🎓 **University of Data Sciences**
Semester Exam: Machine Learning (Theory + Practical Concepts)
Time: 3 Hours | Full Marks: 100
Instructions:

- Attempt all questions from **Section A (Compulsory)**
- Answer **any 3 from Section B** and **any 2 from Section C**
- Show all necessary steps in derivations and numericals
- Use appropriate diagrams/visuals where applicable

🧩 Section A: Compulsory – 30 Marks (Short Answer)

Answer all questions briefly (3–4 lines or as required)

1. Define Machine Learning. Differentiate between Supervised and Unsupervised learning.
2. What is the role of the learning rate in gradient descent?
3. Mention two assumptions of Linear Regression.
4. Define R^2 and Adjusted R^2. Why is Adjusted R^2 preferred in multiple regression?
5. Write the cost function for Ridge Regression.
6. What is the curse of dimensionality in the context of KNN?
7. List any two key differences between Logistic Regression and Linear Regression.
8. Define VIF and its use in detecting multicollinearity.
9. What is data leakage? Give one example.
10. Compare Batch, Stochastic, and Mini-batch Gradient Descent.

📘 Section B: Theory & Derivation – 40 Marks (Any 3 x 13.33)

Q1. Linear Regression & Error Metrics
a) Derive the normal equation for Simple Linear Regression.
b) Define and derive the formula for Mean Squared Error (MSE).
c) Explain the geometrical interpretation of regression line fitting.

Q2. Gradient Descent
a) Derive the update rule for θ_0 and θ_1 in gradient descent for linear regression.
b) Explain how learning rate affects convergence.
c) What happens when the cost function is non-convex?

Q3. Regularization Techniques
a) Explain the bias-variance tradeoff with a diagram.
b) Derive the Ridge Regression cost function and show how it modifies the normal equation.
c) Contrast Lasso and Ridge in terms of feature selection.

Q4. Principal Component Analysis (PCA)
a) Derive PCA from first principles using variance maximization.
b) Explain the role of eigenvectors and eigenvalues in PCA.
c) Why is standardization important before applying PCA?

Q5. Logistic Regression
a) Derive the log-likelihood function for logistic regression.
b) Explain how gradient descent is applied to logistic regression.
c) Interpret the meaning of coefficients in logistic regression.

🔢 Section C: Numericals & Applied Problems – 30 Marks (Any 2 x 15)

Q6. Multiple Linear Regression & Multicollinearity
Given a dataset with 3 predictors and a dependent variable, perform multiple linear regression and compute the R^2 and Adjusted R^2. Also, calculate VIF for each predictor and interpret.

Q7. Naive Bayes Classifier (Numerical)
You are given the following data about email spam classification. Using Naive Bayes, calculate the probability that a new email with specific word counts is spam. Include Laplace smoothing.

Q8. PCA Numerical
Given a 2D dataset:

$$X = \begin{bmatrix} 2 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix}$$

a) Standardize the data
b) Compute covariance matrix
c) Extract eigenvalues and eigenvectors
d) Project the data to 1D using PCA

Q9. Gradient Descent Simulation
Implement gradient descent from scratch (pseudo-code allowed) for the cost function:

$$J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})^2$$

Use initial $\theta = 0$, learning rate = 0.01, data with 3 points. Show 3 iterations manually.

Q10. Classification Metrics & Confusion Matrix
A binary classifier gives the following results:

- TP = 50, FP = 10, TN = 30, FN = 10
  Calculate:
  a) Accuracy
  b) Precision
  c) Recall
  d) F1-score
  e) Plot the confusion matrix diagram

✍️ Bonus Question (5 Marks)

What are the advantages and limitations of using Ensemble Methods like Random Forest and Gradient Boosting in real-world datasets?