Deccan Education Society's
**FERGUSSON COLLEGE (Autonomous) Pune-4**
Affiliated to Savitribai Phule Pune University
**M.Sc – I (Semester : I)**
**Subject : Data Science**

O/C

05 JAN 2023

**Paper Title : Database Management System (2019 pattern)**
**Paper Code : CSD4104**
**2019 Pattern (4 Credits)**

**[Time: 3 Hours]** **[Max. Marks : 50]**

**Instructions to the candidates:**
1) All questions are compulsory.
2) Neat diagrams must be drawn wherever necessary.
3) Figures to the right side indicate full marks.

Q1) **Answer the following questions (Any Five)** **[5 x 2 = 10 Mks]**

a) What do you mean by database and Database Management System? List the names of any two DBMS.

b) List two special characters that are used for pattern matching on strings in SQL?

c) Give the schema representation for the *Student* relation-

| Roll_no | Name | Birth_date |
|---|---|---|
| 1 | Geeta | 2011-01-01 |
| 2 | Seeta | 2009-12-01 |
| 3 | Meeta | 2010-11-11 |

Also, write the degree and cardinality of the *Student* relation

d) Define the terms: Candidate key, Primary key

e) Show the symbols for the following Relational Algebra operations: Select, Project, Set Difference and Cartesian Product.

f) What is a view? Write its importance.

Q2) **Answer the following questions (Any Four)** **[4 x 3 = 12 Mks]**

a) Consider the following schemas-

Book(b_no, bname, publisher, price)
Author(a_no, a_name, address)

Relationship between Book and Author is many to many.
Write Relational Algebra expressions for the following queries:

i. Find the number of books of each author.
ii. Find all details of books written by author 'Sudarshan'.

b) Distinguish between RDBMS and traditional file organization system.

c) What do you mean by integrity constraints? Discuss with suitable example.

d) Consider the following expressions, which use the result of a Relational Algebra operation as the input to another operation:
Transform following relational algebra queries into SQL statements:

i. $\Pi_{customer_name}(borrower) \cup \Pi_{customer_name}(depositor)$
ii. $\sigma_{amount > 1200}(loan)$
iii. $\Pi_{loan_number}(\sigma_{amount > 1200}(loan))$

e) Explain natural join using Relational Algebra.

Q3) **Answer the following questions (Any Three)** [3 x 3 = 9 Mks]

a) Predict the output of the following queries:

i) select concat(left('Data Technology',4), right('Information Science',7));

ii) select ceil(10.76);

iii) select replace('Hot Coffee', 'Hot', 'Cold');

b) Describe various SQL statements.

c) Write a function in SQL to accept project name as input and returns the number of employees working on the project.

The relations are:

Project(PNO, P_NAME, PTYPE,DURATION)

Employee(ENO, E_NAME, QUALIFICATION, JOINDATE )

The relationship is as follows:

PROJECT - EMPLOYEE: M-M Relationship, with descriptive attributes as start_date (date), no_of_hours_worked (integer).

d) Explain SELECT statement with suitable example.

Q4) **Answer the following questions (Any Two)** [2 x 3 = 6 Mks]

a) Discuss various symbols used in E-R diagram.

b) Explain composite, multivalued and derived attributes with suitable example.

c) Draw an E-R diagram to represent Banking Enterprise System(Consider Account and Loan relations)

Q5) **Answer the following questions (Any Two)** [2 x 4 = 8 Mks]

a) Write a short note on Graph based NoSQL database.

b) Give two examples of column based store NoSQL database and key-value pair NoSQL database.

c) Differentiate between RDBMS and NoSQL database.

Q6) **Answer the following questions (Any One)** [1 x 5 = 5 Mks]

a) Consider the following tables-

Teacher(t_no, t_name, college_name, dept_name)

e-test(e_no, test_name)

Teacher and e-test are related with many-to-many relationship.

Solve the following queries-

i. Count the number of teachers who passed NET exam of Computer Science department.

ii. List teacher name and test name exam from 'Fergusson College'.

b) Consider the following tables-

Person (pnumber, pname, birthdate, income),

Area(aname, area_type).

