ELEC946 Intelligent System Design, Spring 2021
Programming Assignment 1

Instructor: Gil-Jin Jang Email: gjang@knu.ac.kr
School of Electronics Engineering, Kyungpook National University

The purpose of programming assignment 1 is for the students to gain hands-on experience on basic python programming. Writing python program source codes, file I/O, performing simple numerical operations, and executing the written codes.

(Note: this assignment is almost identical to ELEC801 Pattern Recognition programming assignment 1, Fall 2020)

1 Data

The Iris dataset is composed of 150 samples of 4-dimensional vectors with 1 class label. There are 3 distinct classes, and there are exactly 50 samples per class. You will be given two files, `iris.txt` and `iris.csv`, with different formats as follows:

```
[iris.txt]
SL SW PL PW CLASS
5.1 3.5 1.4 0.2 Iris-setosa
4.9 3 1.4 0.2 Iris-setosa
4.7 3.2 1.3 0.2 Iris-setosa
...
7 3.2 4.7 1.4 Iris-versicolor
6.4 3.2 4.5 1.5 Iris-versicolor
6.9 3.1 4.9 1.5 Iris-versicolor
...
6.3 2.5 5 1.9 Iris-virginica
6.5 3 5.2 2 Iris-virginica
6.2 3.4 5.4 2.3 Iris-virginica
...
```

```
[iris.csv]
sepal_length,sepal_width,petal_length,petal_width,species
5.1,3.5,1.4,0.2,Iris-setosa
4.9,3.0,1.4,0.2,Iris-setosa
4.7,3.2,1.3,0.2,Iris-setosa
...
7.0,3.2,4.7,1.4,Iris-versicolor
6.4,3.2,4.5,1.5,Iris-versicolor
6.9,3.1,4.9,1.5,Iris-versicolor
...
6.3,3.3,6.0,2.5,Iris-virginica
5.8,2.7,5.1,1.9,Iris-virginica
7.1,3.0,5.9,2.1,Iris-virginica
...
```

These files are downloaded from `https://raw.githubusercontent.com/jbrownlee/Datasets/master/iris.csv` and `https://wiki.opencog.org/w/File:Iris.txt`. `iris.txt` uses space and tab characters to split the values, and `iris.csv` uses comma (,) (csv — comman-separated values).

2 To do

Write python codes for

- Accepting name of the input file as a command-line argument
- Read numeric values (150×4) and text labels (150)
- Find and display mean and standard deviation for each of the 4 dimensions

2.1 Using pandas for file I/O

Write a python code `read_iris_v1.py` that reads `iris.txt` and `iris.csv`, compute mean and standard deviation for each dimension, and display them. Example output is:

```
$ python3 read_iris_v1.py iris.txt
-------------------------------------------------------------------
                  SL             SW             PL             PW
mean            5.84           3.05           3.76           1.20
std             0.83           0.43           1.76           0.76
-------------------------------------------------------------------
$ python3 read_iris_v1.py iris.csv
-------------------------------------------------------------------
        sepal_length    sepal_width   petal_length    petal_width
mean            5.84           3.05           3.76           1.20
std             0.83           0.43           1.76           0.76
-------------------------------------------------------------------
```

Requirements:

1. Use python `sys` package for command line argument.
   - Suggested package variable: `sys.argv`
2. You may use python `pandas` package, or any other packages.
   - Suggested method: `pandas.read_csv`
   - Using any other method is accepted.
3. Use `numpy` package to process and numeric data.
   - Suggested methods: `numpy.asarray, numpy.mean, numpy.std`
   - Using any other method is accepted.
4. The display format may vary, no deduction for format mismatch such as alignments or precision of the outputs (number of digits after the decimal point).
5. The first header line is important. You have to read the name of the dimensions from the input files. For example, the name of the first dimension is `SL` in `iris.txt` but `sepal_length` in `iris.csv`.

2.2 Using basic file I/O

Write a python code `read_iris_v2.py` that does exactly the same things as 2.1, but without any package for file I/O.

Requirements:

1. Use `sys` and `numpy` packages only. No other packages are allowed.
   - Suggested methods: `open, file.readline, str.split, numpy.asarray, numpy.mean, numpy.std`
   - Using any other method is accepted as long as no other packages are imported.
2. The display format may vary, no deduction for format mismatch such as alignments or precision of the outputs (number of digits after the decimal point).
3. The first header line is important. You have to read the name of the dimensions from the input files. For example, the name of the first dimension is `SL` in `iris.txt` but `sepal_length` in `iris.csv`.

3 Submission and Grading Scheme

1. Files to be submitted: `read_iris_v1.py`, `read_iris_v2.py`. No other files should be included. Evaluation score deduction will be given for submitting unnecessary files.
2. Make a zip file of the above two files, `hw1.zip`, and upload it to `lms.knu.ac.kr`
3. First 5 lines should have your student ID, your name, file name, platform, and a list of required package(s). See the template file.
4. Homework 1 is 5% of total score.

10% Basic score for submission

10% Correct in the first 5 lines

50% Executability and correctness of the output

30% Code readibility

Due 3/10 Wednesday 11:59 LMS time

Late submission 3/11 Thursday 09:59 LMS time, 10% deduction per hour.