

SRM VALLIAMMAI ENGINEERING COLLEGE

**(An Autonomous Institution)
SRM Nagar, Kattankulathur – 603 203**

DEPARTMENT OF COMPUTER APPLICATIONS

QUESTION BANK



III SEMESTER M.C.A.

MC4361 - MACHINE LEARNING

Regulation – 2024

Academic Year 2025 – 2026 (ODD Semester)

Prepared by

Ms.D.Revathy

Assistant Professor(O.G)

SRM VALLIAMMAI ENGINEERING COLLEGE

(An Autonomous Institution)

SRM Nagar, Kattankulathur-603 203

DEPARTMENT OF COMPUTER APPLICATIONS

QUESTION BANK

SUBJECT : MC4161 - MC4361 - MACHINE LEARNING

SEM/YEAR: III / II

UNIT I - INTRODUCTION

Human Learning - Types – Machine Learning - Types - Problems not to be solved - Applications- Languages/Tools– Issues. Preparing to Model: Introduction – Machine Learning Activities -Types of data - Exploring structure of data - Data quality and remediation - Data Pre-processing.

UNIT - I: PART – A

Q. No	Question	BT Level	Competence	Course Outcome
1	Define Human Learning.	BTL1	Remember	CO1
2	List the types of Human Learning.	BTL1	Remember	CO1
3	What is Machine Learning?	BTL1	Remember	CO1
4	Name the types of Machine Learning.	BTL1	Remember	CO1
5	What is supervised learning?	BTL2	Understand	CO1
6	What is unsupervised learning?	BTL2	Understand	CO1
7	Define reinforcement learning.	BTL1	Remember	CO1
8	What is the difference between AI and ML?	BTL2	Understand	CO1
9	List any four applications of Machine Learning.	BTL1	Remember	CO1
10	Mention two problems that cannot be solved using ML.	BTL1	Remember	CO1
11	What are some popular ML tools or languages?	BTL1	Remember	CO1
12	What do you mean by 'Preparing to Model'?	BTL2	Understand	CO1
13	State the purpose of data pre-processing.	BTL2	Understand	CO1
14	Differentiate between structured and unstructured data.	BTL2	Understand	CO1
15	What are nominal and ordinal data types?	BTL1	Remember	CO1
16	What is the role of data exploration in ML?	BTL2	Understand	CO1
17	Define data quality in ML.	BTL1	Remember	CO1
18	What is missing data?	BTL1	Remember	CO1

19	List any two data remediation techniques.	BTL1	Remember	CO1
20	Define outliers in a dataset.	BTL1	Remember	CO1
21	What is normalization?	BTL2	Understand	CO1
22	What is meant by scaling data?	BTL2	Understand	CO1
23	List any two data pre-processing methods.	BTL1	Remember	CO1
24	Define data transformation.	BTL1	Remember	CO1
25	Define Data Pre-processing?	BTL1	Remember	CO1

UNIT - I: PART – B

Q. No	Question	Mark	BT Level	Competence	Course Outcome
1	Explain in detail the types of human learning with examples.	16	BTL3	Apply	CO1
2	Discuss the various types of machine learning with real-world examples.	16	BTL3	Apply	CO1
3	Write a detailed note on problems that cannot be solved using machine learning.	16	BTL4	Analyze	CO1
4	Explain at least five key applications of ML in various domains.	16	BTL3	Apply	CO1
5	Compare and contrast the tools and languages used in ML (like Python, R, TensorFlow, etc.).	16	BTL4	Analyze	CO1
6	What are the challenges/issues faced in implementing ML solutions?	16	BTL4	Analyze	CO1
7	With examples, describe all the activities involved in preparing to model.	16	BTL4	Analyze	CO1
8	How do the different types of data impact machine learning modeling?	16	BTL4	Analyze	CO1
9	Explain the concept of data structure exploration. Why is it crucial before modeling?	16	BTL5	Evaluate	CO1
10	Describe the issues in data quality. How are they addressed during the ML pipeline?	16	BTL5	Evaluate	CO1
11	Discuss the importance of data pre-processing in ML.	16	BTL3	Apply	CO1
12	Outline various data pre-processing techniques and their advantages.	16	BTL4	Analyze	CO1
13	Create a case study showing the impact of improper data preprocessing in model accuracy.	16	BTL6	Create	CO1
14	Analyze the structure of a dataset (e.g., Titanic dataset) and explain key insights derived.	16	BTL6	Create	CO1
15	Evaluate different data remediation methods and justify the selection for noisy data.	16	BTL5	Evaluate	CO1

16	Design an end-to-end ML workflow including learning type, data prep, modeling, and application.	16	BTL5	Evaluate	CO1
17	Create a case study showing the impact of improper data preprocessing in model accuracy.	16	BTL3	Apply	CO1

UNIT II - MODEL EVALUATION AND FEATURE ENGINEERING

Model Selection - Training Model - Model Representation and Interpretability – Evaluating Performance of a Model - Improving Performance of a Model - Feature Engineering: Feature Transformation - Feature Subset Selection.

