UNIVERSITY OF RUHUNA

Faculty of Engineering

End-Semester 7 Examination in Engineering: March 2022

Module Number: EE7209 **Module Name: Machine Learning**

[Three Hours]

[Answer **all questions**, each question carries **10** marks]

Please attach page 6 and page 7 to the answer script.

---

a) (i) Briefly explain the difference between testing and validation data set.
(ii) What is overfitting? List two (2) methods to overcome it.

[2.0 Marks]

b) (i) How does Machine Learning differ from Deep Learning?
(ii) A machine learning problem is given as "salary prediction based on years of experience." What type of machine learning problem is this?

[2.0 Marks]

c) (i) Briefly explain what you understand by a confusion matrix.
(ii) Which is worse to have; false negatives or false positives? Briefly explain your answer.

[2.0 Marks]

d) Are each of the following a Regression (R) problem or Classification (C) problem?
(i) Logistic Regression
(ii) Random Forest
(iii) Artificial Neural Network (ANN)
(iv) Support Vector Machine (SVM)

[2.0 Marks]

e) You have been hired as a Data Scientist to a company that has a project with Teaching Hospital, Karapitiya to decide whether a new visiting patient has any of heart problems, asthma, or COVID-19 (a patient can have one or more of these diseases). You have been given data from the past two years (2020 and 2021) regarding patient vitals (e.g., temperature, height, weight, diseases etc) in regard to the above three (3) conditions.
(i) Briefly explain how you would set the machine learning problem and list the steps for solving it.
(ii) If you were to use Artificial Neural Network (ANN) for this, 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.

[2.0 Marks]

a) (i) Briefly explain giving a graphical example how outliers are removed using principal component analysis (PCA).
(ii) State one (1) similarity and one (1) difference between linear discriminant analysis (LDA) and PCA.

[2.0 Marks]

b) (i) What is variance accounted for? How would you use this in PCA?
(ii) When should you NOT use PCA?

[2.0 Marks]

c) (i) What is the first step of PCA?
(ii) Implement the first step of PCA for the data given in Table Q2c.

[2.0 Marks]

d) (i) In a problem for "Customer segmentation using machine learning", the following features have been used. There are no missing values or outliers in this dataset. List two (2) possible pre-processing techniques to be used in this data.

- Customer's Age
- Customer's Gender
- Customer's Annual income
- Customer's Spending score

(ii) Machine learning is used to predict if a client will subscribe to deposit in a bank. Two of the features are age of the client and gender of the client. How would you handle missing values in each of these features?

[2.0 Marks]

e) (i) The ABC water board company uses machine learning to check potability of water (i.e. how safe it is for drinking). There are 3726 labeled data samples which are all numerical data. 10% of data is from class I and remaining from class II. Briefly describe how you would handle the misbalanced classes.

(ii) Combined Cycle Power Plant Output prediction is done using decision tree algorithm. There are some outliers in the data. How should these outliers be handled before executing the machine learning algorithm?

[2.0 Marks]

Q3 a) (i) Briefly explain the 'explore' vs 'exploit' concepts in Reinforcement Learning (RL) and how this affects the final outcome.
(ii) Give two (2) examples of when you should not use RL in an application.

[2.0 Marks]

b) (i) Briefly explain two (2) roles of the activation functions in Neural Networks?
(ii) Give a difference in application for sigmoid vs tanh activation functions.

[2.0 Marks]

c) We want to classify some data using the network shown in Figure Q3c. There are three input parameters and two output classes. The network has one hidden layer and one output layer. There is no bias or activation function in the network (i.e. $\sigma(x) = x$). Derive the general expressions for the output y in terms of x, u and w.

[2.0 Marks]

d) Can you represent the boolean function in Table Q3d with a single unit from a neural network? If yes, show the architecture and calculation. If not, briefly explain why not.

[2.0 Marks]

e) Table Q3e gives symptoms and diagnosis for ten (10) patients.

(i) Would you use Naïve Bayes algorithm to classify this data? Justify your answer.

(ii) Show with justification how a patient with following symptoms will be classified from the above algorithm used in Q3e (i)

- Runny Nose = Yes
- Fever = Yes

[2.0 Marks]

a) Are the following statements TRUE or FALSE? Justify your answer.

(i) When the training data set is small, a model is more likely to overfit.

(ii) MODEL 1: $ax+b+\xi$

MODEL 2: $ax^7+bx^6+cx^5+dx^4+ex^3+fx^2+gx+h+\xi$

If there are only 8 data points, MODEL 2 is likely to overfit the data.

[2.0 Marks]

b) Figure Q4b shows how the error of an algorithm varies with the number of epochs for training and validation datasets.

(i) What are the consequences of stopping the training at each epoch given by A, B and C?

(ii) What is the best place to stop the training A, B, C or a different iteration? If it is a different location than A, B or C, you may mark your solution on Figure Q4b and attach the page with the answer script.

[2.0 Marks]

c) You are given the 2 identical plots shown in Figure Q4c, which illustrates a dataset with two classes as 'o' in a larger circle and '▪' in a smaller circle. Classes have equal number of instances. Draw the decision boundary when you train an SVM classifier with linear and polynomial (order 2) respectively. Mark your solution on Figure Qc and attach the page with the answer script.

[2.0 Marks]

d) Give two (2) advantages and two (2) disadvantages of using decision trees for classification of real world problems?

[2.0 Marks]

e) Table Q4e shows a dataset used to learn a decision tree for predicting if a person is sad (S) or happy (H) based on the colour of the shirt/ blouse (Green, Blue or Red), whether they are wearing a jacket and the number of toes they have. Answer the following questions based on Table Q4d and assume no pruning.

(i) What is H(emotion | Jacket=Yes)?

(ii) What is H(emotion | toes=11)?

(iii) Which attribute would the decision tree building algorithm choose for the root of the tree?

(iv) Draw the full decision tree that would be learnt for this data

[2.0 Marks]

a) (i) In decision trees do you prefer a shorter tree or a longer tree? Justify your answer.

