<div align="center">

Course: Artificial Intelligence

Assignment 1

Deadline:22 January,2026(11:59 pm)

</div>

Task 1: Binary classification using KNN classifier.

Dataset Description:

You are given a CSV file "data.csv" containing the measurements from a digitized image of a fine needle aspirate (FNA) of a breast mass. Divide the 80% data for training and 20% data for testing.

Independent variables/ Attributes/ Features:

1. Radius_mean
2. Texture_mean
3. Perimeter_mean
4. Area_mean
5. Smoothness_mean
6. Compactness_mean
7. Concavity_mean
8. Concavepoints_mean
9. Symmetry_mean
10. Fractal_dimension_mean
11. Radius_se
12. Texture_se
13. Perimeter_se
14. Area_se
15. Smoothness_se
16. Compactness_se
17. Concavity_se
18. Concavepoints_se
19. Symmetry_se
20. Fractal_dimension_se
21. Radius_worst
22. Texture_worst
23. Perimeter_worst
24. Area_worst
25. Smoothness_worst
26. Compactness_worst
27. Concavity_worst
28. concave points_worst

29. Symmetry_worst
30. Fractal_dimension_worst

Dependent variable/ Target Attribute/ Class:
i. "Diagnosis": (M = malignant, B = benign)

Model:

Build KNN based classification model from scratch to be trained on a given dataset and recommend the best **value for K** (number of neighbors) and **distance matrix** that produces the highest accuracy for test data.

Hyperparameters for Experimentation
- Values of K: 3,4,9,20 and 47.
- Distance Matrix: Euclidean Distance, Manhattan Distance, Minkowski Distance, Cosine Similarity, Hamming Distance .

Expected Outcomes:

- Pick the best value of K and distance matrix based on the testing accuracy, and for the best value of K, print confusion matrix, recall, and precision.
- Plot the graph between K values vs Accuracy considering different distance matrices.
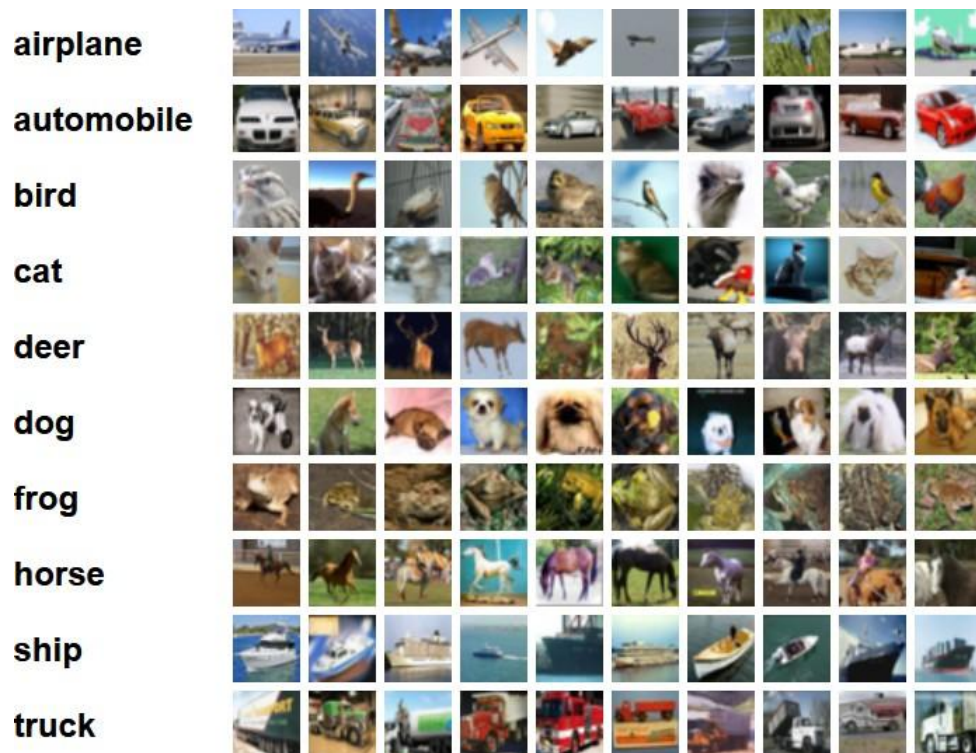- Inferences on the plots and inferences on the results observed.

Bonus : Visualize the decision boundary for the best KNN model according to your results.

Task 2: Multi class classification using KNN classifier CIFAR10 dataset
Dataset Description

The CIFAR-10 dataset is a collection of 60,000 32x32 color images in 10 classes, with 6,000 images per class. The dataset is split into 50,000 training images and 10,000 test images. Download [Link](#) .

Model:
Build KNN based classification model from scratch to classify these images into one of the 10 classes.

Expected Outcomes:

- Build a KNN-based classification model from scratch to classify images in the CIFAR-10 dataset.
- Experiment with different values of K (number of neighbors) and various distance metrics, including Euclidean Distance, Manhattan Distance, Minkowski Distance, Cosine Similarity, and Hamming Distance.
- Pick the best value of K and distance matrix based on the testing accuracy, and for the best value of K, print confusion matrix, recall, and precision.

- Visualize the impact of different distance metrics on the accuracy across varying values of K.
- Insights and inferences based on the experimental results and plots.

Note:

- **You are not supposed to use pre-built functions for models from the libraries like sklearn, pytorch.**
- **Students will also be rewarded extra for solving bonus questions, thorough experimentation and insightful analysis based on the results and graphs they produce.**

Report should be in PDF form and the report by a team should also include the observations about the results of studies.

Instruction:

Upload all your codes(.ipynb file) and reports(.pdf file) in a single zip file.

• Give the name of the folder as Group_Assignment1 Example: Group01_Assignment1

• Give the name of the zip file as Group_Assignment1.zip Example: Group01_Assignment1.zip

We will not accept the submission if you don't follow the above instructions