z are statistically independent. Show that the mean and variance of their sum satisfies [4 Marks]

$$E[x + z] = E[x] + E[z]$$
$$var[x + z] = var[x] + var[z]$$

2. Show that the '$tanh$' function and the logistic sigmoid function $\sigma(a) = \frac{1}{1+\exp(-a)}$ are related by

$$\tanh(a) = 2\sigma(2a) - 1$$

Hence show that a general linear combination of logistic sigmoid functions of the form

$$y(x, w) = w_0 + \sum_{j=1}^{M} w_j \sigma\left(\frac{x - \mu_j}{s}\right)$$

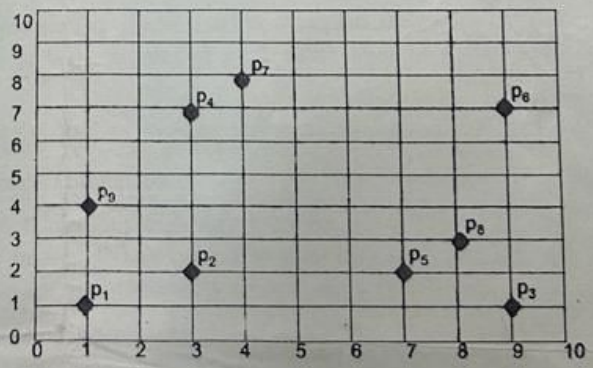
is equivalent to a linear combination of '$tanh$' functions of the form

$$y(x, u) = u_0 + \sum_{j=1}^{M} u_j \tanh\left(\frac{x - \mu_j}{s}\right)$$

and find expressions to relate the new parameters $\{u_1, \ldots\ldots, u_M\}$ to the original parameters $\{w_1, \ldots\ldots, w_M\}$. [4 Marks]

3. Drive the decomposition of expected test error $E_{x,y,D}[(h_D(x) - y)^2]$ in the form of variance, noise and bias. D is the dataset $D = \{(x_1, y_1), \ldots\ldots, (x_n, y_n)\}$ drawn independent and identically distributed data from some distribution $P(X, Y)$, where x is the given input, y is the label and h_D is the hypothesis (or classifier). [4 Marks]

4. Consider the following sample data: [7 Marks]



Point	x	y
p_1	1	1
p_2	3	2
p_3	9	1
p_4	3	7
p_5	7	2
p_6	9	7
p_7	4	8
p_8	8	3
p_9	1	4

Now, create a distance matrix by computing distance between each pair of points. Thereafter, use the concept of agglomerative clustering to create clusters. Use single, complete and average linkage approach to create clusters. Show the final dendrograms of each case.

5. Consider a learning problem with 2D features. How are the decision tree and 1-nearest neighbor decision boundaries related? [2 Marks]

6. Consider a simple classification problem: there is one feature x with values in $\mathbb{R}$, and class y can be 1 or -1. You have 2 data points: $(x_1, y_1) = (1,1)$, $(x_2, y_2) = (-1, -1)$.
For this problem write down the quadratic programming problem for an SVM with slack variables and Hinge loss. Denote the weight for the slack variables C, and let the equation of the decision boundary be

$$wx + b = 0$$

[4 Marks]

Indian Institute of Information Technology Allahabad

Mid / End Sem Question Paper

Course Name: Advanced Programming Practices **Course Instructor/ Co-ordinator:** Dr. B S Sanjeev

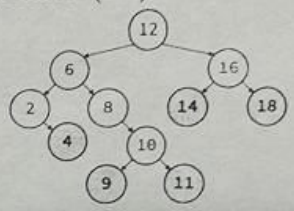
Course Code: PC-IT-APP502 **Program Name(s):** MTech (IT) **Exam Date:** 13.12.2023 **MM:** 20

Answer all the questions

Note: Irrelevant (not wrong) answers carry penalties.

1. Answer the following:
   a. What are the operations supported by the stack ADT? **(1m)**
   b. Using proper figures, explain queue constructed from two stacks. Internals of stack are not available for any operation, only the stack operations can be performed. **(3m)**