(ii) Figure Q5a shows the decision boundaries obtained from three learning algorithms: decision trees, logistic regression, and nearest neighbor classification. '+' and 'o' depict two different classes. Beside each of the three plots, write the name of the learning algorithm and the number of misclassifications from each method.

[2.0 Marks]

b) There are six (6) data points on a one-dimensional (1D) number line: x=1, x=2, x=3, x=7, x=8, x=9. Two cluster centres A and B are initiated as x=0.5 and x=3. Starting with this initialization, graphically show how the k-means algorithm would estimate the points-to-cluster partitioning and the positions of the updated cluster centres (approximately). Repeat the process until convergence. Mark each iteration with t. Example t=1, t=2 etc. Each iteration should be on a new graph.

[2.0 Marks]

c) You are given a scatter plot of a 2D dataset in Figure Q5c. Draw the first and second principal components on the plot.

[2.0 Marks]

d) Table Q5d shows data from 8 different days that the University of Ruhuna Cricket team decided to practice or not based on four (4) different conditions. Answer the following questions using information in Table Q5d.

(i) Calculate the eight (8) conditional probabilities of the attributes.
Eg: $P(Outlook|Practice = Yes)$

(ii) What is the entropy of "Practice"?
Hint: $entropy = H(S) = -\sum_{i=1}^{n} P(x_i) \times \log_2 P(x_i)$

(iii) Which attribute should you choose as the root of a decision tree?

[4.0 Marks]

Table Q2c

Student	Telecommunication	Renewable Power	Computer Engineering
1	90	60	90
2	90	90	30
3	60	60	60
4	60	60	90
5	30	30	30

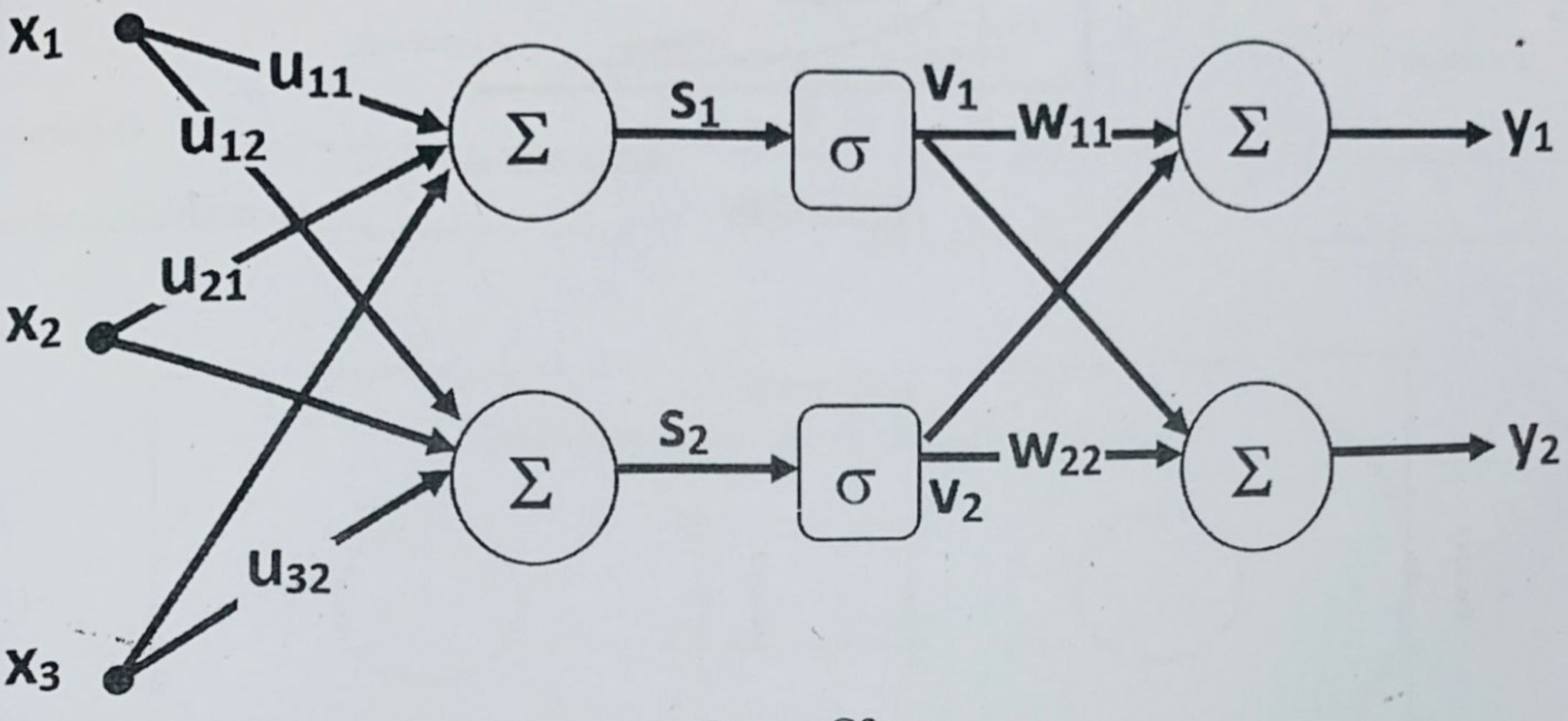


Figure Q3c

Table Q3d

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

Table Q3e

Patient ID	Cold	Runny Nose	Fever	Diagnosis
1	Yes	Yes	No	Flu
2	Yes	Yes	No	Flu
3	No	No	Yes	Covid-19
4	No	No	Yes	Covid-19
5	No	No	Yes	Covid-19
6	Yes	Yes	No	Flu
7	Yes	Yes	No	Flu
8	No	No	Yes	Covid-19
9	Yes	Yes	No	Flu
10	No	No	Yes	Covid-19

