# CS454-554 Homework 2: *k*-Means Clustering for Unsupervised Learning Fall 2023/2024

In this homework, your task is to implement *k*-means clustering to calculate cluster centroids on the Fashion-MNIST dataset. (**https://en.wikipedia.org/wiki/Fashion_MNIST**)

You are provided with a compressed file called **fashion_mnist.zip**. Please unzip it and use the **fashion_mnist.csv** file as your dataset file. The dataset contains 28x28 images from 10 different classes. Each class represents a different type of fashion item but in this homework, you are not going to use the class information. Each row of the file corresponds to one instance and there are 60000 instances. The first column of each row contains the label of the image (which you will ignore), and the remaining 784 columns that represent the grayscale value of each pixel constitute the input image that you will use to calculate the cluster centroids.

Implement *k*-means clustering for *k*=[10, 20, 30], where for each *k*:

- Plot the reconstruction loss as a function of iterations; we should see the error decreasing and then converging.
- Once you get convergence, plot the cluster centroids as 28x28 grayscale images. These different centers should roughly look like different variants of examples from the different classes.

You are **not** allowed to use any library function for statistical calculations; that is, you should code the routines for calculating distances or averaging yourself. As in the previous homework, you can use pandas for **loading the data** and the numpy array as a **data structure** but you are not allowed to use pandas or numpy functions for calculations (np.sum, np.linalg.norm, data_structure.sum etc).

This homework is due **Nov 16th (Thursday), 23:00**.
Your submission should include a short report of your findings, the plots, and your source code.
Upload your report **as a pdf file** to LMS alongside your .py/.m code file.