Index number: .....

UNIVERSITY OF RUHUNA

Faculty of Engineering

End-Semester 5 Examination in Engineering: November 2024

Module Number: EE5253 **Module Name: Machine Learning (C-18)**

Part-A

[30 Minutes]

Instructions for Candidate

- Write your index number on top of every page.
- Question paper contains 20 multiple choice questions.
- Answer all questions in the paper itself.
- Each question carries 0.5 marks.
- Read the question and all answers before making the choice.
- **For each question, put an X mark on the letter: (a), (b), (c), or (d) which corresponds to the correct answer, by using a black or blue pen.**

---

1. What does it mean by the statement "KNN is a non-parametric algorithm"?
   (a) KNN makes assumptions about the underlying data distribution.
   (b) KNN doesn't assume a specific data distribution.
   (c) KNN requires a large amount of training data.
   (d) KNN is computationally expensive.
2. Which of the following is the correct equation for calculating the Entropy in the context of Decision Trees given that S is the dataset, n is the number of classes and, p_i is the probability of class i?
   (a) $Entropy\ (S) = -\sum_i^n p_i * \log_2(p_i)$
   (b) $Entropy\ (S) = \sum_i^n p_i^2 * \log_2(p_i^2)$
   (c) $Entropy\ (S) = -\sum_i^n p_i * ln(p_i)$
   (d) $Entropy\ (S) = \sum_i^n p_i^2 * ln(p_i^2)$
3. What are the two main techniques involved in the random forest algorithm?
   (a) Bagging and Boosting.
   (b) Boosting and Feature Randomness.
   (c) Model Ensemble and Feature Randomness.

(d) Bagging and Feature Randomness.

4. Which of the following is NOT a purpose of adding a bias term in artificial neural networks?
   (a) Shifting the activation function.
   (b) Increasing the model's complexity.
   (c) Acting as an intercept in a linear equation.
   (d) Allowing the model to fit data that doesn't pass through the origin.
5. Which of the following is NOT a Feed-Forward Neural Network (FFNN)?
   (a) Multilayer Perceptron (MLP).
   (b) Convolutional Neural Network (CNN).
   (c) Recurrent Neural Network (RNN).
   (d) All of the above are FFNNs.

Table 1 presents a portion of the COVID-19 infection dataset. Please use this data to answer questions 6 and 7.

Table 1

ID	Fever	Cough	Breathing Issues	Infected
1	No	No	No	No
2	Yes	Yes	Yes	Yes
3	Yes	Yes	No	No
4	Yes	No	Yes	Yes
5	Yes	Yes	Yes	Yes
6	No	Yes	No	No
7	Yes	No	Yes	Yes
8	Yes	No	Yes	Yes
9	No	Yes	Yes	Yes
10	Yes	Yes	No	Yes
11	No	Yes	No	No
12	No	Yes	Yes	Yes
13	No	Yes	Yes	No
14	Yes	Yes	No	No

6. What is the entropy for the given dataset by considering the "Infected" column as the target variable?
   (a) 0.91
   (b) 0.87
   (c) 0.13
   (d) 0.99
7. What is the information gain for the "Fever" attribute in the given dataset?
   (a) 0.91
   (b) 0.87

(c) 0.13

(d) 0.99

8. For a classification problem, a machine learning engineer tests two algorithms, as shown in Table 2. Based on the information provided, which model would be more suitable for this task?

Table 2

Logistic Regression	Random Forest Classifier
• Training Accuracy: 83.32% • Testing Accuracy: 82.22%	• Training Accuracy: 100.00% • Testing Accuracy: 85.53%

(a) Logistic Regression.

(b) Random Forest.

(c) Both Logistic Regression and Random Forest.

(d) An ensemble model of a given techniques.

9. Based on the confusion matrix in Figure 1, what is the most accurate conclusion that can be drawn by interpreting the values provided?

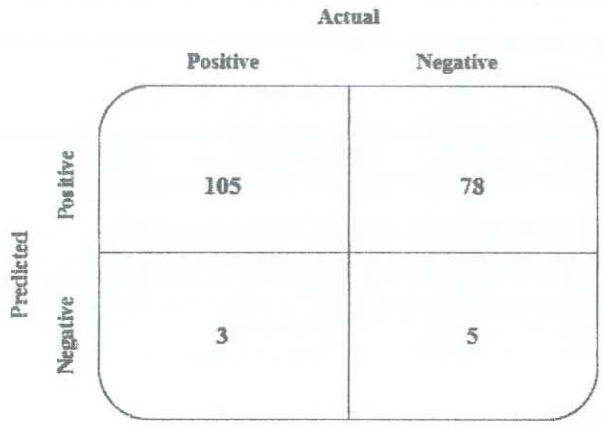

Figure 1

(a) Model indicates satisfactory performance for both positive and negative classes while positive class is performing better.