UNIT II PART – A

Q. No	Questions	BT Level	Competence	Course Outcome
1	Define model selection.	BTL1	Remember	CO2
2	What is the purpose of training a model?	BTL1	Remember	CO2
3	Define underfitting and overfitting.	BTL1	Remember	CO2
4	List any two performance evaluation metrics.	BTL1	Remember	CO2
5	What is cross-validation?	BTL1	Remember	CO2
6	Name two types of model interpretability techniques.	BTL1	Remember	CO2
7	What is the goal of feature engineering?	BTL1	Remember	CO2
8	Define feature transformation.	BTL1	Remember	CO2
9	What is feature selection?	BTL1	Remember	CO2
10	Differentiate between feature selection and feature extraction.	BTL2	Understand	CO2
11	Explain the term ‘bias-variance tradeoff’.	BTL2	Understand	CO2
12	What do you mean by model interpretability?	BTL2	Understand	CO2
13	Differentiate between training and testing datasets.	BTL2	Understand	CO2
14	How does regularization improve model performance?	BTL2	Understand	CO2
15	What is the importance of normalization in feature transformation?	BTL2	Understand	CO2
16	Define precision and recall.	BTL1	Remember	CO2
17	Explain confusion matrix briefly.	BTL2	Understand	CO2
18	Why is feature subset selection required?	BTL2	Understand	CO2
19	Mention any two filter-based feature selection techniques.	BTL1	Remember	CO2
20	What is AUC-ROC in model evaluation?	BTL2	Understand	CO2
21	List two wrapper methods used in feature selection.	BTL1	Remember	CO2
22	What is the role of hyperparameter tuning in model training?	BTL2	Understand	CO2
23	What is the significance of training-validation-test split?	BTL2	Understand	CO2
24	Explain early stopping in training a model.	BTL2	Understand	CO2
25	What is L1 and L2 regularization?	BTL1	Remember	CO2

UNIT II PART – B

Q. No	Question	Mark	BT Level	Competence	Course Outcome
1	Explain the steps involved in model selection. Compare different model selection strategies.	16	BTL3	Apply	CO2
2	Discuss the various challenges in training a machine learning model and how to address them.	16	BTL4	Analyze	CO2
3	Illustrate with examples how model interpretability helps in decision making.	16	BTL4	Analyze	CO2
4	Evaluate different model performance metrics and explain when to use each.	16	BTL5	Evaluate	CO2
5	Explain bias-variance tradeoff. How do they affect model performance?	16	BTL4	Analyze	CO2
6	Design an experiment using k-fold cross-validation to compare two machine learning models.	16	BTL6	Create	CO2
7	Analyze the impact of data preprocessing and feature scaling on model performance with a case study.	16	BTL4	Analyze	CO2
8	Compare filter, wrapper, and embedded methods for feature subset selection.	16	BTL4	Analyze	CO2
9	Propose a feature transformation technique for handling skewed data distribution and justify your approach.	16	BTL6	Create	CO2
10	Discuss how hyperparameter tuning improves the performance of ML models. Explain with examples.	16	BTL5	Evaluate	CO2
11	Apply dimensionality reduction techniques (e.g., PCA) and interpret the results.	16	BTL3	Apply	CO2
12	Develop a classification model pipeline including feature engineering, model training, and evaluation.	16	BTL6	Create	CO2
13	Explain the role of regularization in improving model generalization. Compare L1 and L2 with examples.	16	BTL5	Evaluate	CO2
14	Assess different approaches to improve model performance on imbalanced datasets.	16	BTL5	Evaluate	CO2
15	Justify the need for interpretable models in sensitive domains like healthcare and finance.	16	BTL5	Evaluate	CO2
16	Illustrate the end-to-end process of feature engineering for a regression task.	16	BTL3	Apply	CO2
17	Analyze common mistakes in model evaluation and their consequences in deployment.	16	BTL4	Analyze	CO2
18	Propose an optimal feature subset selection technique for high-dimensional data and validate its effectiveness.	16	BTL6	Create	CO2

UNIT III - BAYESIAN LEARNING

Basic Probability Notation- Inference – Independence - Bayes’ Rule. Bayesian Learning: Maximum Likelihood and Least Squared error hypothesis- Maximum Likelihood hypotheses for predicting probabilities- Minimum description Length principle – Bayes optimal classifier – Naïve Bayes classifier - Bayesian Belief networks -EM algorithm.

