NATIONAL UNIVERSITY OF COMPUTER AND EMERGING SCIENCES
(KARACHI CAMPUS)

Department of Software Engineering

ASSIGNMENT NO 01

Subject: Machine Learning Date: 5th Sep 2025
Total Marks: 100 **Due Date: 20,Sep,25**
Course Instructor: Muhammad Minhal Raza

Instructions to be strictly followed. Attempt All Question's.

Question no 1 (Project MileStone's , Max Group of 3 Person's).

1. Identify the unique data set related to any real-life problem which are not previously explored by machine learning techniques. Use online resources like research papers,etc.. to find the data set. Uniqueness matters here a lot.First Approve the data set from me. Discuss features, background, problem source and problem statement here . (Screen-Short and Data Set File Required Here).

2. Explore your data set by performing EDA and Data prepossessing in python.(Python File Required Here).

3. On same python file apply KNN on your data set and perform Model Evaluation. (Python File Required Here).

4. On same python file implement the KNN in python without using library by doing the custom code and again evaluate the model on your custom algorithm.(Python File Required Here).

5. Compare your Results. (Python File Required Here).

6. Type code and paste output under each question.

Question no 2

Consider the feature x_1 shows the prices in PKR and feature x_2 represents size of the apartment in a certain area of the city. We can show each apartment using these two features values as

$x = [x_1, x_2]^T$

we can show the labels as,

$$l = \begin{cases} 1 & if\ x\ is\ positive \\ 0 & if\ x\ is\ negative \end{cases}$$

Each apartment is represented as (x,*l*) and the training set consists of M such samples.

Perform the followings:

A) Formulate the mathematical equation of the training set from i = 1 to M. where i shows examples in the total M samples.
B) Devise hypothesis *H* for this problem.
C) Formulate the equation of the hypothesis error given the training set.
D) What is the difference between most specific hypothesis and most general hypothesis?

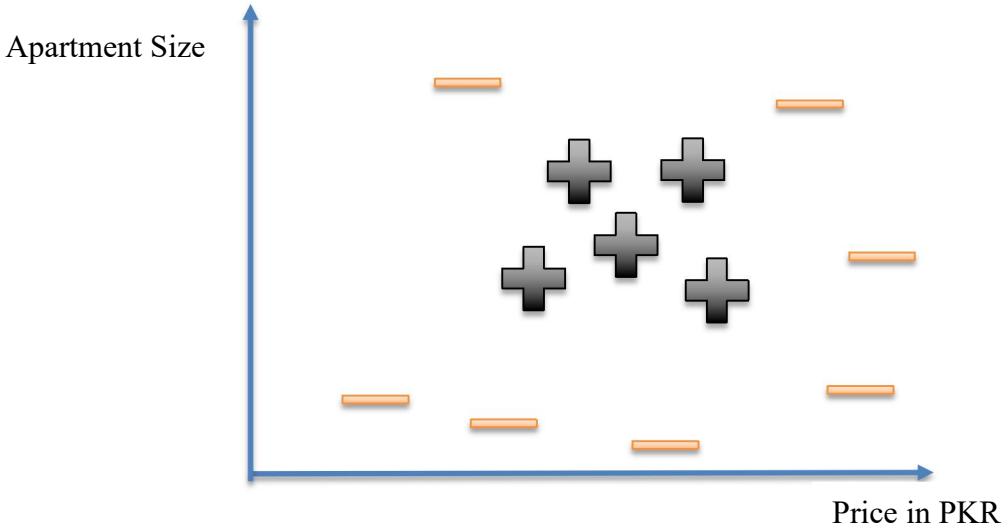

Figure A: Dataset containing training set for the class that represents family apartment. Each data sample shows one apartment, and its relevant coordinates give information about price and size. Subsequently, a positive sample describes family-oriented apartment whereas negative sample shows non-family-oriented apartment i-e bachelors or workers residence.

Question no 3

Write a summarized report and analysis regarding mathematical formulation of the following Research Papers already uploaded on a GCR.

A Comprehensive Survey of Machine Learning Techniques and Models for Object Detection.

Question no 4

Explore the Weighted KNN given below by using different Chebyshev Distance and City Block Distance Function on any example data set. Perform all calculations and formulate the problem mathematically,

$$\hat{y} = \arg\max_{c \in C} \sum_{(x_i, y_i) \in N_k(x)} \underbrace{\frac{1}{d(x, x_i)}}_{\text{weight } w_i} \underbrace{I(y_i = c)}_{\text{vote for class } c}$$

Piece by piece:

- x: the new point we want to label.
- C: the set of classes (e.g., $\{A, O\}$).
- $N_k(x)$: the **k nearest neighbors** of x (by your chosen distance).
- $d(x, x_i)$: distance from x to neighbor x_i (Euclidean, **cityblock/Manhattan**, etc.).
- $I(y_i = c)$: 1 if neighbor's class equals c; otherwise 0.
- $\frac{1}{d(x,x_i)}$: the **weight**—closer neighbors (small d) get **bigger** say.
- $\arg\max$: pick the class c that gets the **largest total weighted vote**.

2. Manhattan Distance (a.k.a. L1 Norm, Taxicab Distance)

$$d(x, y) = \sum_{i=1}^{m} |x_i - y_i|$$

- Measures distance along grid lines (like streets in Manhattan city).
- Useful when movement is restricted in right angles or data has high dimensions.

Example: (2,3) and (5,7):

$$d = |2 - 5| + |3 - 7| = 3 + 4 = 7$$

Chebyshev Distance (L∞ Norm)

$$d(x, y) = \max_i |x_i - y_i|$$

- Takes the largest difference among features.
- Useful when similarity is defined by the maximum deviation in one dimension.

Example: (2,3) and (5,7):

$$d = \max(|2 - 5|, |3 - 7|) = \max(3, 4) = 4$$

Note : Good , Neat and Clean Hand Written Assignment Should Be Submitted By All Students Separately on A4 Sheets/papers. Clearly Visible Diagram's and Calculation's . Use Cam Scanner App to Submit a Single File. Avoid Overwriting and Unnecessary Steps and Details.