2023-August-Mathematics of Network Algorithms

Assignment 2

- Deadline: 5 pm on 18th August, 2023. Please submit your assignment in the specified format here (The form will close at the mentioned time.)
- You **must** submit python file named as: *enrolment-nr-assignment-nr-question-nr*.py For example, for the student with enrolment number 20251010, a solution for the first question should be in the file `20251010-02-01.py`. (Notice the $8+2+2$ digits.)
- Your code will be evaluated with the command `$ python 20251010-02-01.py`, and will be (automatically) cross-checked with known/expected solutions. 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. (10 pts) [Mean Squared Error] The first line of file `asst-2-Q1.txt` contains an integer t that denotes the number of test cases. The next t lines contains two vectors written element-wise and one after another.

   Write a python program that reads these lines, converts the raw data into two vectors of the same size (by considering the first half of the entries as the first vector and second half of the entries as another vector). It then outputs the squared error of these two vectors.

   Recall that for two vectors $\mathbf{x} = \langle x_1, x_2, \ldots x_n \rangle$ and $\mathbf{y} = \langle y_1, y_2, \ldots y_n \rangle$ we defined the squared error as $\frac{1}{n}\sum_i (x_i - y_i)^2$.

   You output should print t lines with your *roll number* followed by 'space' and then the value of mean square error function for the test case (rounded upto 4 digits). For example, the output for the `asst-2-Q1.txt` is as follows:

   20251010 5.8563
   20251010 9.0650

2. (10 pts) [Most Frequent Pair] The first line of file `asst-2-Q2.txt` contains an integer t that denotes the number of test cases. The next t lines contains string of characters.

   Write a python program that reads these lines, and outputs the most frequent pair in the string. A *pair* is a collection of two characters.

   You output should print t lines with your *roll number* followed by 'space', then the pair. For example, the output for the `asst-2-Q2.txt` is as follows:

   20251010 ae
   20251010 xy
   20251010 pq

3. (10 pts) [Approximating Function] The first line of file `asst-2-Q3.txt` contains an integer t that denotes the number of test cases. The next t lines contains $n+1$

numbers for a fixed n. Each line can be interpreted as vector $\mathbf{x}$ with n entries followed by a scalar y.

Write a python program that computes one vector $\mathbf{w}$ with n entries that has the following properties.

- Value of any scalar w_i in $\mathbf{w}$ is of the form $p + 0.05 * q$ where $p \in \{0, 1, 2\}$ and $q \in \{0, 1, \ldots, 9\}$.
- Value of $\mathbf{w}^T \cdot \mathbf{x}_j$ is as close to y_j as possible for every $j \in [t]$.

Note that we want to find $\mathbf{w}^T$ that works well for every data point.

Consider the following two vectors of t dimension each: The first vector consists of values obtained by $\mathbf{w}^T \cdot \mathbf{x}_j$, and the second vector $\mathbf{y}$ consists of values corresponding to y_j for every $j \in [t]$.

You output should print only one line with your *roll number* followed by 'space', then the mean squared error of these two vectors. For example, the output for the `asst-2-Q3.txt` is as follows:

20251010 106.128