Please attach page 6 and page 7 to the answer script. Index No EG/_____/_____

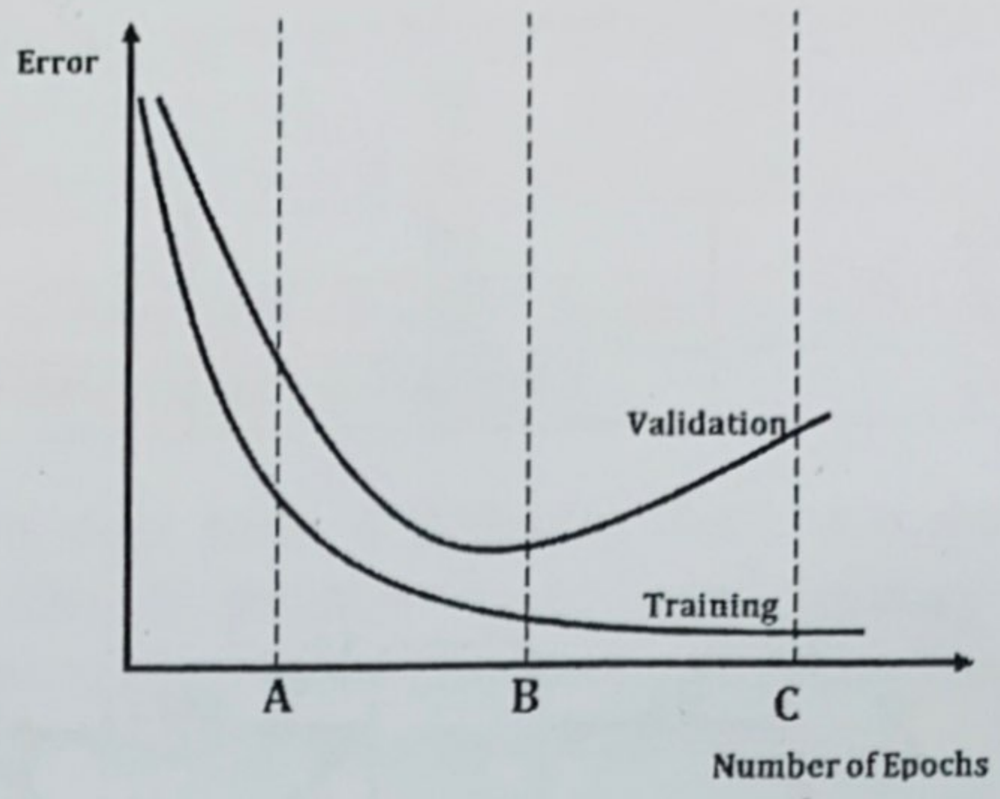
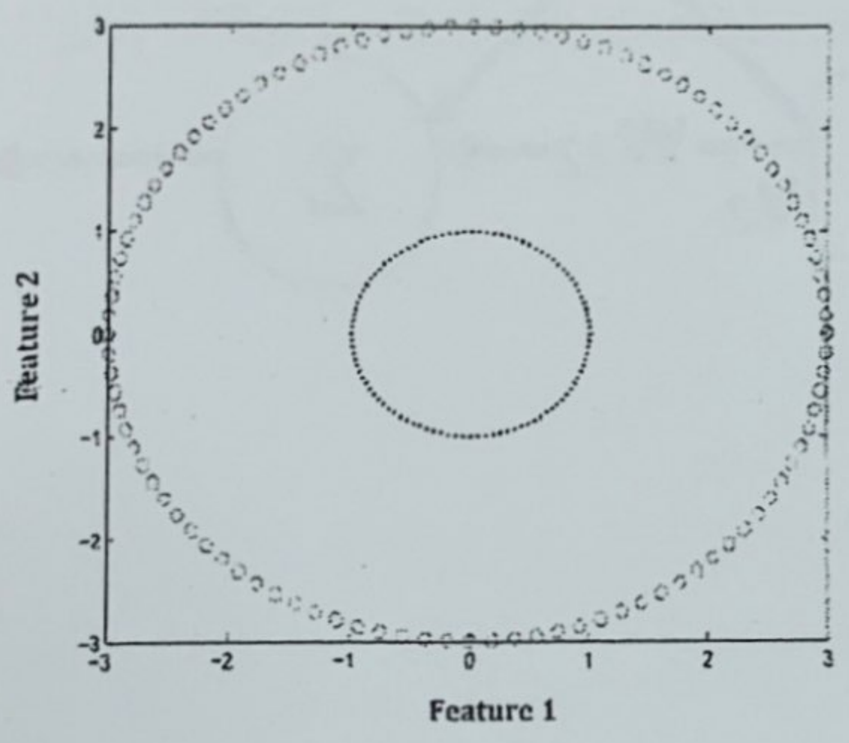
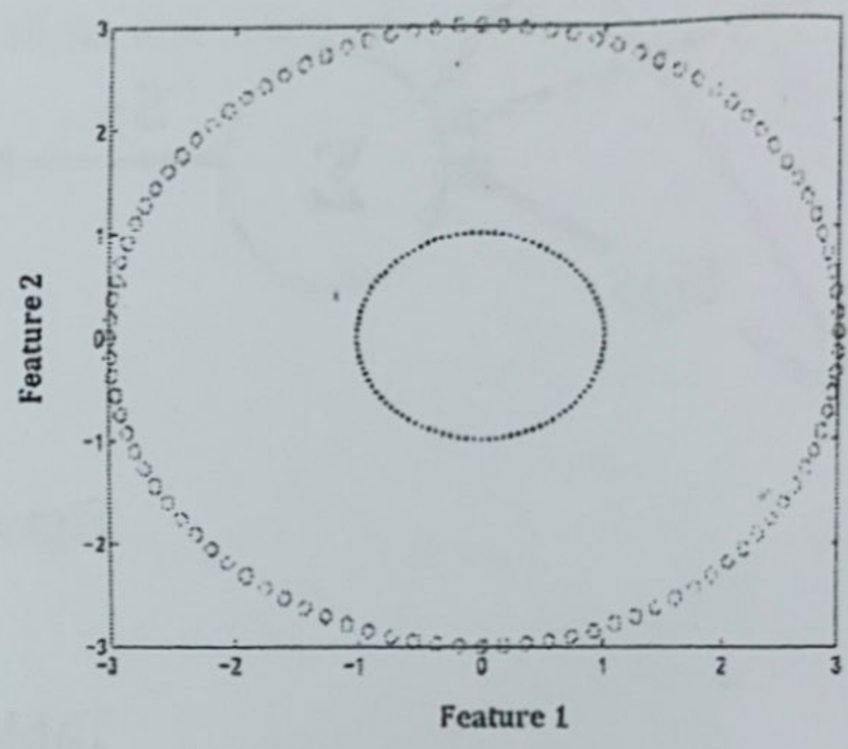


Figure Q4b



linear



polynomial (order 2)

Figure Q4c

(Mark your solution on the given figures.)

