

Tribhuvan University
 Institute of Science and Technology
SCHOOL OF MATHEMATICAL SCIENCES
 First Assessment 2079

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 answer in their own words as far as practicable.

Attempt All Questions

Group A [5 × 3 = 15]

1. Why data normalization is important? Explain Min-max and Z-score data normalization.
2. Discuss the concept of model overfitting and underfitting.
3. Discuss the concept of Bagging in ensemble learning.
4. How K-Means++ differs from K-Means? Explain.
5. Why SVD is used? Discuss its working.

Group B [5 × 6 = 30]

6. Derive weigh update rule for logistic regression using gradient descent.

OR

Discuss the concept of locally weighted linear regression. Consider a query point $x=5$ and let $x^1=6$, $x^2=4$, and $x^3=3$ are three points in the training set. Find Cost function for the locally weighted linear regression.

7. Consider the following dataset. Find Information of gain of each attribute and draw decision tree for first iteration.

Weather	Temperature	Wind Level	Go Out (Class)
Sunny	High	Low	No
Sunny	Normal	Normal	Yes
Cloudy	High	Normal	No
Cloudy	Normal	High	Yes
Sunny	Normal	High	No
Rainy	High	Normal	No
Rainy	Low	High	No

8. Continue question number 7 and construct complete decision tree. Then, predict class label for the tuple {Cloudy, Low, Normal}.
9. When DBSCAN algorithm gives better performance than other clustering algorithms? Discuss its working.

OR

Divide the data points $\{(2,10), (2,5), (8,4), (5,8), (6,4)\}$ into two clusters using agglomerative clustering.

10. What are the features of principle components? Discuss the working of PCA algorithm.

Tribhuvan University
 Institute of Sciences and Technology
SCHOOL OF MATHEMATICAL SCIENCES
First Assessment 2079

Subject: Multivariable Calculus for Data Science

Full Marks: 45

Course No.: MSMT 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. Define scalar triple product of three vectors and give its geometrical meaning. Using scalar triple product, verify that the vectors $\vec{u} = \vec{i} + 5\vec{j} - 2\vec{k}$, $\vec{v} = 3\vec{i} - \vec{j}$, and $\vec{w} = 5\vec{i} + 9\vec{j} - 4\vec{k}$ are coplanar. [1+2]
2. Find the parametric equations and symmetric equation for the lines through (2,1,0) and perpendicular to both the vectors $\vec{i} + \vec{j}$ and $\vec{j} + \vec{k}$. [3]
3. Define curvature of a vector function $\vec{r} = \vec{r}(t)$. Find the curvature of the vector function $\vec{r}(t) = p \cos t \vec{i} + p \sin t \vec{j}$, where p is constant. [1+2]
4. Find the limit, if it exists, or show that the limit does not exist:
 a) $f(x, y) = \frac{5y^4 \cos 2x}{x^4 + y^4}$ b) $f(x, y) = \frac{xy}{\sqrt{x^2 + y^2}}$ [1.5+1.5]
5. Explain why the function $f(x, y) = \sqrt{x + e^{4y}}$ is differentiable at the given point (3, 0). Find the linearization $L(x, y)$ of the function $f(x, y)$ at that point. [1+2]

Group B [5 × 6=30]

6. Prove the Parallelogram Law $|\vec{a} + \vec{b}|^2 + |\vec{a} - \vec{b}|^2 = 2|\vec{a}|^2 + 2|\vec{b}|^2$ for any two vectors \vec{a} and \vec{b} . Give its geometric interpretation. Also if $\vec{a} + \vec{b}$ and $\vec{a} - \vec{b}$ are orthogonal, show that the vectors \vec{a} and \vec{b} must have the same length. [2+1+3]

