# Classification using KNN, Logistic Regression and SVM
# Midterm Project for Math 452

Due: Nov. 8th, 11:59PM

## Project Overview

This project will provide hands-on experience with K-Nearest Neighbors (KNN), Logistic Regression (LR), Support Vector Machines (SVM), K-Means Clustering, and Principal Component Analysis (PCA) using the MNIST dataset. Students will learn how to implement these algorithms, evaluate their performance, and analyze clustering results. The project consists of two main parts:

- Classification using KNN, LR, and SVM
- Clustering using K-Means

### Part 1: Classification of Handwritten Digits

Dataset: MNIST
Tasks:

1. Data Preprocessing:
   - Load the MNIST dataset and normalize the pixel values to the range [0, 1].
   - Flatten the images to create 784-dimensional vectors.
2. Model Implementation:
   - Implement and train KNN, Logistic Regression, and SVM models using the training dataset.
   - Use libraries such as scikit-learn to simplify the implementation.
3. Hyperparameter Tuning:
   - For KNN, determine the optimal value of $k$ using ten-fold cross-validation.
   - Perform this cross-validation process 10 times with different random seeds to ensure robustness in the results.
   - For SVM, experiment with different kernels and regularization parameters, and evaluate their performance using cross-validation.
4. Performance Evaluation:
   - Split the dataset into training and testing sets (e.g., 80/20 split).
   - Evaluate model performance using metrics such as accuracy, precision, recall, F1-score, and confusion matrix.
   - Visualize results using plots (e.g., ROC curves, confusion matrices).
5. Comparative Analysis:
   - Compare the performance of KNN, LR, and SVM based on the evaluation metrics.
   - Discuss the strengths and weaknesses of each model in the context of handwritten digit classification.

### Part 2: Clustering using K-Means Clustering

Tasks:

- Use K-Means clustering to group the MNIST images into clusters based on pixel intensity.
- Experiment with different values of $k$ (number of clusters) and determine the optimal $k$ using the Elbow Method.
- Visualize the clusters and the cluster centers.
- Discuss the effectiveness of K-Means clustering on the MNIST dataset.

## Final Report Requirements

Your final write-up should be structured like a research paper, between 4-10 pages, following the provided template. You can get an overleaf template for SIAM journal though the link `https://www.overleaf.com/latex/templates/template-for-siam-online-only-journals/zsntjqvxdqgp`

The report should include the following sections:

1. Title, Author(s):
   Include all authors who contributed to the project.
2. Abstract (300 words max) (5%): Briefly describe the problem, approach, and key results.
3. Introduction (10%): Describe the problem, its importance, and an overview of results.
4. Related Work (5%): Discuss relevant published work and how your approach differs.
5. Data (10%):
   Describe the MNIST dataset and any preprocessing steps taken.
6. Methods (30%):
   Discuss your approach for classification and clustering, including hyperparameter tuning.
7. Experiments (30%):
   Detail the experiments performed, metrics used for evaluation, and results obtained.
8. Conclusion (5%):
   Summarize key results and suggest ideas for future extensions or applications.
9. Writing / Formatting (5%):
   Ensure the report is clearly written and well-formatted.

### Presentation Requirements

#### Presentation Schedule:

The presentations are scheduled for October 30, 2024 (Wednesday) and November 1, 2024 (Friday). Each group will present their project during this time.

#### Format:

Groups have the option to present their work using either slides or a poster.

#### Time Allocation:

Each group will have 10 minutes to present their work, followed by a brief Q&A session.

### Submission Guidelines

#### Collaboration

You can work in teams of up to 3 people. We do expect that projects done with 3 people have more impressive writeup and results than projects done with fewer people. For example, to get a sense for the scope and expectations for projects. While we encourage that you work in teams, you may also work alone.

**Submissions**

Deadline for submit your report is Nov.8th, 11:59pm.

Submit your final report as a PDF. Include supplementary material (e.g., source code, visualizations) in a separate PDF or ZIP file. List all authors under the title and include footnotes for non-enrolled authors.