UNIVERSITY OF RUHUNA

Faculty of Engineering

End-Semester 7 Examination in Engineering: March 2021

Module Number: EE7206 **Module Name: Machine Learning**

[Three Hours]

[Answer **all questions**, each question carries **10** marks]
ATTACH this Question paper with the answer script

Index No: _____/_____/_____

Q1 a) (i) Briefly explain supervised, unsupervised and reinforcement learning. Give one (1) example for each type.

[1.5 Marks]

(ii) Briefly explain the difference between training and validation data set.

[0.5 Marks]

b) Briefly explain support vector machine (SVM) algorithm.

[1.0 Mark]

c) Briefly explain the differences between principal component analysis (PCA) and support vector machine (SVM).

[2.0 Marks]

d) Briefly describe one (1) difference between linear regression and logistic regression.

[0.5 Marks]

e) Draw an actor-critic reinforcement learning architecture and briefly describe the actor and critic functions.

[1.0 Mark]

f) **Circle one (1) correct answer on the question paper itself.**
As the number of training examples goes to infinity, your model trained on that data will have
(A) Lower variance
(B) Higher variance
(C) Same variance
(D) It depends on the features selected

[0.5 Marks]

g) Show the derivation for the update equations for $J(\theta_1, \theta_2) = {\theta_1}^2 + {\theta_2}^2$ using multivariate gradient descent.

[1.0 Mark]

h) There are a set of data from patients who have visited Karapitiya Teaching Hospital during the year 2020. A set of features (e.g., temperature, height) have been extracted for each patient. Our goal is to decide whether a new visiting patient has any of diabetes, heart disease, or COVID-19 (a patient can have one or more of these diseases).

(i) Would you use a separate neural network for each of the diseases or a single neural network with one output neuron for each disease? Justify your answer.

[1.0 Mark]

(ii) Some patient features are expensive to collect (e.g., PCR) whereas others are not (e.g., temperature). Therefore, we have decided to first ask our classification algorithm to predict whether a patient has a disease, and if the classifier is 80% confident that the patient has a disease, then we will do additional examinations to collect additional patient features. In this case, which classification methods do you recommend: neural networks, decision tree, or naïve Bayes? Justify your answer.

[1.0 Mark]

Q2 a) Convolutional neural network (CNN) and Long short-term memory (LSTM) are two algorithms based on artificial neural networks (ANN) and backpropagation. List the **full names** of four (4) other algorithms based on ANN.

[1.0 Mark]

b) (i) What is 'overfitting'?
(ii) How do you detect 'overfitting'?
(iii) Briefly describe two (2) ways to handle overfitting.

[2.0 Marks]

c) Briefly explain one (1) way to handle mismatched classes.

[1.0 Mark]

d) Briefly describe the optimization algorithm for gradient descent.

[1.0 Mark]

e) Write the pseudo code for k-means algorithm.

[1.0 Marks]

f) List four (4) different types of clustering models and an example algorithm for each type of model.

[1.0 Mark]

g) Briefly explain the importance of two (2) methods used for data pre-processing.

[1.0 Mark]

h) Can you represent the boolean function in Table Q2 with a single unit from a neural network? If yes, show the architecture and calculation. If not, briefly explain why not.

[2.0 Marks]

Table Q2

A	B	output
1	1	0
0	0	0
1	0	1
0	1	0

Q3 a) List four (4) main factors one should consider when selecting a suitable machine learning algorithm for a given application. Briefly describe the importance of each factor.

[2.0 Marks]

b) Describe how to apply naïve Bayes for a spam filter. State any assumptions made.
[2.0 Marks]

c) Table Q3 gives symptoms and diagnosis for eight (8) patients. Find out if patient number 9 has COVID-19 using naïve Bayes algorithm. Show all your workings.
[6.0 Marks]

Table Q3

Patient ID	Cold	Runny Nose	Headache	Fever	COVID-19?
1	Y	N	Mild	Y	**No**
2	Y	Y	No	N	**Yes**
3	Y	N	Strong	Y	**Yes**
4	N	Y	Mild	Y	**Yes**
5	N	N	No	N	**No**
6	N	Y	Strong	Y	**Yes**
7	N	Y	Strong	N	**No**
8	Y	Y	Mild	Y	**Yes**
9	**Y**	**N**	**Mild**	**N**	**?**

Q4 a) **Circle ONE (1) answer** most suitable for each of the following (i) to (viii) multiple choice questions. **Circle on the question paper itself and attach the question paper** to the answer script.

