Nirma University

Institute of Technology

Semester End Examination (IR), December - 2025

M. Tech. in Computer Science and Engineering /

M. Tech. in Computer Science and Engineering (Data Science), Semester-I

6CS203CC25 – APPLIED MACHINE LEARNING

Roll / Exam No.	25MCD005	Supervisor's initial with date	

Time: 3 Hours

Max. Marks: 100

Instructions:
1. All questions are compulsory. (No Optional Questions)
2. Use section-wise separate answer book.
3. Figures to right indicate full marks.
4. Assume necessary data wherever required and specify them.
5. Avoid attempting questions and their sub-questions in a random order.

Section – I

Q.1 Answer following questions: **[20]**

(A) CO1, BL3 Consider a binary classification dataset: [10]

$$\{(x_i, y_i) \mid x_i \in \mathbb{R}^n,\ y_i \in \{-1, +1\},\ i = 1, \ldots, m\}.$$

Derive the hard-margin Support Vector Machine (SVM) formulation by applying Lagrange optimization. Start from the primal problem, construct the Lagrange function and Kuhn–Tucker conditions. Derivation should clearly show the transition from the primal to the dual form.

(B) CO4, BL5 The confusion matrix of a classification model with three classes (A, B, and C) is given below: [10]

Predicted Class	Actual Class: A	Actual Class: B	Actual Class: C
A	45	3	2
B	4	53	1
C	5	6	33

Compute Precision, Recall, Accuracy and F1 score for Class A, B and class C for the evaluation of the classification results.

Q.2 Answer following questions: **[15]**

(A) CO4, BL4 Explain, with justification, why the Normal Equation method is appropriate for model parameter estimation in some situations. In your explanation, compare it with gradient descent optimization method in terms of memory usage, computation time, and suitability for large datasets (in terms of number of features). [05]

(B) CO3, BL2 Compare MAE, MSE, and RMSE as evaluation metrics for regression. Explain how each metric treats errors, why MSE and RMSE penalize larger errors more heavily, and in which situations each metric is preferred? [10]

Q.3 CO2, BL3 You are given the following dataset of two features — Height (in mm) and Weight (in kg) — and a binary class label indicating whether a person is an Athlete (Yes) or Non-Athlete (No). **[15]**

Height (mm)	Weight (kg)	Class
1800	80	Yes
1750	70	Yes
1600	60	No
1650	65	No
1780	75	Yes
1630	62	No

A new person has the following attributes: Height = 1700 mm, Weight = 68 kg.

(i) Compute the mean and variance of each feature for both classes (Yes and No).
(ii) Apply Gaussian normalization to each feature.
(iii) Calculate Priors – P(Yes) and P(No).
(iv) P(Height=1700|Yes), P(Weight=68|Yes) and P(Height=1700|No), P(Weight=68|No).
(v) Predict the class label for the new person using Gaussian Naïve Bayes Classifier.

Section – II

Q.4 Answer following questions: **[20]**

(A) CO3, BL4 Some semi-supervised techniques use more than one classifier trained on different feature perspectives of the same data. Explain how such an approach may help enhance learning performance. Compare this approach with a semi-supervised method that relies on only a single classifier. [10]

(B) CO3, BL3 Using Hierarchical Agglomerative Clustering with complete linkage and using similarity as the proximity measure, cluster the six objects (T1 to T6) given in the table below. [10]

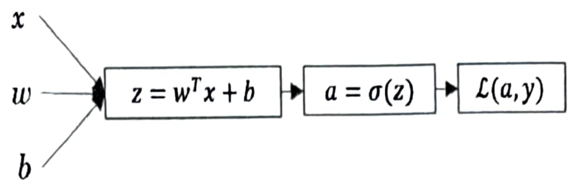
(i) At each iteration, identify which two clusters are merged, and specify the similarity value at which the merge occurs.
(ii) Construct the dendrogram, labelling the objects 1–6 and marking the similarity level for each merge.
(iii) If the dendrogram is cut at similarity = 0.60, identify the resulting clusters and list the objects in each cluster.

Similarity Matrix

	T1	T2	T3	T4	T5	T6
T1	1	0.85	0.40	0.30	0.55	0.20
T2	0.85	1	0.35	0.25	0.60	0.15
T3	0.40	0.35	1	0.90	0.70	0.30
T4	0.30	0.25	0.90	1	0.75	0.45
T5	0.55	0.60	0.70	0.75	1	0.25
T6	0.20	0.15	0.30	0.45	0.25	1

Q.5 Answer following questions: **[20]**

(A) CO1, BL3 — Consider following computation graph for the single layer neural network with sigmoid activation function and loss model as $\sum_{i=1}^{n} -1/n[y_i \log(a_i)+(1-y_i)\log(1-a_i)]$. [10]



Make use of backward propagation steps to derive equations for gradient calculation with respect to w and b. Also derive weight and bias update steps for gradient descent method.

(B) CO3, BL4 — Compare and contrast sigmoid, tanh, ReLU, and SoftMax activation functions in terms formula, range, and typical use-cases. [10]

Q.6 CO4, BL5 — Using the ID3 decision tree method, determine which attribute (Study Hours, Attendance, or Assignment Submission) provides the highest Information Gain for predicting whether a student will score Above Average. **[10]**

Study Hours	**Attendance**	**Assignment Submission**	**Score Level**
High	Good	Yes	Above Avg
High	Poor	Yes	Above Avg
Low	Good	No	Below Avg
Low	Poor	No	Below Avg
High	Good	Yes	Above Avg