Q. No	Questions	BT Level	Competence	Course Outcome
UNIT III PART – A				
1	Define probability.	BTL1	Remember	CO3
2	What is a random variable?	BTL1	Remember	CO3
3	Write the basic probability notation for a conditional probability.	BTL1	Remember	CO3
4	State Bayes’ Theorem.	BTL1	Remember	CO3
5	Define prior and posterior probabilities.	BTL1	Remember	CO3
6	What is the concept of independence in probability?	BTL1	Remember	CO3
7	Distinguish between dependent and independent events.	BTL2	Understand	CO3
8	What is inference in probability theory?	BTL2	Understand	CO3
9	Mention any two assumptions of Naïve Bayes classifier.	BTL1	Remember	CO3
10	What is Maximum Likelihood Estimation (MLE)?	BTL1	Remember	CO3
11	What is Least Squares Error hypothesis?	BTL1	Remember	CO3
12	State the concept of the Minimum Description Length (MDL) principle.	BTL2	Understand	CO3
13	Define likelihood function.	BTL1	Remember	CO3
14	What is the goal of Bayesian Learning?	BTL2	Understand	CO3
15	State two limitations of Naïve Bayes classifier.	BTL2	Understand	CO3
16	Differentiate MAP and MLE estimation.	BTL2	Understand	CO3
17	Write the formula for the Naïve Bayes classification.	BTL1	Remember	CO3
18	What is a Bayesian Belief Network?	BTL1	Remember	CO3
19	What are nodes and edges in Bayesian Networks?	BTL2	Understand	CO3
20	Define EM algorithm in brief.	BTL1	Remember	CO3
21	State the two main steps of the EM algorithm.	BTL2	Understand	CO3
22	What is a hypothesis in machine learning?	BTL1	Remember	CO3
23	Define posterior probability.	BTL1	Remember	CO3
24	Explain the term ‘evidence’ in Bayes’ rule.	BTL2	Understand	CO3
25	What is the purpose of Bayesian networks?	BTL2	Understand	CO3

UNIT III PART – B

Q. No	Question	Mark	BT Level	Competence	Course Outcome
1	Explain the basic probability notation and rules with suitable examples.	16	BTL3	Apply	CO3
2	Derive Bayes’ Rule and explain how it is used in Bayesian Learning.	16	BTL4	Analyze	CO3
3	Explain the concept of conditional probability and independence with an example.	16	BTL4	Analyze	CO3
4	Discuss inference and its importance in probabilistic models.	16	BTL3	Apply	CO3
5	Compare Maximum Likelihood Estimation and Least Squares Error Hypothesis with examples.	16	BTL4	Analyze	CO3

6	Describe the process of estimating probabilities using Maximum Likelihood Hypotheses.	16	BTL3	Apply	CO3
7	Explain the Minimum Description Length (MDL) principle and how it affects hypothesis selection.	16	BTL4	Analyze	CO3
8	Describe the working of Naïve Bayes classifier with a step-by-step example.	16	BTL5	Evaluate	CO3
9	Evaluate the performance of Naïve Bayes classifier on a given dataset.	16	BTL5	Evaluate	CO3
10	Describe the structure and working of Bayesian Belief Networks with an example.	16	BTL4	Analyze	CO3
11	Differentiate between Naïve Bayes and Bayesian Networks with use cases.	16	BTL4	Analyze	CO3
12	Derive the expression for Bayes Optimal Classifier and explain its significance.	16	BTL5	Evaluate	CO3
13	Explain the steps of the Expectation-Maximization (EM) algorithm with a real-time example.	16	BTL3	Apply	CO3
14	Implement Naïve Bayes classifier algorithm and show intermediate steps on a sample dataset.	16	BTL6	Create	CO3
15	Apply EM algorithm to solve a clustering problem involving missing values.	16	BTL6	Create	CO3
16	Analyze the advantages and disadvantages of Bayesian Learning in practical machine learning tasks.	16	BTL5	Evaluate	CO3
17	Illustrate with diagram how Bayesian Networks help in decision making under uncertainty.	16	BTL4	Analyze	CO3
18	Propose a real-world application using Bayesian networks and describe how inference is performed.	16	BTL6	Create	CO3

UNIT IV - PARAMETRIC MACHINE LEARNING

Logistic Regression: Classification and representation – Cost function – Gradient descent – Advanced optimization – Regularization - Solving the problems on overfitting. Perceptron – Neural Networks – Multi – class Classification - Backpropagation – Nonlinearity with activation functions (Tanh, Sigmoid, Relu, PRelu) - Dropout as regularization.

