LAIKIPIA UNIVERSITY

UNIVERSITY EXAMINATIONS

SECOND SEMESTER 2023/2024 ACADEMIC YEAR

FOURTH YEAR EXAMINATION FOR THE DEGREE OF BACHELOR OF COMPUTER SCIENCE

COMP 428: MACHINE LEARNING AND ANALYTICS

STREAM: R ***TIME: 2 HRS***

DAY: MONDAY [2.30 – 4.30 P.M] ***DATE: 15/04/2024***

THIS QUESTION PAPER CONSISTS OF EIGHT (8) PAGES

PLEASE DO NOT OPEN UNTIL THE INVIGILATOR SAYS SO.

SECTION A: (COMPULSORY) TOTAL MARKS FOR THIS SECTION IS 30.

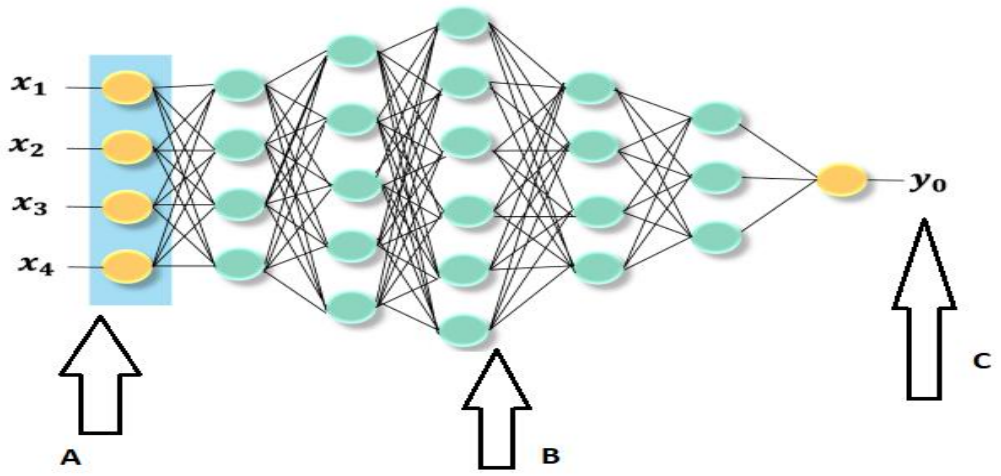
QUESTION ONE (30 MARKS)

a) If we want to visit a new place, we take help of Google Maps, which shows us the correct path with the shortest route and predicts the traffic conditions. It may also predicts the traffic conditions such as whether traffic is cleared, slow-moving, or heavily congested, explain two technologies that makes Google maps achieve this**.** **(4 Marks)**

b) Natural language processing (NLP) refers to the branch of computer science concerned with giving computers the ability to understand text and spoken words in much the same way human beings can ,Explain any **FIVE** challenges of NLP **(5 Marks)**

c) Deep learning is a subset of machine learning which provides the ability to machine to perform

Human-like tasks without human involvement. It provides the ability to an AI agent to mimic the human brain. Explain the label A B C below **(6 Marks)**

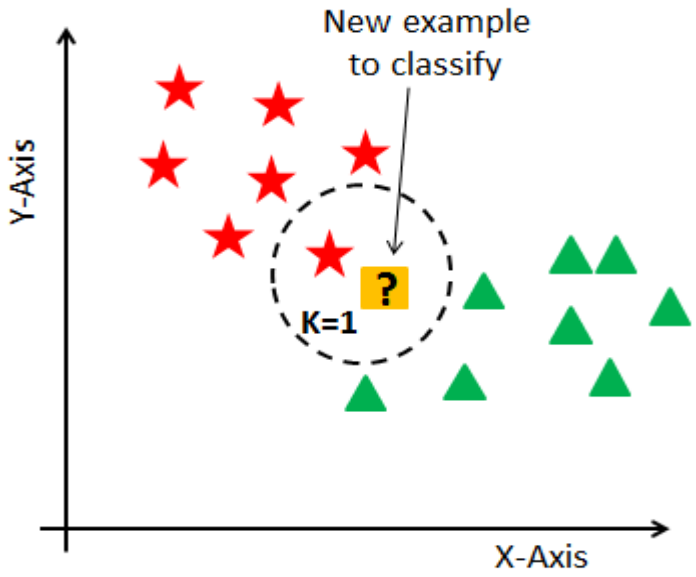


d) A genetic algorithm is a search heuristic that is inspired by Charles Darwin's theory of natural evolution. This algorithm reflects the process of natural selection where the fittest individuals are selected for reproduction in order to produce offspring of the next generation , explain the stages of genetic algorithm (**7 Marks**)

