**Students:** Jose Miguel Logiurato, María Elena Leizaola and Andrés Aguilar.

**Texas Tech University**

**Department of Computer Science**

**Course Name:** Introduction to Artificial Intelligence **Number:** CS3368 **Sem:** Summer 2024

**Instructor:** Dr. Juan Carlos Rojas **Email:** [Juan-Carlos.Rojas@ttu.edu](mailto:Juan-Carlos.Rojas@ttu.edu)

# Project 1

This project is to be completed in teams of 2. Only one submission per team is necessary.

Submit your solutions packed together into a single ZIP file, and a single report document in Word or PDF.

## Project Description

You are asked to develop a classifier for images in the “Fashion MNIST” dataset. This is a set that is composed of 10 types of fashion items (shirts, shoes, pants, etc.)

Background pattern

Description automatically generated

This dataset is largely compatible with the digits MNIST set. Uses the same image size, and the same number of labels. You can find the dataset original files, and a description of the categories at this location: <https://github.com/zalandoresearch/fashion-mnist>

As part of the project, you will need to:

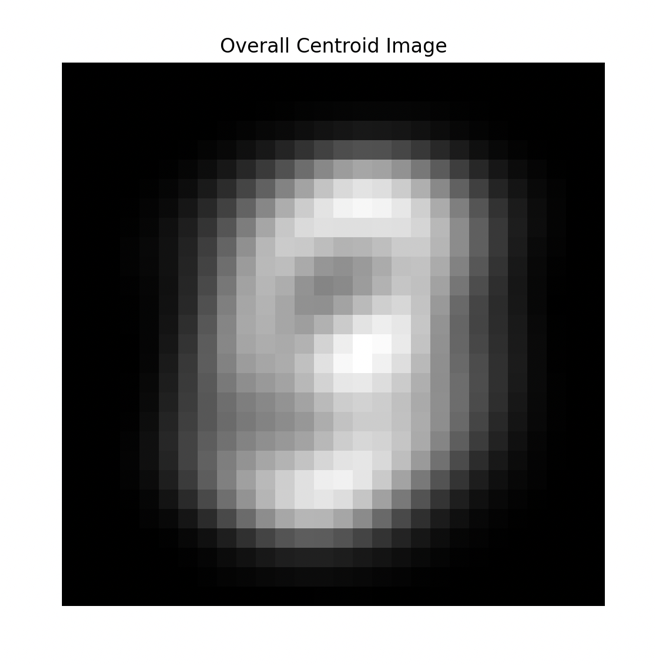
1. **Pre-process the raw data files and convert them into a format suitable for Scikit-Learn classifiers**
2. **Explore the dataset: quantity of examples of each class, distribution of pixel values, centroid images: overall and per-class**

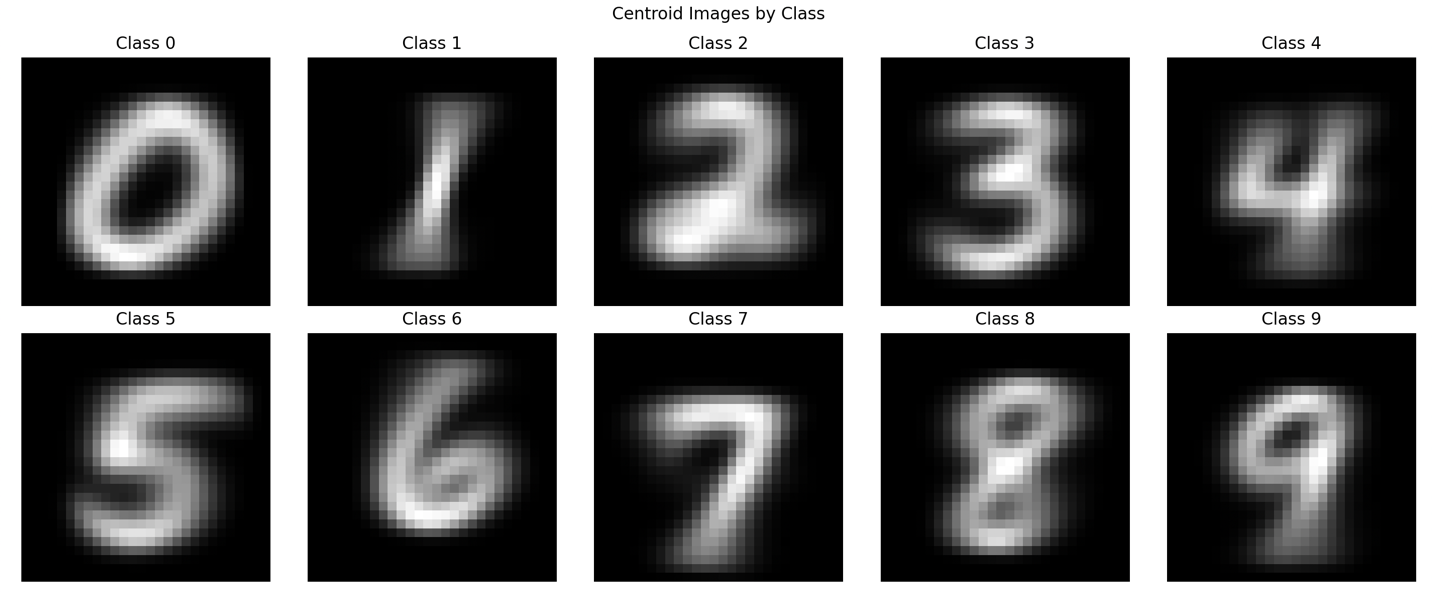
A graph of a number of classes

Description automatically generated with medium confidence

**A graph with numbers and lines

Description automatically generated**





1. **Attempt multiple classifiers (at least 4). You are free to choose which ones. You should attempt to optimize the available parameters of each classifier to get the best results.**
2. **Keep a scorecard of the accuracy of all your results.**

**A graph of different classifiers

Description automatically generated with medium confidence**

|  |  |
| --- | --- |
| **METHOD** | **ACCURACY** |
| **NAÏVE BAYES** | **0.5516** |
| **SOFTMAX REGRESSION** | **0.9215** |
| **RANDOM FOREST** | **0.9686** |
| **kNN** | **0.9729** |

1. **Display several examples of images that were mis-classified by the best classifier.**
2. **Study the pixel importances from the best classifier. Display them as an image.**
3. **Study the effect of image resolution on the accuracy of your best classifier.**
4. **Personal experience with project.**

## Code Expectations

You should submit multiple Python files, one for each type of classifier and/or experiment. The purpose of each file should be clearly explained in the file header.

The programs should run as-is, with no modification necessary to reproduce the results shown in your report.

Your programs should be well commented and use reasonable variable names.

If your program is based on a class example, you need to mention so, keep the copyright notice, and make notes of your modifications.

## Report Document Expectations

Your report document should describe your various experiments and their results. You don’t need to include copies of the source code in the document. But you are expected to mention which files correspond to which results in your report.

You should analyze your results, and not only post them. You should explain why you came up to certain conclusions based on your results.

Also, please provide some insight into your personal experiences developing this project.