[0.5 Marks x 8 = 4.0 Marks]

(i) Predicting chance of employment of students graduating from university is which type of problem?
(A) Classification (B) Clustering (C) Regression (D) None of these

(ii) Using patient body weight and other parameters to predict if patients are diabetic or not is which type of problem?
(A) Classification (B) Clustering (C) Regression (D) None of these

(iii) "Using machine learning to group song genres" is which type of problem?
(A) Classification (B) Clustering (C) Regression (D) None of these

(iv) Using features in text to predict emotion of the sender is which type of problem?
(A) Classification (B) Clustering (C) Regression (D) None of these

(v) Which of the following is **NOT** a "performance" measure in machine learning?
(A) Percentage of identifying hand written digits in given data
(B) Fraction of correctly classified score of a run score of cricket player
(C) Percentage of ECG arrhythmia classified by an algorithm
(D) All of the above are performance measures

(vi) The kernel trick

(A) can be applied to every classification algorithm

(B) can be applied to every regression algorithm

(C) exploits the fact that in many learning algorithms, the weights can be written as a linear combination of input points

(D) is commonly used for dimensionality reduction

(vii) Which of the following can be used as a feature to predict if a candidate is going to win elections or not

(A) Previous winnings (B) Gender (C) Assets (D) All of the above

(viii) A regression model in which more than one independent variable is used to predict the dependent variable is called

(A) Simple linear regression model

(B) Multiple regression model

(C) Either of the above

(D) None of the above

b) Using the dataset in Table Q4, we want to build a decision tree which classifies output as T or F, given the binary variables A, B, C.

Table Q4

A	B	C	output
F	F	F	F
T	F	T	T
T	T	F	T
T	T	T	F

(i) Draw a decision tree starting at 'A' that would be learned by the greedy algorithm with zero training error. You do not need to show any computation.

[1.5 Marks]

(ii) Is this tree optimal? If yes, briefly justify your answer. If no, give the optimal solution.

[1.5 Marks]

c) k-means clustering is applied to data shown in Figure Q4c.

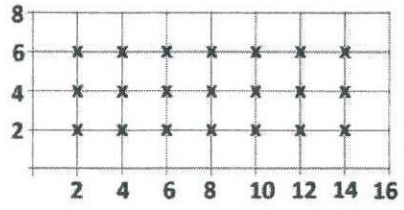

Figure Q4c

(i) In Figure Q4c, suppose k=2. Are (14, 6) and (16, 6) good coordinates for initialization? Briefly justify your answer.

[1.0 Mark]

(ii) If k = 2 and the coordinates of the initialized points are (12, 6) and (14, 6), use the grids given in Figure Q4c (ii) to indicate the k-means **centroids and clusters** obtained until convergence. Show intermediate steps as well.

[2.0 Marks]