Q. No	Questions	BT Level	Competence	Course Outcome
UNIT IV PART – A				
1	What is logistic regression?	BTL1	Remember	CO4
2	Define the sigmoid function.	BTL1	Remember	CO4
3	Differentiate between classification and regression.	BTL2	Understand	CO4
4	What is a cost function in logistic regression?	BTL1	Remember	CO4
5	State the role of gradient descent in logistic regression.	BTL2	Understand	CO4
6	What do you mean by regularization?	BTL1	Remember	CO4
7	List any two types of regularization techniques.	BTL1	Remember	CO4
8	What is overfitting in machine learning?	BTL1	Remember	CO4
9	How can overfitting be avoided in logistic regression?	BTL2	Understand	CO4
10	Define perceptron.	BTL1	Remember	CO4

11	Compare perceptron and logistic regression.	BTL2	Understand	CO4
12	What is the role of weights in a perceptron?	BTL1	Remember	CO4
13	What is a neural network?	BTL1	Remember	CO4
14	What is the function of an activation function?	BTL2	Understand	CO4
15	Define backpropagation.	BTL1	Remember	CO4
16	What is the need for multi-class classification?	BTL2	Understand	CO4
17	What is the difference between one-vs-rest and one-vs-one approaches?	BTL2	Understand	CO4
18	Name any three activation functions used in neural networks.	BTL1	Remember	CO4
19	Write the formula for ReLU activation function.	BTL1	Remember	CO4
20	Compare Tanh and Sigmoid functions.	BTL2	Understand	CO4
21	What is PReLU and how is it different from ReLU?	BTL2	Understand	CO4
22	Define dropout regularization.	BTL1	Remember	CO4
23	What is the significance of using dropout in deep networks?	BTL2	Understand	CO4
24	What is the vanishing gradient problem in deep learning?	BTL2	Understand	CO4
25	How does backpropagation update the weights in a neural network?	BTL2	Understand	CO4

UNIT IV PART – B

Q. No	Question	Mark	BT Level	Competence	Course Outcome
1	Explain the representation of logistic regression and derive the cost function.	16	BTL4	Analyze	CO4
2	Describe how gradient descent is applied in logistic regression with derivations.	16	BTL4	Analyze	CO4
3	Implement a binary classifier using logistic regression for a given dataset and explain your steps.	16	BTL3	Apply	CO4
4	Explain L1 and L2 regularization. How do they help in solving overfitting?	16	BTL4	Analyze	CO4
5	Design a model using logistic regression to classify handwritten digits and evaluate its performance.	16	BTL5	Evaluate	CO4
6	Describe perceptron learning rule and implement it for a binary classification task.	16	BTL3	Apply	CO4
7	Compare perceptron and logistic regression with suitable examples and diagrams.	16	BTL4	Analyze	CO4
8	Explain the architecture and working of a simple feedforward neural network.	16	BTL4	Analyze	CO4
9	Derive the backpropagation algorithm with suitable mathematical expressions.	16	BTL4	Analyze	CO4
10	Illustrate the importance and functionality of activation functions like Tanh, Sigmoid, ReLU, and PReLU.	16	BTL4	Analyze	CO4
11	Compare different activation functions with plots and highlight their advantages and drawbacks.	16	BTL5	Evaluate	CO4
12	Discuss the concept of dropout regularization and explain how it combats overfitting.	16	BTL4	Analyze	CO4
13	Develop a neural network model to solve a multi-class classification problem.	16	BTL6	Create	CO4

14	Illustrate with an example how backpropagation works in a multi-layer network.	16	BTL3	Apply	CO4
15	Explain how advanced optimization techniques (e.g., Adam, RMSprop) improve training of neural networks.	16	BTL4	Analyze	CO4
16	Evaluate the performance of a neural network using different activation functions.	16	BTL5	Evaluate	CO4
17	Propose a hybrid regularization strategy combining dropout and L2 regularization.	16	BTL6	Create	CO4
18	Solve a real-world classification problem using logistic regression and report confusion matrix results.	16	BTL5	Evaluate	CO4

UNIT – V NON PARAMETRIC MACHINE LEARNING

k- Nearest Neighbors- Decision Trees – Branching – Greedy Algorithm – Multiple Branches –Continuous attributes – Pruning. Random Forests: ensemble learning. Boosting – Adaboost algorithm. Support Vector Machines – Large Margin Intuition – Loss Function - Hinge Loss – SVM Kernels.

