**DSCI – 633 Foundations of Data Science and Analytics**

**Assignment 3 (Deadline: Thurs, 09/29/2022, 11:59 pm EST)**

**Objective:**

Code **Gradient Descent from scratch in python**. Follow these steps to achieve the task:

1. Normalize the dataset using the library learned in Tuesday’s class (09/20/2022) from sk-learn.
2. Initialize the bias and the weight theta randomly. You can use the random() function.
3. Compute the predicted value of y with the randomly assigned bias and weight

(Learning Rates = [0.0001,0.001,0.01])

1. Calculate the cost function from predicted and actual values of y. Remember, the goal is to minimize the cost function.
2. Use gradient descent to get updated bias and weights.

Dataset: Use the Wisconsin breast cancer dataset. Code it using Google Colab notebook. You can use the NumPy, Math, and Panda libraries. *Code commenting is required.*

**Link to dataset:** [Wisconsin breast cancer dataset](https://raw.githubusercontent.com/aiforsec/RIT-DSCI-633-FDS/main/Syllabus/Datasets/wisconsin-breastCancer-data.csv)

**Tasks References:**

* Gradient Descent Algo - [Lecture 4:](https://github.com/aiforsec/RIT-DSCI-633-FDS/blob/main/Syllabus/Lecture04/Lec04.pptx) Slide: 18 onwards ([Hands-on Machine learning book](https://www.knowledgeisle.com/wp-content/uploads/2019/12/2-Aur%C3%A9lien-G%C3%A9ron-Hands-On-Machine-Learning-with-Scikit-Learn-Keras-and-Tensorflow_-Concepts-Tools-and-Techniques-to-Build-Intelligent-Systems-O%E2%80%99Reilly-Media-2019.pdf), Chap 4)

**Implementation and Submission:**

Code in a Google Colab notebook and submit an **editor-accessible link** in MyCourses under the Assignment section. Add your official name & RIT email address to the top of the co-lab book.

**Questionnaires:**

For any questions, please post on the slack channel or discuss with TA – Addl Tariq during office hours (Tuesday & Thursday, 11 am to 12 pm).

**Grading Rubric**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Competent  Performance | Acceptable Performance | Mediocre  Performance | Low  Performance | No Performance |
| All  program runs as required with provided accuracy. | A small number of minor problems in one of the functions, e.g., issues in predict of class probabilities. | Coding of function might have error in logic OR Code does not run, but logic is correct. | Basic programming done, but multiple functions contain logical issues | None of the objective is achieved |
| (10 points) | (8 points) | (6 points) | (4 points) | (0 points) |

***Note:*** *No copying or plagiarism of the code from peers or the Internet.*