Hw1 CSC 767 Neural Networks & Deep Learning

Descriptive Statistics and Analysis Using Python and Python Libraries

Some parts of the code are in files Hw1_code_someResults.pdf and Help.zip.

Analyze how the code of Help.zip which contains much more than it is required in Hw1 but it will be useful in next assignments if you want to learn more. The pdf files also contain some analysis. You can see the recommendable format for your submission. You might want to download both zip files.

File 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 should be for classification with numerical attribute values, have more than two classes with not so many attributes and no (explicitly) missing values.

All teams must use different datasets. Send me an e-mail with your choice. The principle will be FIRST COME, FIRST SERVED. (Iris Flowers and Wine Quality datasets are excluded as a choice)

1. Swedish Auto Insurance Dataset.
2. Pima-Indians Diabetes.
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, and 3 shown in file Hw1_code_someResults.zip – **30 points**.
b) Analysis of results of PART1 – **20 points.**

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

Listing 4: Pairwise Pearson Correlation, Skew for Each Attribute, Univariate Density Plot, Correlation Matrix Plot (correlation matrix must be normalized). See Help.zip file.

Listing 5: Rescaling Data, Standardize Data, Normalize Data, Binarization (Help.zip).

Completion of Listings 4 and 5 of Part2 – 50 points.
Analysis of results PART2 – 20 points.

NOTE: For completion of PART2 use Lec3 and the link below

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

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

Listing 6: Complete any 6 calculations and plottings using seaborn package **which are different from implementation of matplotlib – 60 points.**

Analysis of results PART3 – 20 points

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

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

https://jakevdp.github.io/PythonDataScienceHandbook/04.14-visualization-with-seaborn.html

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

Evaluation:

a) **Part1** – max **50 points**

b) Part2 – max **70 points**

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

Total for Hw1 – max 200 points

Submit on the Blackboard:

a) Upload your dataset (csv) with all the details like description and a link to the dataset file.

b) Upload your Python files1 (py, ipynb and pdf) (Solution PART1) which combines your program code, all your outputs and your analysis after each output). Make sure that your pdf file contains all text and graphics. Include the Listing number accompanied by a), b) c) etc. as well as comments.

c) Upload your Python file2 (py, ipynb and pdf) (Solution PART2) which combines your program code, all your outputs and your analysis after each output). Make sure that your pdf file contains all text and graphics. Include the Listing number accompanied by a), b) c) etc. as well as comments.

d) Upload your Python file3 (py, ipynb and pdf) (Solution PART3) which combines your program code, all your outputs and your analysis after each output). Make sure that your pdf file contains all text and graphics. Include the Listing number accompanied by a), b) c) etc. as well as comments.

e) List all members of your team. There is no need for each member of the team to upload the Hw1 on the Blackboard. 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. Please, read the Syllabus.