MDS551-2079

Tribhuvan University
Institute of Science and Technology
2079

Master Level / I Year /IInd Semester/ Science
Data Science (MDS 551)
(Programming with Python)

Full Marks: 45
Pass Marks: 22.5
Time: 2 hours

Candidates are required to give their answers in their own words as for as practicable.

Attempt All Questions

Group A
(5×3=15)

1. Why problem analysis is important? What is algorithm? (1.5+1.5)

2. Define variable and data type. What are different numeric types in python? (1+2)

3. Explain the use of continue statement with example. (3)

4. How do you create an array using numpy? Write a program to find sum of two matrices stored in numpy arrays. (1+2)

5. How can you draw multiple plots in one figure using pyplot? Explain with suitable example. (3)

Group B
(5×6=30)

6. Explain tuple data type with example. Compare tuple with list. Explain different ways of slicing lists with example. (1+2+3)

OR

Explain dictionary data type with example. How can you access dictionary items? Explain nested dictionary with example. (2+2+2)

7. Why do we need selection statements in programming? Explain all selection statements used in python. (2+4)

OR

Define looping statement. Why do we need looping statements in programming? Write a program to count number of odd integers in a list. (1+2+3)

8. What is function? Why do we need functions in programming? Explain different ways of passing arguments to functions with example. (1+1+4)

9. How can we read and write files in Python? Write a program that reads data from a text file, counts number of vowels in it, and writes number of vowels in a separate text file. (2+4)

10. Compare pandas data frame with series. How can you read and write CSV file using pandas? How do you merge and join data frames in pandas? Explain with suitable example. (2+2+2)

IOST,TU

MDS I Year, II Semester

MDS552-2079 ✡

Tribhuvan University
Institute of Science and Technology
2079

✡

Master Level / I Year /IInd Semester/ Science
Data Science (MDS 552)
(Applied Machine Learning)
Candidates are required to give their answers in their own words as for as practicable.
Attempt All Questions

Full Marks: 45
Pass Marks: 22.5
Time: 2 hours

Group A   (5×3=15)

1. Differentiate between supervised and unsupervised learning.
2. Discuss the concept of mini-batch gradient descent algorithm.
3. What are various distance measures used for measuring similarity? Explain briefly.
4. What is Dunn Index? Explain.
5. What is meant by activation function? Discuss ReLU and Leaky ReLU activation functions.

Group B   (5×6=30)

6. Derive weight update rule for logistic regression.
   OR
   Fit the logistic regression model through the following data. Show one epoch of training. Assume $\alpha=0.1$.

X_1	0.8	0.4	0.5	0.6
X_2	0.7	0.3	0.8	0.25
Y	0	1	0	1

7. Use Agglomerative clustering algorithm to divide the data points {(2,10), ((2,5), (8,4), (5,8), (7,5), (6,4)} into two clusters.

8. Derive weigh update rule for BPTT?
   OR
   Consider a 2 bit grey level image with resolution 6x6 and a 3x 3 filter whose all diagonal elements are 1 and non-diagonal elements are zero. Compute feature map and then compute pooled feature map using 3x3 window. Use average pooling.

9. Discuss various measures used in multi-class classification with example.

10. Define the terms value and policy functions? Explain value and policy iteration algorithms.

IOST,TU

Scanned with CamScanner

<div align="center">

Tribhuvan University
Institute of Science and Technology
2079
✿

</div>

Master Level / I Year /IInd Semester/ Science   Full Marks: 45
Data Science (MDS 553)   Pass Marks: 22.5
(Statistical Methods for Data Science)   Time: 2 hours

Candidates are required to give their answers in their own words as for as practicable.
Attempt All Questions

<div align="center">

Group A   [5 × 3 = 15]

</div>

1. What do you mean by multinomial distribution? Obtain the expression for its mean.

2. Write the applications of Extreme Value Distribution.

3. Distinguish between prior and posterior distributions with an illustrative example.

4. Show that the Binomial distribution is a special case of Generalized Power Series Distribution.

5. Differentiate between Null and Alternative hypothesis.

<div align="center">

Group B   [5 × 6 = 30]

</div>

6. Let X follows binomial distribution with parameters n and θ, and θ follows Beta distribution with parameters α and β. Find the posterior distribution of θ given x. Also, find the mean and variance of posterior distribution.

