CIS 4930-002: Spring 2019
Projet 1

Total Points: 200 (plus 50 points extra credit)
Due: Friday 04/26/2019

The purpose of this assignment is to

- Test your familiarity with parallel programming and OpenMP
- Test your familiarity with Classification Problems
- See if you're comfortable working with big (ish) data sets.

Problem 1 OpenMP- 100 points

The Maximum subarray problem is defined as follows:
Given an arrays of 'N' integers, find the conitguous subarray within it that has the largest sum.

For example:
Array: { -2, 1, -3, 4, -1, 2, 1, -5, 4 }
Solution: { 4, -1, 2, 1}

Specifications

1. The C++ code for the algorithm is in thee Examples folder.
2. Edit the C++ code to use a dynamic array (not a vector) to work with arbitrary input.
3. Run the code on linprog and time it.
4. Now, edit the code with OpnMP pragmas to parallelize it.
5. Run the code on linprog again and time it.
6. All I/O to be done with stdin/stdout.
7. Turn in your code and 2 screenshots of linprog, one showing sequential time and one showing parallelized time in a tarball.

Problem 2 Classification Problems - 100 points

Test the code for Baye's classification with a UCI classification dataset of your choice.

1. The Python code for Baye's classification will be posted on the Examples folder
2. Edit the part of the code that calculates variance and standard deviations to ignore the column with the class label (different datatsets have this column in different places).
3. Run the code using an AWS instance, since linprog does not have the required libraries.
4. Turn in your edited code, with the dataset chosen as a comment, and a screenshot of the output in a tarball.

Problem 3 Course Improvement- 50 points (extra credit)

This is the first time this course has been offered. It was also supposed to be offered under the title "Contemporary Programming Techniques and Tools". However, due to administrative requirements, it was changed to "Big Data" with less than 2 months before term started. This, unfortunately combined with my ongoing illness, meant that most of the prepared course material was rendered pretty useless, and the class was taught pretty much on the fly.

That being said, I made an honest effort to turn this into a real Big Data class, while teaching 2 other large classes. However, I realize there is a lot of room for improvement, which is often true for many first time course offerings. Please give me your honest opinions on the class and what you would like to see if this class were offered again. We welcome all forms of criticism.