OR

Derive a vector equation of a straight line

- a) through the given point \vec{a} and parallel to the vector \vec{b}
- b) through two points \vec{a} and \vec{b} . Also, find a vector equation for the line through the point (1, 0, 6) and perpendicular to the plane $x + 3y + z = 5$. [2+2+2]
7. Derive the expression for the derivative of scalar triple product of three vectors. Find the derivative of the scalar triple product of the vectors $t\vec{i} + t^2\vec{j} + t\vec{k}$, $(t+1)\vec{i} + (t+2)\vec{j} - 3t\vec{k}$ and $t^2\vec{i} + 2t\vec{j} + t\vec{k}$ at $t = 2$. [2+4]
8. Find the domain and range of the function $f(x, y) = \sqrt{16 - 4x^2 - y^2}$. Describe the graph of f . Sketch a contour map of this surface using level curves corresponding to $c = 1, 2, 3, 4, 5$. [1.5+1.5+1.5+1.5]
9. a) Let $f(x, y)$ be defined on an open disk D that contains the point (a, b) . Prove that if the functions f_x and f_y are continuous on D , then $f_{xy}(a, b) = f_{yx}(a, b)$ [4]

- b) Show that $z = e^x \sin y$ satisfies the equation $\frac{\partial^2 z}{\partial x^2} + \frac{\partial^2 z}{\partial y^2} = 0$. [2]

10. a) 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

$$\frac{dz}{dt} = \frac{\partial z}{\partial x} \cdot \frac{dx}{dt} + \frac{\partial z}{\partial y} \cdot \frac{dy}{dt},$$

where the ordinary derivatives are evaluated at t and the partial derivatives are evaluated at (x, y) . [3]

- b) Calculate $\frac{\partial z}{\partial u}$ and $\frac{\partial z}{\partial v}$ using the following functions:

$$z = f(x, y) = 3x^2 - 2xy + y^2, x = x(u, v) = 3u + 2v, y = y(u, v) = 4u - v. \quad [3]$$

OR

- a) Prove that if f is a differentiable function of x and y , then f has a directional derivative in the direction of any unit vector $u = (a, b)$ and $D_u f(x, y) = f_x(x, y)a + f_y(x, y)b$. [3]

- b) Find the direction for which the directional derivative of $f(x, y) = 3x^2 - 4xy + 2y^2$ at $(-2, 3)$ is a maximum. What is the maximum value? Find the maximum rate of change of $f(x, y) = \sqrt{x^2 + y^4}$ at $(-2, 3)$ and the direction in which this maximum rate of change occurs. [3]

Tribhuvan University
 Institute of Sciences and Technology
SCHOOL OF MATHEMATICAL SCIENCES
 First Assessment 2079

Subject: Statistical Methods for Data Science

Full Marks: 45

Course No: MDS 553

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.

Attempt ALL Questions.

Group A [5 × 3 = 15]

- There card players play a series of matches, the probability that A will win any game is 20%, the probability that player B will is 30%, and the probability that the player C will win is 50%. If they play 6 games, what is the probability that player A will win 1 game, player B will win 2 games and player C will win 3 games?
- Obtain the mean and variance of multinomial distribution.
- Differentiate between parametric and non parametric test.
- What is hypothesis? Differentiate between simple and composite hypothesis.
- What do you understand by Most Powerful Test (MPT)

Group B [5 × 6 = 30]

- What do you know about multinomial distribution? Obtain the moment generating function of multinomial distribution.
- In a certain computer hardware manufacturing industry six different types of machines are working to cut pieces of wires. The number of wires of unequal length recorded in a day is as follows:

Machine	1	2	3	4	5	6
No. of wire	2	0	4	8	5	11

Do these data provide sufficient evidence that the six machines equally cut the wires of unequal length? Apply Kolmogorov Smirnov test at 5% level of significance.

OR

The heart beating rate of 5 vegetarians and 5 non vegetarians are recorded below:

Vegetarians	56	67	82	60	75
Non vegetarians	53	42	75	58	65

Is the mean heart beating rate of non vegetarians significantly high? Use Mann Whitney U test.

