Course No: CSE490
Full Marks: 60

Time: 03 hours

N.B.
i) Answ
ii) All

Bangabandhu Sheikh Mujibur Rahman Science and Technology University
Department of computer Science & Engineering
4th year 2nd Semester B.Sc. Engineering Final Examination-2018
Course Name: Machine Learning
Full Marks: 60
Course No.: CSE490
N.B.:
Times: 3(Three) Hours

i. Answer SIX questions, taking any THREE from each section.
ii. All questions are of equal values
iii. Use separate answer script for each section.

Section -A

1. a) What are the main goals of machine learning? Write the five applications of machine learning. 3

   b) What is linear regression? Find maximum likelihood for linear regression. 7

2. a) Write down the properties of maximum likelihood estimators. 2

   b) What is the predicted probability of a 25 yr. old having pregnancy success with first ART attempt? 2

   c) Considering three random variables representing color, size and shape of a fruit, each taking three values (color takes red, green and blue, size takes small, large and medium, and shape takes circle, rectangle and triangle). Assume prior probability of your own for each case and then classify of a new instance (green, large and circle) with Naïve Bayes classifier. 6

3. a) Write down the advantages and disadvantages of naïve Bayesian classifier. 3

   b) Write down the K Nearest Neighbor (KNN) algorithm. Briefly explain the KNN for interpolation. 7

4. a) Compare supervised and unsupervised learning. 3

   b) How can you make a decision tree using information gain based on the following table? 7

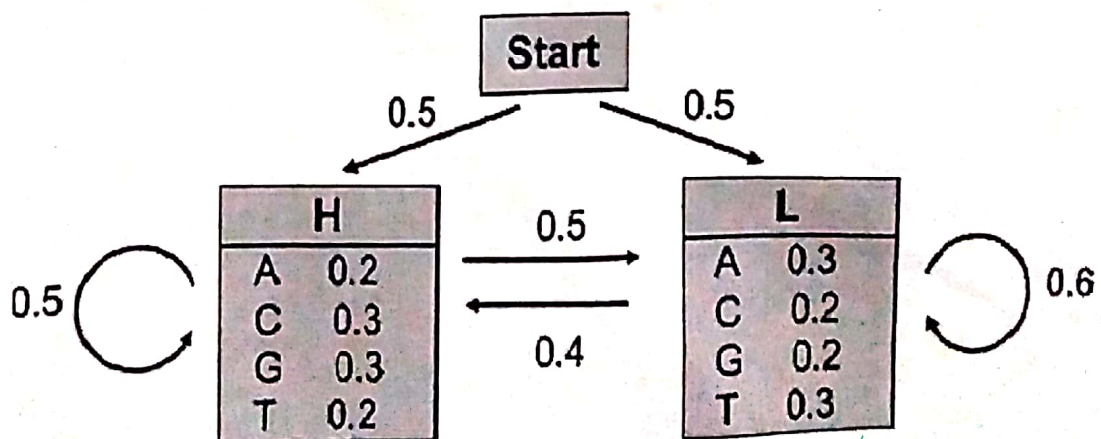| Predictors				Target
Outlook	Temp	Humidity	Windy	Play Golf
Rainy	Hot	High	False	No
Rainy	Hot	High	True	No
Overcast	Hot	High	False	Yes
Sunny	Mild	High	False	Yes
Sunny	Cool	Normal	False	Yes
Sunny	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Rainy	Mild	High	False	No
Rainy	Cool	Normal	False	Yes
Sunny	Mild	Normal	False	Yes
Rainy	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Sunny	Mild	High	True	No

N.B.

i) Answer any **SIX (6)** from the following **EIGHT (8)** questions.

ii) All questions are of equal values.

