Name: **Enrolment No:**	UPES

UNIVERSITY OF PETROLEUM AND ENERGY STUDIES
Online End Semester Examination, May 2021

Course: Machine Learning (AI-Minor) **Semester: IV**
Program: B. Tech. (All Branches) **Time 03 hrs.**
Course Code: MRAI0202 **Max. Marks: 100**
Instructions: Attempt all the questions

SECTION A **(6 X 5 = 30 Marks)**

S. No.		Marks	CO
Q 1	(a) Machine learning is _____ field. (i) Inter-disciplinary (ii) Single (iii) Multi-disciplinary (iv) All of the above (b) If the regression involves three independent variable, it is called as (i) Multiple regression (ii) Simple regression (iii) One regression (iv) None of these (c) Which library in python programming language contains machine learning algorithms? (i) Matplotlib (ii) sklearn (iii) TensorFlow (iv) Pandas (d) In which technique 70-80% of data is used for training and rest for testing? (i) Hold-out method (ii) K-fold cross validation method (iii) Bootstrap sampling (iv) None of these (e) Which regression algorithm predict the likelihood of output to be 1 given certain value of input (i) Linear regression (ii) Polynomial regression (iii) Logistic regression (iv) None of these	**5 M**	**CO1**
Q 2	(a) Which of the following are critical aspects of learning in ANN (i) Number of layers (ii) Number of nodes in each layer (iii) Interconnection weights (iv) All of these (b) In back propagation algorithm, multiple iterations are known as (i) Degree (ii) Epoch (iii) Cardinality (iv) None of these (c) Shallow neural networks generally have maximum of _____ hidden layers. (i) 1 (ii) 2 (iii) 3 (iv) None of these (d) Which of the following activation function is continuous and differentiable (i) Sigmoid (ii) Threshold (iii) Piecewise linear (iv) None of these (e) ANN is made of_____ neuron. (i) Biological (ii) Digital (iii) Logical (iv) None of these	**5 M**	**CO2**
Q 3	(a) The distance between hyperplane and data points is called as (i) Hyper Plane (ii) Margin (iii) Error (iv) Support Vectors (b) Which of the following is not a supervised classification algorithms? (i) Density based algorithm (ii) Random Forest (iii) SVM (iv) Decision Trees (c) In a classification problem if actual values is [1,1,1,0,1,0,1,1] and the predicted	**5 M**	**CO3**

