Hw1 CSC 84020 Neural Networks and Pattern Recognition

Descriptive Statistics, Classification Using Python and Python Libraries
Analysis of Results

All parts of the code are given in the file HELP_Hw1_code_someResults.pdf. Compare the code of this file with the one presented in HELP1.zip and HELP2.zip files. Analyze how the code of zip files is embedded into the complete code (HELP_Hw1_code_someResults.pdf) as this is the only homework in which the complete code will be given. File HELP_Hw1_code_someResults.pdf uses Iris Flowers dataset.

You must replace Iris dataset with one of the datasets given below or any dataset **for classification** you like. All datasets are comprised of tabular data and no (explicitly) missing values.
All teams must use different datasets. Exclude Iris from your choice. Send me an e-mail with your choice. The principle will be FIRST COME, FIRST SERVED.

1. Swedish Auto Insurance Dataset.
2. Wine Quality Dataset.
3. Sonar Dataset.
4. Banknote Dataset.
5. Abalone Dataset.
6. Ionosphere Dataset.
7. Wheat Seeds Dataset.
8. your choice.

PART1: for plotting you **must use** matplotlib

a) Completion of all parts of listings 1, 2, 3, 4 and 5 from file HELP_Hw1_code_someResults.zip – **30 points**.
b) Analysis of results of PART1 – **30 points. For each listing analysis should be more than 100 words.**

PART2: for plotting you **must use** matplotlib

Listing 6: Pairwise Pearson Correaltion, Skew for Each Attribute (or some if you have more than 6), Univariate Density Plot, Correlation Matrix Plot.

Listing 7: Rescaling Data, Standardize Data, Normalize Data, Binarize Data.

c) **Completion of Listings 6 and 7 of Part2 – 30 points.**
d) **Analysis of results PART2 – 30 points. For each listing analysis should be more than 100 words.**

NOTE: For completion of PART2 use Lec3 and resources as

https://scikit-learn.org/stable/user_guide.html

https://scikit-learn.org/stable/modules/preprocessing.html

PART3: for plotting you **must use seaborn**

Listing 8: Complete any 8 calculations and plottings using seaborn package **which are not included into the previous calculations and plottings with matplotlib. Two of these might be related to PCA – 40 points.**

Analysis of results PART3 (For each listing analysis should be more than 100 words) - 40 points

NOTE: you may use the following resources for completion of PART3

The seaborn.pdf presentation on the BB

http://seaborn.pydata.org/tutorial/relational.html

http://seaborn.pydata.org/

https://seaborn.pydata.org/tutorial.html

Evaluation:

a) **Part1** – max **60 points**

b) Part2 – max **60 points**

c) Part3 - - max **80 points**

Submit on the Blackboard:

a) Upload your dataset with all the details like description and a link to it.

b) Upload your Python file1 (.zip) (Solution PART1) which combines your program code (.py, .ipynb and converted to pdf .ipynb), all your outputs and your analysis after each output). Make sure to include the Listing number accompanied by a), b) c) etc. as well as comments.

c) Upload your Python file2 (.zip) (Solution PART2) which combines your program code (.py, .ipynb and converted to pdf .ipynb), all your outputs and your analysis after each output). Make sure to include the Listing number accompanied by a), b) c) etc. as well as comments.

d) Upload your Python file3 (.zip) (Solution PART3) which combines your program code (.py, .ipynb and converted to pdf .ipynb), all your outputs and your analysis after each output). Make sure to include the Listing number accompanied by a), b) c) etc. as well as comments.

e) List all members of your team as well as course number and Hw1 under comments. There is no need for each member of the team to upload the Hw1 on the BB. But make sure that the submission is done before the expiration of due date and time.

The max number of points for Hw1 is 200 points.

Your e-mail submissions will be ignored.