2. Consider the given tree:

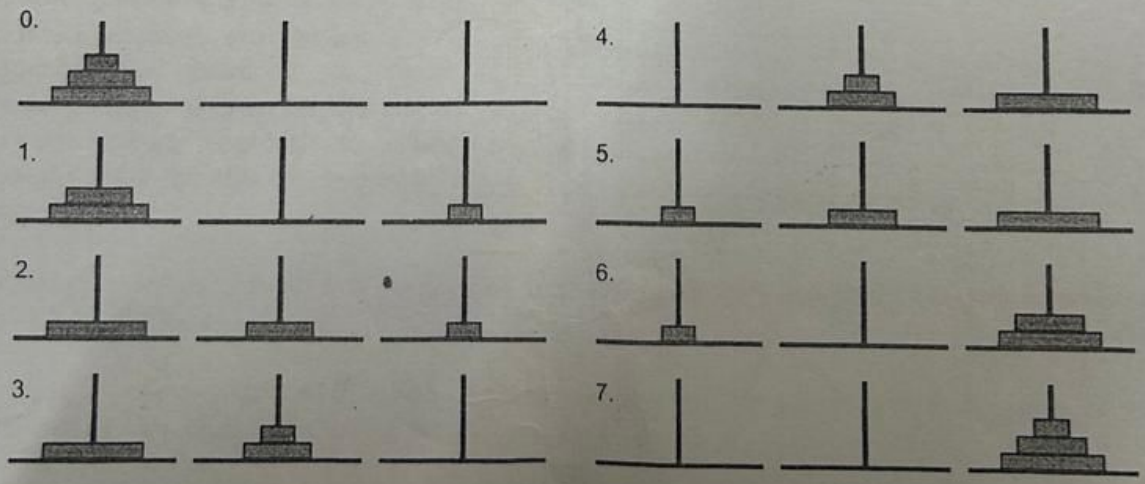


   a. Using the tree, explain how to delete root in BST. **(1m)**
   b. Write a function to print post-order output. **(1m)**
   c. Write the output of post-order search for the given tree. **(2m)**

3. You are given the sequence of integers 10, 9, 8, 7, 6, 1, 2, 3, 4, 5.
   a. Construct a B-Tree that can have at most 3-children with the aid of diagrams. **(2m)**
   b. What is the time complexity of search operation in B-tree? **(1m)**
   c. What is the time complexity of delete operation in B-tree? **(1m)**

4. You are to construct a dynamic table (T) with an initial size of 1.
   a. Show how the table changes when integers from 1, 2, …, 10 are inserted. **(1m)**
   b. Prove that the time complexity for insertion is $O(1)$. **(3m)**

5. Towers of Hanoi is a problem that involves disks of decreasing radii that are placed at the top in a peg (ie, a holder) as shown below. Solution to the problem requires moving only one disk at a time, ***never keeping a larger disk over a smaller one***, and yet shifting all disks from the left peg to right peg. An additional central peg can be used for the transfer. Write a recursive function (only pseudocode) to solve the problem along with useful figures to show that it works. **(4m)**



[Q 3] Image Compression **[8 Marks]**

(a) **[3 Marks]** Explain the concept of spatial redundancy, spectral redundancy and psychovisual redundancy in context of image compression with example.

(b) **[3 Marks]** Explain the role of entropy coding in image compression methods and how it reduces the redundancy in data.

(c) **[2 Marks]** Consider the following probabilities & codes assigned to a 4-symbols based image. Compute the average code word length for above coding scheme.

Symbol	Probability of Occurrence in an Image	Codes Assigned
a_1	0.24	0
a_2	0.46	10
a_3	0.12	110
a_4	0.18	111

[Q 4] Image Segmentation **[8 Marks]**

(a) **[2 Marks]** Explain how second order derivative can be used for edge detection with suitable diagram/example.

(b) **[2 Marks]** Explain the Otsu's thresholding method for image segmentation. In which scenario, the Otsu's thresholding method fails.

(c) **[2 Marks]** Explain when multi-level thresholding is useful for segmentation with diagram/example.

(d) **[2 Marks]** Explain the watershed segmentation method with diagram.

[Q 5] Image Restoration **[8 Marks]**

(a) **[2 Marks]** What is band reject filters? Explain its application in image restoration.

(b) **[2 Marks]** Explain the adaptive filtering and Homomorphic filtering with diagram/example.

(c) **[2 Marks]** What is periodic noise and how to remove periodic noise from image?

(d) **[2 Marks]** How to remove motion blurring? Give justification for your answers.

MRM2023026

Indian Institute of Information Technology Allahabad

End Sem Question Paper

Course Name: Image and Video Processing **Course Coordinator:** Dr. Satish Kumar Singh
Course Code: PC-IT-IVP504 **Program Name(s):** MTech **Exam Date: Dec 16, 2023**

Max. Marks: 40 **Exam Duration:** 3 Hours

Important Instructions: (i) All questions are compulsory, (ii) Students should note that it is a closed book examination and hence using electronic gadgets or any other copying material is not permitted

---

[Q 1] Image Enhancement **[8 Marks]**

(a) [1+1+2 Marks] Consider the following 5×5 input images.

100	92	9	0	78
0	0	180	65	86
67	89	80	0	182
0	198	0	75	0
110	0	170	90	69

(i) The given image is corrupted by which type of noise among salt noise and paper noise?

(ii) Which type of filter is required to remove the noise in the given image among 3×3 max filter and 3×3 min filter?

(iii) Compute the results of filtering using the identified filter in (ii), only for middle 3×3 sub-image.

(b) [1 Mark] Explain the difference between correlation and convolution.