Table Q4e

Colour of Shirt/ Blouse	Wearing Jacket	Number of Toes	Emotion (Output)
G	Yes	10	S
G	Yes	10	S
G	No	10	S
B	No	10	S
B	No	10	H
R	Yes	10	H
R	Yes	10	H
R	No	10	H
R	Yes	11	H

Please attach page 6 and page 7 to the answer script. Index No EG/_____/_____.

gorithm _____ _____ _____

isclassification # _____ _____ _____

Figure Q5a
(Mark your solution on the given figure)

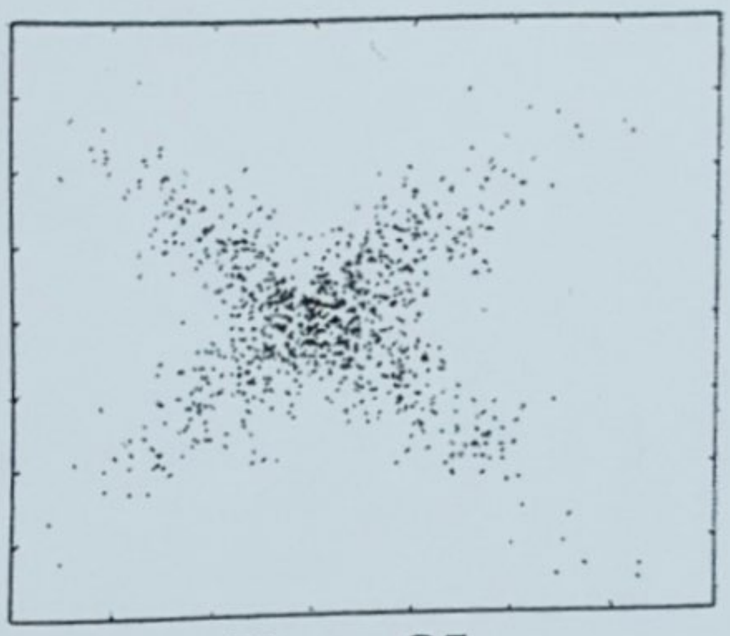


Figure Q5c
(Mark your solution on the given figure)

Table Q5d

Day	*Outlook*	*Humidity*	*Wind*	*Captain Present?*	*Practice?*
1	Sunny	Normal	Weak	No	No
2	Sunny	Normal	Strong	No	No
3	Overcast	High	Weak	Yes	No
4	Overcast	Normal	Weak	Yes	Yes
5	Sunny	High	Strong	No	Yes
6	Sunny	Normal	Strong	Yes	Yes
7	Sunny	Normal	Weak	Yes	Yes
8	Overcast	High	Weak	No	Yes

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]

Index No: ____/____/_____

(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.

