Stats 141C High Performance Statistical Computing Spring 2017

Homework 4

Lecturer: Cho-Jui Hsieh Date Due: June 12, 11:59pm, 2017

Keywords: *Multicore Programming*

For this homework, we will use the data and code downloaded from `http://www.stat.ucdavis.edu/~chohsieh/teaching/STA141C_Spring2017/hw4_code.zip`. In this folder, we provide the code for the nearest neighbor classification algorithm in "go_knn.py" (you can directly run this python code, and it will print out the accuracy on a9a dataset). The main function in this file is "go_nn": for each testing sample, it computes the nearest neighbor in the training set, and checks whether the label predicted by the nearest neighbor matches the testing label.

Problem 1. Multicore Programming [50 pt]

Modify the "go_nn" function to parallelize the computation using multiple cores. We suggest using python "multiprocessing" module. Make sure you get the same accuracy with the original code. Run your new code using 4 cores and report the run time. Compare the run time with the original (single threaded) code.