

JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

TEST -3 EXAMINATION- 2023

M. Tech-I Semester (CSE-Data Science)

COURSE CODE (CREDITS):22M1WC1131 (3)

MAX. MARKS: 35

COURSE NAME: Data Warehousing and Data Mining

COURSE INSTRUCTORS: Dr. Pardeep Kumar

MAX. TIME: 2 Hour

*Note: (a) All questions are compulsory.*

*(b) Marks are indicated against each question in square brackets.*

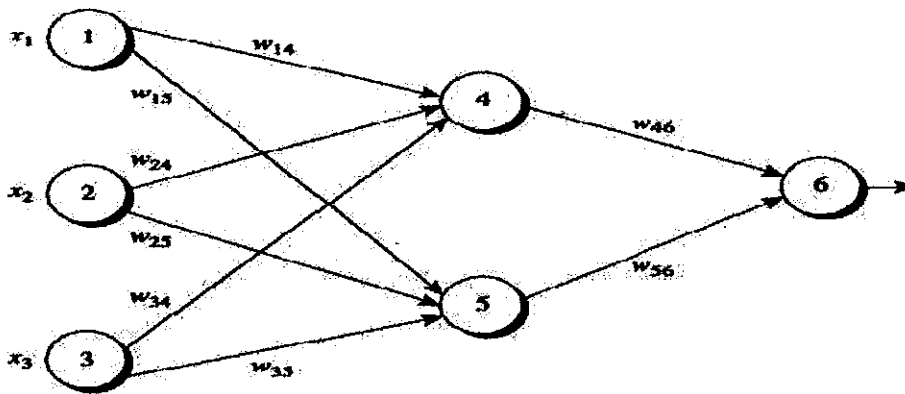
*(c) The candidate is allowed to make Suitable numeric assumptions wherever required for solving problems*

1. Consider the IMDb database of the following movies given as under:

| S.No | Movie Name          | IMDb Rating | Duration (In Minutes) | Genre  |
|------|---------------------|-------------|-----------------------|--------|
| 1    | Mission-2           | 8.0         | 160                   | Action |
| 2    | Gadar-2             | 6.2         | 170                   | Action |
| 3    | To Jhuthi Ma Makkar | 7.2         | 168                   | Comedy |
| 4    | Oh My God           | 8.2         | 155                   | Comedy |
| 5    | Animal              | 8.0         | 190                   | Action |

What would be the genre of the movie Gadar with the following features (IMDb rating=7.0 and duration= 160 Minutes) ? [CO-6] [7]

2. Consider 8 persons showing interest in personal loan from a reputed public sector bank. Let point(x,y,z) represents their age(in years), salary(in Indian rupees) and income from other sources(in Indian rupees). So the data is represented as Ram(20,35K,10K), Sita(30,25K,5K),Laxman(50,20K,12K),Anil(40,20K,2K),Prem(45,50K,15K),Sunil(21,46K,15K), Sourabh(23,32K,11K) and Meenu(32,27K,13K). The bank manager's task is to check whether the person is rich, medium earning or poor based on the given data. Use k-means clustering algorithm technique based on Euclidean distance to ease manager's task. [CO-5] [10]
3. Consider the given below feed forward neural network:



The initial input, weights and biases are given as follows:

**Initial Input, Weight, and Bias Values**

| $x_1$ | $x_2$ | $x_3$ | $w_{14}$ | $w_{15}$ | $w_{24}$ | $w_{25}$ | $w_{34}$ | $w_{35}$ | $w_{46}$ | $w_{56}$ | $\theta_4$ | $\theta_5$ | $\theta_6$ |
|-------|-------|-------|----------|----------|----------|----------|----------|----------|----------|----------|------------|------------|------------|
| 1     | 0     | 1     | 0.2      | -0.3     | 0.4      | 0.1      | -0.5     | 0.2      | -0.3     | -0.2     | -0.4       | 0.2        | 0.1        |

The learning rate is 0.9. The target output is 1 for the given tuple {1,0,1}. Do the dry run for back propagation learning mechanism by considering the above ANN. Also write the pseudo code of back propagation algorithm and discuss its complexity. [CO-4] [10]

4. Consider the weather data set given below:

| Outlook  | Temperature | Humidity | Windy | Play |
|----------|-------------|----------|-------|------|
| Sunny    | Hot         | High     | False | No   |
| Sunny    | Hot         | High     | True  | No   |
| Overcast | Hot         | High     | False | Yes  |
| Rainy    | Mild        | High     | False | Yes  |
| Rainy    | Cool        | Normal   | False | Yes  |
| Rainy    | Cool        | Normal   | True  | No   |
| Overcast | Cool        | Normal   | True  | Yes  |
| Sunny    | Mild        | High     | False | No   |
| Sunny    | Cool        | Normal   | False | Yes  |
| Rainy    | Mild        | Normal   | False | Yes  |
| Sunny    | Mild        | Normal   | True  | Yes  |
| Overcast | Mild        | High     | True  | Yes  |
| Overcast | Hot         | Normal   | False | Yes  |
| Rainy    | Mild        | High     | True  | No   |

Predict the humidity for the information {Outlook= Sunny, Temperature=Hot, Windy=False, and Play=No}

[CO- 4] [8]