An area can have one or more persons living in it, but a person belongs to exactly one area.

The attribute 'area_type' can have values either 'urban' or 'rural'.

Solve the following queries-

i. List the names of all people whose birthday falls in the month of '*March*'.

ii. Count the people who are born on '*2012-01-06*'.

Deccan Education Society's
**FERGUSSON COLLEGE (Autonomous) Pune-4**
Affiliated to Savitribai Phule Pune University

O/C

04 JAN 2023

**M.Sc.–I (Semester: I)**
**Subject: Data Science**
**Paper Title: Data Structures (2019 pattern)**
**Paper Code: CSD4103**
**2019 Pattern (4 Credits)**

**[Time: 3 Hours]** **[Max. Marks : 50]**

**Instructions to the candidates:**
1) All questions are compulsory.
2) Neat diagrams must be drawn wherever necessary.
3) Figures to the right side indicate full marks.

Q1) **Answer the following questions (Any Five)** [5 x 2 = 10 Mks]
a) Discuss the advantages of using comments in programming.
b) What is an IDE? Name any two popular IDEs.
c) Explain the role of Exception Handling in programming.
d) When do we use frozen binaries? Name its contents.
e) Demonstrate use of join functions with suitable example.
f) Give example of "use of % operator" in string formatting.

Q2) **Answer the following questions (Any Four)** [4 x 3 = 12 Mks]
a) Python variables do not have specific types. Justify this statement with suitable examples.
b) Compare "for" vs "while" loop with suitable examples.
c) Explain lambda functions with suitable example.
d) Discuss in brief the concept of modules and packages in Python programming.
e) Explain scope of Global variables and Local variables.

Q3) **Answer the following questions (Any Three)** [3 x 3 = 9 Mks]
a) Write a short note on Set data structure and explain its usage in real world applications.
b) Differentiate between del and remove functions supported on List.
c) Write a program that creates dictionary of cubes of even numbers in the range 1 – 10.
Output: {2:8, 4:64, ....}
d) Write a function Max_Min that accepts a tuple and returns both max and min values from the function. Call the above function and display the output.

Q4) **Answer the following questions (Any Two)** [2 x 3 = 6 Mks]
a) Write a Python program to print every 10th character from the given file.
b) Define serialization. Discuss how Python supports serialization.
c) Write a short note on Cursor.

PTO

Q5) **Answer the following questions (Any Two)** **[2 x 4 = 8 Mks]**

a) Discuss the components of Web Scraping.

b) Write a program to perform following

   i. Define a function Calculate, which accepts a string and returns the length of string and word count as output. (Note: use of regular expression is must in both cases.)

   ii. In main script: accept the sentence from end user, call the above Calculate function and display the output.

c) Write a program to extract an email address and date from the sentence given by end user.

e.g., "M.Sc. Data Science program is starting from **26/09/2022** For queries you can reach us at msc.datascience@fergusson.edu".

Q6) **Answer the following questions (Any One)** **[1 x 5 = 5 Mks]**

a) Explain Inheritance and its types with suitable example.

b) Demonstrate the difference between class methods, instance methods and static methods with suitable example. Python code snippet expected.

02 JAN 2023

Deccan Education Society's

**FERGUSSON COLLEGE (Autonomous) Pune-4**

Affiliated to Savitribai Phule Pune University

**M.Sc. Data Science (Semester : I)**

**Subject: Data Science**

**Probability and Statistics**

**CSD4101**

**2019 Pattern (4 Credits)**

**[Time: 3 Hours]** **[Max. Marks : 50]**

---

**Instruction to candidates:**

i) **All questions are compulsory**
ii) **Figures to the right indicate full marks**
iii) **Use of scientific calculators is allowed.**

**Q1. Attempt any <u>five</u> questions:** **[5x2=10]**

a. Define skewness and state the 'R' command to obtain coefficient of skewness for raw data.
b. A and B are two independent events defined on $\Omega$ such that $P(A) = 0.5$ and $P(B) = 0.4$ Tell the value of P(AUB)'
c. Obtain all possible distinct arrangements of the letters of the word 'STATISTICS'.
d. $X \rightarrow N(4, 25)$.State the distribution of $2X - 5$.
e. Define correlation for a bivariate data. Also state the 'R' command to obtain correlation coefficient for a bivariate data.
f. Define odds ratio in logistic regression.

