CSCI 4336– Introduction to Machine Learning

Term Project.

Due: Thursday April 21, 2022 by 11:59pm

You may do this project either on your own or with a partner from the class. Here are the requirements:

1. Pick your favorite data set (Kaggle, UCI, etc.) and figure out how you want to analyze it. You may choose either linear regression gradient descent or a logistic regression gradient descent algorithm to be used with the data set. Once you decide what you want to do, you need to email the instructor by April 1st your plan of action on what you are going to implement and what type of information are you going to extract and conclude out of the data. These are some example projects that students have worked on at Stanford University: <http://cs229.stanford.edu/projects.html> .

This step is your first deliverable in writing via email to the instructor.

1. You must implement the code for this project using a high-level programming language: Java, C, C++. That means the machine leaning algorithms must be implemented from scratch using your choice of language. Keep in mind that your results must be clear and concise to be able to graph them using your favorite graphing tool. Your code must also be interactive as much as possible to be able to allow the user to interact with it.
2. You must be creative. Creativity and usefulness of your project will be part of the grade. This means also that you should find data that can and needs to be cleaned.
3. Visualization and analysis of the data is important, you may use Jupyter Notebooks to do so to help you visualize the data.
4. You need to provide the code, and a completely and cleanly written report to the instructor by the due date as well as schedule an appointment to demo your code and your project to both the instructor and the TA. This is also part of the grade.
5. When in doubt ask the instructor about any details or if you have any questions.

Best of Luck