Section-B

5. a) How does the perceptron's training algorithm work in case of the logical AND operation? 2

   b) Why is multi-layer learning better than two layer learning in ANN? 2

   c) Write down the back propagation by explicitly mentioning the input, hidden and output layers computations. 6

6. a) What is support vector machine? Write down the application of SVM. 2

   b) How does support vector machine work? Explain with example. 3

   c) Write down the strength and weakness of k-means clustering. 2

   d) What is clustering? Discuss different types of clustering. 3

7. a) What is dimension reduction? 1

   b) Discuss the steps of PAC learning model. 3

   c) Calculate F-score for the following table. 3

Actual Class\Predicted class	cancer = yes	cancer = no	Total
cancer = yes	90	210	300
cancer = no	140	9560	9700
Total	230	9770	10000

   d) Explain the reinforcement learning. 3

8. a) What is hidden markov model? Explain with example. 1

   b) Consider the sequence S=GGCACTGAA from the following figure. Calculate the most probable path using the Viterbi algorithm. 6

   c) Define prior probability, posterior probability and likelihood. 3

Bangabandhu Sheikh Mujibur Rahman Science and Technology University
Department of Computer Science & Engineering
4th Year 2nd Semester B.Sc. Engineering Examination-2019

Course No: CSE490
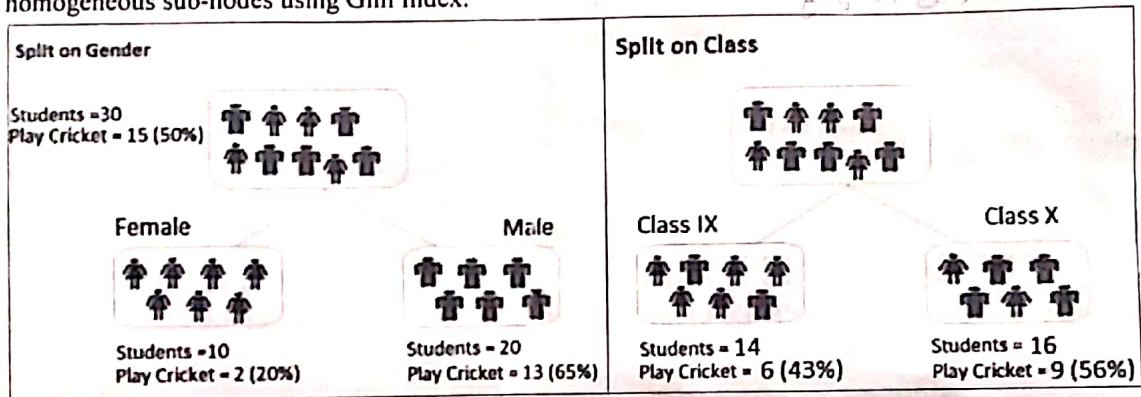Full Marks: 60

Course Title: **Machine Learning**
Time: **03 hours**

N.B.
i) Answer any **SIX (6)** from the following **EIGHT (8)** questions.
ii) All questions are of equal values.

Q.1 (a) What is Machine Learning? Show the workflow of Machine Learning with a diagram? 1+3
(b) Classify different types of Machine Learning. 3
(c) Distinguish between clustering and classification with an example. 3

Q.2 (a) Decision tree splits the nodes on all available variables and then selects the split which results in most 4
homogeneous sub-nodes. Referring to example in the snapshot below, where we want to segregate the
students based on target variable (playing cricket or not). In the snapshot, we split the population using
two input variables Gender and Class. Now, you need to identify which split is producing more
homogeneous sub-nodes using Gini Index.

Split on Gender

Students =30
Play Cricket = 15 (50%)

Female

Students =10
Play Cricket = 2 (20%)

Male

Students = 20
Play Cricket = 13 (65%)

Split on Class

Class IX

Students = 14
Play Cricket = 6 (43%)

Class X

Students = 16
Play Cricket = 9 (56%)

(b) What is the core idea of k Nearest Neighbors (kNN) learning? How to decide the class label in kNN? 2+2
(c) How do you choose the number of neighbors in kNN? 2

Q.3 (a) Let us consider the following weather data. You need to calculate the probability of playing cricket 4
when the weather is overcast using Naïve Bayes Classifier.

Weather	Play	Weather	Play	Weather	Play	Weather	Play
Sunny	Yes	Overcast	Yes	Sunny	Yes	Overcast	No
Sunny	Yes	Rainy	No	Overcast	Yes	Rainy	Yes
Overcast	Yes	Rainy	No	Sunny	Yes		
Rainy	No	Sunny	Yes	Rainy	No		

(b) When classification accuracy is not enough to evaluate a model? 2
(c) Consider the following confusion matrix for a binary classification model. Find accuracy, precision, 4
recall and F1 score of it.

