BITS Pilani
Pilani | Dubai | Goa | Hyderabad

BITS F464 – MACHINE LEARNING

ASSIGNMENT 1 – LINEAR MODELS FOR REGRESSION AND CLASSIFICATION

Due Date: Saturday, September 30th, 2023, at 23:59 IST

Marks: 15 marks (7.5%)

Assessment type: Written evaluation with model results in a ipynb Notebook.

Overview:

In this assessment, you will have to showcase your knowledge of machine learning techniques learnt so far in the course and their applications. This assessment will allow you to apply your learning and use what you learnt thus far to compare two machine learning models.

Your answers must be submitted in the same format as that of the supplied ipynb notebook template. You will have to generate your own dataset using the supplied code and sample dataset.

Assessment criteria:

This assessment will measure your ability to:

- Analyse data quality and quantity of a generated dataset to determine if the data is sufficient.
- Apply the appropriate machine learning method to the generated dataset.
- Compare, interpret and communicate outputs from machine learning techniques.

Assessment details:

This assessment aims to help gain hands-on experience by developing and applying appropriate machine learning models. Each scenario carries equal weightage (5 marks) and assesses the correct application of the respective machine learning technique and the interpretation of the results obtained. Highlight the difference between the below Machine Learning techniques by generating your own unique dataset with the supplied dataset and code. All implementations need to be done from scratch. Use of ML library functions is not permitted.

1. Stochastic Gradient Descent and Batch Gradient Descent using Linear Regression
2. Lasso and Ridge Regression using Polynomial Regression
3. Logistic Regression and Least Squares Classification

Dataset:

The Diabetes dataset given contains the following features. It is used to predict which patient is diabetic based on the analysis of the patient's history and key medical factors.

1. No. of Pregnancies: To express the number of pregnancies
2. Glucose level: To express the Glucose level in blood
3. Blood Pressure: To express the Blood pressure measurement
4. Thickness of Skin: To express the thickness of the skin
5. Insulin level: To express the Insulin level in blood

6. BMI: To express the Body mass index
7. Diabetes Pedigree Function: To express the Diabetes percentage. DPF calculates diabetes likelihood depending on the subject's age and their diabetic family history.
8. Age: To express the age
9. Outcome: To express if the person is diabetic; 1 is Yes and 0 is No

Notebook requirements:

You are given a sample notebook to complete this assignment. Please enter your code in the appropriate sections of the notebook. For each scenario, there are three main sections: implementation of each machine learning method and your insights after comparing the two methods.

Evaluation requirements:

For each scenario, your notebook should clearly display the relevant data pre-processing (if any), data modelling with both techniques and an interpretation of the results obtained.

Submission format:

Use the supplied ipynb notebook for adding your code and insights. Please edit the header markdown text to add your team information (full names and id numbers of each team member).

Things to Remember

1. Add your team no to the name of the files (For example, **TeamXX_Assignment1.pdf** and **TeamXX_Assignment1.ipynb**). Only one submission per team needs to be submitted, ***DO NOT MAKE MULTIPLE SUBMISSIONS.***
2. Please submit the **generated dataset** and both **ipynb and PDF version of your code** after running and testing the entire code thoroughly. We will run the notebooks and we should be able to run them seamlessly.
3. Files to be submitted in single folder (named as **TeamXX_Assignment1**) on google classroom:
    a. Generated Dataset
    b. ipynb notebook
    c. pdf version of ipynb notebook

IMPORTANT NOTE:

- ***NO MAKEUPS AND LATE SUBMISSIONS WILL BE ACCEPTED AND MARKED.***
- ***ANY KIND OF PLAGIARISM WILL LEAD TO SEVERE PENALIZATION.***

Contact for clarifications:

In case of any queries, please contact the course's Teaching Assistants (TAs) by email, and any other communication is invalid. It would be best if you wrote a mail to all the following TAs for clarification.

1. Pranjali Attarde, p20220018@hyderabad.bits-pilani.ac.in
2. Priya Baju, p20230804@hyderabad.bits-pilani.ac.in
3. S Shashank, p20210412@hyderabad.bits-pilani.ac.in