**Programming Assignment-2 Multivariate Linear Regression with**

**First Name1: \_\_\_\_\_\_\_\_\_\_\_\_ Last Name1: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**First Name2:\_\_\_\_\_\_\_\_\_\_\_\_ Last Name 2: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Part-I:**

The Jupyter notebook file, LR-Multivariate-incomplete.ipynb, is provided.

Please fill in the blanks, where you can find the indications -- “### START YOUR CODE HERE ###” until the end of “Part-I” in the given Jupyter Notebook.

**As you can see, the methods of the class are spread over several cells. In this case, you can concentrate on each method separately. An additional package “jdc” is needed to support this feature. Please read** [**https://alexhagen.github.io/jdc/**](https://alexhagen.github.io/jdc/) **and install this package in your Anaconda or whatever IDE you use.**

**Experiment:**

Two files, “prj2data1.csv” and “prj2data1\_test.csv” are provided. Please put down the following information after training (learning rate = 0.01 and epochs=1000):

*w*0 =

*w*1 =

*w*2 =

Train accuracy:

Test accuracy:

**Part-II:**

**Background:**

Health insurance is insurance that covers the whole or a part of the risk of a person incurring medical expenses, spreading the risk over a large number of persons. In this study, you will build a model to predict individual medical costs billed by health insurance.

Two files, “insurance.csv” and “encoded\_insurance.csv” are provided. “insurance.csv” is the original data file that shows the raw data, which consists of 6 independent variables and 1 dependent variable (i.e., charges):

1. Age: age of primary beneficiary.
2. Sex: insurance contractor gender, female, male.
3. BMI: Body mass index, providing an understanding of body, weights that are relatively high or low relative to height, objective index of body weight (kg/m^2) using the ratio of height to weight, ideally 18.5 to 24.9.
4. Children: Number of children covered by health insurance/Number of dependents.
5. Smoker: Is the person a smoker or not.
6. Region: the beneficiary’s residential area in the US, northeast, southeast, southwest, northwest.
7. Charges: Individual medical costs billed by health insurance.

**Fill in the blanks in Part-II of the experiment.** The “encoded\_insurance.csv” file contains the data that has been preprocessed. For example, female/male is encoded to be 0/1 and number of children is encoded in terms of one-hot vector. Please use “encoded\_insurance.csv” to train your model in Part II of the Jupyter notebook.

**Q1:** What is the output if you don’t perform feature scaling for the training and testing sets?

**Q2:** Where did you put the codes for data normalization (i.e., feature scaling)?

**Q3:** After normalizing your features, what are the train and test accuracies if you don’t normalize the dependent variable “charges”? (learning rate = 0.01 and epochs=1000):

Train accuracy:

Test accuracy:

**Q4:** After normalizing your dependent variable, what are the train and test accuracies? (learning rate = 0.01 and epochs=1000):

Train accuracy:

Test accuracy:

**Submission:**

* **Rule1:**
  + If you work with a partner, please name your zipped file as follows:

PA2\_LNAME1\_LNAME2.Zip for folder and PA2\_LNAME1\_LNAME2.docx for a word document, i.e., the file names should include both LAST NAMEs.

* + If you work on your own, the format should be

PA2\_LNAME.Zip for folder and PA2\_LNAME.docx for a word document.

* **Rule2:**
  + Put your FULL names whether working in a group or individual in the word document that answers all the questions.
* **Rule3:**
  + **EVERYONE** in the class should submit this Assignment, which should provide all files (like test excel files etc.. ) that are necessary for the execution of code in the submission folder.
* **Rule4:**
  + Please submit the completed Jupyter file as well as this Word document.