Lab 3 *(due: Feb 07)*
Machine Learning - COSC 4360

Department of Computer Science and Electrical Engineering

Spring 2025

Exercises

*Create a **New Project** for every exercise. Take a screenshot of the source code along with its output and place the **source code** and the **screenshot** in a **zipped folder** named **LastNameFirstName_Lab3***

Exercise 1

Given the dataset: *recipes_muffins_cupcakes_scones.csv*, print the **variance ratio** and plot the **cumulative sum** of the **variance ratio** for all *8 features*. In addition, using as a reference the plots in slides 110-115, create a **scatter** plot between PC1 and PC2. In addition, create a **histogram** of features and plot a **heatmap** with the features with the largest variation in PC1 and PC2. Furthermore, **find** and **print** the features with the **highest** (max) and **lowest** (min) variation both in PC1 and PC2 (positively and negatively correlated). Finally, plot a **correlation heatmap**.

Note: You will need to *standardize* your data points.

Exercise 2 *(Optional)*

In continuation of Ex. 1 and given the following data point: [38, 18, 23, 20, 9, 3, 1, 0], insert it into the **scatter** with the data points above. Using **KNN** with **K=3**, perform **classification**.

Note: Submit through **Canvas**