Assignment 1 - Linear Regression

Homework assignments will be done individually: each student must hand in their own answers. Use of partial or entire solutions obtained from others or online is strictly prohibited. Electronic submission on Canvas is mandatory.

Please follow the below instructions when you submit the assignment.

- You are NOT allowed to use packages for implementing the code required in this assignment. You can use packages for data processing and data split (e.g., k-fold cross validation).
- Your submission should consist of a zip file named A1_Firstname_LastName.zip which contains:
  - a jupyter notebook file (.ipynb) or a python file (.py). The file should contain the code and the output after execution (in comments if you use python). You should also include detailed comments.

1. **Linear regression** (20 points)

   a) California Housing Dataset (10 pts)

   Please implement a simple linear regression model using the closed-form solution (normal equation) and stochastic gradient descent to train the model on this dataset for predicting the house price (in millions). Split the data into training, validation, and test. Report the mean squared error (MSE) on the test data. Draw the training and validation loss curves. You need to write down every step in your experiment.

   b) UCI Machine Learning: Bike Sharing Data Set (10 pts)

   Please write a Ridge regression model and use mini-batch gradient descent to train the model on this dataset for predicting the count of total rental bikes including both casual and registered. You do not need to use all the features. Use K-fold cross validation and report the mean squared error (MSE) on the test data. You need to write down every step in your experiment.