(c) [1 Mark] Explain aliasing in images with diagram.

(d) [2 Marks] How Gaussian filtering tackles the problem of Box filtering? Explain with examples.

[Q 2] Morphological Operations **[8 Marks]**

(a) [2 Marks] Explain what would happen in binary erosion and dilation if the structuring element is a single point, valued 1, with reasons.

(b) [2 Marks] Explain opening and closing operations with examples.

(c) [4 Marks] You have given a binary image A, and a structuring element B, perform boundary extraction on the image using the structuring element. Show the final result B(A) with all intermediate solutions. (You can perform the padding by 0. And assume the shaded pixel as the center of the structure element).

A									
0	1	1	1	1	0	0	0	0	0
1	1	1	1	1	1	1	1	0	0
1	1	1	1	1	1	1	1	1	1
1	1	0	1	1	1	1	1	1	1
1	1	0	0	0	1	1	1	1	1

B		
1	1	1
1	1	1
1	1	1

Indian Institute of Information Technology Allahabad

Mid / End Sem Question Paper

Course Name: Advanced Programming Pracces **Course Instructor/ Co-ordinator:** Dr. B S Sanjeev

Course Code: PC-IT-APP502 **Program Name(s):** MTech (IT) **Exam Date:** 18.10.2023 **MM:** 25

Answer all the questions

~~(All questions carry equal marks)~~

Note: Irrelevant (not wrong) answers carry penalties.

1. You are given n integers stored in a linked list ($n > 2$). Write a C function to delete all the middle element(s) till only a maximum of 2 elements remain and print output after every deletion.

   (Note: If the list has odd number of elements, delete the middle element. Otherwise delete the two middle elements.) **(7.5m)**

2. Explain the following terms with an example each: **(5x1.5m=7.5m)**
   a. Time complexity
   b. Space complexity
   c. Big O (O())
   d. Big Omega (Ω())
   e. Big Theta (Θ())

3. You are given n unique integers. The task is to do the following -- while reading each number from the input, create one disjoint set for every number. The apply n-1 union operations with between sets. The purpose of this exercise is to generate a sorted sequence easily at the end with some additional work as required. **(10m)**
   a. Explain how you would approach the problem with clear example with the aid of diagrams.
   b. Write pseudocode to solve the above problem.
   c. What is the total time complexity of your approach to find the sorted sequence?

   (Note: As union operation has to be done only after input is completely read and every element is in a separate set. No C code is to be used.)

additional work apart from union ops should be minimum

MRH2023026

Indian Institute of Information Technology Allahabad

Mid Sem Question Paper

Course Name: Mathematics for IT **Course Co-ordinator:** Dr. Mohammed Javed **Exam Date:** October 18, 2023

Course Code: PC-IT-MIT501 **Program Name(s):** M.Tech IT (MRH, MSD, MNS), PhD (IT) MRH, WP

[Maximum Marks is 25] [Duration 2 Hours]

Answer All the Questions

[Q1] Find the Eigenvalues and Eigenvectors for the matrix given below, and show the relationship between the eigenvectors. [5 Marks]

$$\begin{pmatrix} 2 & -3 & 0 \\ 2 & -5 & 0 \\ 0 & 0 & 3 \end{pmatrix}$$

[Q2] Find the SVD for the matrix given below. Give any 2 significant observations made during the SVD computation. [5 Marks]

$$A = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}$$

[Q3] Solve the following system of equations using LU decomposition. Write a brief Pseudocode for computing LU decomposition. [5 Marks]

$$x_1+x_2+x_3 = 1$$
$$4x_1+3x_2-x_3 = 6$$
$$3x_1+5x_2+3x_3 = 4$$

[Q4.(A)] Let A, B, and C be independent events, with P(C)> 0. Prove that A and B are conditionally independent given C. [3 Marks]

[Q4.(B)] A four-sided die is rolled repeatedly, until the first time (if ever) that an even number is obtained. What is the sample space for this experiment? [2 Marks]

[Q. 5] Two boxes containing candies are placed on a table. The boxes are labelled B1 and B2. Box B1 contains 7 cinnamon candies and 4 ginger candies. Box B2 contains 3 cinnamon candies and 10 pepper candies. The boxes are arranged so that the probability of selecting box B1 is 1/3 and the probability of selecting box B2 is 2/3. Suresh is blindfolded and asked to select a candy. He will win a colour TV if he selects a cinnamon candy. What is the probability that Suresh will win the TV (that is, he will select a cinnamon candy)? [5 Marks]

Introduction to Machine Learning

Quiz 1
Max Marks: 20

Q1. A pair of fair dice is rolled. Let X denote the sum of the number of dots on the top faces. **[4 Marks]**

1. Construct the probability distribution of X for a pair of fair dice.
2. Find P(X≥9).
3. Find the probability that X takes an even value.
4. Draw the histogram that can graphically illustrates the probability distribution.

