ourdiu.com | Shared by SIDDHARTHO SEN

Daffodil International University
Faculty of Science & Information Technology
Department of Software Engineering
Midterm Examination, Fall 2026
Course Code: DS421; Course Title: Machine Learning Driven Data Analysis I
Sections & Teachers: A,B[FA]; C,D,E[FH]; F,G[MRN]

Time: 1:30 Hours | Marks: 25

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 university wants to develop a student performance analysis system using three academic indicators: assignment score, quiz score, and examination score. Since these indicators may contain overlapping information, the university applies Principal Component Analysis (PCA) to transform the original features into a smaller set of uncorrelated components. The scores of four students are given below. **Apply** Principal Component Analysis (PCA) to the marks obtained in the first two quizzes to reduce the original two features to a single feature. Obtain the reduced-dimensional representation of the dataset and briefly explain how PCA helps preserve important information while reducing the number of features.	[Marks-10]	CO-1 *Level-3*
2.	A company wants to predict product prices based on their sizes using a linear regression model, $\hat{y} = wx + b$, where w is the weight and b is the bias. The company collects the following data: a product with size 1.0 has a price of 2.5, a product with size 2.0 has a price of 4.2, and a product with size 3.0 has a price of 6.8. Here, x represents the product size, and y represents its actual price in arbitrary units. Initially, the model has a weight $w = 1.0$ and a bias $b = 0.0$. The learning rate is 0.01, and Mean Squared Error (MSE) is used as the loss function.	[Marks-5+5=10]	CO-2 *Level-4*

Student	Quiz 1	Quiz 2	Quiz 3
S1	8	9	11
S2	10	11	12
S3	12	13	14
S4	14	14	13

ourdiu.com | Shared by SIDDHARTHO SEN

	a) Analyze the Batch Gradient Descent process using the entire dataset to determine the gradients for the weight and bias, identify their updated values after the first iteration, and examine the predicted prices for all three data points using the updated parameters.		
	A hospital develops a machine-learning model to classify patients into three categories based on the urgency of their medical conditions: Critical, Urgent, and Routine. The system is evaluated on 5,000 patient cases. Of these, 500 patients are classified as Critical, 1,000 as Urgent, and the remaining 3,500 as Routine. The model correctly classifies 400 Critical cases, 700 Urgent cases, and 3,000 Routine cases. Among the 500 actual Critical cases, 60 are classified as Urgent, while the remaining misclassified cases are predicted as Routine. Among the 1,000 actual Urgent cases, 200 are classified as Critical, while the remaining misclassified cases are predicted as Routine. Among the 3,500 actual Routine cases, 100 are classified as Critical and 400 are classified as Urgent. **b)** Construct a 3-class confusion matrix for the given patient classification scenario, considering Critical as the positive class. Calculate its F1-score and **examine** how the balance between precision and recall influences the model's performance compared with accuracy and precision alone.		
3.	A financial institution develops a loan approval system using a Decision Tree model. Although the model achieves high accuracy on the training data, its performance decreases significantly on unseen data. Furthermore, an evaluation reveals that applicants from rural areas are rejected more frequently than applicants from urban areas, even when they have similar financial profiles. These findings raise concerns about the model's generalization ability, fairness, and ethical implications. **Interpret** the model performance issue associated with the difference between training and unseen data in relation to the bias–variance trade-off. Also, **Assess** the ethical concern arising from the difference in loan approval rates and suggest appropriate measures to address it.	[Marks-5]	CO-3 *Level-5*