7. Let X has a mixed distribution with DF; F(x) defined as follows. Obtain the mean and variance of X.

$$F(x) = \begin{cases} 0 & if \quad x < 0 \\ \dfrac{x^2}{4} & if \quad 0 \le x < 1 \\ \dfrac{x+1}{4} & if \quad 1 \le x < 2 \\ 1 & if \quad x \ge 2. \end{cases}$$

8. An experiment designed to compare three preventive methods against corrosion yielded the following maximum depths of pits (in thousands of an inch) in pieces of wire subjected to the respective treatments:

IOST,TU

Method I	77	54	67	74	71	66	
Method II	60	41	59	65	62	64	52
Method III	49	52	69	47	56		

Use the 0.05 level of significance to test the hypothesis that the three samples come from identical population. Use Kruskal Wallis H test.

9. An agricultural experiment was conducted to compare the yields of paddy at 4 plots in Godawari by using the three different chemical fertilizers Nitrogen (N), Phosphorus (P), and Potash (K). The yields of paddy in (Qtl) were given in the following table:

Chemical Fertilizer	Plot			
	I	II	III	IV
N	122	83	138	121
P	81	89	79	65
K	80	82	65	58

Do the data provide sufficient evidence to support the null hypothesis that the population of yields of paddy corresponding to the three types of fertilizers do not differ in location? Use Friedman two-way ANOVA test.

OR

Two groups of data managers, one group consisting of trained ones, another groups are not trained have the following number of correction required.

| Trained | 78 | 64 | 75 | 45 | 82 |
| Untrained | 110 | 70 | 53 | 51 | |

Use Mann Whitney U test to test if there is a significantly difference between the two average number of correction of trained and untrained data manager.

10. State and Prove Neymann- Pearson's Lemma (N-P Lemma).

OR

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. $\alpha \leq 1 - \beta$.

Tribhuvan University
Institute of Sciences and Technology
2079

Subject: Multivariable Calculus for Data Science
Course No: MDS 554
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.

Attempt ALL questions.

Group A [5×3 =15]

1. Find the normal vector of the space curve $\vec{r} = x\vec{i} + y\vec{j} + z\vec{k}$,where $x = t^2$, $y = t^2, z = t^3$ at point (1, 1, 1).

2. Find the equation of the tangent plane to $z = x^2 \cos(\pi y) - \dfrac{6}{xy^2}$ at point $(2, -1)$.

3. Find and classify all the critical points of the function: $f(x, y) = (y - 2)\, x^2 - y^2$.

4. 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 $0 \le x \le 2, 0 \le y \le 1$.

5. With the help of Gauss's divergence theorem, show that

$$\iint_{S} \vec{F} \cdot \hat{n}\, ds = \frac{4}{3}\pi\,(a + b + c)$$

where $\vec{F} = ax\,\vec{i} + by\,\vec{j} + cz\,\vec{k}$ and S is the surface of the sphere $x^2 + y^2 + z^2 = 1$.

Group B [5 ×6 =30]

6. Establish the vector equation of a straight line through two points $\vec{a}$ and $\vec{b}$. Find the vector equation of the line through the point (2, 1,0) and perpendicular to both the vectors $\vec{k} - 2\vec{j}$ and $\vec{j} + 2\vec{k}$. Also find the scalar and vector projections of $\vec{q} = \vec{i} - \vec{j} + \vec{k}$ onto $\vec{p} = \vec{i} + \vec{j} + \vec{k}$. [2+2+2]

7. Derive the expression for the derivative of vector triple product of three vectors. Find the derivative of the scalar triple product of the vectors
$\vec{p} = (a \cos t, b \sin t, 0)$, $\vec{q} = (-a \sin t, b \cos t, t)$ and $\vec{r} = (1, 2, 3)$ at $t = 0$. [3+3]

