

# CS 250 Spring 2025

## Homework Assignment 2

**Due on Moodle:** Before class on Friday, February 7. Mail submissions will not be accepted. Submit one single Jupyter Notebook for all questions together.

- 1) From the MNIST dataset introduced in class, write code in Python that
  - a) generates a frequency distribution chart (histogram or simple bar chart) of all the digits in the dataset. This means, this chart will show how many times each of the digits '0', '1', etc. occur in the dataset. [5 points]
  - b) Generates the ten “average” digits and displays them on the screen. An “average zero” for instance, is the pixel-wise sum of all “zero” images divided by the number of such images. This is similar to what we did in class, but it has to be done programmatically for the ten digits and plotted as subplots in the same figure. Please don't just manually write the code from class ten times to generate ten separate images. [10 points]
- 2) Write code in Python that
  - a) Eliminates all data except the classes '0', '1' and '5' from the MNIST dataset. [5 points]
  - b) Uses k-means to cluster this data into three clusters [5 points]
  - c) Displays a few images from each of the three clusters. You can use one single, or three different figures for this part. [5 points]