Homework 3
APPM 5650 Fall 2021
Randomized Algorithms

Due date: Monday, Sept 13 2021 at 10:20 AM
Theme: Linear Algebra and Probability Background

Instructions Collaboration with your fellow students is allowed and in fact recommended, although direct copying is not allowed. Please write down the names of the students that you worked with. The internet is allowed for basic tasks only, not for directly looking for solutions.

An arbitrary subset of these questions will be graded.

Problem 1: [READING] The reading is Petros Drineas and Michael W. Mahoney, *Lectures on Randomized Numerical Linear Algebra*, The Mathematics of Data, IAS/Park City Math. Ser., vol. 25, Amer. Math. Soc., Providence, RI, 2018. These are the notes used in that summer school. An electronic copy is available for free at http://arxiv.org/abs/1712.08880.

Read the short introduction (section 1), and then skim section 2 on "Linear Algebra" and section 3 on "Discrete Probability." If any of that material looks new to you, or you are rusty on it, read it in more detail.

Deliverable: Do the reading. You do not need to turn in anything.

Problem 2: [PROGRAMMING] Let $U \in \mathbb{R}^{n\times 2}$ be a matrix where the first column is $\ln(1), \ln(2), \ldots, \ln(n)$ and the second column is $\ln(n+1), \ln(n+2), \ldots, \ln(2n)$. This is the same matrix as in last week's homework, so you can re-use your code.

Deliverable: For $n = 10^5$, compute $\|UU^T\|$ to 4 significant digits (where $\|\cdot\|$ is the spectral norm)

Hints: You might find it helpful to compare with the naive implementation for a small n to make sure your code is correct.

On my laptop in 2019, the norm computation takes roughly 0.05 seconds for $n = 10^5$.

Problem 3: [MATH] *Prosecutor's Fallacy aka Base Rate Fallacy.* Suppose 3% of CU students have COVID-19, and COVID-19 tests have a 1% false positive rate (meaning that a healthy individual has a 1% chance of the test returning a "positive"/"infected" result; aka, 99% specificity)[1]. The test also has a 2% false negative rate (a sick individual has a 2% chance of the test returning a "negative" result; aka, 98% sensitivity/power).

You go in for a COVID-19 test and test positive. In the absence of other information (you may or may not be showing symptoms), what is the chance that you are actually sick with COVID-19?

Problem 4: [PROGRAMMING] Request a research computing (RC) account from https://rcamp.rc.colorado.edu/accounts/account-request/create/organization or go to the main page https://www.colorado.edu/rc/ and click "Request an account to use RC Resources". **This may take a few days, so do not leave this until the end of the week!** If it asks you for

[1]For many diseases and tests, this rate can be quite high, but for COVID-19 tests, either PCR or antigen, this rate is nearly 0 and "Most false-positive results are thought to be due to lab contamination or other problems with how the lab has performed the test, not limitations of the test itself" https://www.health.harvard.edu/blog/which-test-is-best-for-covid-19-2020081020734.

a reason you are requesting the account, tell them it is for this class, and that you expect to use one of the applied math nodes on the "Blanca" cluster. Once you have the account, login to a login node, and look at the file `/rc_scratch/stbe1590/secretMessage.txt`

Deliverable: what is the message in the file `secretMessage.txt`?

Hints: you'll need to get an account, learn what the login node is, know very basic `ssh` commands, use the "duo" two-factor authentication app, do basic command line usage in a linux shell, and view the contents of a file (with `cat` or a text editor). Depending on your background, this ranges from trivial to overwhelming. If you are knew to these tools, I can give hands-on help in office hours, but please **read the documentation on the rc website first!** They have worked very hard to document every step of the process, so please take the time to read their instructions before you ask for help.