Assignment 1: Data Lab

- Due by Friday 10th September, 2021 by 5pm IST.
- To be submitted to the following email address: office.of.gr@gmail.com
- The subject of the email should be: Assignment Number [1,2,3,4,5 or 6]: Data Lab, 2021
- Please clearly mention your name and roll number.
- Submit your work as a single pdf file. Additional material, code, etc can/should also be submitted, but there should be atleast 1 pdf, which has the entire assignment.
- Wherever there is code, in the assignments, the code should be well documented and easy to understand / follow.

The objective of the assignments is three fold. One is to be able to develop expertise in writing and communicating about technical topics. This will be done by using the IEEE conference style format for all assignments. The other is to explain, in your own way, the mathematical ideas that are embedded within the technical topic of interest. For example, in this case it is Linear regression. The third is to use the topic, in this case of linear regression, to understand a problem from the real world. So in a sense the objective is to write what one may call a mathematical essay on Linear Regression.

Title could be: Assignment 1: a mathematical essay on linear regression.
Abstract. Give a brief overview of your assignment.
Author: Name, Department, Institution, Email

Section 1: Introduction

In this section, the 1st paragraph should be on a broad overview of the topic. The 2nd paragraph should be an overview of the technical aspects (i.e. in this case it is linear regression). The 3rd paragraph should be about the problem that you are aiming to solve/understand using linear regression. Finally, the 4th paragraph should give an overview of the paper.

Section 2: Linear regression

This section should outline the key principles underlying Linear regression.

Section 3: The problem

(a) Outline the problem, and plot/visualise the data.
(b) Make progress on the problem, by applying the techniques of linear regression to the problem at hand.
(c) Discuss any insights and observations.

Imagine that we're data scientists/ engineers employed by a consulting company. Our consulting firm has been hired by a nonprofit organization whose mission is to advocate for better health outcomes for low income populations in the United States. We've been asked to examine whether low income groups are at greater risk for being diagnosed and dying from cancer. If successful, our analysis will help the nonprofit with lobbying and fundraising.

Goals:

1. Demonstrate whether or not cancer incidence and mortality are correlated with socioeconomic status.
2. Provide both quantitative and visual evidence that the nonprofit can take and use to further their mission.

Objectives:

1. Gather, Clean and prepare data
2. Exploratory analysis

3. Statistical model
4. Visualizations

Section 4. Conclusions

Write about 1 paragraph on the key insights that were obtained from your study and also outline any further avenues of possible investigation.

References

Please put in all the references that you have used during the assignment. The format should be the same as the IEEE conference format.