Actual Values

	Positive (1)	Negative (0)
Predicted Values Positive (1)	55	2
Predicted Values Negative (0)	5	38

Q.4 (a) Write the Random Forests algorithm. What are the differences between Random Forests and Decision
Trees? 4
(b) How does Support Vector Machines (SVMs) work? How to identify the right hyper-plane in SVM? 4
(c) How can SVM classify non-linearly separable data? 2

Q.5 (a) What is regression analysis? Why do you need to use regression analysis?
   (b) Explain simple linear regression and multiple linear regression.
   (c) Suppose you are given two list of values for a regression model (as follow):

Original values:	-2	1	-3	2	3	5	4	6	5	6	7
Predicted values:	-1	1	-2	2	3	4	4	5	5	7	7

Find MAE, MSE, and RMSE to evaluate the model.

Q.6 (a) Write down the properties of maximum likelihood estimators.
   (b) Consider the data in the following table:

ID	Height	Age	Weight
1	5	45	77
2	5.11	26	47
3	5.6	30	55
4	5.9	34	59
5	4.8	40	72
6	5.8	36	60
7	5.3	19	40
8	5.8	28	60
9	5.5	23	45
10	5.6	32	58
11	5.5	38	??

Predict the weight of ID 11 based on their height and age using KNN.
   (c) What do you mean by feed forward network?

Q.7 (a) What is the criteria for good clustering?
   (b) Write the k-Means clustering algorithm. What are the strength and weakness of k-Means clustering
   (c) Consider the following proximity matrix. Find the clusters by applying single link and complete lin
   hierarchical clustering. Also show your results by drawing a dendrogram.

Point	P1	P2	P3	P4	P5
P1	0	0.10	0.41	0.55	0.35
P2	0.10	0	0.64	0.47	0.98
P3	0.41	0.64	0	0.44	0.85
P4	0.55	0.47	0.44	0	0.76
P5	0.35	0.98	0.85	0.76	0

Diff... ___ ... ___ Whi.. ___ more desirable and why?

	0.55	0.47	0.44	0	0.76
P4					
P5	0.35	0.98	0.85	0.76	0

Q.8 (a) Differentiate learning and generalization. Which one is more desirable and why?   3

(b).   5

A=0.35

B=0.9

Input
Node

0.1

0.8

0.4

0.6

0.3

0.9

Output

Output
Node

Hidden
Node
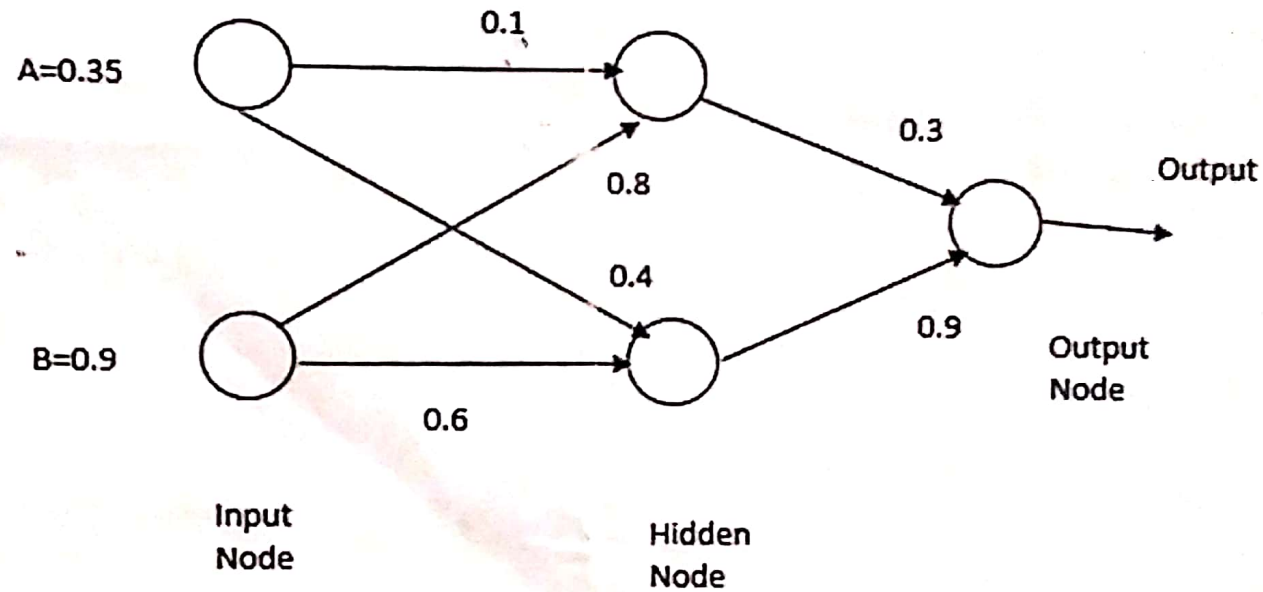
i. Perform a forward pass on the network.
ii. Perform a backward pass on the network.
iii. Perform a further forward pass and comment on the result.
Use sigmoid function as activation function.