Q. No	Questions	BT Level	Competence	Course Outcome
UNIT V PART – A				
1	What is k-Nearest Neighbors (k-NN)?	BTL1	Remember	CO5
2	State any two applications of k-NN.	BTL1	Remember	CO5
3	Define Decision Tree.	BTL1	Remember	CO5
4	What is meant by Greedy Algorithm in decision trees?	BTL2	Understand	CO5
5	Define entropy in the context of decision trees.	BTL1	Remember	CO5
6	What is branching in decision trees?	BTL1	Remember	CO5
7	How are continuous attributes handled in decision trees?	BTL2	Understand	CO5
8	What is pruning in decision trees?	BTL2	Understand	CO5
9	Mention two disadvantages of decision trees.	BTL1	Remember	CO5
10	Define ensemble learning.	BTL1	Remember	CO5
11	What is a Random Forest?	BTL1	Remember	CO5
12	Differentiate between bagging and boosting.	BTL2	Understand	CO5
13	What is AdaBoost?	BTL1	Remember	CO5
14	How does AdaBoost improve weak learners?	BTL2	Understand	CO5
15	What is the principle behind boosting algorithms?	BTL2	Understand	CO5
16	Define Support Vector Machine (SVM).	BTL1	Remember	CO5
17	What is the objective of a large margin classifier?	BTL2	Understand	CO5
18	Define Hinge Loss in SVM.	BTL1	Remember	CO5
19	What is the role of the kernel function in SVM?	BTL2	Understand	CO5
20	What are the types of SVM kernels?	BTL1	Remember	CO5
21	State any two advantages of SVM.	BTL1	Remember	CO5
22	What is the decision boundary in SVM?	BTL2	Understand	CO5
23	What happens when data is not linearly separable in SVM?	BTL2	Understand	CO5
24	What is a support vector?	BTL1	Remember	CO5
25	Write two differences between Decision Trees and SVM.	BTL2	Understand	CO5

UNIT V PART – B

Q. No	Question	Mark	BT Level	Competence	Course Outcome
1	Explain the working of the k-Nearest Neighbor algorithm with a suitable example.	16	BTL3	Apply	CO5
2	Discuss the advantages, disadvantages, and limitations of k-NN.	16	BTL4	Analyze	CO5
3	Construct a Decision Tree using the ID3 algorithm for a given dataset.	16	BTL4	Analyze	CO5
4	Compare Gini Index and Information Gain used in decision trees.	16	BTL4	Analyze	CO5
5	Explain the concept of pruning in decision trees with examples.	16	BTL3	Apply	CO5
6	How do decision trees handle continuous and categorical attributes?	16	BTL4	Analyze	CO5
7	Write and explain the Greedy approach followed in decision tree induction.	16	BTL3	Apply	CO5
8	Explain Random Forest algorithm and discuss its benefits over single decision trees.	16	BTL4	Analyze	CO5
9	With neat diagram, explain how AdaBoost improves model performance.	16	BTL4	Analyze	CO5
10	Design an AdaBoost ensemble with a weak learner on a sample dataset.	16	BTL5	Evaluate	CO5
11	Explain the working of the Support Vector Machine (SVM) algorithm.	16	BTL3	Apply	CO5
12	Derive the SVM optimization problem and explain the concept of margin maximization.	16	BTL5	Evaluate	CO5
13	Explain the role of hinge loss in SVM and its influence on the cost function.	16	BTL4	Analyze	CO5
14	Analyze the use of kernel trick in SVM and explain different types of kernels with examples.	16	BTL4	Analyze	CO5
15	Construct an SVM classifier for a binary classification problem using polynomial kernel.	16	BTL5	Evaluate	CO5
16	Compare and contrast AdaBoost and Random Forests with suitable examples.	16	BTL5	Evaluate	CO5
17	Critically evaluate the performance of k-NN, Decision Trees, and SVM on high-dimensional data.	16	BTL6	Create	CO5
18	Propose a hybrid model combining Random Forest and AdaBoost for medical data classification.	16	BTL6	Create	CO5