# Assignment 8

CS 750/850 Machine Learning

- **Due**: April 20th at 11:59PM
- **Submisssion**: Turn in both a **PDF** and the **source code** on MyCourses
- **Questions**: Piazza
- **Office hours**:
  - Marek: Wed 1:30-2:30pm (email for an appointment)
  - Soheil: Mon 2pm-4pm
  - Xihong: Thu 1:30-3:30pm https://unh.zoom.us/j/587656408

## Problem 1 [33%]

Describe, in words, the results that you would expect if you performed K-means clustering of the eight shoppers in Figure 10.14 in ISL, on the basis of their sock and computer purchases, with K = 2. Give three answers, one for each of the variable scalings displayed. Explain.

## Problem 2 [33%]

In this problem, you will compute principal components for the `Auto` dataset. First remove qualitative features, which cannot be handled by PCA. Then:

1. Compute principal components without scaling features. Plot the result (you can use `ggfortify::autoplot`).
2. Compute principal components after scaling features to have unit variance. Plot the result (you can use `ggfortify::autoplot`).
3. How do the principal components computed in parts 1 and 2 compare?

## Problem 3 [33%]

1. Apply PCA to a subset of the MNIST dataset. Plot the first few principal vectors. Describe what you observe and interpret the results.
2. **Optional**(+10 points): Apply k-means clustering to the MNIST dataset both with the original features, and using at least two different subsets of the principal components. Which approach works best? Explain.