- Poverty Alleviation fund has provided grants on income generation program to different districts. A sample of 5 districts from Terai region, 9 from Hilly region and 6 from Himali region are selected and grants on income generation activities (in million Rs.) on different districts were recorded as follows:-

Himali	7	33	111	39	72	128			
Hilly	176	266	213	135	95	54	86	75	45
Terai	138	290	66	98	208				

Use Kruskal Wallis H test to test whether there is any significant difference in the mean grants in three geographical regions

9. A survey was conducted in four hospitals in a Kathmandu to obtain the number of babies born over a 12 months period. This time period was divided into four seasons to test the hypothesis that the birth rate is constant over all the four seasons. The results of the survey were as follows:

Hospital	No. of births			
	Winter	Spring	Summer	Fall
A	92	72	94	77
B	15	16	10	17
C	58	71	51	62
D	19	26	20	18

Analyze the data using Friedman two way ANOVA test.

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

OR

Find BCR in a normal distribution to test $H_0: \mu = \mu_0$ vs $H_1: \mu = \mu_1$. Given, probability of type I error = α and $\sigma = 1$.

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. What is problem analysis? Why do we need it?
2. Define variable. Explain rules for defining python variables with example. What is global variable?
3. Write a program to test whether a number is even or odd.
4. Write a program to count number of vowels in a string.
5. Explain recursive function with example.

Group B [5 × 6 = 30]

6. Explain selection statements in python with suitable example.

OR

- ✓ 7. What is looping statement? Compare for loop with while loop.
- ✓ 7. Explain list data type with example. How list is different from tuple? What is list comprehension?

OR

- ✓ 8. Explain dictionary data type with example. What is nested dictionary?
- ✓ 8. Why do we need function in programming? Explain different ways of passing arguments in functions.
- ✓ 9. How do you read and write CSV files in python? Explain.
10. Explain the use of break and continue statements in programming. What is nested control statement?

Tribhuvan University
 Institute of Sciences and Technology
SCHOOL OF MATHEMATICAL SCIENCES
 First Assessment 2079

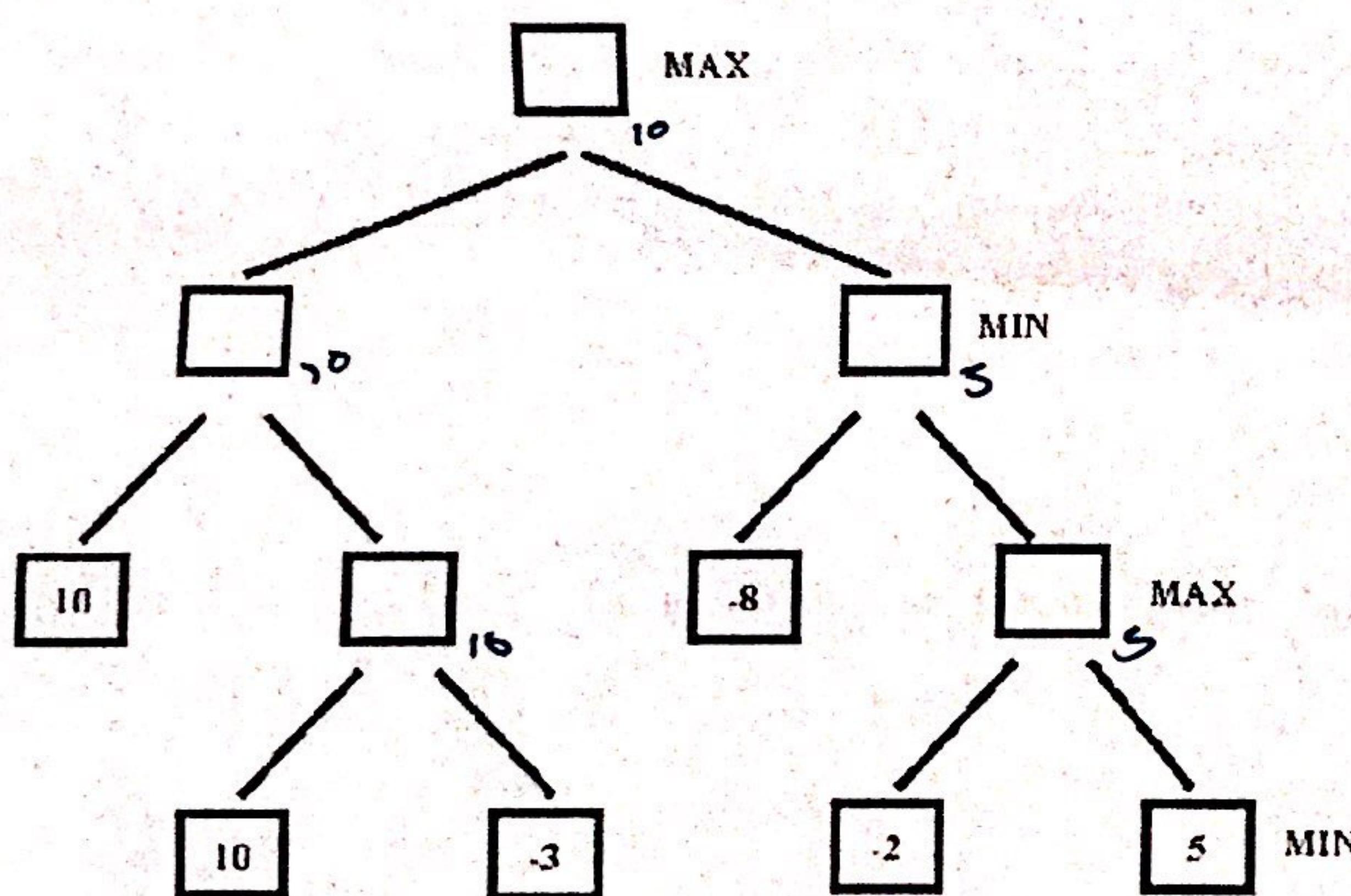
Subject: Artificial Intelligence
Course No. MDS 556
Level: MDS / I Year/II Semester