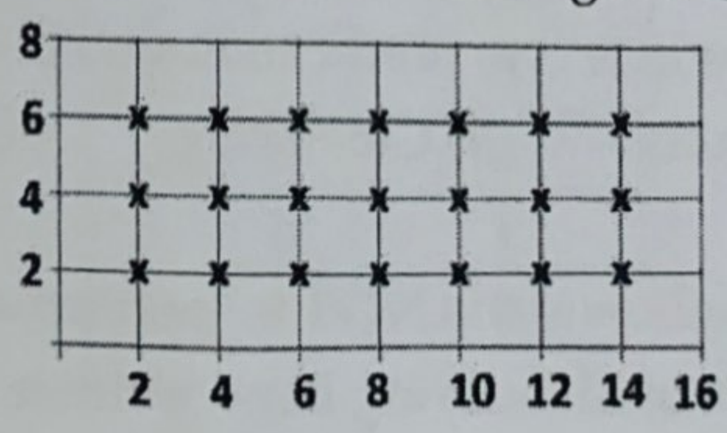


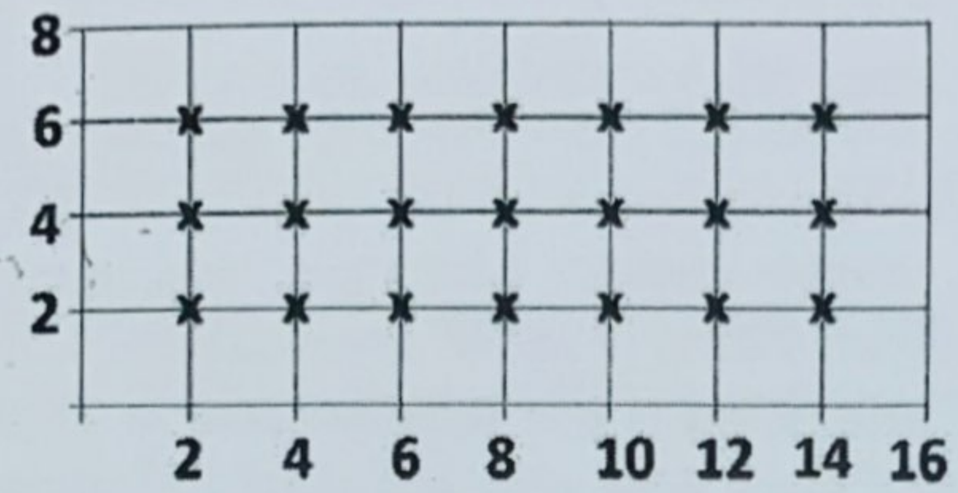
Figure Q4c

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

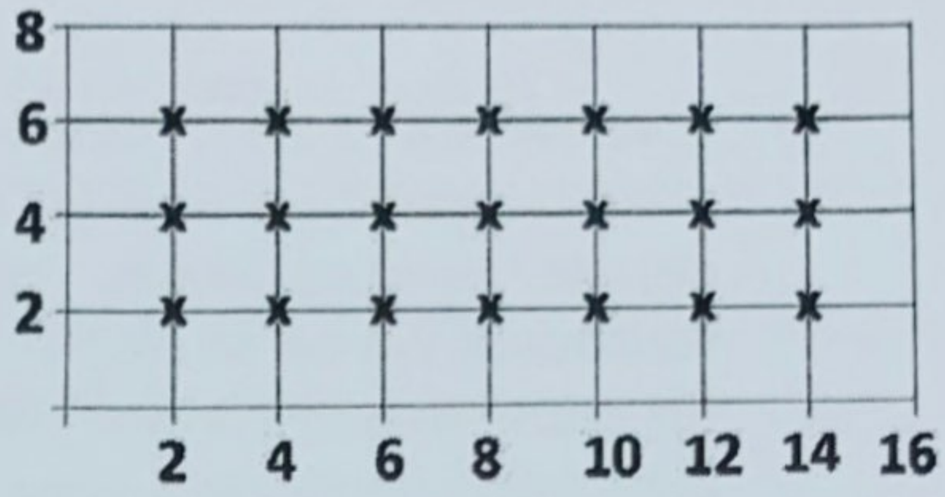
[1.0 Mark]

(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.

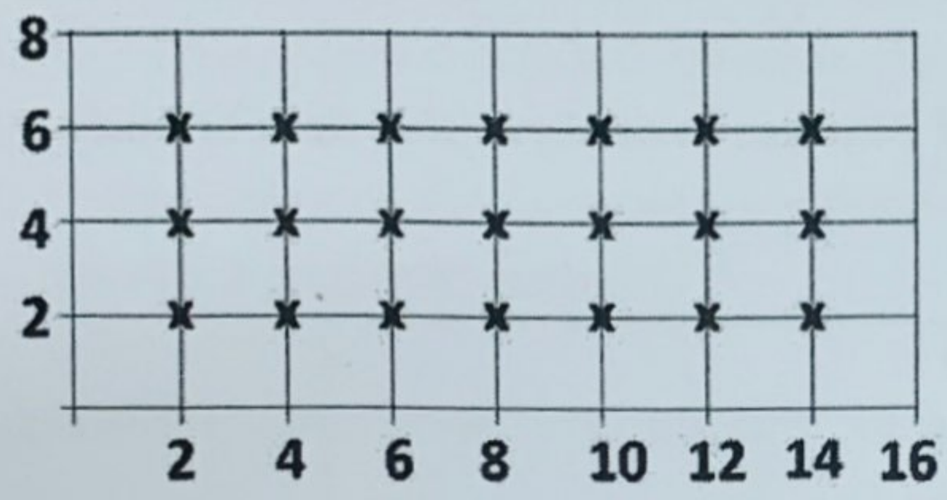
[2.0 Marks]



