Department of Computer Science and Engineering
University of Barishal
Final Examination-2021
Course Title: Machine Learning and Data Mining
Course Code: CSE-4213
4th Year 2nd Semester
Admission Session: 2017-18

Time: 03 Hours

Marks: 60

Answer any five Questions from the followings.

1. a) What are the differences between supervised and unsupervised methods? [2]

   b) How does supervised learning work? [4]

   c) Describe the steps involved in ML life cycle. [6]

2. a) Describe the methods used to split the dataset in machine learning. [3]

   b) Why data preprocessing is important in machine learning? Describe three popular data preprocessing techniques in machine learning. [4]

   c) Describe the different performance metrics to evaluate the algorithms. Explain with numerical examples. [5]

3. a) What do you mean by Convolutional Neural Network? [3]

   b) Why do we prefer Convolutional Neural networks (CNN) over Artificial Neural networks (ANN) for image data as input? [3]

   c) Explain the different layers in CNN. [6]

4. a) What are Support Vectors in SVMs? Why SVM is an example of a large margin classifier? [2]

   b) What are *Hard-Margin* and *Soft-Margin* SVMs? [3]

   c) Use a support vector machine to classify the following dataset: [7]

Class	X1	X2
+	1	1
+	2	2
+	2	0
-	0	0
-	1	0
-	0	1

   i) Plot the six training sets on the X1-X2 axis.

   ii) Separate the classes with maximum margin separator and give your choice for vector *w* and intercept *b*.

5. a) What is K-Nearest Neighbor (KNN) Algorithm for Machine Learning? How is KNN different from k-means clustering? [3]

   b) Discuss the maximizing the margin process in Support Vector machine (SVM) with necessary figures and equations. [4]

c) For a SunBurn dataset given below, construct a decision tree. [5]

Name	Hair	Height	Weight	Location	Class
Amir	Blonde	Average	Light	No	No
Babar	Brown	Tall	Average	Yes	Yes
Sakib	Blonde	Short	Heavy	Yes	No
Mushfik	Red	Short	Average	No	Yes
Hadi	Brown	Average	Heavy	No	No
Jeleous	Brown	Tall	Light	No	Yes
Helal	Blonde	Tall	Heavy	Yes	No
Akbar	Red	Short	Heavy	No	No

6. a) What is the basic difference of Recurrent Neural Network (RNN) from Artificial Neural Network? Give some examples of RNN application. [4]

b) What is the vanishing Gradient Problem of RNN? [3]

c) Shortly describe Long-Term Memory (LSTM) networks and explain how it can avoid long term dependency problem. [5]

7. a) What is the difference between Random Forest and Decision Trees? [4]

b) Consider the below sample data set. In this data set, we have four predictor variables, namely: Weight, Blood flow, Blocked Arteries, Chest Pain [8]

Blood Flow	Blocked Arteries	Chest Pain	Weight	Heart Disease
Abnormal	No	No	130	No
Normal	Yes	Yes	195	Yes
Normal	No	Yes	218	No
Abnormal	Yes	Yes	180	Yes

Creating A Random Forest for the above data and check the following query.

Blood Flow	Blocked Arteries	Chest Pain	Weight	Heart Disease
Normal	No	Yes	218	No

8. a) Compute the squared distance matrix on given the data from Table-1. [4]

	X_1	X_2
A	0	0
B	0	1
C	-1	2
D	2	0
E	3	0
F	4	-1

Table-1

b) Perform K-means clustering on the dataset from Table 1. Use the first and last data points as initial centers (K = 2). Given the final parameters, which cluster would belong to $x\binom{1}{1}$? [8]

"Education is the most powerful weapon which you can use to change the world."
- Nelson Mandela