NARSIMHA REDDY ENGINEERING COLLEGE
An Autonomous Institution| Affiliated to JNTUH | Approved by AICTE
Accredited by NBA & NAAC with 'A' Grade

Department of Computer Science and Engineering

MACHINE LEARNING AND ITS APPLICATIONS

UNIT WISE QUESTION BANK

UNIT–I Introduction and Feature Engineering				
S.No	**Questions**	**BT**	**CO**	**PO**
Part – A (Short Answer Questions)				
1	What is the fundamental difference between Supervised Learning and Unsupervised Learning?	L1	CO1	**PO1**
2	Define Machine Learning and state its goal.	L1	CO1	**PO1**
3	What is Semi-supervised Learning?	L1	CO1	**PO1**
4	State the objective of the Find-S algorithm in concept learning.	L1	CO1	**PO2**
5	What is Feature Engineering?	L1	CO1	**PO1**
6	Differentiate between Feature Selection and Feature Extraction.	L1	CO1	**PO3**
7	Define Min-Max Normalization	L1	CO1	**PO1**
8	Briefly explain the purpose of z-score normalization	L1	CO1	**PO5**
9	What is the Curse of Dimensionality?	L1	CO1	**PO5**
10	Name two popular Filter Methods used for Feature Selection	L1	CO1	**PO2**
Part – B (Long Answer Questions)				
11	Explain the different types of **Machine Learning** (Supervised, Unsupervised, Semi-supervised, and Reinforcement Learning) with a suitable real-world application for each.	L2	CO1	**PO2**
12	Describe the steps involved in the **Concept Learning** process using the **Find-S algorithm**.	L2	CO1	**PO3**
13	Explain **Feature Normalization** techniques (**Min-Max normalization, z-score normalization, and constant factor normalization**) and discuss when each should be applied.	L2	CO1	**PO2**
14	Detail the importance of **Feature Engineering** in the Machine Learning pipeline.	L2	CO1	**PO1**
15	Describe the three categories of **Feature Selection** techniques: **Filter, Wrapper, and Embedded methods**.	L2	CO1	**PO2**
16	Explain the principle and working of **Principal Component Analysis (PCA)** as a dimensionality reduction technique.	L2	CO1	**PO5**
17	Explain the difference between **Classification** and **Regression** problems in Supervised Learning, providing two examples.	L2	CO1	**PO3**
18	Discuss the role of **Dimensionality Reduction**. When is it necessary, and what are its potential benefits?	L2	CO1	**PO1**
19	Describe the procedure for **Linear Discriminant Analysis (LDA)** and its primary application in a classification context.	L3	CO1	**PO1**
20	Explain the concept of **Deep Learning** and how it relates to	L3	CO1	**PO1**

		Machine Learning and **Artificial Intelligence**.			
UNIT–II **Supervised Learning – I (Regression and Classification)**					
S.No		**Questions**	**BT**	**CO**	**PO**
Part – A (Short Answer Questions)					
1		What is the **Cost Function** in the context of regression, and what is its role?	L1	CO2	PO2
2		What is the primary objective of a **Regression** model?	L1	CO2	PO2
3		Define **Overfitting** in a machine learning model.	L1	CO2	PO2
4		What is the 'Naive' assumption in the **Naive Bayes Classifier**?	L1	CO2	PO2
5		What are the **Support Vectors** in a Support Vector Machine (SVM)?	L1	CO2	PO1
6		Define the **F1-Score** and explain why it is used instead of simple accuracy for imbalanced datasets.	L1	CO2	PO1
7		How is the "K" determined in the **K-Nearest Neighbors (KNN)** algorithm?	L1	CO2	PO2
8		Differentiate between **Precision** and **Recall**.	L1	CO2	PO1
9		What is the main difference between **Simple Linear Regression** and **Multiple Linear Regression**?	L1	CO2	PO1
10		Briefly explain the concept of **Decision Tree Pruning.**	L1	CO2	PO1
Part – B (Long Answer Questions)					
11	a)	Describe the process of building a predictive model using **Simple Linear Regression**. Include the equation and the role of the **Cost Function** (like MSE).	L2	CO2	PO2
	b)	Explain the working of **Logistic Regression**. Why is it considered a **Classification** algorithm even though its name contains "Regression"?	L2	CO2	PO2
12	a)	Explain the **K-Nearest Neighbors (KNN)** algorithm for classification. What is a major disadvantage, and how can it be mitigated?	L2	CO2	PO1
	b)	Describe the working of **Decision Tree** algorithms like **ID3** or **C4.5**. How do they select the best feature to split on?	L2	CO2	PO2
13	a)	Explain the **Support Vector Machine (SVM)** algorithm. How does it find the **optimal hyperplane** for classification?	L2	CO2	PO1
	b)	Explain how a **Confusion Matrix** is used to evaluate the performance of a classification model. Define **True Positive, True Negative, False Positive, and False Negative** in this context.	L2	CO2	PO1
14	a)	Discuss the concepts of **Bias and Variance** in Machine Learning and explain the **Bias-Variance Tradeoff**. 8. How does **Naive Bayes** work? Explain the process of classification using Bayes' Theorem.	L2	CO2	PO3
	b)	Discuss the evaluation metrics for **Regression** models, such as **Mean Squared Error (MSE)** and **Mean Absolute Error (MAE)**, and their relative advantages.	L3	CO2	PO3
15	a)	Explain **Polynomial Regression**. When is it used, and what is its main risk in modeling?	L2	CO2	PO1
UNIT–III **Supervised Learning – II (Neural Networks)**					