**Q2. Attempt any <u>four</u> questions:** **[4x3=12]**

a. The following data set gives the nicotine content in mg. per cigarette for 23 non menthol, regular type US brands:
1.2 , 1.0 , 0.8 , 0.1 , 0.8 , 0.9 , 0.2 , 1.4 , 1.1 , 0.8 , 1.0 , 1.0 , 1.1 , 1.4 , 1.2 , 0.7 , 1.2 , 1.1 , 1.0 , 1.3 , 0.8 , 0.5 , 0.6
i. Compute the third quartile for the above data.
ii. Explain the 'R' command for finding the third quartile.
b. Explain mean and variance of a continuous random variable.
c. In a certain city, the daily consumption of electric power in millions of kilowatt-hours, is a r.v following Gamma distribution with mean 6 and variance 12. Compute the parameters of the distribution and write the 'R' command for finding the probability that the daily power consumption will exceed 12 million kilowatt – hours.
d. Explain multiple correlation coefficient $R_{3.12}$. State the association between variables when
i. $R_{3.12} = 1$ ii. $R_{3.12} = 0$
e. The probability that a target is destroyed on a shot is 0.7. Assuming that the target will be destroyed, explain the distribution of failed attempts required before the target is destroyed. Also compute its parameters.

**Q3. Attempt any <u>three</u> questions:** **[3x3=9]**

a. Explain multicollinearity. How is it solved?
b. A box of 20 mangoes contains 4 rotten mangoes. 2 mangoes are selected at random without replacement from this box. Obtain the probability distribution of the number of rotten mangoes in the sample.
c. Explain variance and coefficient of variation as a absolute measure and relative measure of dispersion.
d. $X \rightarrow N(7, 81)$. State the 'R' commands to find the following probabilities:
   i. $P(X > 9)$ ii. $P(X < 5)$ iii. $P(4 < X < 7)$

**Q4. Attempt any <u>two</u> questions:** **[2x3=6]**

a. A student is selected randomly from a class of 30 students with roll numbers 1 to 30. Identify the probability distribution of the roll number of student selected and also find the probability that the roll number of the student selected is less than 4.
b. Explain in detail the difference between linear and logistic regression
c. Explain sensitivity of test procedure.

**Q5. Attempt any <u>two</u> questions:** **[2x4=8]**

a. The average number of field mice per acre in wheat field is 12. Determine the probability distribution of number of field mice per acre. Write the 'R' command to find the probability that
   i. less than 7 field mice are found on two acres.
   ii. more than 15 field mice are found on an acre.
b. Explain the concept of residual plot and mean residual sum of squares.
c. A factory has two machines. Past record shows that machine I produces 30% of the items and machine II produces 70% of the items. Further 5% of the items produced by machine I are defective and 1% of the items produced by machine II are defective. If a defective item is drawn at random, determine the probability that it was produced by machine II ?

**Q6. Attempt any <u>one</u> question:** **[1x5=5]**

a. The probability that a married man watches a certain television show is 0.4 and the probability that a married women watches that show is 0.5. The probability that a man watches the show given that his wife does is 0.7. Write the probability that
   i. a married couple watches the show
   ii. a wife watches the show given that her husband also watches it.
   iii. at least one person of the married couple watches the show.
b. $X \rightarrow B(n, p)$. Write the mean and variance of X. Also write additive property of Binomial distribution. Specify one real life example where Binomial distribution can be used.