Q2. A discrete random variable X has the following probability distribution: **[4 Marks]**

x	-1	0	1	4
P(x)	0.2	0.5	a	0.1

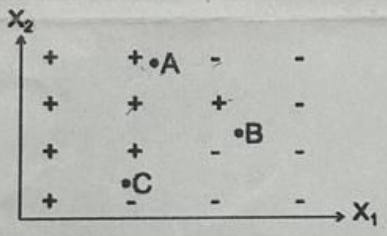
Compute each of the following quantities:

1. a
2. P(0) and P(X>0)
3. The mean μ of X.
4. The variance σ^2 of X and the standard deviation σ of X.

Q3. Answer the following:

1. Output of 1NN for A?
2. Output of 3NN for A?
3. Output of 3NN for B?

[3 Marks]



Q4. Explain in 1-2 sentences the difference between a "lazy" learner (such as nearest neighbour classifier) and an "eager" learner (such as logistic regression classifier). **[2 Marks]**

Q5. Write the cost function of linear regression and its gradient descent function. **[1 Marks]**

Q6. What is RBF-Kernel? Write its formula. **[1 Marks]**

Q7. **True or False**

a) Linear SVMs have no hyperparameters that need to be set by cross-validation. **[1 Marks]**

b) Given the hyperplane defined by the line **[2 Marks]**

$$y = x_1 - 2x_2$$
$$y = (1,-2)^T X = W^T X$$

Is this point correctly predicted?

1) y = 1, X = (1,0) ?

2) y = 1, X= (1,1) ?

c) Maximizing the likelihood of logistic regression model yields multiple local optimums. **[1 Marks]**

d) In the context of Decision Trees, entropy is a measure of disorder or impurity in a node. **[1 Marks]**

Indian Institute of Information Technology Allahabad

End Sem Question Paper

Course Name: Mathematics for IT **Course Co-ordinator:** Dr. Mohammed Javed **Exam Date:** December 11, 2023

Course Code: PC-IT-MIT501 **Program Name(s):** M.Tech IT (MRH, MSD, MNS), PhD (IT) MRH, WP

[Maximum Marks is 40] **[Duration 3.00 Hours]**

Answer All the Questions

[Q1] (a) Obtain the equation of projection ***p*** of point ***b*** onto a line ***a*** and also discuss its projection matrix P. **[5 Marks]**

(b) Show the equation of projection ***p*** onto a subspace ***A*** having two independent basis vectors a_1 *and* a_2 **[5 Marks]**

[Q2] (a) Briefly discuss any five tests for a positive definite matrix. **[5 Marks]**

(b) Fit a quadratic curve through the following points - **(1,1) , (2,5) , (-1, -2)** using linear algebra approach **[5 Marks]**

[Q3] **(a)** As an advertising campaign, a chocolate factory places golden tickets in some of its candy bars, with the promise that a golden ticket is worth a trip through the chocolate factory, and all the chocolate you can eat for life. If the probability of finding a golden ticket is p, find the mean and the variance of the number of candy bars you need to eat to find a ticket. **[5 Marks]**

(b) State and Prove Chebsheve Inequality. **[5 Marks]**

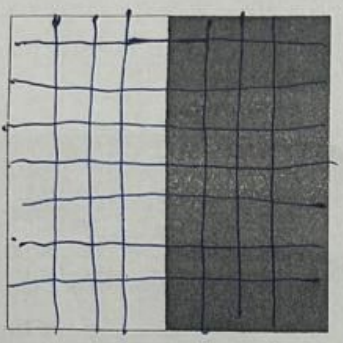
[Q4] (a) You just rented a large house and the realtor gave you 5 keys, one for each of the 5 doors of the house. Unfortunately, all keys look identical. so to open the front door, you try them at random. Find the PMF of the number of trials you will need to open the door, under the following alternative assumptions: (1) after an unsuccessful trial. you mark the corresponding key. so that you never try it again. and (2) at each trial you are equally likely to choose any key. **[5 Marks]**

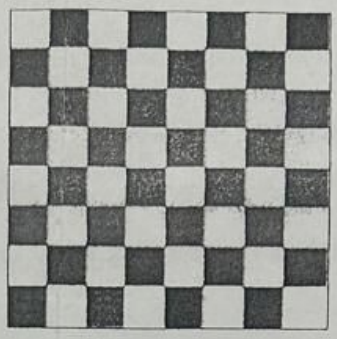
(b) The release of two out of three prisoners has been announced. but their identity is kept secret. One of the prisoners considers asking a friendly guard to tell him who is the prisoner other than himself that will be released, but hesitates based on the following rationale: at the prisoner's present state of knowledge, the probability of being released is 2/3, but after he knows the answer, the probability of being released will become 1/2, since there will be two prisoners (including himself) whose fate is unknown and exactly one of the two will be released. What is wrong with this line of reasoning? **[5 Marks]**

Q9. [4 Marks] The two images shown in the below figure are having only white and black pixels. They are quite different, but their histograms are the same. Suppose that each image is blurred using a 3 × 3 box kernel.