S.No		Questions	BT	CO	PO
Part – A (Short Answer Questions)					
1		What is an **Artificial Neuron** or **Perceptron**?	L1	CO3	PO4
2		What is the role of **Activation Functions** in a Neural Network?	L1	CO3	PO4
3		Name two common **Activation Functions** and their output range.	L1	CO3	PO5
4		What is the primary difference between **Feedforward** and **Recurrent Neural Networks (RNN)**?	L1	CO3	PO5
5		Explain the concept of **Forward Propagation** in an ANN.	L2	CO3	PO3
6		What is the **Learning Rate** hyperparameter, and how does it affect the training process?	L1	CO3	PO4
7		What is a **Cost Function (or Loss Function)** in the context of Neural Networks?	L1	CO3	PO3
8		What is the purpose of the **Convolutional Layer** in a CNN?	L1	CO3	PO2
9		Explain the goal of **Pooling Layers** (e.g., Max Pooling) in a CNN.	L1	CO3	PO4
10		What is **Batch Gradient Descent**?	L1	CO3	PO3
Part – B (Long Answer Questions)					
11	a)	Describe the architecture of a **Multi-Layer Perceptron (MLP) / Artificial Neural Network (ANN)**. Explain the function of the input, hidden, and output layers.	L3	CO3	PO3
	b)	Explain the **Backpropagation Algorithm** in detail. What is its main purpose, and how does it use the chain rule?	L2	CO3	PO3
12	a)	Explain the need for **Activation Functions**. Discuss the properties and drawbacks of the **Sigmoid** and **ReLU (Rectified Linear Unit)** functions.	L2	CO3	PO4
	b)	Describe the different types of **Gradient Descent** (**Stochastic, Mini-Batch, and Batch**) and compare their computational efficiency and stability during training.	L3	CO3	PO4
13	a)	Illustrate the architecture of a **Convolutional Neural Network (CNN)**. Explain the functions of the **Convolution layer** and the **Pooling layer** with respect to image processing.	L2	CO3	PO3
	b)	Explain the **Perceptron Learning Process**. Why can a single-layer perceptron only solve linearly separable problems?	L3	CO3	PO3
14	a)	Discuss the challenges that arise when training Deep Neural Networks, such as the **Vanishing Gradient Problem**.	L2	CO3	PO3
	b)	Explain the role of the **Softmax** function in the output layer of an ANN, especially for multi-class classification problems.	L3	CO3	PO3
15	a)	What is **Dropout** in Neural Networks, and what is its role?	L2	CO3	PO4
	b)	Describe how the **weights and biases** are initialized and updated during the learning process in an ANN.	L3	CO3	PO4
UNIT–IV **Model Validation and Ensemble Methods**					
S.No		**Questions**	**BT**	**CO**	**PO**
Part – A (Short Answer Questions)					
1		What is the difference between **Overfitting** and **Underfitting**?	L1	CO4	PO3
2		Define the **Holdout Method** for model validation.	L1	CO4	PO3
3		What is the primary goal of **Regularization** in machine learning?	L1	CO4	PO3

