Subject: **Multivariable Calculus for Data Science**   *Full Marks*: 45
Course No: MDS 554   *Pass Marks*: 22.5
Level: MDS /I Year /II Semester   *Time*: 2 hrs

Candidates are required to give their answer in their own words as far as practicable.

Group A [5×3=15]

1. Use a Riemann sum with $m = 2$, $n = 3$ to estimate the volume under $f(x,y) = 1 - xy^2$ above the Rectangle $R = [0, 4] \times [-1, 2]$. Take the sample points to be (a) the lower right corners and (b) the upper left corners of the rectangles.

2. Evaluate $\int_2^3 \int_{-1}^4 \int_1^0 (4x^2 y - z^3)\, dz\, dy\, dx$.

3. Define Gradient of a vector point function. If $\vec{r} = x\vec{i} + y\vec{j} + z\vec{k}$ and $r = |\vec{r}|$, then prove that
$\nabla r^m = m r^{m-2} \vec{r}$.

4. Define line integral along a curve C. Evaluate the line integral $\int_C \vec{F} \cdot d\vec{r}$ where
$\vec{F} = x^2 y^2 \vec{i} + y\vec{j}$ and C is the curve from $(0, 0)$ to $(4, 4)$ along the parabola $y^2 = 4ax$.

5. State Gauss's divergence theorem. If a closed surface S encloses a volume V and
$\vec{F} = x\vec{i} + 2y\vec{j} + 3z\vec{k}$, Using Gauss's divergence theorem, find the value of $\iint_S \vec{F} \cdot \vec{n}\, ds$.

Group B [5 ×6=30]

6. a) Evaluate $\iint_D 5x^3 \cos y^3\, dA$ where D is the region bounded by $y = 2$, $y = \frac{1}{4} x^2$ and the y-axis.

   b) A lamina occupies the part of the disk $x^2 + y^2 \le 1$. Find its center of mass if the density at any point is proportional to the square of its distance from the origin.

OR

   a) Use a double integral to determine the volume of the solid that is bounded by $z = 8 - x^2 - y^2$ and $z = 3x^2 + 3y^2 - 4$.

   b) Evaluate $\iint_D xy^3\, dA$ where D is the region bounded by $xy = 1, xy = 3, y = 2$ and
   $y = 6$ using the transformation $x = \frac{v}{6u}, y = 2u$.

7. a) Evaluate $\iiint_E (12y - 8x)\,dV$ where E is the region behind $y = 10 - 2z$ and in front of the region in the xz- plane bounded by $z = 2x,\ z = 5$ and $x = 0$.

   b) Use a triple integral to determine the volume of the region below $z = 4 - xy$ and above the region in the xy-plane defined by $0 \le x \le 2,\ 0 \le y \le 1$.

8. a) Evaluate $\iiint_E dV$, where E is the solid enclosed by the ellipsoid $x^2/a^2 + y^2/b^2 + z^2/c^2 = 1$.
   Use the transformation $x = au,\ y = bv,\ z = cw$.

   b) Determine the surface area region formed by the intersection of the two cylinders $x^2 + y^2 = 9$ and $x^2 + z^2 = 9$.

9. Define Surface integral. Let $\vec{G} = (x^2 - yz)\,\vec{i} + (y^2 - zx)\,\vec{j} + (z^2 - xy)\,\vec{k}$ be a vector field.

   Is $\vec{G}$ a irrotational? Justify. Also find a scalar function Ψ such that $\vec{G} = \nabla \Psi$. [1+2+3]

10. a) State Green's theorem in the XY- plane. Use Green's theorem to find the area bounded by the curve $4x^2 + 9y^2 = 36$. [1+2]

    b) Verify Green's theorem in the plane for $\int_C [(3xy - 3x^2)\,dx + (2x + y^2)\,dy]$

    where C is the closed curve given by the line $y = x$ and parabola $x = y^2$. [3]

OR

a) Find the equation of the tangent plane to the surface with parametric equation

   $x = u^2,\ y = v^2$, and $z = u + 2v$. [3]

b) State Stokes' theorem in a surface S. Verify Stokes' theorem for the vector function

   $\vec{F} = x\,\vec{i} + y\,\vec{j}$ around the square boundary $x = 0,\ y = 0,\ x = 2,\ y = 3$. [3]

***

Subject: **Programming with Python**
Course No: **MDS 551**
Level: **MDS /I Year /II Semester**

Full Marks: **45**
Pass Marks: **22.5**
Time: **2 hrs**

Candidates are required to give their answer in their own words as far as practicable.