(a) Would the histograms of the blurred images still be equal? Explain.

(b) Either sketch the two histograms or give two tables detailing the histogram components.





Q10. [5 Marks] Perform Histogram equalization on the given image of 9x9 dimension. The grey scale values lie from 0 to 7.

4	4	4	4	4	4	4	0	4
3	3	5	1	4	4	7	4	1
3	2	1	6	5	3	3	4	5
7	2	1	0	3	3	0	0	0
6	3	7	4	6	1	4	5	3
3	6	3	4	5	4	7	1	2
5	6	4	7	1	0	0	4	4
5	5	5	5	5	5	5	5	5
0	5	4	6	2	1	2	1	2

MRH2023026

Indian Institute of Information Technology, Allahabad
MTech 1st Sem, Aug-Dec 2023 – Mid Sem Exam
Image and Video Processing (IVP)

Duration: 2 Hours
Note:
Max. Marks: 25

- The Electronic Materials such as **calculator, phone, laptop,** etc. are not allowed.
- Write your answers very precisely and to the point only.
- Attempt all questions. All parts of a question should be attempted at the same place.
- Avoid the malpractices during test. It may lead to severe penalty.

Q1. [2 Marks] What is the need of image sampling and quantization? Illustrate its working using diagram?

Q2. [2 Marks] Which is better among Box filter and Gaussian filter? Give reasons for your answer.

Q3. [2 Marks] Explain how median filter can be useful to remove the Salt and Pepper noise using example.

Q4. [2 Marks] When aliasing occurs? How Nyquist–Shannon sampling theorem is useful to avoid aliasing?

Q5. [2 Marks] For the matrix given below, consider the two highlighted pixels and find
A. City-block distance
B. Chessboard distance

	0	1	2	3	4	5
0	4	5	6	5	4	4
1	3	3	5	1		4
2	3	2	1	6	8	3
3	4	5	3	5	3	3
4	6		7	4	6	1
5	3	6	3	4	5	4

Q6. [2 Marks] Explain 4-adjacency and 8-adjacency with examples.

Q7. [2 Marks] Explain Gamma and Piecewise-Linear Transformations with examples. Give examples where these transformations can be applied.

Q8. [2 Marks] Consider the following 5×5 input image in spatial domain. Compute the output values to the nearest integer by using the 3×3 averaging filter in spatial domain. You can leave the computation for border pixels.

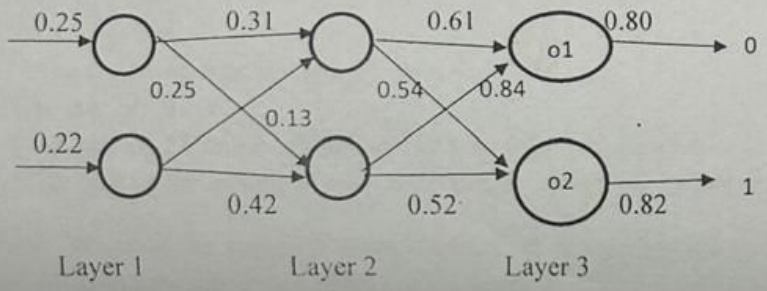
	0	1	2	3	4
0	10	20	30	30	70
1	20	30	30	40	40
2	30	40	40	40	50
3	20	70	70	70	80
4	90	10	70	70	60

Q6. Consider the following sample data points

X	1	3	9	3	7	9	4	8	1
Y	2	2	4	7	2	7	8	3	4

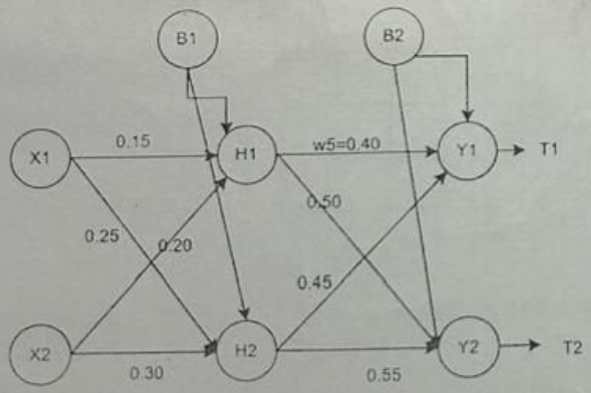
Now, create a distance matrix by computing distance between each pair of points. Use the concept of agglomerative clustering to create clusters. Use single, complete and average linkage approach to create clusters. Show the final dendrograms of each case. [6]

Q7. Compute the total error in this feed forward neural network [2]



Here, bias at layer 2 is 0.42 and for layer 3 it is 0.67, in layer 3 the names of node are given as *o1 and o2*, where the actual output for o1 should be 0 and for o2 it should be 1.

