**Lab Assignment #1 – Apply Linear Regression Algorithm to model the relationship between independent variables and the dependent variable for a real-world problem.**

Due Date: By the dropbox deadline.

Purpose: The purpose of this Lab assignment is to:

* Practice the use of linear regression and polynomial regression
* Design and implement Python applications that incorporate linear regression solutions

References: Read reference books, Stanford tutorial on linear regression (http://ufldl.stanford.edu/tutorial/supervised/LinearRegression/), and the lecture slides. This material provides the necessary information that you need to complete the exercises.

* Your submission **MUST** include the **dataset files (if any)** that you may have used. Instead of using a dataset file, **if you need call some function to fetch a dataset**, then that should be fine too.
* Your submission **must** be a **zip** file **if compressed**. Any other compression **is not accepted**.
* You MUST run the program of an exercise in the relevant **.ipynb** file and **MUST retain** **the output that gets generated** (Note that the textual output of a program in an .ipynb file stays in that .ipynb file if you do not delete the output).
* You **MUST** create a **demo video** of less than 3 minutes of your solution. **Do not** **show yourself** in the demo video. Upload your video in your personal youtube account or google-drive account and share its link with the instructor through the **Comments** **box** of submission pageas mentioned next (Do not share the video publicly).
* During submission at the dropbox, you should see a **Comments** **box** present near the bottom of the submission page. **Write the link of your video** in this **Comments** **box**.
* Next, upload your solution and submit.

**Exercise 1: Linear Regression**

Write an application using scikit-learn to train/test the real estate data. Use **Linear Regression** model. Use the dataset from UCI repository: <https://archive.ics.uci.edu/ml/datasets/Real+estate+valuation+data+set>.

The target in the dataset is “**Y house price of unit area**”. Also, determine the coefficient of determination (R2) of the model.

(5 marks)

**Exercise 2: Polynomial Regression**

Write an application using scikit-learn to train/test the real estate data. Use **Polynomial Regression** model. The dataset is from California housing data: <https://scikit-learn.org/stable/modules/generated/sklearn.datasets.fetch_california_housing.html>. Use only the two features **AveRooms** and **AveBedrms** (out of the eight features). You should invoke the PolynomialFeatures constructor as **PolynomialFeatures(degree=2, include\_bias=False)**.

(To understand this way of calling the PolynomialFeatures constructor, the website <https://realpython.com/linear-regression-in-python/#polynomial-regression-with-scikit-learn> may be helpful.)

Also, determine the coefficient of determination (R2) of the model.

(Hint: **X, y = fetch\_california\_housing(return\_X\_y=True, as\_frame=True)** can be used to return the feature data in **X** (in form of a DataFrame object) and target data in **y**. Check the type of **y**. If the type of **y** is found to be *DataFrame*, then convert the type of **y** to *Series* using the DataFrame method **squeeze**.)

(5 marks)

**Evaluation:**

|  |  |
| --- | --- |
| **Design and Functionality:**  Correct design and implementation of requirements  Code explanation if asked | 90% |
| **Documentation of code using comments:**  At least a single-line comment for each functionality | 10% |
| **Total** | 100% |

You must name your Jupyter notebook file(s) according to the following rule:

**YourFullname\_COMP377Labnumber\_Exercisenumber**.ipynb

Example: **JohnSmith\_COMP377Lab1\_Ex1**.ipynb

**Submission rules:**

Submit your solution as a **zip file** that is named according to the following rule:

**YourFullname\_COMP377Labnumber.zip**

Example: **JohnSmith\_COMP377Lab1.zip**

Use 7-zip to compress files (https://www.7-zip.org/download.html).