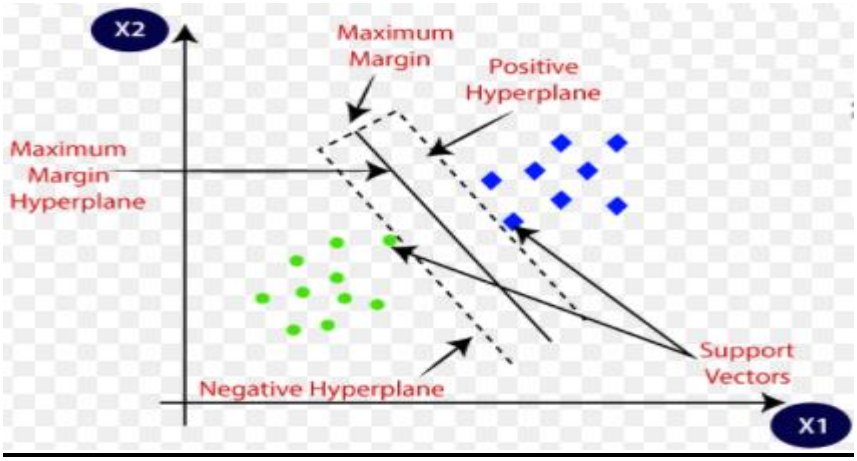
e) A 4 - input neuron has weight 1 2 3 4 , The transfer function is linear with a constant of prortionality of 2 , the inputs are , 4, 10 , 5 and 20, Calculate the output **(4 Marks)**

f) Identify the machine learning technique and explain how they work **(4 Marks)**

i.


New example
to classify
Y-Axis
?
K=1
X-Axis

ii.


X2
Maximum
Margin
Positive
Hyperplane
Maximum
Margin
Hyperplane
Support
Vectors
Negative Hyperplane
X1

SECTION B. TOTAL MARKS FOR THIS SECTION IS 40.

ANSWER ANY TWO QUESTIONS FROM THIS SECTION. EACH QUESTION IN THIS SECTION CARRIES 20 MARKS.

QUESTION TWO (20 MARKS)

a) Suppose you are given the following data set in CSV format

Index	Country	Age	Salary	Purchased
0	India	38	68000	No
1	France	43	45000	Yes
2	Germany	30	54000	No
3	France	48	65000	No
4	Germany	40	nan	Yes
5	India	35	58000	Yes
6	Germany	nan	53000	No
7	France	49	79000	Yes
8	India	50	88000	No
9	France	37	77000	Yes

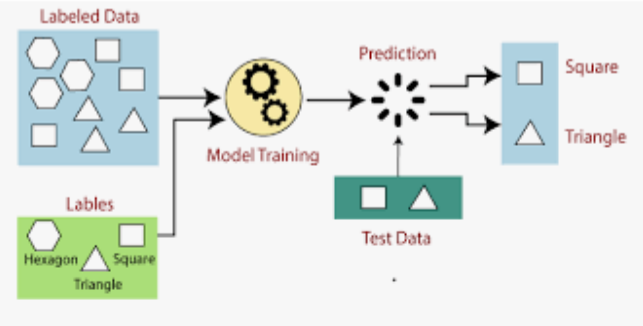
i) Write a python program to load the data set make assumption where necessary **(4 Marks)**

ii) Write a program to extract dependent and independent variable **(2 Marks)**

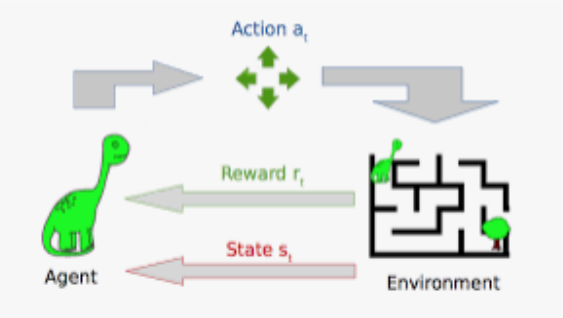
iii) Write a program to handle missing data **(2 Marks)**

b) One of the application areas of naïve bayes algorithm is sentiment analysis, using an example explain the meaning of the term sentimental analysis. **(4 Marks)**