10 DEC 2022

O/C

Deccan Education Society's

**FERGUSSON COLLEGE (Autonomous) Pune-4**

Affiliated to Savitribai Phule Pune University

**M.Sc.- II (Semester: III)**

**Subject: Data Science**

**Paper Title: Optimization Techniques**

**Paper Code:CSD5301**

**2019 Pattern (4 Credits)**

**[Time: 3Hours]** **[Max. Marks: 50]**

***Instructions to the candidates:***

1) *All questions are compulsory.*
2) *Neat diagrams must be drawn wherever necessary.*
3) *Figures to the right side indicate full marks.*

Q1) **Answer the following questions (Any Five). Two Marks Each** **10 M**

a) State the importance of Shadow Price in Sensitivity Analysis.

b) Compute the optimal solution to given Assignment problem using Hungarian's Algorithm.

| | Job I | Job II | Job III | Job IV |
|---|---|---|---|---|
| Riya | 10 | 5 | 13 | 15 |
| Anil | 3 | 9 | 18 | 3 |
| Jay | 10 | 7 | 3 | 2 |
| Suneet a | 5 | 11 | 9 | 7 |

c) Describe a network diagram for the following set of activities:

| Activity | A | B | C | D | E | F |
|---|---|---|---|---|---|---|
| Immediate predecessor | - | A | A | B , C | - | E |
| Duration | 2 | 3 | 4 | 6 | 2 | 8 |

d) Compute the dual of the given L.P.P:

$$Max\ Z = 6x_1 + 7x_2\ ,subject\ to$$
$$5x_1 - 2x_2 \leq 5\,,$$
$$3x_1 - 2x_2 \geq 4\,, \qquad x_1, x_2 \geq 0$$

e) State two applications of PERT.

f) Identify the type of solution this L.P.P has, when the last optimal table is given below: Comment on whether alternate solution exists or not. If yes, then which variable will enter the alternate optimal solution

| | | 3 | 4 | 5 | 0 | 0 | |
|---|---|---|---|---|---|---|---|
| Cb | Xb | X1 | X2 | X3 | S1 | S2 | b |
| 0 | S1 | -1 | 0 | -2 | 1 | 2 | 7 |
| 4 | X2 | 3/4 | 1 | 2 | 0 | 4 | 9 |

Q2) **Answer the following questions (Any Four). Three Marks Each** **12 M**

a) Predict the number of critical points this function will possess and state them.

$$f(x_1, x_2, x_3) = 7x - 8y + 2xy - x^2 + y^3$$

b) Explain thedegeneracy in solving Linear Programming Problem with respect to following :

i) When does degeneracy exists ?

ii) What does it indicates ?

c) Describe the range which is possible for cost coefficient of $x_1$ so as the optimal solution of the given L.P.P remains same.

| | | c1 | c2 | c3 | c4 | c5 | |
|---|---|---|---|---|---|---|---|
| Table I | Cj | 4 | 6 | 2 | 0 | 0 | |
| **CB** | XB | **x1** | **x2** | **x3** | **S1** | **s2** | **b** |
| 4 | x 1 | 1 | 0 | -1 | 1.33 | -0.33 | 1 |
| 6 | x2 | 0 | 1 | 2 | -0.33 | 0.33 | 2 |
| | Zj | 4 | 6 | 8 | 3.33 | 0.67 | |
| | Zj-Cj | 0 | 0 | 6 | 3.33 | 0.67 | |

d) Interpret the point of minima for the given function using steepest descent method.Starting point is given as (-1,1)

$$f(x, y) = x^2 + 2x + y^2 + 4$$

e) Classify the given function as convex or concaveusing the derivatives.Justify

$$h(x, y) = 4x^2 + 2x + 2y^2 + y + 13$$

Q3) **Answer the following questions (Any Three) Three Marks Each** **9 M**

a) Determine the optimal solution of given L.P.P

$$Maximize\ Z = -4x_1 - 3x_2 - 9x_3\ ,subject\ to$$

$$2x_1 + 4x_2 + 6x_3 \geq 15$$

$$6x_1 + x_2 + 6x_3 \geq 12\ , \quad x_1, x_2, x_3 \geq 0$$

using Two Phase method, where the optimal table for Phase-I is already given below:

