Homework 1 *(due: Jan 28)*
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_HW1***

Exercise 1

Download the *wine* dataset from the following Machine Learning repository: ML Repository. Rename *wine.data* to *wine.data.csv*. In *wine.names*, you can read the descriptions of all the attributes.

The dataset contains 13 features (columns 2-14) that contribute to the quality of wine. The dataset contains data for three types, i.e., labels, of wines, identified by the category values: 1, 2, and 3 (column 1). The dataset contains 178 records. Plot a graph of the **testing accuracy** of the data set for different values of K (use **KNN** with **K=1-10**), using a **test size** of 20%.

Note: The names of all columns are:

names = ['class', 'Alcohol','Malic Acid','Ash','Acadlinity','Magnisium','Total Phenols','Flavanoids', 'NonFlavanoid Phenols', 'Proanthocyanins', 'Color Intensity', 'Hue', 'OD280/OD315', 'Proline']

Exercise 2

In the given the dataset (Files->Datasets): *breast-cancer-wisconsin.zip*, perform classification using the **KNN** supervised learning algorithm with **K=5** and **test_size=0.30**. Print the **accuracy** of your model as well as the **Visual Confusion Matrix**.

Note 1: In: *breast-cancer-wisconsin.names*, you can read the descriptions of all the attributes.

Note 2: Remove all rows that contain a **?** (this should reduce the number of rows from 699 to 683). Ignore the first column (ID number) and assign the remaining columns to X (features) and Y (label).

Exercise 3 *(Optional)*

Solve the **locker riddle** which reads as follows:

Imagine 100 lockers numbered 1 to 100 with 100 students lined up in front of those 100 lockers:

The first student opens every locker.

The second student closes every 2nd locker.

The 3rd student changes every 3rd locker; if it's closed, she opens it; if it's open, she closes it.

The 4th student changes every fourth locker (e.g., 4th, 8th, etc.).

The 5th student changes every 5th locker (e.g., 5th, 10th, etc.).

That same pattern continues for all 100 students.

Here's the question: "Which lockers are left open after all 100 students have walked the row of lockers?"

Exercise 4 *(Optional)*

From the textbook **Elementary Linear Algebra** by Larson, Edwards, Falvo (6th Ed.), read the section on **Cryptography**, pages 102-105 and, using Python, replicate the example shown, i.e., **encode** a message and then **decode** it. Your algorithm should be able to work for any message of any length (assume only alphabetic characters and use a whitespace instead of an underscore).

Note 1: Use the same numbers that correspond to characters with the only exception being the replacement of the underscore with the whitespace.

Note 2: You can use the **np.matmul()** function to perform matrix *multiplication*.

Note 3: You can use the **np.linalg.inv()** function to calculate the *inverse* of a matrix.

Note: Submit through **Canvas**