INDIAN INSTITUTE OF TECHNOLOGY KHARAGPUR
End-Autumn Semester Examination 2022-23
Date of Examination: 23.11.22 Session: AN Duration: 3 hrs Full Marks: 70
Subject No.: AI61003 Subject: LINEAR ALGEBRA FOR AI AND ML
Dept/Center/School: Centre of Excellence in Artificial Intelligence
Specific charts, graph paper, log book required: NO,
Special Instructions: Use two separate answer sheets for Part A and Part B

Part A (40 marks)

[5+5]

1. a) You are given m objects (samples) on n dimensional space (n features). Your goal is to transform these objects on lower dimensional space (less than n) such that the variance is preserved as much as possible in the new space. Prove that the eigenvector corresponding to the maximum eigenvalue of the variance-covariance matrix of the data will retain the maximum variance, if the data is to be projected on 1 dimensional space.

   b) The eigenvalues of a variance-covariance matrix of a standardized (0,1) data with 3 features are 2.5, 1.5 and 1. The corresponding eigenvectors are (1,-1,0), (1,1,0) and (0,0,1). You want to project the data point P (0.5, 0.6, 0.3) to new space such that minimum 60% and maximum 85% variance is preserved in the new space. Compute the projected vector for P that satisfies the above criteria simultaneously.

[5+5]

2. a) What is the main function of an auto-encoder? Define the reconstruction error term for a one-layer auto-encoder that transforms 10 dimensional points to 5 dimension. You must explain every notation in your definition.

   b) You have an auto-encoder with 10 input neurons, 10 output neurons and 5 hidden layers each having 5 neurons. The activation function is relu (f(x) = max(0,x)). You are dealing with a dataset D (10 features) with all positive values. If the network parameters are all initialized to non-negative values, prove that the auto-encoder only linearly transforms the data after the first forward pass.

[4+6]

3. a) Define singular value decomposition (SVD) of a matrix A as a product of matrices. You must add description of each of the notations as well as the properties of the matrices in SVD. State and explain Eckart-Young theorem. What is its connection with SVD?

   b) Find the singular values, left singular vectors and right singular vectors for the matrix S $= \begin{bmatrix} 3 & 0 \\ 4 & 5 \end{bmatrix}$.

[5+5]

4. a) Explain the geometry of SVD with a diagram. You must mention the roles played by each of the matrices in SVD.

   b) A movie recommender system collects the data in the form of a matrix, where the rows represent the users and the columns represent movies. The ij-th entry (a_{ij}) of the matrix denotes the rating (0 to 5) of the j-th movie by the i-th user. How do you compute a missing entry (say a_{mn}) in this matrix using a matrix completion technique. You must specify the steps you would need to perform and the objective function that you will use.

Part B (30 marks)

1. a) Which of the following low-rank matrices are suitable for completion? Justify.

$$M_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}, \quad M_2 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}.$$

b) What will be the maximum Coherence value of a n×n matrix with rank r. Justify your answer from the basic definition of Coherence.

Total : 2+3 =5

2. a) What do you understand by Degree of freedom (D.O.F) of a matrix. Compute the D.O.F of a n1 x n2 matrix of rank r. 4

b) Check and justify whether the following matrix is semi-definite or not:

7	-2	4
-2	3	6
4	6	3

4

c) What is the purpose of replacing the rank function with neuclear norm in rank minimization problem? 4

d)

Consider the dictionary

$$D = \begin{bmatrix} 1 & 1 & 0 & -1 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & -1 \end{bmatrix}.$$

Consider the vector $y = [2 \ 1 \ 1]^T$. Comment on what will be a sparse solution to y. 3

Total : 4+4+4+3=15

3. a) Find the eigenvalues and corresponding eigenvectors of:

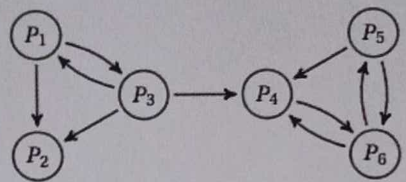
$$A = \begin{bmatrix} 3 & -1 \\ -2 & 4 \end{bmatrix}.$$

2

b) Calculate two iterations of the power method with scaling, starting with $(x\_0 = 1,1)$. 2

(c) Explain why the method does not seem to converge to a dominant eigenvector. 2

d) Compute the page rank vector of the following miniweb:
[Make necessary assumptions if required and state them clearly]



4

Total: 2+2+2+4 = 10