(c) How do you differentiate regression and classification problems in Machine Learning? Give one   2
practical example.

i. Answer **SIX** questions, taking any **THREE** from each section.
ii. All questions are of equal values
iii. Use a **separate answering script** for each section.

Section- A

1. a) Define supervised and unsupervised classification.
   b) Classification is a two step process? Do you agree or not? Explain briefly.
   c) What is information gain? How can you make a decision tree based on the following table

| age	income	student	credit_rating	buys_computer
<=30	high	no	fair	no
<=30	high	no	excellent	no
31...40	high	no	fair	yes
>40	medium	no	fair	yes
>40	low	yes	fair	yes
>40	low	yes	excellent	no
31...40	low	yes	excellent	yes
<=30	medium	no	fair	no
<=30	low	yes	fair	yes
>40	medium	yes	fair	yes
<=30	medium	yes	excellent	yes
31...40	medium	no	excellent	yes
31...40	high	yes	fair	yes
>40	medium	no	excellent	no

2. a) Define the terms loss, risk, decision rule and bayes risk.
   b) What is probability density function? Define minimum error rate classification.
   c) What is discriminant function? Explain how Bayesian can help for multi-classification problem?

3. a) What do you understand by overfitting and tree pruning?
   b) State the *Bayes Rule* and explain how it is applied to pattern classification problems.
   c) Define naive bayes classifier? For the above table 1(c) find class label if data sample X = (age <=30, Income = high, Student = yes).

c

4. a) How can we measure accuracy, error rate, sensitivity and specificity, precision and recall for the following table?

Actual Class\Predicted class	cancer = yes	cancer = no	Total
cancer = yes	90	210	300
cancer = no	140	9560	9700
Total	230	9770	10000

   b) What do you understand by Bagging, Boostin, Adaboost and Random Forest? Explain with appropriate example and equation.

Section - B

5. a) What is pattern clustering? How it differs fromclassification?
   b) Write a short note on General theory of Bayesian Parameter estimation.
   c) Discuss different attribute selection measures.

6. a) How a multi-layer neural network works? Explain with figure.
   b) What is SVM? Why is SVM effective on high dimensional data?
   c) What are the difference between SVM and Neural Network?

7. a) What is Hidden Markov Model? For the following figure calculate P({'Dry','Rain'})
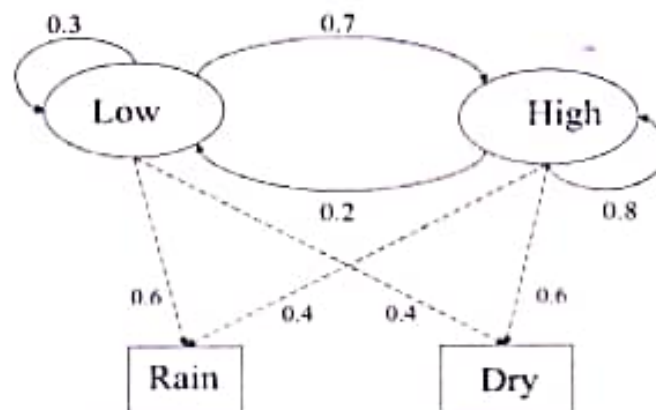
   b) What is Baum-Welch algorithm?
   c) What is the problem of the K-Means method? How can we solve this problem?

8. a) Explain Principal Component Analysis (PCA) with analytical treatment.
   b) Discuss Strength and Weakness of CLIQUE.
   c) What is the basic concept of partition algorithms?
   (d) Clustering is an unsupervised learning method. Do you agree or not. Why?