(b) Some outliers can be detected in the negative class.

(c) Model most probably have trained with an imbalanced dataset.

(d) Information provided is insufficient to get into any conclusion.

10. In hierarchical clustering, what does a dendrogram primarily represent?

(a) The distance between clusters at each stage of the clustering process.

(b) The number of data points in each cluster after clustering is complete.

(c) The average silhouette score of clusters obtained during hierarchical clustering.

(d) The distribution of data points within each cluster across different dimensions.

11. Which one is the default kernel function in Support Vector Machine?

(a) Linear Kernel.

(b) Polynomial Kernel.

(c) Gaussian Kernel.

(d) Radial Bias Function Kernel.

12. Considering the bias-variance trade-off, name the A, B and C curves in Figure 2 appropriately.

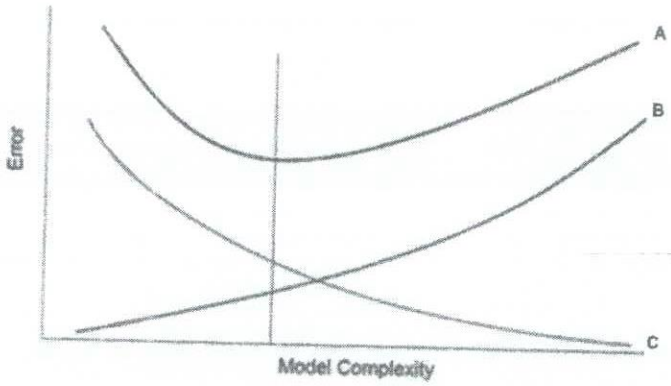

Figure 2

(a) Test Error, Variance, Bias.
(b) Test Error, Bias^2, Variance.
(c) Test Error, Bias, Variance^2.
(d) Test Error, Variance, Bias^2.

13. What is an adversarial attack in the context of AI?

(a) An approach to improve the transparency of AI models.
(b) A technique that ensures privacy in AI data processing.
(c) A method of slightly modifying input data to mislead AI models.
(d) A process of training AI on only unbiased data.

14. Which type of Naive Bayes classifier is best suited for a dataset where the features are continuous values?

(a) Bernoulli Naive Bayes.
(b) Multinomial Naive Bayes.
(c) Gaussian Naive Bayes.
(d) All of the above.

15. In clustering, the silhouette value of a data point can range from -1 to 1. Which of the following correctly describes the interpretation of these values?

(a) A silhouette value close to 1 indicates poor clustering, and close to -1 indicates well-separated clusters.
(b) A silhouette value close to 1 indicates that the data point is well-clustered, while a value close to -1 indicates possible misclassification.
(c) A silhouette value close to 0 suggests that the data point is far from other points, regardless of clusters.
(d) A negative silhouette value implies that the point is near the centroid of its cluster.

16. When dealing with linearly inseparable data, which of the following method can be used in SVM to create a better decision boundary?

(a) Using a linear kernel to separate the data directly.

(b) Adding more data points to make the data linearly separable.

(c) Transforming the data into a higher-dimensional space using a kernel function.

(d) Removing the outliers to make the data linearly separable.

17. You are using a Linear SVM classifier for a 2-class classification problem. Select the correct answer based on the following 2 scenarios in Table 3.

Table 3

Scenario 1	Scenario 2
• If any of the support vectors are removed from the dataset.	• If one of the non-support vector points is removed from the dataset.

(a) The decision boundary will remain the same in both cases.

(b) The decision boundary will shift if a support vector is removed, but remain unchanged if an any other non-support vector point is removed.

(c) The decision boundary will completely change in both cases.

(d) The decision boundary will change only if a non-support vector point is removed.

18. Which of the following are well-known privacy-preserving techniques used in data analysis and machine learning?

a) Homomorphic Encryption, Differential Privacy, Federated Learning.

b) Local Interpretable Model-Agnostic Explanations, Shapley Additive Explanations, Homomorphic Encryption.

c) Homomorphic Encryption, Hybrid Approaches, Local Interpretable Model-Agnostic Explanations.

d) Shapley Additive Explanations, Federated Learning, Differential Privacy.

19. In the K-means clustering algorithm, after initializing the centroids, the algorithm proceeds by assigning each data point to the nearest centroid. Suppose you have the following dataset with 2D points.

P1(2, 3), P2(3, 4), P3(6, 5), P4(7, 8)

The initial centroids are set to C1(2, 3) and C2(7, 8).

After the first assignment step, which data points would be assigned to centroid C1?

(a) P1 and P4.

(b) P1 and P2.

(c) P2 and P3.

(d) P3 and P4.