4		What is the key difference between **L1 Regularization (Lasso)** and **L2 Regularization (Ridge)**?	L1	CO4	PO3
5		What is the main advantage of **Stratified K-Fold Cross-Validation** over simple K-Fold?	L1	CO4	PO3
6		Define an **Ensemble Method**.	L1	CO4	PO3
7		What is the concept of **Diversity** in the context of Ensemble Methods?	L1	CO4	PO3
8		What is **Bootstrap Aggregating (Bagging)**?	L1	CO4	PO3
9		Which ensemble method is **Random Forest** an example of (Bagging or Boosting)?	L1	CO4	PO3
10		What is the main idea behind the **Boosting** technique?	L1	CO4	PO3
Part – B (Long Answer Questions)					
11	a)	Explain the **K-Fold Cross-Validation** technique. Why is it preferred over the simple Holdout method for model evaluation?	L2	CO4	PO3
	b)	Describe the **Bias-Variance Tradeoff** in detail. How do model complexity, overfitting, and underfitting relate to this tradeoff?	L3	CO4	PO3
12	a)	Discuss the causes of **Overfitting** and list three different techniques to prevent it.	L2	CO4	PO3
	b)	Explain **Lasso (L1) and Ridge (L2) Regularization**. Discuss how L1 regularization can be used for **Feature Selection**.	L3	CO4	PO3
13	a)	Explain the working of **Bagging**. How does it help to reduce the **variance** of the model?	L3	CO4	PO3
	b)	Describe the working of the **Random Forest** algorithm. Explain how it leverages both bagging and feature randomness to improve performance	Ll3	CO4	PO3
14	a)	Explain the concept of **Boosting** (e.g., AdaBoost or Gradient Boosting). How does it sequentially combine weak learners to create a strong learner?	L2	CO4	PO3
	b)	Compare and contrast **Bagging** and **Boosting** techniques based on their goal (reducing bias/variance), how they train models (parallel/sequential), and their base learners.	L3	CO4	PO3
15	a)	Explain **Leave-One-Out Cross Validation (LOOCV)**. When might this method be computationally expensive or impractical?	L2	CO4	PO3
	b)	Describe the concept of **Stacking** (Stacked Generalization). How does it use a **meta-model** to combine the predictions of diverse base models?	L2	CO4	PO3
UNIT–V **Unsupervised and Reinforcement Learning**					
S.No		**Questions**	**BT**	**CO**	**PO**
Part – A (Short Answer Questions)					
1		Give two real-world applications of **Unsupervised Learning**.	L1	CO5	PO3
2		State the primary difference between **K-Means** and **K-Nearest Neighbors (KNN)**.	L1	CO5	PO3
3		What are the disadvantages of using Euclidean distance in the **K-Means** algorithm?	L2	CO5	PO3
4		What kind of data is the **K-Modes** clustering algorithm suitable for?	L2	CO5	PO3
5		What is the key distinction between **Agglomerative** and	L2	CO5	PO3

		Divisive Hierarchical Clustering?			
6		Define the terms **Exploration** and **Exploitation** in **Reinforcement Learning (RL)**.	L1	CO5	PO3
7		What is a **Markov Decision Process (MDP)**?	L1	CO5	PO3
8		What is the main advantage of **DBSCAN** over K-Means clustering?	L1	CO5	PO3
9		Explain the concept of **non-associative learning** in RL.	L2	CO5	PO3
10		What is a **Q-Table** in **Q-learning**?	L1	CO5	PO3
Part – B (Long Answer Questions)					
11	a)	Compare and contrast **Supervised Learning** and **Unsupervised Learning** based on data type, objective, and common algorithms used.	L2	CO5	PO3
	b)	Explain the working principle of the **K-Means Clustering** algorithm. Detail the steps involved in the process.	L2	CO5	PO3
12	a)	Explain **Hierarchical Clustering**. Describe the procedures for both **Agglomerative** (Bottom-Up) and **Divisive** (Top-Down) methods.	L2	CO5	PO3
	b)	Describe the clustering methods **K-Modes** and **K-Prototypes**. Explain their suitability for different types of data (categorical and mixed).	L3	CO5	PO3
13	a)	Explain **DBSCAN (Density-Based Spatial Clustering of Applications with Noise)**. Define the core, border, and noise points in this algorithm.	L2	CO5	PO3
	b)	Describe the key components of a **Reinforcement Learning** model: **Agent, Environment, State, Action, and Reward**.	L3	CO5	PO3
14	a)	Explain the fundamental **Exploration vs. Exploitation Trade-off** in Reinforcement Learning. Why is balancing this trade-off crucial?	L3	CO5	PO3
	b)	Describe the working of the **Q-Learning** algorithm. Write down the **Bellman equation (Q-function update rule)** and explain the role of the learning rate and discount factor	L2	CO5	PO3
15	a)	Discuss the various applications of **Clustering** techniques in different domains, such as market segmentation, image analysis, and anomaly detection.	L2	CO5	PO3
	b)	Compare the learning mechanism of **Reinforcement Learning** with that of Supervised and Unsupervised Learning.	L3	CO5	PO3

* **Blooms Taxonomy Level (BT)** (L1 – Remembering; L2 – Understanding; L3 – Applying;L4 – Analyzing; L5 – Evaluating; L6 – Creating)