ourdiu.com | Shared by Md. Tuhid Hasan Sakib

Daffodil International University
Faculty of Science & Information Technology
Department of Software Engineering
Final Examination, Spring 2026
Course Code: SE 544; Course Title: Introduction to Machine Learning
Sections & Teachers: (A,C,D)(ZH); (B,G,H)(FA); (E,F,I,K)(MSA); (J,L,M)(NAJ); N(FF)
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 healthcare organization wants to estimate the probability that a patient will develop a certain disease based on features such as age, blood pressure, and cholesterol level. Doctors require a model that is simple, interpretable, and provides probability scores between 0 and 1 to support clinical decision-making. The outcome is binary: At Risk (1) or Not At Risk (0). Based on the above scenario, identify the most suitable machine learning algorithm for this task. **Justify** your choice. Write the mathematical equation of the model, specify the activation function used, and explain how it works in this context.	[Marks-5]	CO-1 *Level-5*
2.	"Silicon Delta VC" is evaluating 9 early-stage startups to decide whether to provide funding (Decision: Yes/No). The decision-making process is based on three key indicators: Product Stage (MVP, Prototype, or Idea), Market Size (Large or Niche), and Team Experience (Experienced or Newbie).	[Marks-2+8 = 10]	CO-2 *Level-4*

Startup	Product Stage	Market Size	Team Exp.	Funded
1	MVP	Large	Experienced	Yes
2	MVP	Niche	Experienced	Yes
3	Idea	Large	Newbie	No
4	Prototype	Niche	Newbie	No
5	Idea	Niche	Experienced	No
6	MVP	Niche	Newbie	Yes
7	Prototype	Large	Newbie	Yes
8	Idea	Large	Experienced	No
9	Prototype	Niche	Experienced	No

a) Define Gini Impurity and **explain** how it is used to determine the best split in a decision tree.

ourdiu.com | Shared by Md. Tuhid Hasan Sakib

	b) Construct a CART Decision Tree for the dataset above. Use Gini Impurity to determine the best splits at every level. Using your constructed tree, predict the funding decision for a startup with the following profile: [Product Stage = Prototype and Market Size = Niche]. **Analyze** your answer based on the tree path.		
3.	A healthcare organization is performing patient behavior analysis using clinical features. They collected a small sample dataset of 5 patients, each represented using two features: Number of Hospital Visits per Year (x) and Average Medical Expenditure (y). Five sample data points are given below: P1 (5, 30), P2 (15, 25), P3 (40, 60), P4 (55, 65), P5 (70, 80). **a) Construct** the Agglomerative Hierarchical Clustering tree using single linkage to group similar patients. Determine and present the cluster merging order step by step and draw the corresponding dendrogram.	[Marks-10+10=20]	CO-3 *Level-3*
	A machine learning engineer at a pharmaceutical company is developing a drug dosage optimization model. During the training process, the model's loss is represented by the following cost function: $f(x) = (x + 3)^2$ where x represents the dosage adjustment parameter. The engineer initializes the parameter at $x_0 = 8$ with a learning rate of $\alpha = 0.1$.		
	b) Apply gradient descent for 5 iterations to find the optimal dosage parameter that minimizes the cost function. Show all intermediate steps clearly. Based on your results, identify the value that x is converging towards and explain the significance of this value in relation to the cost function f(x).		
4.	A healthcare organization is developing a machine learning model to predict whether patients are at risk of developing a disease based on their medical records. They trained the model using the entire dataset and achieved very high accuracy. However, when the model was deployed, its performance dropped significantly on new patient data. Further analysis revealed that the model performs well for urban patients but poorly for rural patients. Based on the above scenario, evaluate how the data splitting strategy may have contributed to this issue. **Explain** how proper data splitting techniques could improve the model's performance and reliability.	[Marks- 5]	CO-4 *Level-4*