Course Title: Machine Learning   Course No: CSE 440
Total Marks: 60   Time: 3 (three) Hours

N.B.
i) Answer SIX questions taking any THREE from each Section
ii) All questions are of equal values.
iii) Use separate answer script for each section

Section-A

1. a) What is Machine Learning (ML)? ML is an isolated topic. Do you agree or not? Explain   3
       briefly.
    b) What is information gain? How can you make a decision tree based on the following   7
       table

age	income	student	credit_rating	buys_computer
<=30	high	no	fair	no
<=30	high	no	excellent	no
31...40	high	no	fair	yes
>40	medium	no	fair	yes
>40	low	yes	fair	yes
>40	low	yes	excellent	no
31...40	low	yes	excellent	yes
<=30	medium	no	fair	no
<=30	low	yes	fair	yes
>40	medium	yes	fair	yes
<=30	medium	yes	excellent	yes
31...40	medium	no	excellent	yes

2. a) What are supervised and unsupervised learning?   3
    b) Discuss different kinds of attribute selection measurements technique.   4
    c) How can we enhance basic decision tree induction?   3

3. a) What is the quality for good clustering? What are the requirements and challenges of 3
       clustering?
    b) Define naive bayes classifier? For the above table 1(b) find class label if data sample 6
       X = (age <=30, Income = high, Student = yes, Credit rating= excellent).
    c) What is zero-probability problem?   1

4. a) Define Hidden Markov models.   6
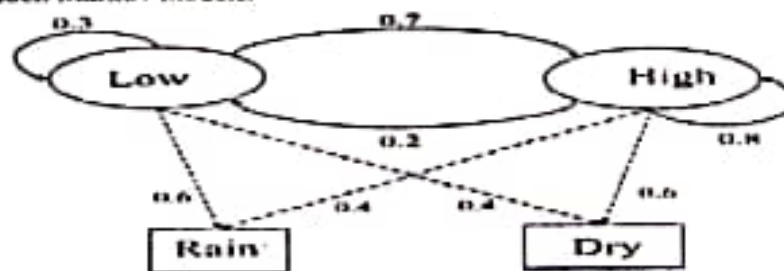

Figure-4(a)

       From figure-4(a),Calculate a probability of a sequence of observations is ('Rain','Rain').
       Where initial probability of two state 'Low' and 'High' is equal.

    b) What do you mean by SVM?   4

Section-B

5. a) How can we measure accuracy, error rate, sensitivity and specificity, precision and recall for the following table?   6

Actual Class\Predicted class	cancer = yes	cancer = no	Total
cancer = yes	110	190	300
cancer = no	240	9460	9700
Total	350	9650	10000

b) Describe k-nearest neighbor method.   4

6. a) Naive Bayes allows computing $P(Y|X)$ by learning $P(Y)$ and $P(X|Y)$. Why not learn $P(Y|X)$ directly? How can we learn $P(Y|X)$ directly? Explain with appropriate equations.   5

b) Define back propagation rule.   5

7. a) What do you mean by sample dimension? How reduce dimension of sample?   2
   b) What is a Networked Virtual Environment (NVE)? What are the elements and components of NVE?   3
   c) What is ensemble learning? Discuss different kinds of ensemble learning?   5

8. a) Describe reinforcement learning.   4
   b) What do you mean by conditional probability? How Bayes probability can be used as classification.   4
   c) What are the advantages and disadvantages of Naive Bayesian Classifier?   2

Course Title: Machine Learning
Course Code: CSE 490
Full Marks: 60
Time: 3(Three) Hours
N.B.
i) Answer SIX questions, taking any THREE from each section,
ii) All questions are of equal values,
iii) Use separate answer script for each section.

SECTION-A

1. (a) What is SVM? Why is SVM effective on high dimensional data?   2

   (b) Following is a data set that contains two attributes, X and Y , and two class labels, "+" and "-". Each   8 attribute can take three different values: 0,1, or 2. The concept for the "+" class is Y=1 and the concept for the "-" class is X=0 v X=2.

