Assignment 1

SRN –PES2UG22EC049

NAME SB DANUSH VIKRAMAN

MCQ Questions (1 mark each)

Q1. Which of the following Python data structures is **immutable**?
a) List
b) Tuple
c) Dictionary
d) Set

Answer: b) Tuple

Q2. Which of the following machine learning techniques is most appropriate for predicting a **continuous output variable**?
a) Logistic Regression
b) Decision Tree Regressor
c) K-Means Clustering
d) Naive Bayes

Answer: b) Decision Tree Regressor

Descriptive Questions (5/7 marks)

Q3. (5 Marks) Differentiate between **classification** and **regression** in machine learning. Explain the types of problems solved by each with suitable examples.

Answer:

- **Classification:**
  - Predicts discrete labels or categories.
  - Example: Email spam detection (Spam / Not Spam), medical diagnosis (Disease / No Disease).
  - Algorithms: Logistic Regression, Decision Trees, Random Forest, SVM.
- **Regression:**
  - Predicts continuous numerical values.
  - Example: Predicting house prices, forecasting temperature.
  - Algorithms: Linear Regression, Ridge/Lasso Regression, Random Forest Regressor.

Key Difference: Classification deals with **categories**, regression deals with **quantities**.

Q4. (7 Marks) Explain the concept of **overfitting** and **underfitting** in machine learning. How can they be identified and reduced?

Answer:

- **Overfitting:**
  - Model learns both patterns and noise in training data, performs very well on training but poorly on unseen data.
  - Identified by **high training accuracy but low test accuracy**.
  - Reduction methods: Cross-validation, regularization (L1/L2), pruning in trees, using simpler models.
- **Underfitting:**
  - Model is too simple to capture underlying patterns, performs poorly on both training and test data.
  - Identified by **low training and low test accuracy**.
  - Reduction methods: Adding more features, using more complex models, reducing bias.

Summary:

- Good models balance **bias and variance** (Bias-Variance Tradeoff).
- Techniques like **cross-validation and regularization** help achieve generalization.