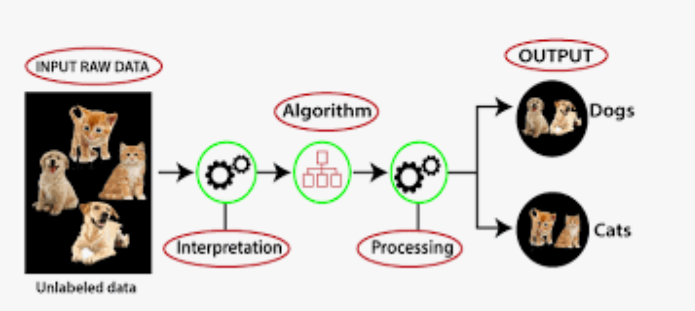
c) Identify the machine learning techniques **A B C** **(6 Marks)**



A



B



C

d) Describe Convolution Neuron Networks **(2 Marks)**

QUESTION THREE (20 MARKS)

a) A recruitment agency faces a challenge of determining the salary of an employee after several years of experience, write a program in python that will help the agency achieve the following tasks based on the following data set

Years of experience	Salary($)
1	110,000
2	120,000
3	130,000
4	140,000
5	150,000
6	160,000
7	170,000
8	180,000
9	190,000
10	200,000

i) Import the necessary libraries **(2 Marks)**

ii) Load data training set **(2 Marks)**

iii) Find out how many rows and columns are in the dataset **(2 Marks)**

iv) Find out the statistical summary **(2 Marks)**

v) Visualize the data set **(2 Marks)**

b) Differentiate between overfitting and under fitting in machine learning **(4 Marks)**

c) Suppose there is a candidate who has a job offer and wants to decide whether he should accept the offer or Not. Using decision tree illustration help the candidate to solve this problem with the following attributes; Salary attribute as Attribute Selection Measure-ASM, distance from the office and availability of Cab facility. **(4 Marks)**

d) Explain the use of Tensorflow in machine learning **(2 Marks)**

QUESTION FOUR (20 MARKS)

a) A patient has been suffering from shortness of breath (called dyspnea) and visits the doctor, worried that he has lung cancer. The doctor knows that other diseases, such as tuberculosis and bronchitis are possible causes, as well as lung cancer. She also knows that other relevant information includes whether or not the patient is a smoker (increasing the

chances of cancer and bronchitis) and what sort of air pollution he has been exposed to. A positive X-Ray would indicate either TB or lung cancer.

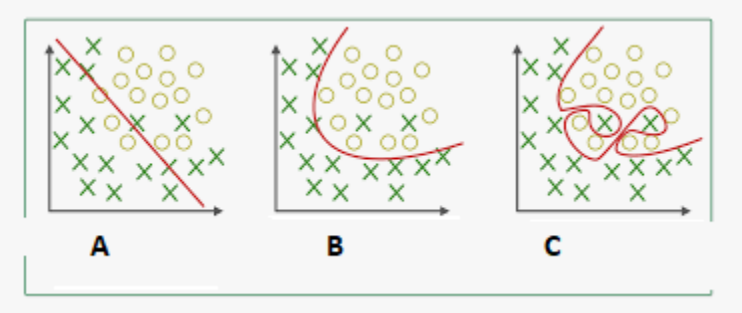
i) What are the nodes, names to represent and what values can they take? **(4 Marks)**

ii) Prepare a conditional probability table (CPT) for the above scenario **(4 Marks)**

b) Describe the use of ReLu activation function **(2 Marks**

c) Using python programming demonstrate how you can Split the dataset into training and Testing set **(4 Marks)**

d) Data fitting is the process of fitting models to data and analyzing the accuracy of the fit. Engineers and scientists use data fitting techniques, including mathematical equations and nonparametric methods, to model acquired data. Identify and explain the following data fitting models **(6 Marks)**



QUESTION FIVE (20 MARKS)

a) Suppose, we have an image of an object that looks similar to cat and dog as shown below , and we want to predict whether it is a cat or dog, which machine learning algorithm is suitable for this scenario. Explain your choice **(4 Marks)**

b. TensorFlow is a machine learning framework ,developed by **Google Brain Team**, list and explain the components of TensorFlow **(6 Marks)**

c. A human being can quickly identify objects without much effort. It is an effortless task for us, but it is a difficult task for a computer, explain the difference between face recognition and detection as used in image processing. **(4 Marks)**

d. An environment is everything in the world which surrounds the agent, but it is not a part of an agent itself. Explain any two agent environments. **(4 Marks)**

e. Using an illustration explain an intelligent agent **(2 Marks)**