CIS*6030 Information Systems

Fall 2022
Instructor: Fangju Wang

Assignment 3 (100%)

File Admission_Predict_Ver1.1.csv is created for prediction of graduate admissions [1][2]. The dataset contains 499 rows. Each row represents a graduate admission case. The dataset contains several parameters which are considered important during the application for Masters Programs. The parameters included are :

1. GRE Scores (out of 340)
2. TOEFL Scores (out of 120)
3. University Rating (out of 5)
4. Statement of Purpose and Letter of Recommendation Strength (out of 5)
5. Undergraduate GPA (out of 10)
6. Research Experience (either 0 or 1)
7. Chance of Admit (ranging from 0 to 1)

Assignment requirements:

1. Write a Python program that reads the data in the CSV file, creates a table in a PostgreSQL database, and loads the data into the table so that individual cases can be searched by using SELECT-FROM-WHERE SQL queries. (20%)

2. Write a Python program that uses an SQL query to read the data from the database, and performs *multiple variable linear regression* on the data for prediction of graduate admissions. Use a Python module to do the analysis. After the regression model has been created, the program should be able to take an unknown case and predict the chance of admission (from 0 to 1). (40%)

3. Write a Python program that uses an SQL query to read the data from the database, and performs *logistic regression* on the data for prediction of graduate admissions. Use a Python module to do the analysis. After the regression model has been created, the program should be able to take an unknown case and predict the chance of admission (0 or 1). (40%)

Training and Testing:

Separately, for 2. and 3., randomly split the given data set into training subset and testing subset, train and test you programs, and very briefly report the testing in your README file.

Also, in the README file please include the parameters your programs calculate and instruction for running your programs.

Submission:

Submit your work as a tar file by the end of Monday, Oct 31, 2022. Don't submit the data file.

Sources:

1 https://www.kaggle.com/mohansacharya/graduate-admissions

2 Mohan S Acharya, Asfia Armaan, Aneeta S Antony, A Comparison of Regression Models for Prediction of Graduate Admissions, IEEE International Conference on Computational Intelligence in Data Science 2019