<table>
<tr><td></td><td>values is [1,0,1,1,0,0,1,0]. Jaccard Index is?
(i) 0.6 (ii) 0.23 (iii) 0.33 (iv) 0.42
(d) Which of the following will be Euclidean distance between the two data points A (3, 3) and B (2, 3)?
(i) 5 (ii) 4 (iii) 2 (iv) 1
(e) Which of the following is a performance measure for classification problem?
(i) R^2 (ii) RAE (iii) RMSE (iv) Confusion matrix</td><td></td><td></td></tr>
<tr><td>Q 4</td><td>(a) Agglomerative clustering algorithm is an example of which type of clustering method?
(i) Hierarchical (ii) Partitioning (iii) Density-based (iv) None of these
(b) When the hierarchical clustering algorithm uses the maximum distance to measure the distance between clusters then it is called as
(i) Single linkage algorithm (ii) Complete linkage algorithm
(iii) Double linkage algorithm (iv) None of these
(c) Which clustering technique identify arbitrary shaped clusters.
(i) Hierarchical (ii) k-means (iii) Density-based (iv) None of these
(d) DBSCAN algorithm stands for_____
(e) Which clustering algorithm calculate the sum square error to measure the quality of clusters?
(i) Hierarchical (ii) k-means (iii) Density-based (iv) None of these</td><td>5 M</td><td>CO4</td></tr>
<tr><td>Q 5</td><td>Differentiate supervised and unsupervised machine learning algorithm with examples.</td><td>5 M</td><td>CO3</td></tr>
<tr><td>Q 6</td><td>List the steps carried out to implement machine learning problem in python programming language.</td><td>5 M</td><td>CO1</td></tr>
<tr><td colspan="4">SECTION B (5 X 10 = 50 Marks)</td></tr>
<tr><td>Q 7</td><td>Differentiate the following algorithms:
(a) Agglomerative and Divisive clustering
(b) k-means and density based clustering</td><td>10 M</td><td>CO4</td></tr>
<tr><td>Q 8</td><td>What are the different types of activation function in neural network? Explain its importance. Enumerate and explain the steps in the back propagation algorithm used to train the network.</td><td>10 M</td><td>CO2</td></tr>
<tr><td>Q 9</td><td>Following is the training data for a group of athletes. Based on this data, use k-NN algorithm and classify Sayan (Weight=56kg; Speed=10kmph) as a Good, Average or Poor sprinter.
<table>
<tr><th>Name</th><th>Weight (kg)</th><th>Speed (kmph)</th><th>Class</th></tr>
<tr><td>Nitesh</td><td>55</td><td>9</td><td>Average</td></tr>
<tr><td>Gurpreet</td><td>58</td><td>8</td><td>Poor</td></tr>
<tr><td>Goutam</td><td>60</td><td>7.5</td><td>Poor</td></tr>
<tr><td>Gulshan</td><td>59</td><td>8.5</td><td>Average</td></tr>
<tr><td>Mohit</td><td>57</td><td>10</td><td>Good</td></tr>
<tr><td>Sahil</td><td>53</td><td>10.5</td><td>Good</td></tr>
<tr><td>Samyak</td><td>53</td><td>10</td><td>Good</td></tr>
</table></td><td>10 M</td><td>CO3</td></tr>
<tr><td>Q 10</td><td>Differentiate the following terms:
(a) Qualitative data and Quantitative data</td><td>10 M</td><td>CO1</td></tr>
</table>

	(b) k-fold cross validation and Bootstrap sampling		
Q 11	Explain the McCulloch-Pitts model of neuron. Design a multi-layer perceptron to implement A XOR B.	**10 M**	**CO2**

SECTION C **(1 X 20 = 20 Marks)**

Q 12

(a) Following is the training data for a group of athletes. Based on this data, use k-NN algorithm and classify Sayan (Weight=56kg; Speed=10kmph) as a Good, Average or Poor sprinter.

Name	Weight (kg)	Speed (kmph)	Class
Nitesh	55	9	Average
Gurpreet	58	8	Poor
Goutam	60	7.5	Poor
Gulshan	59	8.5	Average
Mohit	57	10	Good
Sahil	53	10.5	Good
Samyak	53	10	Good

(b) A leading software company INFOSYS is coming in UPES for hiring B. Tech. students. This company last year selected the students based on certain criteria and is given in table 1. It is to be evaluated using decision tree algorithm that Karan, a student of UPES, wants to find out if he may be offered a job in INFOSYS. His CGPA is low, Communication is Good, Aptitude-High, Programming skills-Bad.

CGPA	Communication	Aptitude	Programming Skills	Job Offered?
High	Good	High	Good	Yes
Medium	Good	High	Good	Yes
Low	Bad	Low	Good	No
Low	Good	Low	Bad	No
High	Good	High	Bad	Yes
High	Good	High	Good	Yes
Medium	Bad	Low	Bad	No
Medium	Bad	Low	Good	No
High	Bad	High	Good	Yes
Medium	Good	High	Good	Yes
Low	Bad	High	Bad	No
Low	Bad	High	Bad	No
Medium	Good	High	Bad	Yes
Low	Good	Low	Good	No
High	Bad	Low	Bad	No
Medium	Bad	High	Good	No
High	Bad	Low	Bad	No
Medium	Good	High	Bad	Yes

Table 1

20 M **CO3**