20. Suppose a Naive Bayes classifier is used to classify emails as "Spam" or "Not Spam." The probability of an email being spam, P(Spam), is 0.3, and the probability of it being not spam, P(Not Spam), is 0.7. Given that a certain word "Offer" appears in an email, the probability of this word appearing in spam emails, P(Offer|Spam), is 0.8, and in not spam emails, P(Offer|Not Spam), is 0.1. Using the Naive Bayes formula, what is the probability that an email containing the word "Offer" is spam?

(a) 0.5

(b) 0.63

(c) 0.77

(d) 0.84

UNIVERSITY OF RUHUNA

Faculty of Engineering

End-Semester 5 Examination in Engineering: November 2024

Module Number: EE5253 **Module Name: Machine Learning (C-18)**

Part-B

[1 Hour and 30 Minutes]

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

Q1 a) As a data engineer you have been assigned to design and implement a deep Feed-Forward Neural Network (FFNN) for a binary classification problem. Furthermore, assume that there are 10 layers in total excluding the input layer and you are restricted to use only ReLU and Sigmoid activation functions throughout the design.

i) Describe the purpose of using non-linear activation functions inside neurons in FFNNs.

[0.5 Marks]

ii) In the given scenario above, select the most suitable activation functions for the hidden layers and output layer. Justify your choices.

[1.0 Mark]

b) Confusion matrix is commonly utilized in classification problems to evaluate prediction models. The confusion matrix provides an overall idea on how well the classification models perform with different classes involved. Assume a situation where a "Random Forest" model has been used in a multi-class classification problem and obtained confusion matrix as indicated in Figure 1.

		Actual			
		Class1	Class2	Class3	Class4
Predicted	Class1	**900**	50	20	80
	Class2	10	**200**	15	40
	Class3	30	20	**100**	30
	Class4	60	30	15	**50**

Figure 1

i) Calculate the overall accuracy for the scenario presented in Figure 1?

[0.5 Marks]

ii) Describe the meaning of precision, recall and F1-score with relevant equations.

[1.0 Mark]

iii) Calculate the F1 Score for all the four classes indicated in confusion matrix in Figure 1. Round the answers to four decimal places and make sure to show all the intermediate steps.

[2.0 Marks]

iv) What would be the best metric to evaluate the provided scenario from accuracy, precision, recall and F1-score? Justify your choice.

[0.5 Marks]

c) Receiver Operating Characteristic (ROC) curves are another evaluation metric which is commonly used in binary classification tasks. It plots "True Positive Rate (TPR or Sensitivity)" against the "False Positive Rate (FPR)".

i) Evaluate the A, (B & C), and D curves indicated in Figure 2 and interpret their behavior considering a binary classification problem.

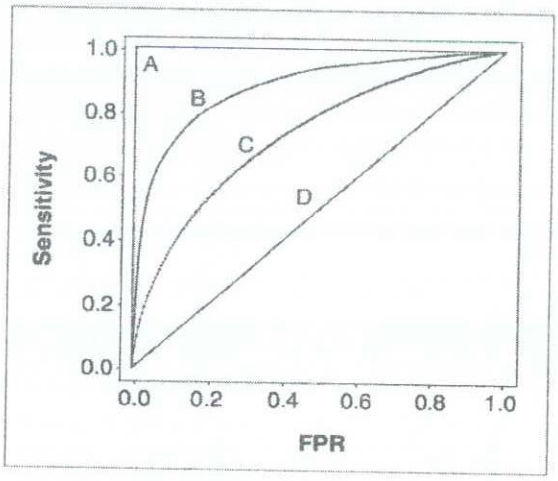

Figure 2

[0.5 Marks]

ii) In a binary classification problem, which used a Logistic Regression classifier, the following results were obtained for four decision thresholds (refer Table 1). Construct the ROC curve for the given scenario by calculating the TPR and FPR for each threshold (round the calculated values for one decimal place and use appropriate scales for each axis).

Table 1

Threshold	True Positives	True Negatives	False Positives	False Negatives
0.1	100	10	80	10
0.5	80	50	40	30
0.8	50	80	10	60
0.95	20	88	2	90

[2.5 Marks]

d) Correlation is a measure of the linear relationship between 2 or more variables. The correlation matrix is a commonly used approach in machine learning for feature selection.

i) Evaluate and interpret the given correlation matrix in Table 2 and select the best **two** features (out of A, B, C, D, and E) to proceed with the model implementation. Justify your selection.

Table 2

Features	A	B	C	D	E	Target
A	1.00	0.90	0.20	-0.30	0.05	0.95
B	0.90	1.00	0.15	-0.25	0.10	0.85
C	0.20	0.15	1.00	0.60	0.35	-0.50
D	-0.30	-0.25	0.60	1.00	0.50	-0.01
E	0.05	0.10	0.35	0.50	1.00	0.02
Target	0.95	0.85	-0.50	-0.01	0.02	1.00

[1.0 Mark]

ii) If there are some categorical variables as well, which technique/algorithm you will employ for the categorical feature section?

