

Lab 4 (due: Feb 14)

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_Lab4**

Exercise 1

From `sklearn.datasets` import the **California housing dataset** (please see link below). Read the 8 features that are the independent, i.e., predictor variables. The last column is the target, i.e., response variable. Read **every 10th row** from all columns -if you read all rows the 3D plot, needed for Ex. 2, will take time to complete. Using **Multiple Linear Regression**, compute the linear regression **coefficients** and **y-intercept**. **Predict** the *Median House Value* given:

8.3153, 41.0, 6.894423, 1.053714, 323.0, 2.533576, 37.88, -122.23

Note 1: **California housing dataset**

Note 2: **Linear Regression method**

Exercise 2

Using the same dataset and, having kept only the **first** two features from the dataset, perform **Multiple Linear Regression** and plot the **3D graph** between the two features and the target, i.e., response variable. In addition, on the same graph, plot a **3D scatter** plot between the same features and target variable.

Exercise 3

Using the *entire* dataset from Ex. 1, *scale* your data using **Standardization**. Find which **coefficient** carries the most weight, irrespective of its sign, and print its name.

Exercise 4

Using the *entire* dataset from Ex. 1, drop the **Longitude** and **Latitude** features, and then create a **pairplot** for the six remaining features.

Note: Submit through **Canvas**