8. Prove that if $x = x(t)$ and $y = y(t)$ are differentiable functions of t and $z = f(x, y)$ is a differentiable function of x and y, then $z = f(x(t), y(t))$ is a differentiable function of t and $\dfrac{dz}{dt} = \dfrac{\partial z}{\partial x} \cdot \dfrac{dx}{dt} + \dfrac{\partial z}{\partial y} \cdot \dfrac{dy}{dt}$,

where the ordinary derivatives are evaluated at t and the partial derivatives are evaluated at (x, y). Also, find the maximum rate of change of $f(s, t) = te^{st}$ at point (0, 2) in the direction at which this maximum rate of change occurs. [3+3]

OR

[3+3]

Calculate $\partial z/\partial s$ and $\partial z/\partial t$ using the following functions: $z = e^{x+2y}$, $x = s/t$, $y = t/s$. Also Show that $u = e^{-x}\cos y - e^{-y}\cos x$ satisfies the Laplace equation $u_{xx} + u_{yy} = 0$.

[3+3]

9. Evaluate $\iint_R xy^3\, dA$ where R is the region bounded by $xy = 1$, $xy = 3$, $y = 2$, $y = 6$

using the transformation $x = \dfrac{y}{6u}$, $y = 2u$. Using polar coordinates, find the area of the part of the surface $z = xy$ that lies within the cylinder $x^2 + y^2 = 1$. [3+3]

10. State Green's theorem in the plane and use it to find the area of the circle of radius 4 unit. Verify Green's theorem in the plane for $\int_C (2xy - x^2)\, dx + (x + y^2)\, dy$ where C is the closed curve given by the line $y = x$ and parabola $x = y^2$.

[1+2+3]

OR

Find the equation of the tangent plane to the surface with parametric equation $x = u^2$, $y = v^2$, and $z = u + 2v$. Verify Stokes' theorem for the vector function

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

***

Tribhuvan University
Institute of Science and Technology
2079
✡

Master Level / I Year /IInd Semester/ Science
<u>Data Science (MDS 556)</u>
<u>(Artificial Intelligence)</u>

Full Marks: 45
Pass Marks: 22.5
Time: 2 hours

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

<div align="center">

Group A
</div>

[5 × 3 = 15]

1. How Total Turing Test is used to assess artificial intelligence?

2. How simulated annealing handles the incompleteness problem of hill climbing search?

3. How knowledge is represented in semantic networks? Represent following knowledge using semantic network; [1+2]
   Gunnu is girl and she has fair complexion. Gunnu studies in class one at Gems school. Gems school is a type of private school. The default fee at all the schools is 1000. Alita has obtained A grade which is less than the grade obtained by Gunnu.

4. What is the role of activation function in artificial neural networks? List any two activation functions. [2+1]

5. Describe machine vision with its steps.

<div align="center">

Group B
</div>

[5 × 6 = 30]

6. Differentiate informed search from uninformed search. Describe with an example how uniform cost search algorithm works? [2+4]

<div align="center">OR</div>

   How performance of search algorithm is determined? Describe how searching is done in And-OR search trees? [2+4]

7. How knowledge is inferred in propositional logic? Using resolution, try to infer "Ram is not happy" from the following knowledge base; [2+4]
   If Ram is happy then Ram is healthy. Ram is either athlete or he is student. An athlete is not happy. Ram will go for shopping if he is healthy.

<div align="center">OR</div>

   What is resolution algorithm in predicate logic? Given following knowledge base; [2+4]

IOST,TU

Everyone who has disease is sick. Ram has a disease. Covid is disease. Everyone who have Covid have fever. Now try to infer "Ram is sick and Ram has fever" using resolution.

8. Define intelligent agent. Given following, classify to which agent type they belong   [1+5]
   a. A route finding agent which selects path having best path cost.
   b. A COVID prediction agent based on medical history
   c. A Grade calculator based on formula

9. How genetic algorithm works? Given a function $f(x)=x^3$ where x ranges from 0 to 15. Now show a single iteration of genetic algorithm to maximize f(x). Use your own required assumptions.
                                                                                        [3+3]

10. How syntax and semantic analysis are done in natural language processing? Support your answer with examples.   [4+2]