| | | | | | Phase I | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | cj | 0 | 0 | 0 | 0 | 0 | -1 | -1 | |
| Cb | xb | x1 | x2 | x3 | s1 | s2 | R1 | R2 | b |
| 0 | x2 | -4/3 | 1 | 0 | -1/3 | 1/3 | 1/3 | -1/3 | 1 |
| 0 | x3 | 11/9 | 0 | 1 | 1/18 | -2/9 | -1/18 | 2/9 | 11/6 |
| | zj-cj | 0 | 0 | 0 | 0 | 0 | 1 | 1 | |

b) Evaluate the next approximation for the given function using Newton's method.Starting point is given as (1,1)

$$f(x, y) = 4x^2 - 2xy + y^2$$

c) Determine the role of gradient of a function in finding the extrema of the function. Evaluate the gradient of the given function$f(x, y) = 3x^3 - 3x^2 + 4y^2$ at point(-2,3).

d) Determine the critical path of a project having the tasks as given below:

| JOB | Time | JOB | Time |
|---|---|---|---|
| (1,2) | 2 | (5,6) | 7 |
| (1,3) | 3 | (4,6) | 3 |
| (2,4) | 2 | (4,7) | 2 |
| (3,4) | 3 | (6,7) | 6 |
| (3,5) | 2 | (5,7) | 5 |
| (4,5) | Dummy | | |
| | | | |

Q4) **Answer the following questions (Any Two). Three Marks Each** **6 M**

a) Analyse the given optimal table and hence discuss the change in optimal solution when a constraint $x_2 \leq 8$ is added.

| | cj | 3 | 5 | 0 | 0 | |
|---|---|---|---|---|---|---|
| Cb | xb | x1 | x2 | s1 | s2 | b |
| 0 | s1 | 1 | 0 | 1 | 0 | 4 |
| 5 | x2 | 3/2 | 1 | 0 | 1/2 | 9 |
| | zj-cj | 9/2 | 0 | 0 | 5/2 | |

b) Apply Hungarian method on given Travelling Salesman Problem and other tools to compute the optimal solution.

| | C1 | C2 | C3 | C4 | C5 |
|---|---|---|---|---|---|
| **C1** | - | 2 | 5 | 7 | 1 |
| **C2** | 6 | - | 3 | 8 | 2 |
| **C3** | 8 | 7 | - | 4 | 7 |
| **C4** | 12 | 4 | 6 | - | 5 |
| **C5** | 1 | 3 | 2 | 8 | - |

c) Evaluate the best approximation(2 iterations) for minimum of given function $f(x,y) = x^2 + y^2$ using Gradient descent method, where $P_0 = (1,1)$

Q5) **Answer the following questions (Any Two) Four Marks Each** **8 M**

a) In modification of a plant layout of a factory , 4 new machines are to be installed and there are 5 vacant places are available A,B,C,D and E. The cost of placing machine I at place j is given below:

| | A | B | C | D | E |
|---|---|---|---|---|---|
| **M1** | 9 | 11 | 15 | 10 | 11 |
| **M2** | 12 | 9 | -- | 10 | 9 |
| **M3** | -- | 11 | 14 | 11 | 7 |
| **M4** | 14 | 8 | 12 | 7 | 8 |

Compute the optimal assignment schedule.

b) Evaluate the Given Integer programming for optimal solution using the branch and bound method :

Maximize Z= $2x_1 + 2x_2$

Subject to,

$5x_1 + 3x_2 \leq 8 \quad x_1 + 2x_2 \leq 4 \ , x_1, x_2 \geq 0$ and $x_1$integer .