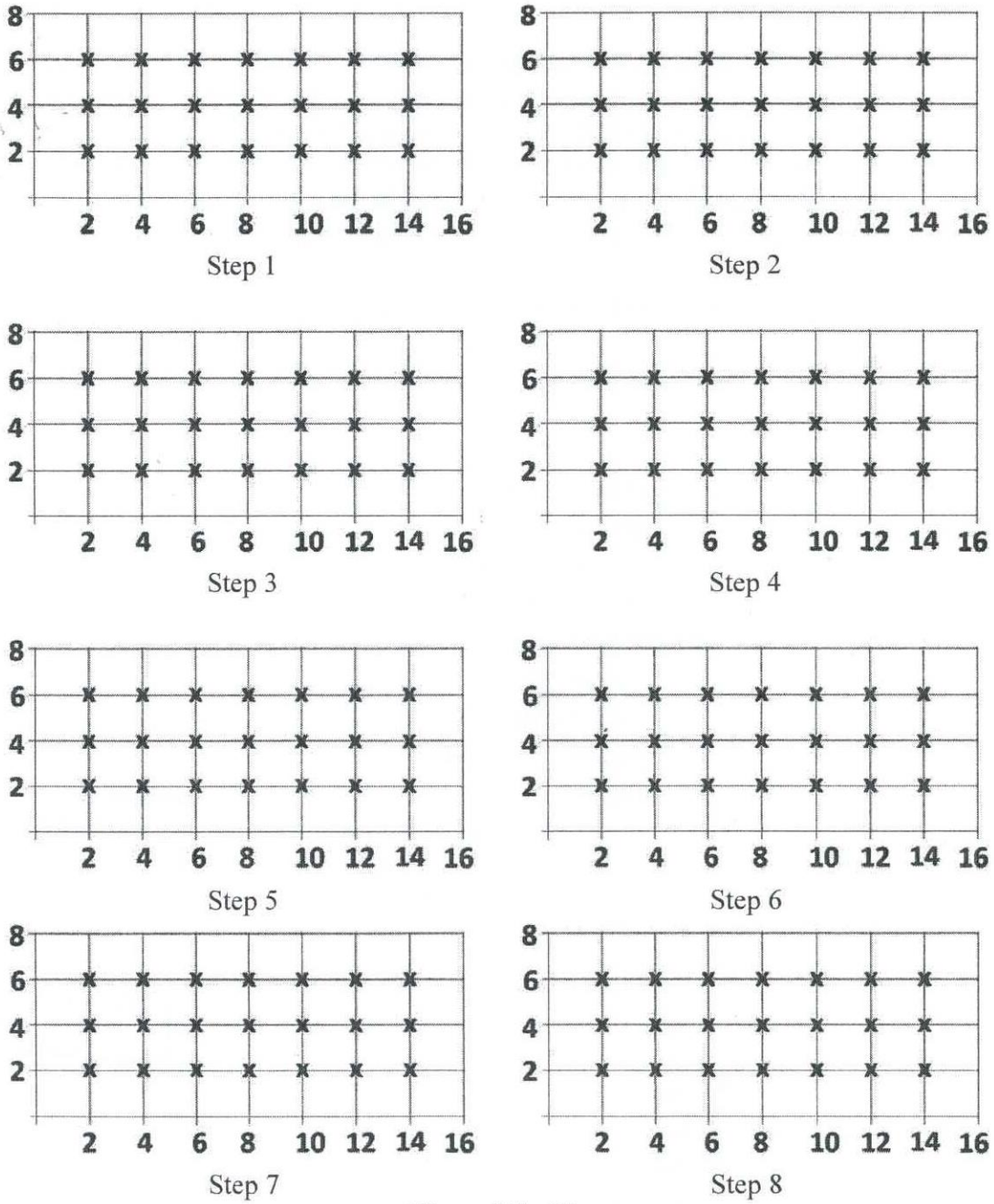

Figure Q4c (ii)

Q5 a) "Sigmoid function is preferred over a tanh function as the activation function of the output layer of a neural network used for binary classification". Do you agree with this statement? Justify your answer.

[2.0 Marks]

Index No: ____/____/_____

b) Suppose we are given eight (8) data points in two classes as below.
Class 1: (3,1), (3,-1), (6,1), (6,-1)
Class 2: (1,0), (0,1), (0,-1), (-1,0)
Using support vector machine (SVM) find the equation of the hyperplane between the two classes. Show all your workings.

[5.0 Marks]

c) A survey was done to compile data from 329 cities about the quality of life. Ratings were compiled for nine (9) different indicators of the quality of life. These are climate, housing, health, crime, transportation, education, arts, recreation, and economics. For each category, a higher rating is better. For example, a higher rating for crime means a lower crime rate. Results are summarized in Figure Q5c (i). Figure Q5c (ii) shows a scree plot of same data after principal component analysis (PCA).

(i) Give a possible reason for applying PCA to this data.

[1.0 Mark]

(ii) How many PCs would you use to analyze this data? Justify your answer.

[2.0 Marks]

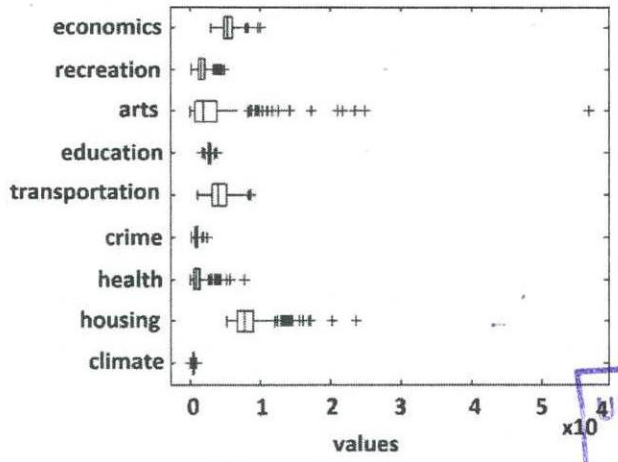

Figure Q5c (i)

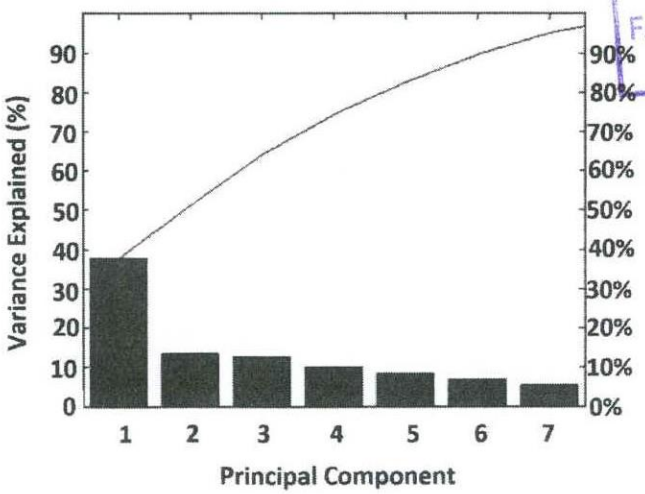

Figure Q5c (ii)