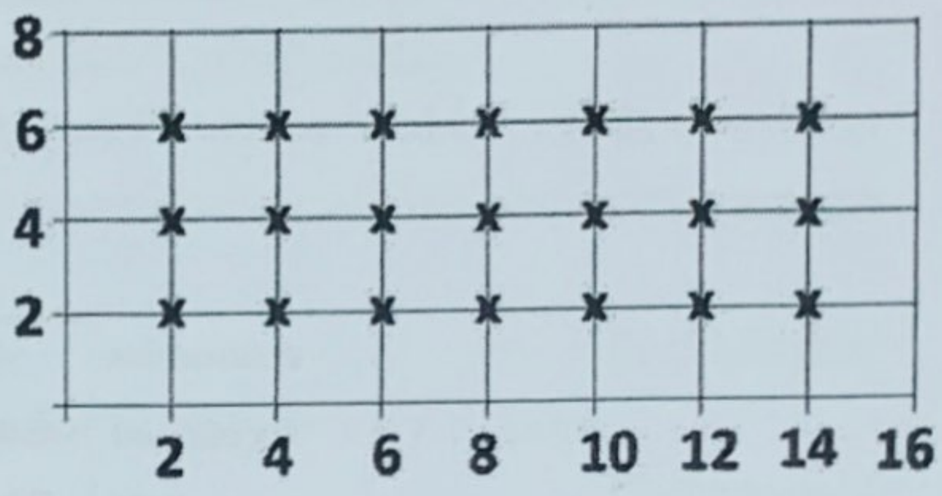
Step 1



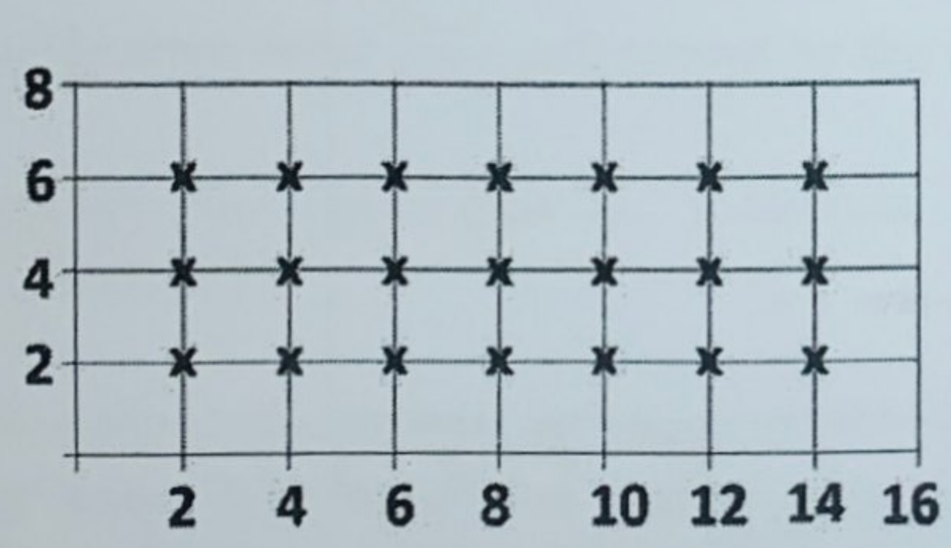
Step 2



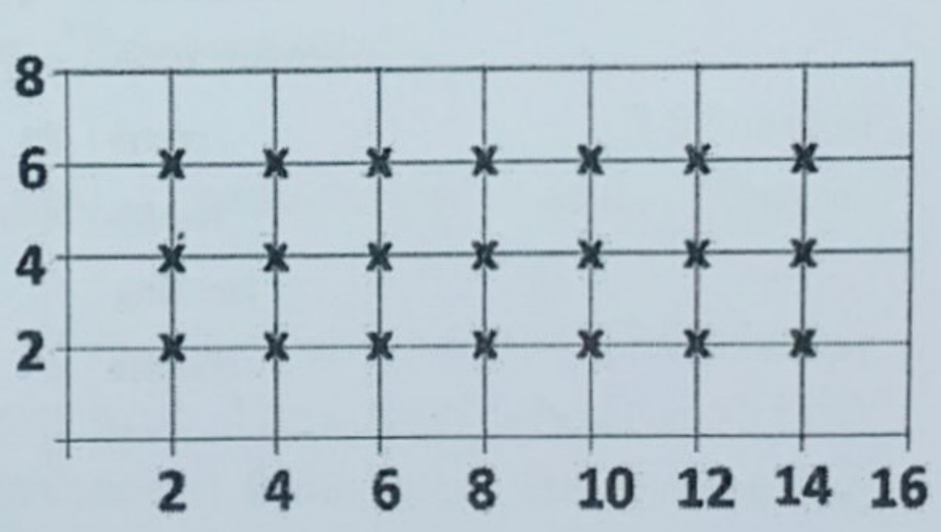
Step 3



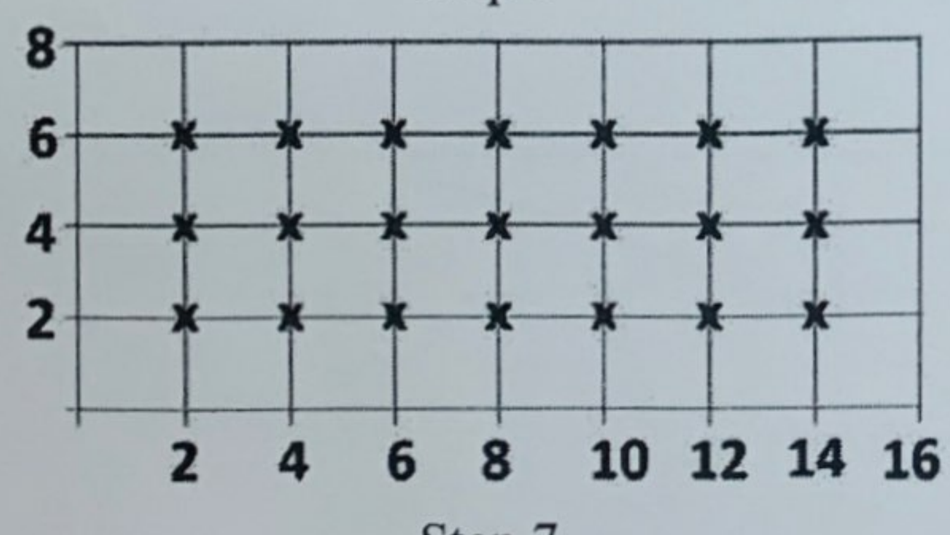
Step 4



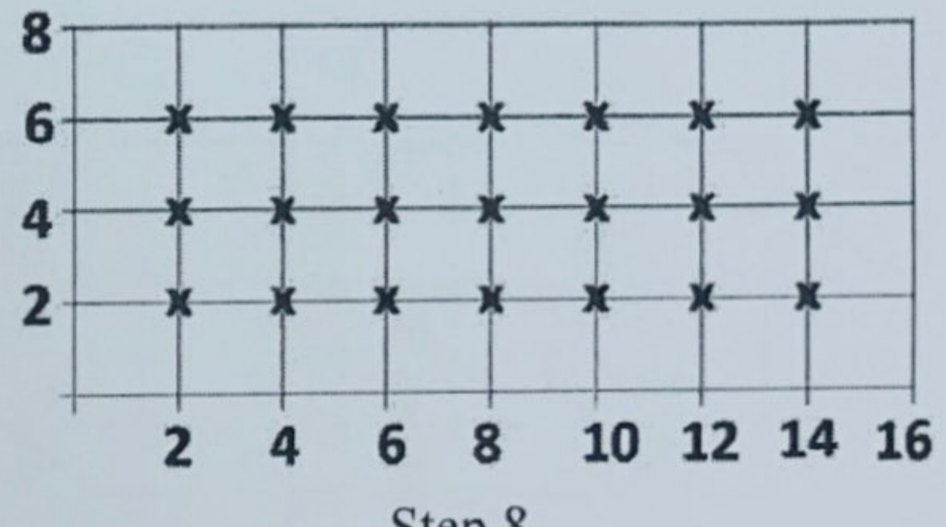
Step 5



Step 6



Step 7



Step 8

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]

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]