Where current non integral optimal solution is $x_1 = 0.57 \, , x_2 = 1.71 \, , \max Z = 4.57$

c) Evaluate the probability of completing the project in 35 days, when critical path is 1-3-4-7-10 .

| Activity | 1-2 | 1-3 | 2-5 | 3-4 | 4-5 | 5-8 | 4-6 | 4-7 | 6-9 |
|---|---|---|---|---|---|---|---|---|---|
| T0 | 3 | 1 | 6 | 8 | 0 | 5 | 6 | 3 | 1 |
| Tm | 5 | 2 | 8 | 12 | 0 | 7 | 9 | 6 | 2 |
| tp | 7 | 3 | 12 | 17 | 0 | 9 | 12 | 8 | 3 |
| | 8-9 | 7-10 | 9-10 | | | | | | |
| T0 | 3 | 8 | 2 | | | | | | |
| Tm | 5 | 15 | 5 | | | | | | |
| tp | 8 | 20 | 6 | | | | | | |

Q6) **Answer the following questions (Any One) Five Marks Each** **5 M**

a) Evaluate the maximum and minimum values of $f(x, y) = 81x^2 + y^2$ subject to the constraint $4x^2 + y^2 = 9$ using Lagrange's Multipliers.

b) Evaluate the non-linear programming problem (NLPP) using Kuhn -Tucker conditions

$$Max \ f(x_1, x_2) = -x_1^2 + 2x_1 + x_2$$
$$subject \ to \ 2x_1 + 3x_2 \leq 6$$
$$2x_1 + x_2 \leq 4 \ , x_1, x_2 \geq 0$$

03 JAN 2023

Deccan Education Society's

**FERGUSSON COLLEGE (Autonomous) Pune-4**

Affiliated to Savitribai Phule Pune University

**M.Sc. (Data Science) (Semester : I)**

**Paper Title: Applied Linear Algebra**

**Paper Code:CSD4102**

**2019 Pattern (4 Credits)**

**[Time: 3 Hours]** **[Max. Marks : 50]**

**Instructions to the candidates:**

1) All questions are compulsory.
2) Neat diagrams must be drawn wherever necessary.
3) Figures to the right side indicate full marks.

Q1) **Answer the following questions (Any Five). Two Marks Each** **10 M**

a) Define, 'Angle between vectors. Find angle between vectors u = (-1, 2 ), v = (2, 3)

b) Find standard deviation of vector a =(1, 2, 4, 1,1)

c) Define an 'orthogonal set'. Is a set S = {(0, 0, 2), (1, -2, 0), (-2, -1, 0)} an orthogonal?

d) Form the convolution vector of vectors u =(1,2,1,4) and v = (2,1,3)

e) Is a real valued function $f(x_1, x_2, x_3) = 2x_1 + x_2 + 3X_3 + 4$ an affine function?

f) Define "Sketch of a data."

Q2) **Answer the following questions (Any Four). Three Marks Each** **12 M**

a) Does vector u = (3,5) belong to linear span of vectors u1 = (1, 1), u2 = ( 2, 1)?

b) If a vector X contains sugar count of 50 patients, where rms(x) = 110. Predict at the most how many patients can have sugar count greater than 150.

c) Define, "Left inverse of a matrix". Also explain which matrices can have left inverse.

d) Compute basis for the row space of matrix $A = \begin{bmatrix} 1 & -1 & 3 \\ 2 & 1 & 6 \\ 3 & 1 & 9 \\ 0 & 4 & 0 \end{bmatrix}$. Hence write rank of A.

e) Consider a vector u = (1, 2, 3, 1, 1).

   a) Express sum of 2nd, 4th components in u as inner product of u with an appropriate vector.

   b) Compute ||-2.x||

Q3) **Answer the following questions (Any Three) Three Marks Each** **9 M**

a) Explain Principal Component Analysis.

b) Calculate convolution of following matrices :

$$A = \begin{bmatrix} 2 & 2 & 1 \\ 1 & -1 & 0 \\ 2 & 1 & 2 \end{bmatrix} \text{ and } B = \begin{bmatrix} 1 & 1 \\ 2 & 0 \end{bmatrix}$$

c) Explain method to select the best possible K for K-means algorithm.

d) Calculate singular values of matrix $A = \begin{bmatrix} 1 & -1 \\ -2 & 2 \\ 2 & -2 \end{bmatrix}$

Q4) **Answer the following questions (Any Two). Three Marks Each** **6 M**

a) Define a running sum function from $R^3$ to $R^3$ and express f as f(X) = AX.

b) Let B = $\{(0, -1, 0), (\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}), (-\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}})\}$ be an orthonormal basis set of $R^3$. Express vector v = (2, 2, 3) as linear combination of vectors in B.

c) Explain, whether set S = {(1, 2, 1), (2, 2, 1), (2, 3, 1)} is linearly independent or dependent in $R^3$.

