Lab 10: Coding Assignment

13 October 2022

Lecturer: Abir De

Important: Please read the instructions mentioned in the questions carefully. We have NOT provided boilerplate code for any question.

1 Problem 1: Find Share Price (Regression)

A log book is given to you where each entry consist of timestamp and share price of a particular company on one trading day.Some of the entries for share price is missing. Your task is to find those missing price by analyzing the data.

Input Format
The first line contains an integer **N**, which is the number of rows of data to follow. This is followed by **N** rows of data, each of which contains a time-stamp(consist of date and time) in the first column and the day's share price in the second column. Two columns of data are separated by tab delimiter.
There are exactly 20 rows in each input file, where the day's share price is missing. The missing price are marked as **"Missing_1", "Missing_2", "Missing_3", ... , "Missing_20"**. These missing records have been randomly dispersed in the rows of data.

Output Format
The output should contain exactly 20 rows, each containing your predicted value, for each of the missing values **(Missing_1, Missing_2,...,Missing_20)** in that order.
Your task is to read your input from respective inputxx.txt and write your output in respective output.txt.

Sample input
250
1/3/2021 16:00:00   Missing_1
1/4/2021 16:00:00   27.47
1/5/2021 16:00:00   27.728
1/6/2021 16:00:00   28.19
1/9/2021 16:00:00   28.1
....
....
....

12/13/2021 16:00:00	27.52
12/14/2021 16:00:00	Missing_19
12/17/2021 16:00:00	27.215
12/18/2021 16:00:00	27.63
12/19/2021 16:00:00	27.73
12/20/2021 16:00:00	Missing_20
12/21/2021 16:00:00	27.49
12/24/2021 13:00:00	27.25
12/26/2021 16:00:00	27.2
12/27/2021 16:00:00	27.09
12/28/2021 16:00:00	26.9
12/31/2021 16:00:00	26.77

Sample Output
26.96
31.98
32.69
32.41
32.32
30.5
29.18
30.8
30.46
30.63
30.96
30.4
28.2
28.2
27.3
27.1666
27.58
26.82
27.13
27.68

2 Problem 3: Predicting Exam scores (Classification Problem)

In an exam, a student has to take four mandatory subjects A, B, C and D. Along with this, he or she has to take one elective subject, say X.

Given, the grades of students for all the four subjects (A, B, C, D) and X is a particular elective

subject which we assume all students opted for, we want to predict the grade for each student for X.

The set of mandatory subjects include $S = $ *Physics, Chemistry, English, Biology, Physical Education, Accountancy and Business Studies, Computer Science, Economics.*

Hence $A, B, C, D \in S$ s.t. $A \neq B \neq C \neq D$ and $X = Mathematics$

The records provided to you are the grades obtained by students who had opted for the following combinations of subjects or courses and **obtained a passing grade in each subject**.

2.1 Examples

Student 1: Chemistry, Accountancy and Business Studies, Physical Education, Physics, Mathematics
.....
Student 4: Physics, English, Mathematics, Economics, Chemistry
.....

2.2 Grading System Criteria

The student is first assessed on a scale of 100. (S)He needs a score of at least 33% to pass in the subject. Among those who pass:

- *Grade 1 is assigned to the top one-eighth of students who pass the course.*

- *Grade 2 is assigned to the next one-eighth of students who pass the course.*

- *.....*

- *Grade 8 is assigned to the last one-eighth of students who pass the course.*

If more than 1 student share the same score and lie in the margin, they share the higher grade.

2.3 Input Format

The first line $\implies N$.
N lines follow each line being a valid JSON object.
The following fields of raw data are given in json.

- *SerialNumber (Numeric): The student identifier*

- *English (numeric): Grade score*

- *Three more numeric fields from among the other compulsory subjects*

2.4 Constraints

- $1 <= N <= 10^5$

- Grade score $\in [0, 8]$

2.5 Output

For each student record that is given as a JSON object, containing the grade obtained in four subjects, output the predicted grade in the elective subject X in a new line (score $\in$ [0,8])

2.6 Training File and Sample Tests

The training file with sample test datahttps://s3.amazonaws.com/hr-testcases/1256/assets/trainingAndTest.zip is available here. The training file is also present in the current directory in which your code is executed by the name of "training.json".

The three files in this package are: training.json sample-test.in.json sample-test.out.json

Training data as well as sample testcases have been provided in the above file for offline training and to help you build your prediction model. When you submit your solution to us, you can assume that the training file can be accessed by reading "training.json" which will be placed in the same folder as the one in which your program is being executed.

2.7 Scoring

For each of the N records in the input file, we will compute:

p = $abs(PredictedGrade - ActualGrade)$

If p = 0 or 1 your answer for that particular student record will be considered correct. i.e, we allow a tolerance of one grade point away from the correct answer, to take into consideration the marginal errors which might occur during the testing or grading process.

Score = $100 * ((C - W)/N)$ Where C = Number of Correct predictions, not more than one grade point away from the actual grade point assigned. W = Number of wrong (incorrect) predictions and N = Total number of records in the input.

While the contest is in progress, only the score based on the sample test case will be displayed to you. After the contest is completed, we will revise the scores based on performance on a hidden test set only.

However, when you make submissions, you will be able to see whether your program attains a positive score on both the sample and the hidden test cases (to avoid a situation where unexpected errors occur on the hidden test set at the end).

3 Submission guidelines

Strictly follow the output format in each question. Do not write any extra space or new line char or anything else but required output; anything else written will result in deduction in marks. You have to submit inputxx.txt and outputxx.txt files both.

Finally, place the 2 folders, namely `Q1` and `Q2` inside a folder named `<ROLL_NUMBER>_L10` and compress it to a tar file named `<ROLL_NUMBER>_L10.tar.gz` using the command
`tar -zcvf <ROLL_NUMBER>_L10.tar.gz <ROLL_NUMBER>_L10`
Follow the submission directory structure and submit the tar file on Moodle.

Write all your outputs for a particular question and sample input to the corresponding output.txt files The directory structure should be -

```
<ROLL_NUMBER>_L10
| - - - - Q1
| - - - - -|- - - - Q1.py
| - - - - -|- - - - Testcases
| - - - - Q2
| - - - - -|- - - - Q2.py
| - - - - -|- - - - Testcases
```

Replace `ROLL_NUMBER` with your own roll number. If your Roll number has alphabets, they should be in *"small"* letters.