2023-August-Mathematics of Network Algorithms

Assignment 4

- Deadline: 10th Nov, 5pm. Please submit your assignment in the specified format here (The form will close at the mentioned time with respect to each of the problems.)
- 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 `20251010-03-01-XYZ.py`.
- Your code will be evaluated with the command `$ python3 20251010-03-01-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.

The objective of this assignment is to build a artificial neural network that predicts whether an individual's income will be greater than $\$50,000$ per year based on several attributes from the census data. For all the following problems, we will use the data `asst-4-data.txt`, whose information can be bound in `asst-4-data-info.txt`.

1. (10 pts) [Cleaning Data] Write a python program that reads `asst-4-data.txt` and converts it into a useful format. The data contains 15 attributes. Your output should be a single line containing your roll number, total number of data points, followed by 15 numbers denoting the number of *valid entries*[1] in each columns.

2. (20 pts) [Building Neural Networks] Write a python code that creates a neural network and fits the above data.

   The output should be a single line containing your roll number, the number of perceptrons in each layer, the number of data points in `asst-4-test.xlsx`[2] on which the neural network correctly predicts output, total number of data points in `asst-4-test.xlsx`, and the time take by entire program.

   For example, if your neural network uses four layers with 20, 5, 10, and 1 perceptrons in these layers, and it solves 955 out of 1000 cases in 125 seconds, then the output should be

   20251010 20 5 10 1 955 1000 125

   You can assume that files `asst-4-data.txt` and `asst-4-test.txt` are present in the same folder as your program.

[1] Use your judgment to define what constitutes a valid entry.

[2] Since you do not have this file, you need to split data in `asst-4-test.xlsx` into training set and test set.