ourdiu.com | Shared by SUPAN ROY

Daffodil International University

Department of Software Engineering
Faculty of Science & Information Technology
Final Examination, Fall 2025
Course Code: SE 544; Course Title: Introduction to Machine Learning
Sections & Teachers: 40(B,D,G,F)(JJS), 40(E,H)(FA), 40(A,C)(NS), 40I(DB)

Time: 2:00 Hours **Marks: 40**

Answer ALL Questions

[The figures in the right margin indicate the full marks and corresponding course outcomes. All portions of each question must be answered sequentially.]

1.	***a)***	A doctor is designing a simple prediction tool that takes a patient's health score from blood sugar levels and estimates the probability that the patient has a certain disease. The model gives z=1.4 for a patient's blood sugar reading. **Compute** the predicted probability of diabetes using the sigmoid function. Round to 3 decimal places. **Interpret** what advice might the doctor give if the threshold is 0.5?	**[Marks-2]**	**CLO-1** ***Level-1***
	b)	**Describe** the fundamental difference with examples between sigmoid function and SoftMax function.	**[Marks-3]**	**CLO-1** ***Level-2***
2.	***a)***	A data scientist is evaluating the performance of a binary classifier on an imbalanced medical dataset where the number of positive cases is very small compared to negative cases. **Explain** what a Precision–Recall (PR) curve represents and how it helps in evaluating classifier performance in such imbalanced settings.	**[Marks-4]**	**CLO-2** ***Level-2***
	b)	A machine learning company is developing a binary classifier to detect whether patients have a particular disease (Positive) or not (Negative) based on medical test data. The classifier outputs probabilities for each patient. To evaluate its performance, the company computes True Positive Rate (TPR) and False Positive Rate (FPR). (FPR, TPR): (0.0, 0.0), (0.37, 0.17), (0.59, 0.35), (0.73, 0.69), (1.0, 1.0) **Calculate** and Plot the ROC curve using the data above clearly and interpret the Curve. Also, **Calculate** the Area Under the Curve (AUC) using the trapezoidal rule. *Interpret the AUC value.*	**[Marks-6]**	**CLO-2** ***Level-3***
3.	***a)***	Daffodil International University cricket team decides whether to play a match (Yes) or cancel (No) based on four weather attributes: Outlook (Sunny / Rain / Overcast), Temperature (Hot / Mild / Cool), and Humidity (High / Normal). The dataset (09 days) is: **a. Construct** a CART decision tree (use Gini impurity to choose splits). Using your tree, predict the decision and justify it for ***Outlook = Sunny and Humidity = Normal***. *(Check table on page 2)*	**[Marks-6]**	**CLO-3** ***Level-3***

ourdiu.com | Shared by SUPAN ROY

Day	Outlook	Temp.	Humidity	Decision
1	Sunny	Hot	High	No
2	Sunny	Hot	High	No
3	Rain	Mild	High	No
4	Rain	Cool	Normal	No
5	Overcast	Cool	Normal	Yes
6	Sunny	Mild	High	No
7	Sunny	Cool	Normal	Yes
8	Rain	Mild	Normal	No
9	Sunny	Mild	Normal	Yes

		b. Calculate the Gini impurity of Decision attribute. **Explain** what Gini Impurity is and how it helps the decision tree choose the best feature to split on.	**[Marks-4]**	**CLO-3** ***Level-3***
	b)	A startup company is performing customer behavior analysis using location-based features. They collected a small sample dataset of 5 customers, each represented using two features: Annual Visits to Store (x) and Average Purchase Amount (y). The company wants to group similar customers using Agglomerative Hierarchical Clustering (Single Linkage). Five Sample Data Points are given below: C1(10,20), C2(20,10), C3(50,50), C4(60,50), C5 (80,70). a. **Construct** the **Agglomerative Hierarchical Clustering tree** (dendrogram) using **Euclidean distance**. Show the cluster merge order.	**[Marks-6]**	**CLO-3** ***Level-3***
		b. If you **cut the dendrogram at height = 30**, determine the **number of clusters**. **Explain** what height (distance) means in hierarchical intra and inter clustering. At height = 30, clusters are formed. **Calculate the total inertia** for clusters at that height.	**[Marks-4]**	**CLO-3** ***Level-2***
4.	***a)***	A health-analytics startup is building a model to classify patients into risk groups based on 20 medical indicators. Since collecting all 20 indicators is expensive, the team wants to reduce the number of input features. **Describe** the main types of feature selection and their sub-types and describe how each method could be applied in this medical data scenario.	**[Marks-5]**	**CLO-4** ***Level-1***

Good Luck