Assignment: Homework 1

How to Hand It In

1. Put all your solutions (report, code, etc.) in one folder. Compress this folder and name it <firstname>_<lastname>_HW1.zip. For example, “Zhiyao_Duan_HW1.zip”.
2. Submit it to the corresponding entry on Blackboard.

When to Hand It In

It is due at 11:59 PM on the date specified on the course calendar. **Late assignments will receive a 20% deduction of the full grade each day.**

Problems (10 points in total)

1. (1 point) What is supervised learning, unsupervised learning and semi-supervised learning? Explain with examples.

2. (1 point) What are the differences between a numerical variable and a categorical variable? Explain with examples.

3. (1 point) What are the similarities and differences between classification and regression? Explain with examples.

4. (2 points) Collect or make up a multi-class classification dataset. There should be more than 15 data examples. There should be more than 6 features, where some features are numerical and others are categorical. The features should have physical meanings (e.g., length, color, voltage, salary, blood type, ratings) and units if applicable. This dataset will be used in later assignments. Organize the dataset as a CSV file where each row is one data example.

5. (5 points) Learn to use Jupyter Notebook. In this assignment, you will practice using Jupyter Notebook to write and execute Python code interactively. Compared to terminal tools, Jupyter Notebook is a convenient and flexible tool for working with Python code and documenting your work. We will be using Jupyter Notebook for all your coding assignments throughout the semester.
   Note: If you are not comfortable with Python programming, it is recommended that you review a Python tutorial before beginning your Jupyter journey.
   a. **Setting up Jupyter:** Install Python and Jupyter. You can install Python and Jupyter directly on your system but using a virtual environment such as Anaconda is recommended.
   b. (0.5 points) **Creating a notebook:** Launch the Jupyter server, create a new notebook in the directory of your choice, and select a Python kernel for the notebook.
   c. Creating and running cells:
      i. (0.5 points) Create and run a cell to print your name.

ii. (0.5 points) Create a cell to define a function that computes the sum of the reciprocals of the first n natural numbers, where n is the function argument. In a new cell, call the function using a given value of n and print the result.
iii. (1 point) Create a cell to install scikit-learn and matplotlib/seaborn packages. In a new cell, import necessary packages, load the Iris dataset from scikit-learn, and visualize one feature of your choice with a scatter plot.
iv. (1.5 point) Matrix Operations using NumPy: Create functions to perform matrix operations using the NumPy (np) library, implementing addition (np.add()), matrix multiplication (np.matmul() or @), elementwise multiplication (*), transpose operation (np.transpose() or .T), eigenvalue computation (np.linalg.eigvals()), determinant calculation (np.linalg.det()), and matrix inversion (np.linalg.inv()). Demonstrate each operation with at least one example using appropriate matrix dimensions. Document each function with input requirements and expected outputs.

d. (0.5 points) **Adding rich text:** Add a markdown cell before c.ii to explain what the function in c.ii does with plain text.
e. (0.5 points) **Exporting notebook:** Export your notebook as .ipynb and .pdf, and submit both files.