[0.5 Marks]

Q2 a) There is a training set consisting of samples and their labels. Each sample comes from one of two classes, 0 and 1. The samples are two-dimensional vectors, and the input data is in the form **{X1, X2, Y}** where X1 and X2 represent the two values for each input vector, and Y is the label for this sample.

After training a Naïve Bayes classifier, we arrived at the following conditional probabilities.

Table 3

	Y = 0	**Y = 1**
X1	$P(X1 = 1 \mid Y = 0) = {}^{1}/_{4}$	$P(X1 = 1 \mid Y = 1) = {}^{1}/_{2}$
X2	$P(X2 = 1 \mid Y = 0) = {}^{1}/_{2}$	$P(X2 = 1 \mid Y = 1) = {}^{2}/_{3}$

Let w1 represents the probability of class 1.

$$\mathbf{w1 = P(Y = 1)}$$

Given that the combined likelihood of the following two samples, {1,0,1} and {0,1,0} is 1/160.

(i) Calculate the P(X1 = 1, X2 = 0) for the sample {1,0,1}.

[1.5 Marks]

(ii) Calculate the P(X1 = 0, X2 = 1) for the sample {0,1,0} .

[1.5 Marks]

iii) Using the provided information that the combined likelihood of these two samples is 1/160, create an equation that represents this condition with w1 as the only unknown term.

[0.5 Marks]

iv) Simplify the expression as much as possible. (Important: Do not want to find the value of w1)

[0.5 Marks]

b) Imagine you are a school administrator who wants to group students based on their study habits and participation to tailor study resources and extracurricular activities. You collected data on five students based on their scores (from 1 to 10) for **Study Time** and **Class Participation**.

Table 4

Student	Study Time (x)	Class Participation (y)
1	8	3
2	2	9
3	7	5
4	1	10
5	6	4

Use **Student 1** and **Student 4** as the initial centroids for **Cluster A** and **Cluster B** respectively.

(i) Assign each student to the suitable cluster using KMeans Clustering. (Important: State all the intermediate steps and calculations properly)

[3.0 Marks]

(ii) Explain the general characteristics of each clusters based on the results seperately.

[1.0 Mark]

c) Clustering is a machine learning technique used to group similar data points together based on their characteristics, allowing for pattern discovery and data segmentation without prior labels.

(i) Describe the terms "Agglomerative Clustering" and "Divisive Clustering" in Hierarchical Clustering using suitable illustrations.

[1.0 Mark]

(ii) Based on the following results in figure 3, briefly explain what should be the best choice for the number of clusters?

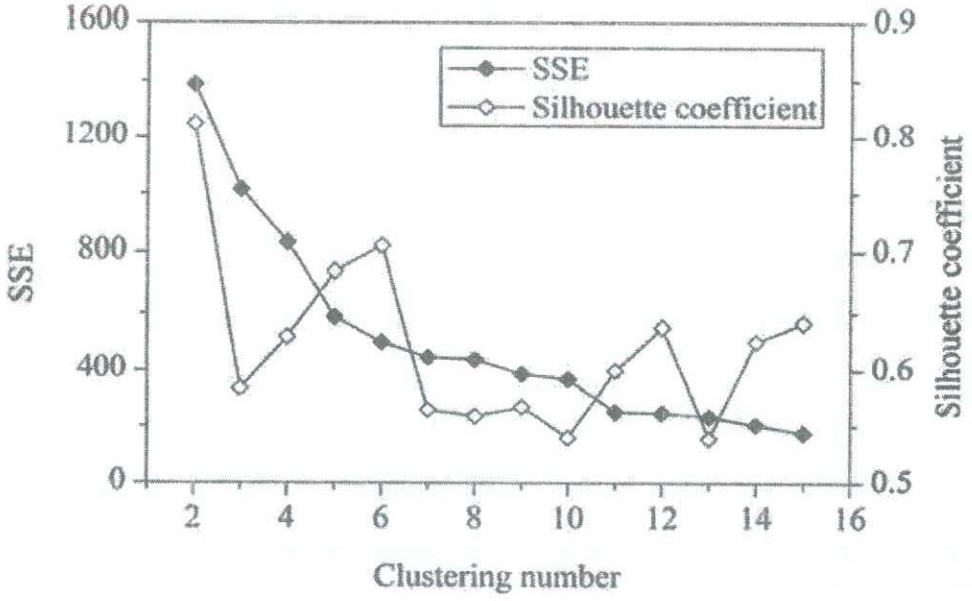

Figure 3

[1.0 Mark]