Q8. Suppose we have three layer neural network including input layer as mentioned below



where x1=0.05 and x2=0.20 represents the input variables, B1=0.35 and B2=0.5 represents the bias, T1=0.2 and T2=0.8. represents the target values.

In the forward pass, find the output values of Y1 and Y2. Thereafter, what would be the total error in the first pass? Then, using backpropagation approach what would be the new weights of w5. *Note:* use sigmoid as the activation function, the values on the arrows represents the weights.

[4]

$\sigma\left(\begin{bmatrix}0.63 & 0.64 & 1\end{bmatrix}\begin{bmatrix}0.61\\0.84\\0.67\end{bmatrix}\right) = 0.83$

$\sigma\left(\begin{bmatrix}0.63 & 0.64 & 1\end{bmatrix}\begin{bmatrix}0.54\\0.52\\0.67\end{bmatrix}\right) = 0.79$

$\sigma\left(\begin{bmatrix}0.25 & 0.22 & 1\end{bmatrix}\begin{bmatrix}0.31\\0.13\\0.42\end{bmatrix}\right) = 0.63$ Layer 2 output o_1

$\sigma\left(\begin{bmatrix}0.25 & 0.22 & 1\end{bmatrix}\begin{bmatrix}0.25\\0.42\\0.42\end{bmatrix}\right) = 0.64$ Layer 2 output down

Ans -

MRH2023026

Indian Institute of Information Technology Allahabad

End Sem Question Paper

Course Name: Introduction to Machine Learning **Course Instructor/ Co-ordinator:** Dr. Anjali Gautam

Course Code: PC-IT-IML503 **Program Name(s):** M.Tech(IT) MRH, MSD, DD M.Tech, PhD(IT) MRH, PhD(IT) including WP

Exam Date: ~~13/12/2023~~ 15 DEC 2023 **MM:** 40

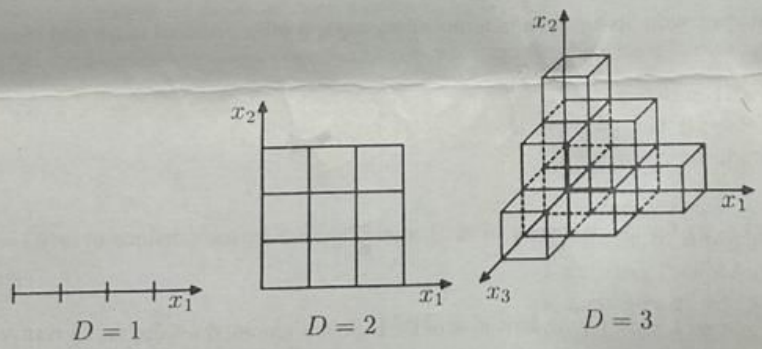
Note: Calculator is allowed and all questions are compulsory.

[5*2=10]

Q1. Short answers expected

A. Why there is a need of non-linear activation function in neural network, clearly write by giving example.
B. Write the formula of Gini impurity and Gini split index.
C. How Stochastic Gradient Descent is different from Gradient Descent? What is the magnitude and direction of the weight update in respect to cost gradient?
D. Why are SVMs fast?
E. Consider k-fold cross-validation. Let's consider the tradeoffs of larger or smaller k (the number of folds). What will be the estimated error on average with a higher number of folds and why?

Q2. Figure illustrate the curse of dimensionality, showing how the number of regions of a regular grid grows exponentially with the dimensionality D of the space. For clarity, only a subset of the cubical regions are shown for D = 3.



Now suppose, if we have D input variables, then how would a general polynomial with coefficients up to order 3 is written? [2]

Q3. Consider the generalization of the squared loss function $\mathbb{E}[L] = \iint \{y(x) - t\}^2 p(x,t)dx\,dt$ for a single target variable t to the case of multiple target variables described by the vector $\boldsymbol{t}$ given by

$$\mathbb{E}[L(\mathbf{t}, \mathbf{y}(\mathbf{x}))] = \iint \|\mathbf{y}(\mathbf{x}) - \mathbf{t}\|^2 p(\mathbf{x}, \mathbf{t})\,\mathrm{d}\mathbf{x}\,\mathrm{d}\mathbf{t}.$$

Using the calculus of variations, show that the function $y(x)$ for which this expected loss is minimized is given by $y(x) = \mathbb{E}_t[t|x]$. Show that this result reduces to $y(x) = \frac{\int tp(x,t)dt}{p(x)} = \int tp(t|x)dt = \mathbb{E}_t[t|x]$ for the case of a single target variable t. [3]

Anjali

Indian Institute of Information Technology Allahabad

Mid Sem Question Paper

Course Name: HSMC Course **Course Instructor/ Co-ordinator:** Prof. Archana Chahal