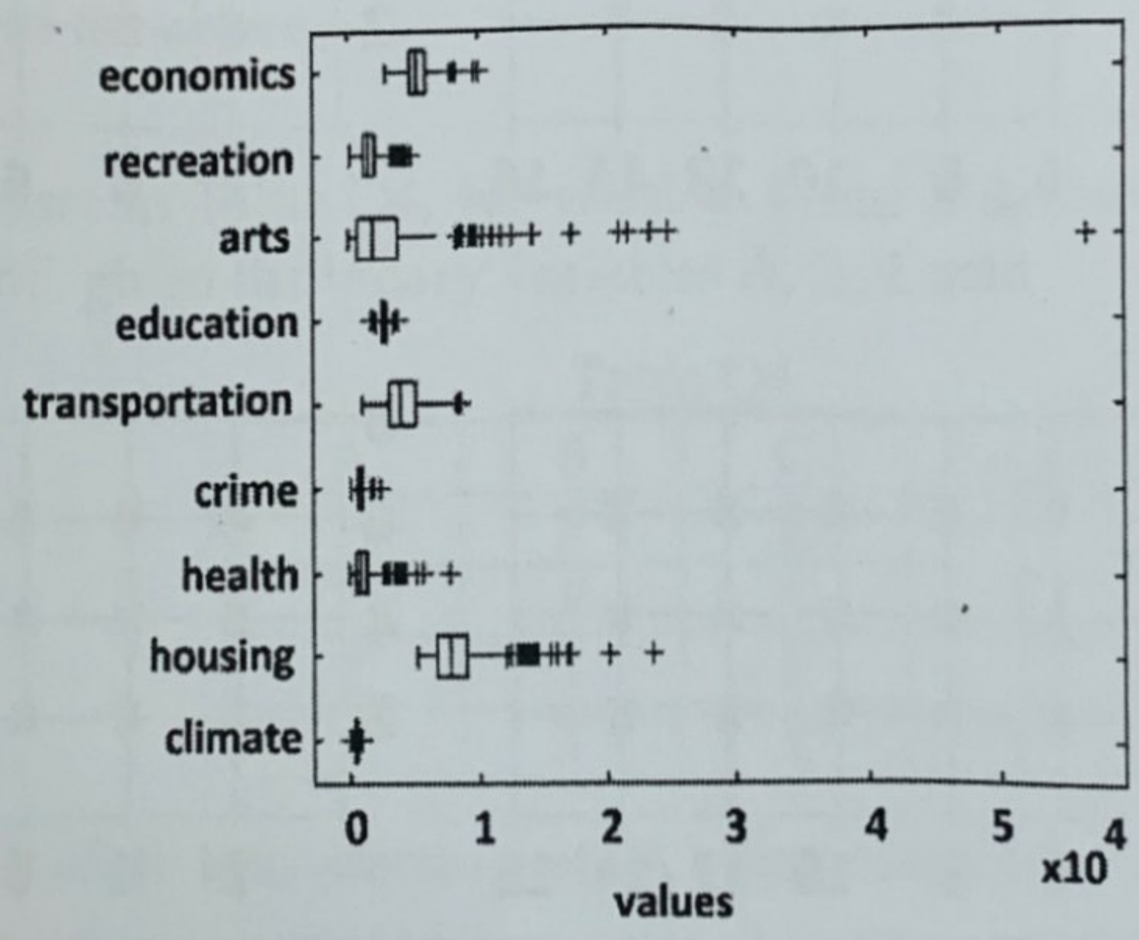


Figure Q5c (i)

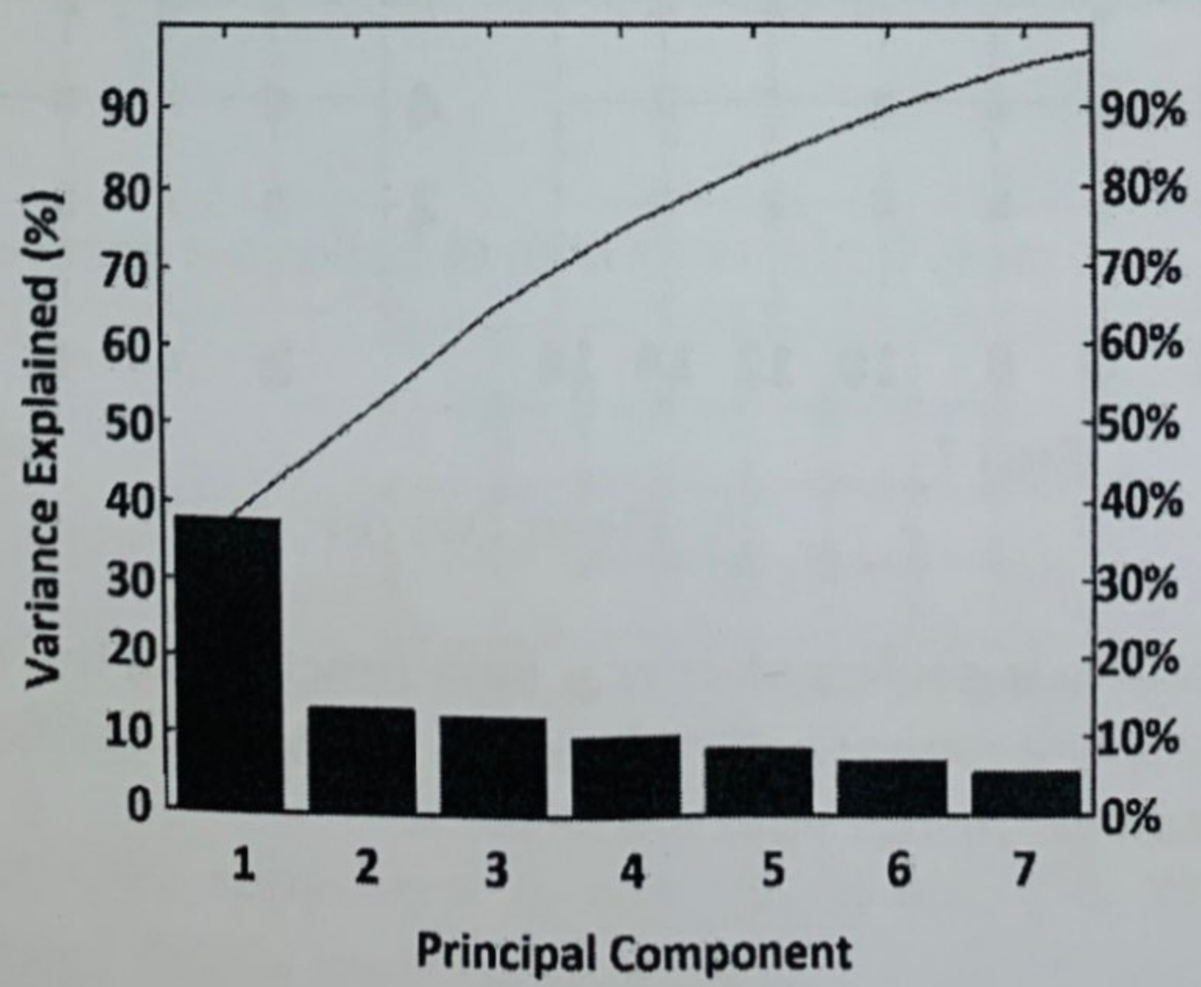


Figure Q5c (ii)

UNIVERSITY OF RUHUNA

Faculty of Engineering

End-Semester 7 Examination in Engineering: October 2019

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

[Three Hours]

[Answer **all questions**, each question carries **10** marks]

---

a) Briefly explain three main uses of Machine Learning (ML). Name a real world example for each of those uses.

