📘 CS209: Machine Learning – Question Bank

🔹 1. Machine Learning Concepts

📖 Theory Questions:

1. Define Machine Learning. How does it differ from traditional programming?
2. Explain the steps involved in designing a learning system.
3. Compare and contrast supervised, unsupervised, and semi-supervised learning.
4. What are the key challenges in machine learning?
5. Explain the concept of overfitting and underfitting with examples.
6. Describe the bias-variance tradeoff in machine learning.
7. What is the role of a hypothesis space in learning?

🧮 Derivation-Based Questions:

1. Derive the cost function for linear regression using the least squares method.
2. Derive the gradient descent update rule for linear regression.
3. Derive the log-likelihood function for logistic regression.
4. Show how the sigmoid function is used in logistic regression and derive its gradient.

🔢 Numerical Problems:

1. Given a dataset, fit a linear regression model and compute the mean squared error.
2. Implement logistic regression on a binary classification dataset and compute accuracy.
3. Given a dataset, perform feature scaling and normalization.
4. Given a confusion matrix, compute precision, recall, F1-score, and accuracy.

🔹 2. Bayesian Learning

📖 Theory Questions:

1. Explain the concepts of prior, likelihood, and posterior with examples.
2. What is the difference between frequentist and Bayesian approaches?
3. Describe the Naïve Bayes classifier. Why is the assumption of conditional independence important?
4. What are the advantages and limitations of Bayesian learning?
5. Explain the concept of marginal likelihood and its role in model selection.

🧮 Derivation-Based Questions:

1. Derive Bayes' Theorem from the definition of conditional probability.

2. Derive the Maximum Likelihood Estimation (MLE) for a Gaussian distribution.
3. Derive the posterior distribution for a Bernoulli likelihood with a Beta prior.
4. Show how the Naïve Bayes classifier is derived from Bayes' Theorem.

🔢 Numerical Problems:

1. Given a dataset, compute the posterior probability using Bayes' Theorem.
2. Implement a Naïve Bayes classifier on a text classification problem.
3. Given a dataset with missing values, estimate the parameters using Expectation-Maximization (EM).
4. Compute the MAP estimate for a Gaussian distribution with a known prior.

🔹 3. Classification & Clustering

📖 Theory Questions:

1. Compare Naïve Bayes, k-NN, and linear classifiers in terms of accuracy and complexity.
2. Explain the working of the k-Nearest Neighbour algorithm.
3. Describe the K-means clustering algorithm. What are its limitations?
4. What is the Expectation-Maximization algorithm? How is it used in clustering?
5. Differentiate between flat and hierarchical clustering.
6. What are mixture models? How do they relate to clustering?
7. Discuss real-world applications of classification and clustering.

🧮 Derivation-Based Questions:

1. Derive the decision boundary for a linear classifier.
2. Derive the update rules for the K-means algorithm.
3. Derive the E-step and M-step for the EM algorithm in Gaussian Mixture Models.
4. Show how the distance metric affects the performance of k-NN.

🔢 Numerical Problems:

1. Given a dataset, implement k-NN and evaluate its performance using cross-validation.
2. Apply K-means clustering to a dataset and compute the within-cluster sum of squares.
3. Given a dataset, perform hierarchical clustering and draw the dendrogram.
4. Fit a Gaussian Mixture Model using EM and compute the log-likelihood.
5. Given a confusion matrix for a classifier, compute the ROC curve and AUC.

✅ Bonus: Mixed Conceptual Questions

1. How does logistic regression relate to Naïve Bayes?
2. Compare EM algorithm and K-means in terms of convergence and assumptions.
3. How can clustering be used to improve supervised learning?
4. Discuss the role of information theory in machine learning.
5. How does semi-supervised learning bridge the gap between supervised and unsupervised learning?