National University of Modern Languages, Karachi
CS Department - Fall 2025
AI Lab 06 Tasks - CS IV

Course Code: CSAI-226L	**Course Name:** Artificial Intelligence-Lab
Instructor: Parshant Vijay	

Submission Instructions:

You must submit the following items:

1. **Source Code Files:**
   - YourName_Task1_KNNClassification_WineQuality.py
   - YourName_Task2_KNNClassification_BreastCancerDetection.py
2. Submit **one single Word/PDF document** named *Lab06_Report_YourName* containing a brief explanation, **one screenshot of the code**, and **one screenshot of the output for each of the 2 tasks**.

LAB # 06
KNN CLASSIFICATION

LAB TASKS

Task 1: Wine Quality Classification

For this task, you'll use the Wine dataset (sklearn.datasets.load_wine) to classify three types of wine based on four chemical features. First, load the data and separate the features (X) from the target (y). Next, split the data into training and testing sets, using a test size of 30%. Then, train a K-Nearest Neighbors (KNN) model using a fixed value of k=5. Evaluate the model by calculating and reporting the accuracy score on the test set. Finally, select two relevant features (e.g., 'alcohol' and 'malic_acid') and create a scatter plot to visually inspect how well the three wine classes are separated by those features.

Task 2: Breast Cancer Detection

This task highlights the importance of feature scaling for distance-based algorithms like KNN, using the Breast Cancer dataset (sklearn.datasets.load_breast_cancer), which is a binary classification problem.

Your first step is to establish a baseline model: split the unscaled data (using a test size of 20\%), train a KNN model with k=7, and record the resulting accuracy score.

Next, you must standardize the features (X) using StandardScaler from sklearn.preprocessing , which ensures all features contribute equally to the distance calculation regardless of their original scale.

Then, repeat the modeling process: split the scaled data, train a new KNN model with k=7, and record its accuracy score.

Conclude by comparing the two accuracy scores (unscaled vs. scaled) and writing a brief explanation of why scaling is a critical pre-processing step for KNN.