Course Code: HM-MS-PHE **Program Name(s):** Physical Education (Sports)

Exam Date: 29.02. 2024 **MM:** 25

Note: All the questions are compulsory:

Ques. 1. Answer the following multiple-choice questions: Marks-2.5 x 2 = 05

a. T.H.R. stands for

i. Testing Heart Rate ii. Target Heart Rate
iii. Target Heart Ratio iv. Testing Health Ratio

b. In 'Ashtang Yoga', at which position the step of *Niyam* is placed:

i. First ii. Second
iii. Third iv. Forth

Ques. 2. Write down the definition of Health given by World Health Organization. Marks- 05

OR

Define Health / *Swasthay* according to Shrimad Bhagwad Gita.

Ques. 3. Define the meaning of Physical Fitness.

Marks-05

OR

Explain the dimensions of Wellness.

Ques. 4. What is the meaning of Lifestyle and explain the role of sports and physical education in the development of healthy lifestyle?

Marks-15

OR

fake one or we need something more. Can the model work by learning to produce realistic samples and all the possible realistic samples? [10]

a. In light of the above, discuss the two player GAN game
b. What are the conditions to achieve optimality for GAN.

[Q4] Given an input x, the posterior probability distribution over outputs y is too complicated to work with or Given a training corpus x, the posterior probability distribution over parameters y is too complicated to work with. Derive the expression for variational inference and use the same for VAE implementation. [12]

MRH2023026

Indian Institute of Information Technology Allahabad

End Sem Question Paper, December 2023

Course name: Research Methodology-1 Course instructor: Dr. Ratan K. Saha

Course code: RM-1, Program name: MTech, PhD, Exam date: ~~15/12/2023~~ 17 DEC 2023

Maximum Marks: 40

Note: Use of scientific calculator is allowed.

1. Answer all the questions [10 × 1 = 10]

   (a) Write down the expression for the 1D Guassian/normal distribution function.

   (b) Name two file formats supported by vector graphics.

   (c) Give two examples of bivariate function.

   (d) Name two popular scientific databases.

   (e) What is the formula for calculating the impact factor of a journal?

   (f) Name two graph-centrality measures.

   (g) What is an "Index journal"?

   (h) What is the closeness centrality?

   (i) What is the Geodesic distance between nodes A and B when no path exists?

   (j) Draw the Venn diagram for $A \cap B'$.

2. Answer all the questions [5 × 3 = 15]

   (a) What are the 'Radar chart' and 'Bubble chart'?

   (b) What are the characteristics of adjacency matrix for a non-directed graph?

   (c) Give the Taylor's series expansion of the following function up to 2nd order differential. Find the value of the function at x=1 using that of x=0 where h=1.

   $$f(x) = 10\, e^{-\frac{(x^2+5x+13)}{2}}. \quad (1)$$

   Comment on the error for such truncation.

   (d) Consider an image filled with random numbers between 0 to 255. Plot the ideal histogram for such an image. What type of probability distribution will follow the entries? Name two other probability distribution functions.

   (e) While fitting a histogram with a probability density function what parameter do we minimize to find out the best fitting? What are the tuning parameters? Explain it with an example.

Indian Institute of Information Technology Allahabad

Mid-Semester Examination

Course Name: Research Methodology – I
Course Code: PC-IT-RMA505 & PC-AS-RMA504
Exam Date: 20.10.2023

Couse Instructor/Coordinator: Dr. Ashutosh Mishra
Prog. Name(s): IT (M. Tech 1, DD), BI (M. Tech 1, DD)
MM: 50

Answer all questions legibly and clearly. State any assumptions made. Use of Scientific/Programmable Calculators is ALLOWED. By taking this test – you agree to abide by the Student Honor code. **For your reference, the student-t table is provided.**

t Table

cum. prob	$t_{.50}$	$t_{.75}$	$t_{.80}$	$t_{.85}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$	$t_{.999}$	$t_{.9995}$
one-tail	0.50	0.25	0.20	0.15	0.10	0.05	0.025	0.01	0.005	0.001	0.0005
two-tails	1.00	0.50	0.40	0.30	0.20	0.10	0.05	0.02	0.01	0.002	0.001
df											
1	0.000	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.31	636.62
2	0.000	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.000	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.000	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.000	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.000	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.000	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.000	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.000	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.000	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.000	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.000	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.000	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.000	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.000	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.000	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.000	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.000	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.000	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.000	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.000	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.000	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.000	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.000	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.000	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.000	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.000	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.000	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.000	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.000	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	0.000	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	0.000	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.232	3.460
80	0.000	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.195	3.416
100	0.000	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.174	3.390
1000	0.000	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581	3.098	3.300
z	0.000	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.090	3.291
	0%	50%	60%	70%	80%	90%	95%	98%	99%	99.8%	99.9%
	Confidence Level										