Q5) **Answer the following questions (Any Two) Four Marks Each** **8 M**

a) Write steps to construct Singular value Decomposition of a matrix.

b) Apply Gram Schmidt process on set{a1 =(1, 1, 1), a2 = (-1, 1, 0), a3 = (-1, 3, 1)} and check whether it is linearly independent or dependent.

c) Compute Eigen values of matrix A and find Eigen space for the smallest Eigen value of it, where $A = \begin{bmatrix} 3 & -2 & 0 \\ -2 & 3 & 0 \\ 0 & 0 & 5 \end{bmatrix}$

Q6) **Answer the following questions (Any One) Five Marks Each** **5 M**

a) Estimate greatest Eigen value and corresponding eigen vector of a matrix $\begin{bmatrix} 6 & 5 \\ 1 & 2 \end{bmatrix}$ by using Power method.

b) Develop QR factorization of matrix $A = \begin{bmatrix} 2 & 3 \\ 2 & 4 \\ 1 & 1 \end{bmatrix}$

17-12-22 2:30 O/C

Deccan Education Society's
**FERGUSSON COLLEGE (Autonomous) Pune-4**
Affiliated to Savitribai Phule Pune University
**M.Sc.–II (Semester:III)**
**Subject: Data Science**
**Paper Title: Deep Learning**
**Paper Code: CSD 5303**
**2019 Pattern (4 Credits)**

**[Time: 3 Hours]** **[Max. Marks: 50]**

**Instructions to the candidates:**

1) All questions are compulsory.
2) Neat diagrams must be drawn wherever necessary.
3) Figures to the right side indicate full marks.

Q1) **Answer the following questions (Any Five).** **[5 x 2 = 10 M]**

a) Name the different applications of Restricted Boltzmann Machine.

b) Match the following:

| Network | Application |
|---|---|
| RNN | Object Localization |
| CNN | Dimensionality reduction |
| RBM | Row Index to Row Value Prediction |
| Autoencoder | Temporal Recommender Systems |

c) Describe the early stopping method.

d) Define: Invariance tolocal translation

e) Describe the various padding strategies in brief.

f) List the component of LSTM.

Q2) **Answer the following questions (Any Four).** **[4 x 3 = 12 M]**

a) Defend the following statement: 'A constant learning rate is not desirable'

b) Discuss how dropout method can be used as regularizer.

c) Explain the concept of parameter sharing.

d) Explain how initialization of weights can affect the performance of ANN

e) Explain semi supervised learning.

Q3) **Answer the following questions (Any Three).** **[3 x 3 = 9 M]**

a) Read the following image

| | | | |
|---|---|---|---|
| 1 | 3 | 2 | 9 |
| 7 | 4 | 1 | 5 |
| 8 | 5 | 2 | 3 |
| 4 | 2 | 1 | 4 |

If the filter size is 2 x 2, compute the output after max pooling operation for:

i. Stride =1

ii. Stride =2

b) Given the truth table for XOR logic gate, use a simple RNN with random weights and compute the output of the hidden layer. (Don't consider the activation function)

c) Infer how GRU is simplified variant of LSTM.

d) Infer the relation between momentum factor and weight adjustment.

Q4) **Answer the following questions (Any Two).** **[2 x 3 = 6 M]**

a) Explain the architecture of LSTM.

b) Differentiate shallow learning with deep learning.

c) Compare feed forward network with RNN.

Q5) **Answer the following questions (Any Two).** **[2 x 4 = 8 M]**

a) Unsupervised pretraining can help to reduce overfitting. Justify

b) Review the concept of backpropagation through time.

c) Justify. Echo state networks simplify RNN.

Q6) **Answer the following questions (Any One).** **[1 x 5 = 5 M]**

a) Write in detail the operations performed in LSTM.

b) Propose how RBM can be used for collaborative filtering.