1. **Machine Learning Concepts**

A. Designing a Learning System

Theory:

1. What are the main components of a machine learning system?
2. Explain the steps involved in designing a machine learning system.
3. Discuss the difference between training, validation, and testing datasets.

Derivation:

1. Derive the expression for empirical risk minimization in supervised learning.

Numerical:

1. Given a dataset, split it into training, validation, and test sets in a 60:20 ratio.

B. Styles of Learning

Theory:

1. Define supervised, unsupervised, and semi-supervised learning with examples.
2. Compare and contrast supervised and unsupervised learning.
3. What are the advantages of semi-supervised learning?

Numerical:

1. Given a dataset with 1000 labeled and 9000 unlabeled samples, suggest a suitable learning style and justify your answer.

C. Supervised Learning

Theory:

1. What is supervised learning? List its applications.
2. Explain the bias-variance tradeoff in supervised learning.

Numerical:

1. Given a set of input-output pairs, fit a simple linear regression model.

D. Unsupervised Learning

Theory:

1. What is unsupervised learning? Give two real-world examples.
2. Discuss the challenges in evaluating unsupervised learning algorithms.

E. Semi-Supervised Learning

Theory:

1. What is semi-supervised learning? How does it differ from supervised and unsupervised learning?
2. Discuss one algorithm for semi-supervised learning.

F. Basics of Decision Theory

Theory:

1. What is decision theory? How is it applied in machine learning?
2. Define loss function and risk function.

Derivation:

1. Derive the Bayes optimal classifier for a two-class problem.

Numerical:

1. Given a loss matrix and class probabilities, compute the expected risk for each action.

G. Information Theory and Probability Distributions

Theory:

1. Define entropy and mutual information.
2. Explain the importance of probability distributions in machine learning.

Derivation:

1. Derive the formula for entropy for a discrete random variable.

Numerical:

1. Calculate the entropy of a binary variable with probabilities 0.7 and 0.3.

H. Linear and Logistic Regression

Theory:

1. What is the difference between linear and logistic regression?
2. List the assumptions of linear regression.

Derivation:

1. Derive the normal equations for linear regression using least squares.
2. Derive the gradient descent update rule for logistic regression.

Numerical:

1. Given a small dataset, fit a linear regression model and compute the coefficients.
2. For a binary classification dataset, fit a logistic regression model and predict the class for a new sample.

---

2. **Bayesian Learning**

A. Notion of Prior, Likelihood, and Posterior

Theory:

1. Define prior, likelihood, and posterior in Bayesian learning.
2. Explain the significance of the prior in Bayesian inference.

Derivation:

1. Derive Bayes' theorem for parameter estimation.

Numerical:

1. Given prior and likelihood values, compute the posterior probability.

B. Naïve Bayes and Conditional Independence

Theory:

1. What is the Naïve Bayes assumption?
2. Explain conditional independence with an example.

Derivation:

1. Derive the Naïve Bayes classifier formula.

Numerical:

1. Given a dataset with features and class labels, compute the class probabilities using Naïve Bayes.

C. Estimation using Maximum Likelihood

Theory:

1. What is the principle of maximum likelihood estimation (MLE)?

2. Discuss the difference between MLE and MAP estimation.

Derivation:

1. Derive the MLE for the mean of a Gaussian distribution.

Numerical:

1. Given a set of data points, estimate the parameters of a normal distribution using MLE.

D. Hidden Variables and Missing Data

Theory:

1. What are hidden variables? How do they affect learning?
2. Discuss strategies for handling missing data in machine learning.

Numerical:

1. Given a dataset with missing values, impute the missing entries using mean imputation.

E. Bayesian Models

Theory:

1. What are Bayesian models? List their advantages and disadvantages.
2. Explain the concept of model evidence in Bayesian learning.

---

3. **Classification & Clustering**

A. Naïve Bayes

Theory:

1. List the steps involved in training and testing a Naïve Bayes classifier.
2. Discuss the limitations of Naïve Bayes.

Numerical:

1. Given a small dataset, classify a new instance using Naïve Bayes.

B. Nearest Neighbour and Linear Classification Models

Theory:

1. Explain the k-Nearest Neighbour (k-NN) algorithm.
2. What are linear classification models? Give examples.

Numerical:

1. Given a set of points, classify a new point using 3-NN.
2. Fit a perceptron model to a linearly separable dataset.

C. K-means and Expectation Maximization for Clustering

Theory:

1. Describe the K-means clustering algorithm.
2. What is the Expectation Maximization (EM) algorithm?

Derivation:

1. Derive the update rules for the K-means algorithm.
2. Derive the E-step and M-step for a Gaussian Mixture Model.

Numerical:

1. Cluster a set of points into 2 clusters using K-means (show all steps).
2. Given a simple dataset, perform one iteration of the EM algorithm.

D. Mixture Models

Theory:

1. What are mixture models?
2. Explain the difference between a mixture model and a single distribution model.

Numerical:

1. Given data generated from two Gaussians, estimate the parameters using the EM algorithm.

E. Flat and Hierarchical Clustering

Theory:

1. Compare flat and hierarchical clustering.
2. Describe the agglomerative hierarchical clustering process.

Numerical:

1. Given a distance matrix, perform one step of agglomerative clustering.

F. Applications of Classification and Clustering

Theory:

1. List real-world applications of classification and clustering.
2. Discuss the use of clustering in image segmentation.

How to Use This Question Bank

- **Theory questions** are for conceptual understanding and short/long answers.
- **Derivation questions** test mathematical and algorithmic understanding.
- **Numerical questions** are for hands-on problem-solving and application.