CS545 2023 – PROBLEM SET 1

This problem set is due on September 13th. Submit your answers as a typeset PDF document. Scanned/handwritten PDFs will not be accepted. For each day that your submission is late your grade will drop by 10%, after 3 days you get a zero grade. Do the problem set by yourself, although it's fine to discuss solution approaches with your classmates. Do not use solutions that you find from past years.

Problem 1. Some simple linear algebra

Show that:

a. $\mathrm{vec}(\mathbf{a} \cdot \mathbf{b}^\top) = \mathbf{b} \otimes \mathbf{a}$, where **a** and **b** are both vectors

b. $\mathrm{vec}(\mathbf{A})^\top \cdot \mathrm{vec}(\mathbf{B}) = \mathrm{tr}(\mathbf{A}^\top \cdot \mathbf{B})$ for two matrices of the same order

c. If $\mathbf{X}$ is a real matrix, then the matrix $\mathbf{X} \cdot \mathbf{X}^\top$ is positive semi-definite

Problem 2. A probability problem

We, the engineers, are much smarter than doctors. What better way to demonstrate your intellectual superiority than by answering something that stumps doctors. A few years back there was an enlightening study on how doctors miscalculate. Doctors were asked to answer a question that they would have to answer in their everyday routine: "If a woman has a positive mammogram, what is the probability that she has breast cancer?" The relevant statistics you need to know are as follows:

- The probability that a woman has breast cancer is 0.6%
- If a woman has breast cancer, there is a probability of 90% that the mammogram will be positive.
- If she has no breast cancer the probability of a positive mammogram is 7%

German doctors estimated that probability to be from 1% to 90%. In the US 95% of doctors estimated the probability to be around 75%. Find the exact number.

Problem 3. Stats operations using linear algebra

Most often you will need to take a derivative, or perform some other horrible operation. Such things are much easier to perform when you are dealing with simple linear algebra operations as opposed to code functions (e.g. `mean()`, `cov()`, `var()`, etc). In this exercise you will gain some familiarity with how to transform common statistical operations to linear algebra expressions.

1. You are given K grayscale images of size M by N. Each image is represented as a column vector and they are all presented to you as an $(M \times N)$ by K matrix.
    a. Use a linear algebra expression that computes the mean image as an M by N matrix. Don't use summations, don't use for loops. You can freely construct any arbitrary matrix to use in any expression that you need. Hint: *vec-transpose*
    b. Obtain a 2×2 covariance matrix of the average of the top half of all the images (not top half of the vector representation) with the average of the bottom half.
2. Now your images are colored, and packed in a M by N by 3 by K tensor. The dimension of size 3 contains the red, green and blue channel.
    a. Like above, show the expression for the mean image over all channels.
    b. Do that again, but compute the mean image of only the red channel.

Make sure you use proper math notation (don't use code shortcuts). See the notes from the first lecture on ideas on how to solve these problems. I also strongly recommend that you implement the above operations in code so that you can verify that they indeed work (we don't need to see that code in your submission).

Problem 4. Signal Processing is Linear Algebra

As promised in class, matrix multiplies can perform all kinds of linear operations. Here you will have to implement a spectrogram using only a matrix multiplication. Your input sound is a column vector $\mathbf{x}$. Describe exactly how you would construct a matrix $\mathbf{A}$, such that the product $\mathbf{A} \cdot \mathbf{x}$ will produce the $\mathrm{vec}(\cdot)$ of the complex spectrogram coefficients. Your transform should have a DFT size of 64, a hop size of 32, and will use a Hann window. Make an image plot of the real part of the matrix $\mathbf{A}$ and allow me to marvel at its beauty.

Extra extra credit: Show me how to transform the resulting frequency coefficient vector to a matrix that I can plot as a spectrogram. Do that with a sound that you record yourself (and use an appropriately larger DFT size).

Remember that we have a TA. Also remember that the more you procrastinate, the TA will have less time to talk to you. Have fun!