Shahjalal University of Science and Technology
Department of Computer Science and Engineering
4th Year 1st Semester Final Examination June 2021 (Session: 2017-18)
Course Code: CSE 475 Credits: 3 Course Title: Machine Learning
Time: 3 hrs Total Marks: 100

Group A

[Answer all the questions]

1. Answer any FIVE **5x2=10**

a) A Perceptron gives a linear decision boundary – Justify.
b) Using a kernel we get the benefit of higher dimension, without projecting data to higher dimension – Justify.
c) Differentiate between classification and regression.
d) Define eigen values and eigen vectors.
e) What is regularization?
f) What is representation learning?
g) What is the difference between classical machine learning and deep learning?
h) What criteria should an activation function have?

2. Answer any FOUR **4x5=20**

a) Briefly explain how the Q-learning algorithm works.
b) Briefly describe the gradient descent optimization and discuss the effect of different learning rate.
c) What is a kernel function? With an appropriate example explain how using a kernel makes a non-linearly separable dataset linearly separable.
d) What are the common practices to avoid Over-fitting?
e) Explain what bias-variance trade-off is. How to choose the right model?
f) i. What is convolution? Apply convolution on the following using 'valid' padding and stride 1.

10	20	30
40	50	60
70	80	20

Input

1	-1
-0.5	1

Kernel

ii. What is pooling? Apply max pooling on the result of the previous convolution.

3. Answer any TWO **2x10=20**

a) Assume you own a video-sharing platform WeTube and here's the watch-and-action history of the user SaradinWeTubePoreThaki.

Video	Type	Length (Min)	#Views	Recommended by a friend	Action
1	Drama	40	1000	No	Liked
2	Vlog	10	500	No	Disliked
3	Cat Video	5	50000	Yes	Liked
4	Tech News	15	2000	Yes	Disliked
5	Food Recipe	10	400	No	No-Reaction
6	Drama	75	2000	No	Liked
...	...	...	...	...	...
9999	Pottery	8	10	No	?
10000	Music	5	1200	Yes	?

You value her time, so you want to predict her action given the history using a neural network to recommend her good videos in the future.

i. Draw a suitable neural network for predicting the actions. Use appropriate labels and biases etc. 3
ii. Show the weight matrices with proper initialization. 2
iii. For each layer write what activation function(s) are suitable and why. 1
iv. What loss function can be used here? 1
v. How much training data are needed for making a reasonable prediction? 1
vi. What techniques you can apply to optimize the training? 2

b) Consider the classification task of the animal kingdom given below using Neural Networks and answer the following questions.

	Animal	Give Birth	Can Fly	Live in Water	Have Legs	Class
1	Human	Yes	No	No	Yes	Mammal
2	Python	No	No	No	No	Non-Mammal
3	Whale	Yes	No	Yes	No	Mammal
4	Frog	No	No	Sometimes	Yes	Non-Mammal
5	Bat	Yes	Yes	No	Yes	Mammal

i. How many learning parameters are there in the simplest possible network? 2
ii. How different types of feature values are fed into the network? 2
iii. How to initialize the network? 2
iv. How much training data are needed for a fair classification? 2
v. How long do you need to train the network?

c) 8

i. Draw the architecture of a CNN where:
- Input is an 8x8 gray scale image.
- There are two convolution layers-
  - In the 1st convolution layer a 2x2 kernel is used with 'valid' padding and stride 2.
  - In the 2nd convolution layer, two 2x2 kernels are used with 'same' padding and stride 1.
- Then max pooling is done by a 2x2 filter.
- The output of the pooling is flattened and then a dense layer (mention what number of neurons you used) is applied with two hidden layers.
- There are three output classes.

ii. What activation function(s) are used? 1
iii. What is the total number of learning parameters in the entire network? 1

Group B

[Answer all the questions]

4. Answer any FIVE 5x2=10

a) What are the major aspects of learning?
b) What is machine learning?
c) Name the steps of a machine learning process.
d) Define parametric vs. non-parametric models with example.
e) What is over-fitting?
f) What is the limitation of a single layer Neural Network (Perceptron)?
g) What is a loss function? Name some common loss functions.
h) Increasing the number of hidden layers increases the prediction accuracy – justify.

5. Answer any FOUR 4x5=20

a) What is Clustering? Write the pros and cons of the K-means algorithm.
b) Define training set, features, response variable, generalization, and curse of dimensionality.
c) Derive the expression for a Naive Bayes' Classifier from conditional probability.
d) What is simple linear regression? Consider the following points for linear regression: {(0, 1), (2, 1), (2, -7), (4, -7)}. Find the equation of the regression line, show all the calculations.
e) Draw a simple neural network for classifying the OR function.
f) What are support vectors? Plot the data points (0, 1, Red), (1, 0, Red), (1, 1, Red), (0, 2, Blue), (2, 0, Blue), (2, 2, Blue). Are the classes linearly separable? What kernel(s) are suitable here for classifying the dataset using support vector machine?

6. Answer any TWO 2x10=20

a) Consider your activities of last ten days:

Deadline	Weather	Mood	Activity
Urgent	Good	Off	Play
Urgent	Bad	Off	Study
Near	Good	Off	Play
None	Good	On	Play
None	Bad	Off	Sleep
None	Good	On	Play
Near	Bad	On	Study
Near	Bad	Off	Sleep
Near	Good	Off	Play
Urgent	Bad	On	Study

i. Calculate the prior probabilities: P(Play), P(Study), P(Sleep)
ii. Given that there is a deadline near, the weather is bad and your mood is off, find out what you will do today using the Naïve Bayes' classifier. 1.5 8.5

b) Suppose you have written a classifier for email spam filtering. You have tested your classifier with some data and got the following predictions. Compute the Confusion Matrix from the output and calculate **Accuracy, Sensitivity, Specificity, Precision** and **f1 score** from that matrix. 10

	Target	Prediction
1	Spam	Not Spam
2	Spam	Spam
3	Not Spam	Not Spam
4	Spam	Spam
5	Not Spam	Spam
6	Not Spam	Not Spam
7	Not Spam	Not Spam
8	Spam	Spam
9	Spam	Not Spam
10	Not Spam	Not Spam

c) i. Explain what the curse of dimensionality is. 2

ii. Assume there are 7 features in a dataset. We want to reduce it to a 2D dataset using PCA. Write down all the mathematical steps to do that. In each step show the dimensions of all the intermediate vectors and matrices. 6

iii. Write the steps if we want to restore the original dataset. 2