[3.0 Marks]

b) i) Briefly explain the main ML types.
ii) Identify the most suitable ML type for the following applications. Justify your selection.

1. Predicting the rainfall on a particular day, given historical data.
2. Categorize an unlabeled news article dataset according to the type of news article.
3. An obstacle avoiding robot.
4. A profit maximizing trading agent.

[4.0 Marks]

c) State two differences between the feature selection and the feature reduction.

[1.0 Marks]

d) Show how a derived feature can reduce the complexity of a classification problem using a graphical example.

[2.0 Marks]

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

[2.0 Marks]

b) Suppose you are given 1000 data samples to categorize into 10 classes. Would you select "k-fold cross validation" or "validation based on train-test set division"? Briefly explain the reasons for your selection.

[2.0 Marks]

c) An accuracy result obtained from a ML algorithm is shown in Figure Q2 c).
i) What is the problem with the result?
ii) Briefly describe three techniques that can be tried out to overcome this problem.

[3.0 Marks]

d) In order to solve a classification problem, suppose you are given a training data set of 1,000 columns and 1,000,000 rows. You are asked to reduce the dimension of this data set so that the model computation time can be reduced. Your machine has memory constraints. Stating any practical assumptions you make, explain how you reduce the dimensionality of the given data set.

[3.0 Marks]

Q3 a) Consider the two-dimensional data set given in Table Q3 a).

i) Plot Antenna Length Vs. Abdomen Length for the data set and draw the decision boundary for the two classes; Grasshopper and Katydid.

ii) Compute and mark the means for the two classes in part a) i).

iii) Classify the new instance (5.1, 7) into Grasshopper or Katydid using the k-Nearest Neighbors (kNN) classification method where, k = 2.

Hint: Use Euclidean distance measurement. Clearly show your workings.

iv) What is the practical limitation of selecting k = 2 to this problem? What k would you select to overcome that limitation? Justify your answer.

v) How do you mitigate the sensitivity of nearest neighbor algorithms to irrelevant features?

[5.0 Marks]

b) Using naïve Bayes theorem, deduce whether the person called Drew is a Male or a Female based on the previously seen data given in Table Q3 b).

[2.0 Marks]

c) i) Explain the two main types of clustering.

ii) What is the use of a dendriogram in clustering?

iii) Write a psudo code for k means clustering.

[2.0 Marks]

d) When should you use classification over regression?

[1.0 Mark]

Q4 a) Consider a perceptron, the simplest possible Neural Network (NN). Its trainable parameters are the weights and the bias.

i) What is defined by the weights in the NN?

ii) What is the purpose of using the bias parameter?

[2.0 Marks]

b) Consider a perceptron that takes in a 4 dimensional feature vector. Its weight vector is randomly initialized to [0.2, 0.6, 0.4, 0], and the bias is initialized to 1. Assume you are given a training data sample having the feature vector [1, 5, 8, 4].

i) Show how the output value is calculated for this feature vector through forward propagation.

ii) Explain how to convert this simple perceptron into a sigmoid perceptron.

[3.0 Marks]

c) A commonly used error function for binary classification with a NN is,

$$-[y^{(i)} \log(y'^{(i)}) + (1-y^{(i)}) \log(1-y'^{(i)})]$$

where, $y^{(i)}$ is the output for the i^{th} data sample, and $y'^{(i)}$ is the predicted output. Show that the given error function is optimal.

[2.0 marks]

d) i) "tanh function is better than the sigmoid function to be used as the activation function of the output layer of a NN used for binary classification". Do you agree with this statement? Justify your answer.

ii) "During the training process of a NN, the input layer should never be trained". Do you agree with this statement? Justify your answer.

iii) Discuss the uses of Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs).

[3.0 marks]

a) i) Briefly explain two main uses of Association Rule Mining (ARM).

ii) Explain the Apriori algorithm used in ARM using the example given below. State any assumptions you make.

Example

order 1: apple, egg, milk, carrot

order 2: milk

order 3: apple, egg, carrot

order 4: apple, egg

order 5: apple

iii) Define the following terms in the context of frequent item set mining.

1. Support
2. Confidence
3. Lift

iv) Using the example given in part a) ii), compute the following support, confidence and lift percentages.

1. support {apple,egg}
2. confidence{apple->egg}
3. confidence{egg->apple}
4. lift {apple,egg}
5. lift{egg,apple}

[6.0 marks]

b) Briefly explain the concept of Support Vector Machines (SVM).

[1.0 mark]

c) i) What does it mean by Genetic Algorithms (GA)?

ii) Explain the GA process using a pseudo code.

[2.0 marks]

d) Explain reproduction process (i.e. selection, crossover, and mutation) using an example.

[1.0 mark]

Figure Q2 c)

Table Q3 a)

Insect ID	Abdomen Length (cm)	Antenna Length (cm)	Insect Class
1	2.7	5.5	Grasshopper
2	8.0	9.1	Katydid
3	0.9	4.7	Grasshopper
4	1.1	3.1	Grasshopper
5	5.4	8.5	Katydid
6	2.9	1.9	Grasshopper
7	6.1	6.6	Katydid
8	0.5	1.0	Grasshopper
9	8.3	6.6	Katydid
10	8.1	4.7	Katydids

Table Q3 b)

Name	Sex
Drew	Male
Claudia	Female
Drew	Female
Drew	Female
Alberto	Male
Karin	Female
Nina	Female
Sergio	Male