Group A [5 ×3=15]

1. Why do we need problem analysis before writing programs?

2. Explain identity operators in Python with example.

3. What is list type in python? Compare list with tuple.?

4. Write a program in Python to test whether a number entered is palindrome or not.

5. Why do we need functions in programming? Explain recursive function with example.

Group B [5 × 6 = 30]

6. Explain each type of if statement in Python with example.

OR

How can you read and write files in Python? Explain reading and writing files with example.

7. Explain array broadcasting in NumPy with example. How do you search and sort arrays using NumPy?

OR

What is dictionary data type? Explain its uses with example. What is nested dictionary?

8. Why looping is important in programming? Explain both for and while loop with example.

9. Why pandas is important in data analysis? What is data frame? How do you merge two data frames in Pandas?

10. How can you draw multiple plots in one figure using pyplot? Explain with example, how do you generate bar graphs and pie charts using pyplot?

***

Subject: **Applied Machine Learning**   Full Marks: **45**
Course No: **MDS 552**   Pass Marks: **22.5**
Level: **MDS /I Year / II Semester**   Time: **2hrs**

Candidates are required to give their answers in their own words as far as practicable.
**Attempt ALL questions**

Group A   [5×3=15]

1. Differentiate between K-Means and DBSCAN algorithm.
2. Explain the mathematical model of neuron.
3. What do you mean by exploration and exploitation in Reinforcement Learning? How are they important for agent?
4. Considering grey scale image of size 6 x 6, apply convolution filter of edge detection as below:

$$\begin{bmatrix} 1 & 0 & 1 \\ 0 & 4 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

5. Explain the Markov Decision Process used in RL.

Group B   [5×6=30]

6. What do you mean by dimensionality reduction? How does PCA works? Explain.

OR

7. Explain the Expectation Maximization algorithm for Gaussian Mixture Models.

8. Given a neural network, perform forward propagation considering a classification problem and compute the loss. Consider a sigmoid activation function and learning rate 0.1.
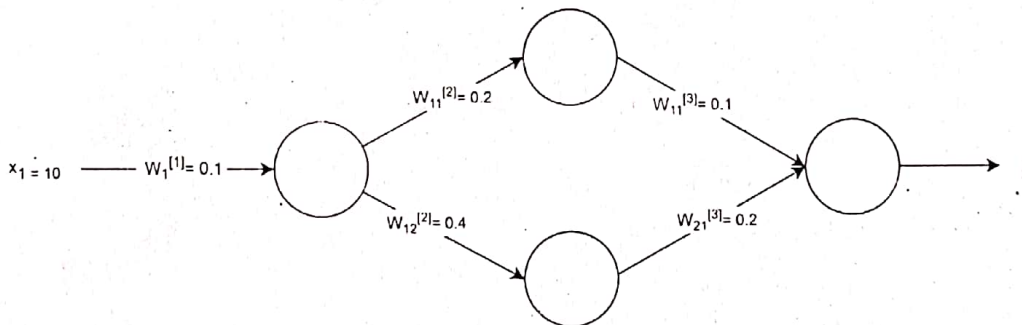
9. Explain the concept of back propagation through time.

OR

Design a convolutional neural network that takes input of 64 x 64 size and the hidden layers are as follows:

Layer 1: 10 filters of size 3 x 3 with padding = 1, Max Pooling 2 x 2
Layer 2: 32 filters of size 5 x 5 with padding = 0, Max Pooling 2 x 2
Layer 3: 5 filters of size 3 x 3 with padding = 0, Average Pooling 2 x 2
Dense Layer: 120 units   Dense Layer: 5 units

10. What do you mean by Policy? Explain the objective and algorithm of Policy Iteration in RL.

***

10. Describe the discretization mridly app in actions mdp with its softenings.

Natural Language Processing   MDS /I Year/II Semester   Full Mark: 45
MDS505   Pass Mark: 22.5
                                                                Time: 2 Hours

Candidates are required to give their answer in their own words as far as practicable.

Attempt ALL question.