IMRH2023026

Indian Institute of Information Technology Allahabad

Mid-Semester Examination

Course Name: **Research Methodology – I**
Course Code: **PC-IT-RMA505 & PC-AS-RMA504**
Exam Date: **20.10.2023** (Session – I)

Couse Instructor/Coordinator: **Dr. Ashutosh Mishra**
Prog. Name(s): **IT (M. Tech 1, DD), BI (M. Tech 1, DD)**
MM: **50**

Answer all questions legibly and clearly. State any assumptions made. Use of Scientific/Programmable Calculators is ALLOWED. By taking this test – you agree to abide by the Student Honor code. **For your reference, the student-t table is provided.**

1. Given the following data: Compute the Mean, Median, Mode and the Standard Deviation. **[5]**

Value	1	2	3	4	5	6	7	8
Frequency	4	6	4	4	3	2	1	1

2. Draw the box plot for the data presented in question 1. Ensure that you mark all the relevant metrics. **[5]**

3. For the samples given in question 1: compute the excess Kurtosis. **[10]**

4. For the samples given in question 1: Compute the Adjusted Skewness Coefficient. **[10]**

5. Two groups of students were tutored separately using an Intensive (covering larger content in fixed amount of time) or Paced (Covering lower content in the fixed time) approach. The two groups have the following metrics. Assuming normal distributions for both groups: Check if paced tutoring yields higher student performance. Use a significance level of $\alpha \leq 0.05$. **[10]**

Group	Tutoring	N	Mean	Std. Deviation
1	Intensive	12	46.31	6.44
2	Paced	10	42.79	7.52

6. For the tutoring data provided in question 5, compute the 90% confidence interval for the difference in the mean scores of the two groups. **[10]**

Q4. Consider the case of two classes, and suppose we take the D-dimensional input vector x and project it down to one dimension using $y = w^T x$. If we place a threshold on y and classify $y \geq -w_0$ as class C1, otherwise class C2. We can select a projection that maximizes the class separation by adjusting the components of the weight vector w. Let's consider there are N1, N2 points of class C1 and class C2 respectively, and the mean vectors of the two classes are given by

$$\mathbf{m}_1 = \frac{1}{N_1} \sum_{n \in C_1} \mathbf{x}_n, \qquad \mathbf{m}_2 = \frac{1}{N_2} \sum_{n \in C_2} \mathbf{x}_n$$

The simplest measure of the separation of the classes, when projected onto w, is the separation of the projected class means. This suggests that we might choose w so as to maximize

$$m_2 - m_1 = \mathbf{w}^{\mathrm{T}}(\mathbf{m}_2 - \mathbf{m}_1) \qquad \textbf{where}$$

$$m_k = \mathbf{w}^{\mathrm{T}}\mathbf{m}_k$$

is the mean of the projected data from class C_k. The two classes are well separated in the original 2D space (x_1, x_2) but have considerable overlap when projected onto the line joining their means.

A. What was the idea proposed by Fisher to solve the issue mentioned above. Write the terms used by them using above equations. [2]

B. What is the Fisher criterion $J(w)$, write its equation in terms of m_1, m_2 and other required terms and show that it can be rewritten as: [3]

$$J(\mathbf{w}) = \frac{\mathbf{w}^{\mathrm{T}}\mathbf{S}_{\mathrm{B}}\mathbf{w}}{\mathbf{w}^{\mathrm{T}}\mathbf{S}_{\mathrm{W}}\mathbf{w}}$$

Q5. Solve this

A. Verify the relation $\frac{d\sigma}{da} = \sigma(1-\sigma)$ for the derivative of the logistic sigmoid function defined by $\sigma(a) = \frac{1}{1+\exp(-a)}$. [3]

B. By making use of the result $\frac{d\sigma}{da} = \sigma(1-\sigma)$ for the derivative of the logistic sigmoid, show that the derivative of the cross-entropy error function $E(w) = -\ln p(t|w) = -\sum_{n=1}^{N}\{t_n \ln y_n + (1-t_n)\ln(1-y_n)\}$ for the logistic regression model where $y_n = \sigma(a_n)$ and $a_n = w^T \emptyset_n$, weight vector w, $\emptyset$ is M-dimensional feature space for the dataset $\{\emptyset_n, t_n\}$ where $t_n \in \{0,1\}$, $\emptyset_n = \emptyset(x_n)$ with n=1,…., N, is given by

$$\nabla E(\mathbf{w}) = \sum_{n=1}^{N}(y_n - t_n)\phi_n$$

Note: For two class problem, the posterior probability of class C1 can be written as a logistic sigmoid acting on a linear function of the feature vector $\emptyset$ so that $p(C_1|\phi) = y(\phi) = \sigma\left(\mathbf{w}^{\mathrm{T}}\phi\right)$ with $p(C_2|\phi) = 1 - p(C_1|\phi)$. [5]