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. How Turing test is used to measure intelligence from the dimension of acting humanly?
2. Classify which of following is model based, simple reflex, learning, utility based agent. Give a very brief description to your answer.
 - a) A disease prediction system
 - b) A jurisdiction agent with repository of cases
 - c) A game agent with rewards
3. Given following state space, show how Minimax search can be used;



4. How knowledge is inferred in semantic networks? Represent following using semantic network;

All humans are mammal
All humans have eyes.
Weight of Ram, who is human, is larger than weight of Jelly.
Fishes have red color
Jelly is a fish and has weight 1pounds
Ram belongs to QA team from May 1st to May 21st

5. Describe the mathematical model of ANN.

Group B [5 × 6 = 30]

6. How A* search guarantees better solution than greedy best first search? Support your answer by constructing a state space with proper heuristics and actual costs and illustrate that A* returns better path than greedy best first.

OR

Justify A* search guarantees to give solution while greedy best first search does not. Support your answer by constructing a state space with proper heuristics and actual costs and illustrate that A* returns goal but greedy best first does not.

7. What is skolemization? Construct FOPL for following statements;

All students are talent person. Every talent person are smart. Someone who is not smart is dumb. Ram is dumb but he is talent person. Try to infer Ram is student using resolution.

OR

Write resolution algorithm for propositional logic. Construct propositional statements for given knowledge;

MDS is best course in Nepal. If MDS is best course in Nepal then employment in Nepal is high. Employment in Nepal is high and Salary in Nepal is best.

Now using resolution algorithm try to infer that "If MDS is not best course in Nepal and employment in Nepal is high then salary in Nepal is not best".

8. How uncertain knowledge is encoded using Bayesian networks? Consider following scenario, construct equivalent Bayesian network.

During a medical checkup of a patient, doctor suspects three diseases as the cause of the condition. The suspected diseases are covid, viral, typhoid which are independent of each other. There are four symptoms fever, stiff neck, body pain, cough which the doctor wants to check for presence in order to find the most probable cause of the condition. The symptoms are conditionally dependent to the three diseases as fever depends on all of three. stiff neck depends on typhoid. body pain depends on covid and typhoid, whereas cough depends only on viral. Assume all random variables are Boolean, they are either 'true' or 'false'.

9. Using your own assumptions, configure PEAS framework for,

- a) Weather forecasting agent
- b) Vote counting agent

10. What does learning in ANN means? How perceptron learning is used to train ANN?