Group A [5 × 3 = 15]

1. What is NLP? List out the challenges of NLP?

2. What is Stemmer? Discuss which problem of stemmer is solved by Lemmatizer?

3. What is pragmatics? Why Pragmatics is important in NLP?

4. What is Parsing in NLP? Explain with example.

5. What is WordNet? Discuss the usability of the WordNet.

Group B [5 × 6 = 30]

6. Write note on Part-of-Speech (POS) Tagging.
   OR
   Write a note on Name Entity Recognition (NER).

7. Explain Probabilistic CFG with its application.
   OR
   Define Finite State Machines (FSM). How can we use FSM to analyze presence of the postfix? Explain with examples.

8. Explain the steps involved in NLP based problem solving.

9. List the elements of Semantic Analysis. What is Semantic Analysis? What are the difference between Polysemy and Homonymy? Explain with examples.

10. What is n-gram language model? How can we use this to capture the context of language? Example with example.

***

Tribhuvan University
Institute of Science and Technology
SCHOOL OF MATHEMATICAL SCIENCE
Second Assessment 2080

Subject: **Statistical Method for Data Science**
Course No: **MDS 553**
Level: **MDS / I year/ II Semester**

Full Marks: 45
Pass Mark: 22.5
Time: 2hrs.

Candidates are required to give their answers in their own word as far as practicable

Attempt ALL questions.
Group A   [5 x 3 =15]
1. A random sample of size 14 from a Poisson distribution with mean 'μ' yields the following data values:

10	5	7	7	5	5	9
8	3	7	8	6	8	4

   If the prior distribution of 'μ' is Exponential (0.5), find the posterior distribution of μ.
2. A manufacturer took a sample of 9. The life of the batteries in hours until failure was
      30   40   30   50   42   30   48   46   44
   At the 0.05 level of significance, is there evidence that the mean life of batteries is more than 40 hours?
3. Define compound distribution.
   If X has an exponential distribution with parameter $\alpha=12$ and N has a Poisson distribution with $\lambda=55$.
   Calculate the expected value and variance of S.

4. Prove that for testing of hypothesis $H_0:\theta=\theta_0$ vs $H_1:\theta=\theta_1$, its power is never less than its size
   i.e. $\alpha \leq 1-\beta$.

5. Shoes are produced by two machines A and B. 50% of the shoes were produced by machine A with an estimate of 10% of them being defective. On machine B, 20% of the shoes produced are defective. If a shoe taken at random is found to be defective, what is the probability that the shoe was produced by machine (i) A.? (ii) B.?

Group B   [5×6=30]
6. A continuous random variable X has pdf. $f(x)$ where,
$$f(x)= K x; \quad 0\leq x<1$$
$$\quad\quad K (2-x); \quad 1\leq x<2$$
$$\quad\quad = 0; \text{ otherwise.}$$

   a) Find the value of K
   b) Mean and variance of X
   c) $P(0.75\leq X\leq 1.5)$
OR
Let X has a mixed distribution with distribution function; $F(x)$ defined as follows:
$$F(x)= 0 \text{ if } x\leq 0$$
$$\quad = (2x/3) \text{ if } 0\leq x<1$$
$$\quad = (x+1)/3 \text{ if } 1\leq x<2$$
$$\quad = 1 \text{ if } x\geq 2$$

Find the mean and variance of variable X
7. Let X is distributed as a Poisson distribution with parameter λ and λ itself is distributed as a Gamma distribution with parameter α and β. Find the posterior distribution of λ given x. Also, calculate its mean and variance.

8. Define extreme value distribution in brief and also write the applications of the distribution.

9. Let a random variable has normal distribution with unknown 'μ' and known variance '3' i.e. X~N (μ, 3). Derive the Likelihood Ratio Test (LRT) for testing the null hypothesis H_0: μ=11 against H_1: μ≠11 at a 5% level of significance.

n= 20
s =12 ; x̄
α = 0.01

10. The scores of three matched groups under the six conditions are given below

Group	Conditions					
	I	II	III	IV	V	VI
A	9	5	2	5	6	7
B	6	4	3	4	6	5
C	5	1	3	3	6	5

Analyze the data using Friedman's test to identify if there is significant difference in variation between matched groups. Use 5% level of significance.

OR

A researcher company has designed three different systems to clean up oil spills. The following table contains the results, measured by how much surface area (in square meters) is cleared in 1 hour. The data were found by testing each method in several trials. Are the three systems equally effective? Use Kruskal Wallis H test at α=5%.

System A	55	60	63	56	59	55
System B	57	53	64	49	62	
System B	66	52	61	57		

***