X	Y	Number of Instances	
		+	-
0	0	0	100
1	0	0	0
2	0	0	100
0	1	10	100
1	1	10	0
2	1	10	100
0	2	0	100
1	2	0	0
2	2	0	100

   I. Build a decision tree on the data set. Does the tree capture the "+" and "-" concept?

   II. What are the accuracy, precision, recall, and F_1 measure of the decision tree? (Note that precision, recall, and F_1 measure are defined with respect to the "+" class).

2. (a) What do you understand by Bagging and Boosting?   2

   (b) Consider the deterministic reinforcement environment drawn below (let $\gamma = 0.5$). The numbers on the   8 arcs indicate the immediate rewards.
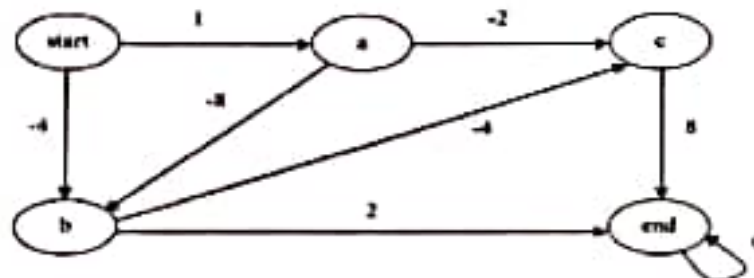
   Assume we use a Q table for this task and initialize all of its entries to the immediate rewards obtained from each state-action pair.

   i. A learner follows the path $start \rightarrow a \rightarrow b \rightarrow end$. Using one-step, regular Q learning, which entries in the Q table change? Show the calculations that produce the new Q table entries.

   ii. Which of the learner's steps in Part I are exploration steps? Explain.

3. (a) Give an example of a dataset to which you could apply linear regression. State what the goal of applying   3 linear regression to that dataset is, and make up a few numerical values for data points to make it clear what the data could look like.

   (b) What is VC-dimension? Briefly explain PAC learning model.   3

   (c) Calculate precision, recall and accuracy for the following table.   4

Actual Class\Predicted class	cancer = yes	cancer = no	Total
cancer = yes	90	210	300
cancer = no	140	9560	9700
Total	230	9770	10000

4. (a) Consider the following examples and the kernel function, K that counts the number of features t, which two examples have the same value.

	F1	F2	F3	F4	Output
Ex1	T	F	F	F	T
Ex2	T	T	T	F	F
Ex3	F	F	F	F	T

I. What is the kernel matrix for this situation? Be sure to label and explain the axes.

II. Assuming that Ex2 and Ex3 turn out to be the only support vectors for this task, write down the learned decision function, using variables where insufficient information is provided. Be sure to explain your answer.

III. Explain how support vector machines can be viewed as following the minimal description-length principle.

(b) Calculate the probability of a sequence of observation, {'Dry','Rain'} where initial probability of two states are equal. **4**

5. (a) If Pr(response|trt)=0.50 and Pr(response|placebo)=0.25, then calculate OR(trt vs. placebo). **2**

(b) You are given m data points, Is it possible to train error converges to the true error as $m \to \infty$? If yes how? If not why? **3**

(c) Suppose you have the following training set with three Boolean input x, y and z, and a Boolean output U. Suppose you have to predict U using a naive Bayes classifier, **3**

x	y	z	U
1	0	0	0
0	1	1	0
0	0	1	0
1	0	0	1
0	0	1	1
0	1	0	1
1	1	0	1

I. After learning is complete what would be the predicted probability $P(U=0|x=0, y=1, z=0)$?

II. Using the probabilities obtained during the Bayes Classifier training, what would be the predicted probability $P(U=0|x=0)$?

(d) How a multi-layer neural network works? Explain with figure. **2**

6. (a) You have a dataset that involves four features. Feature D's values are in [0,100]. For the other three features, all of their possible values appear in this dataset. **4**

	A	B	C	D	Category
Ex1	T	X	F	75	true
Ex2	F	Y	T	20	false
Ex3	T	Z	T	10	true
Ex4	F	Y	T	35	false
Ex5	F	X	T	90	false
Ex6	T	Z	F	50	false

What are the three nearest-neighbors to Example 6? Explain. If this example was in the test set instead of the training set, would k-NN predict it correctly (using k=3)?

(b) What loss (objective) function should we use to judge the fit? **2**