2023-August-Mathematics of Network Algorithms

Assignment 3

(Last Updated August 29, 2023)

- Deadline: 5 pm on 8th September, 2023. Please submit your assignment in the specified format here (The form will close at the mentioned time.)
- You can only use `numpy` python library for math related functions.
- You **must** submit python file named as: *enrolment-nr-assignment-nr-question-nr-student-name*.py
  For example, for the student XYZ with enrolment number 20251010, a solution for the first question should be in the file `20251010-02-01-XYZ.py`.
- Your code will be evaluated with the command `$ python3 20251010-02-01-XYZ.py`.
- Any deviation from these instructions related to submission will adversely affect the number of test cases your algorithm can solve.
- The points for each question will be determined by the quality of the output.
- Some test cases for the problem are available on the web-page.

---

1. (5 pts) [Cleaning Data] We have two files viz `asst-3-Q1.txt` that describes the data and `asst-3-Q1.xlsx` file that contains actual data. The first line of file `asst-3-Q1.txt` contains two integers m (denoting the number of data points), and n (denoting dimension of each data point). The next n lines has three entries: type of data (int or float), lower range, upper range. The file `asst-3-Q1.xlsx` contains m many rows, each containing n entries.

   Write a python program that reads these two files and cleans the data. Your program should be able to do the following cleaning:

   - Fix wrong types of values if they are fixable (i.e. covert float to int is such conversion is valid. For example, it should convert 4.0000 to 4 but it should *not* convert 4.0001 to 4)
   - Omit the data point (i.e. the row) if it contains missing values, wrong types of values (which are not fixable), or values that are outside the range.
   - Remove duplicate rows.

   The output should be a single line containing your roll number and the nr of useful rows. For example, the output for the `asst-3-Q1.txt` is as follows:

   20251010 3

2. (10 pts) File `asst-3-Q2.xlsx` contains matrix A of dimension $m \times n$ where each row corresponds to a data point. Write a program to compress A to another matrix B of dimension $m \times 2$[1]. Your program should also be able to recover the data and store it as matrix A' (of the dimension $m \times n$).

   The output should be a single line containing your roll number, followed by $||A - A'||_F/(m * n)$ and then $||B||_F$. For example, the output for the `asst-3-Q1.txt` is as follows:

   20251010 0.6313 5.9326

3. *To be added...*

[1]Here, you can hardcode 2 in your program