Assignment # 02 (CLO-2)

Python Coding Assignment

Due date: Monday, May 27, 2024

Total Marks = 20: Q1=10, Q2=10

The assignment submission should include a word file with code (pasted from the two .py files), output (figures/results), and your response/comments. In addition, you need to submit two code .py files of question 1 and 2 as well. All the three files should be zipped and uploaded as a single zip file.

1. In class, we saw the implementation of linear regression to predict the salary of a person who works on a certain position in a company. In that example, we used polynomial with degree 4 and implementation was done using gradient descent. Starting with the same code of the class as a base (you **cannot** start with a different code), you should now solve the same problem by writing your own normal equations code. You will need inverse or pseudoinverse *inv/ pinv* function of the *numpy* module to solve this question. Position level vs salary data is given in the ***posal.csv*** file. You should confirm your results by comparing with built-in version of linear regression (*from sklearn.linear_model import LinearRegression*) and the class gradient descent version.

   Now using the above code, predict the chance of student getting an admission in a graduate program based on the GRE score, TOEFL score, university rating, SOP, LOR, CGPA, and research. The data file ***admit.csv*** is included. You can randomly split the data using *train_test_split* function, with 80% for training and 20% for testing. You can report your accuracy using *r2_score* function of the *sklearn* module. **(10)**

2. Starting with the same code of the class as a base (you **cannot** start with a different code), solve the problem where you classify whether a human spine is normal or abnormal. You need to modify the code, so that now it performs classification task using logistic regression. You cannot use the built-in logistic regression classifier to solve this problem. You will need *exp* and *log* functions of the *numpy* module to solve this question. 12 extracted image features including

various angles, tilts, slopes, radii along with the class labels (0 for normal, 1 for abnormal) are given in the ***spine.csv*** file. You can randomly split the data using *train_test_split* function, with 80% for training and 20% for testing. You can report your accuracy using *confusion_matrix* function of the *sklearn* module. You should also confirm your results by comparing with built-in version of logistic regression classifier (*from sklearn.linear_model import LogisticRegression*). **(10)**