WRITTEN ASSIGNMENT

Tasks for Course: DLMDSPWP01 – Programming with Python

CONTENT

1. TASK

The starting point for your term paper will be the course book, the contents of which will serve as the basis for an in-depth examination of one of the following questions. You are expected to research and cite from sources corresponding to your chosen topic.

1.1 Description of the Task

You get (A) 4 training datasets and (B) one test dataset, as well as (C) datasets for 50 ideal functions. All data respectively consists of x-y-pairs of values.

Structure of all CSV-files provided:

X	Y
x1	y1
...	...
xn	yn

Your task is to write a Python-program that uses training data to choose the four ideal functions which are the best fit out of the fifty provided (C) *.

    i)   Afterwards, the program must use the test data provided (B) to determine for each and every x-y-pair of values whether or not they can be assigned to the four chosen ideal functions**; if so, the program also needs to execute the mapping and save it together with the deviation at hand

    ii)   All data must be visualized logically

    iii)   Where possible, create/ compile suitable unit-test

* The criterion for choosing the ideal functions for the training function is how they minimize the sum of all y-deviations squared (Least-Square)
** The criterion for mapping the individual test case to the four ideal functions is that the existing maximum deviation of the calculated regression does not exceed the largest deviation between training dataset (A) and the ideal function (C) chosen for it by more than factor sqrt(2)

In order to give proof of your skills in Python related to this course, you need to adhere to certain criteria when solving the exercise; these criteria are subsequently described under 'Details.'

1.2 Details

You are given four training datasets in the form of csv-files. Your Python program needs to be able to independently compile a SQLite database (file) ideally via sqlalchemy and load the training data into a single five-column spreadsheet / table in the file. Its first column depicts the x-values of all functions. Table 1, at the end of this subsection, shows you which structure your table is expected to have. The fifty ideal functions, which are also provided via a CSV-file, must be loaded into another table. Likewise, the first column depicts the x-values, meaning there will be 51 columns overall. Table 2, at end of this subsection, schematically describes what structure is expected.

After the training data and the ideal functions have been loaded into the database, the test data (B) must be loaded line-by-line from another CSV-file and – if it complies with the compiling criterion – matched to one of the four functions chosen under i (subsection above). Afterwards, the results need to be saved into another four-column-table in the SQLite database. In accordance with table 3 at end of this subsection, this table contains four columns with x- and y-values as well as the corresponding chosen ideal function and the related deviation.

Finally, the training data, the test data, the chosen ideal functions as well as the corresponding / assigned datasets are visualized under an appropriately chosen representation of the deviation.

Please create a Python-program which also fulfills the following criteria:

- Its design is sensibly object-oriented

- It includes at least one inheritance

- It includes standard- und user-defined exception handlings

- For logical reasons, it makes use of Pandas' packages as well as data visualization via Bokeh, sqlalchemy, as well as others

- Write unit-tests for all useful elements

- Your code needs to be documented in its entirety and also include Documentation Strings, known as "docstrings"

Table 1: The training data's database table:

X	Y1 (training func)	Y2(training func)	Y3(training func)	Y4(training func)
x1	y11	y21	y31	y41
...	...	...	...	...
xn	y1n	y2n	y3n	y4n

Table 2: The ideal functions' database table:

X	Y1 (ideal func)	Y2 (ideal func)	...	Ym (ideal func)	...	Y50 (ideal func)
x1	y11	y21	...	ym1	...	y50_1
...	...	...	...	...	...	...
xn	y1n	y2n	...	ymn	...	y50_n

The database table of the test-data, with mapping and y-deviation

X (test func)	Y (test func)	Delta Y (test func)	No. of ideal func
x1	y11	y21	N1
...	...	...	...
xn	y1n	y2n	y3n

1.3 Additional Task

Assume that your successfully created project is on the Version Control System Git and has a Branch called develop. On this Branch, all operations of the developer team are combined. Write the Git-commands necessary to clone the branch and? develop on your local PC. Imagine that you have added a new function. Write all necessary Git-commands to introduce this project to the team's develop Branch.

Please note: You need the commands for commit, push. Afterwards, you would make a Pull-request and your contribution would be added "merged" to the develop Branch after one or several of your team has reviewed your changes.

1.4 Remarks

The Dataset for this task will be made available on request for each particular student. Therefore, a ticket for the tutor should be opened, whereupon the access to the data will be granted. A copy will be sent to the responsible persons – thus a later manipulation by the students is prevented.

Your entire source code is expected to be contained in the appendix of your written assignment, so that we are able to test your entire program, including the outputs. Your input data is not required.

The aim is to fully reconstruct your work, your decisions and your assessment of the task result through your assignment.

2. ADDITIONAL INFORMATION FOR THE EVALUATION OF THE WRITTEN ASSIGNMENT

When conceptualizing and writing the written assignment, the evaluation criteria and explanations given in the writing guidelines should be considered.

3. TUTORIAL SUPPORT

In this written assignment task, several support channels are open; as the student, it is your responsibility to select your preferred support channel. The tutor is available for technical consultations and for formal and general questions regarding the procedure for processing the research essay. However, the tutor is not required to approve outlines or parts of texts and drafts. Independent preparation is part of the examination work and is included in the overall evaluation. However, general editing tips and instructions are given in order to help you get started with the written assignment.