ECE/CS 434 : Real-World Algorithms for IoT and Data Science : Homework 2 : Due 11 :59pm, Wed, Feb 28, 2024

Please start each of the 7 questions on a new page, then scan and upload these pages onto Gradescope.

Problem 1 [15 points]

Consider the 2^{nd} column of the Fourier matrix, which is $[e^{j0}\ e^{j\theta}\ e^{j2\theta}\ \ldots\ e^{j(N-1)\theta}]^T$.
(a) Prove that the 3^{rd} column is orthogonal to the 2^{nd} column.
(b) Prove that any column is orthogonal to the 2^{nd} column.
(c) Prove that any two columns are orthogonal.

Problem 2 [24 points]

Consider a N dimensional vector $\bar{v}$ expressed in the identity basis.

(a) Express the vector $\bar{v}$ in an orthonormal basis F, where F is a $N \times N$ matrix.
(*Hint : See class notes on how we express a signal in different basis.*)

(b) Let's call the above vector $\bar{w}$. Now create a matrix B such that $B\bar{w}$ scales the i^{th} element of $\bar{w}$ by a scalar b_i. What should be the matrix B ?

(c) Let's denote the vector $B\bar{w}$ as vector $\bar{z}$. Now convert vector $\bar{z}$ back into the original identity basis.

(d) Now write all the above operations on vector $\bar{v}$ in one equation in terms of F and B.

(e) Write the Eigen-decomposition equation of a matrix A, where S contains the eigenvectors of A and Λ is the diagonal matrix containing the eigenvalues.

(f) Given the above exercise you have done, explain in plain English what Eigen-decomposition does to a vector (in other words, what happens when matrix A is multiplied to vector x) ?

Problem 3 [4 points]

In class, we discussed the analogy of expressing a painting "lonely queen" in two different bases ; one was ⟨ red, green, blue ⟩ and the other was ⟨ purple, brown, white ⟩.

Can you come up with another analogy from the real world where the same "thing" can be expressed in 2 different "bases". Write the "thing" and the 2 "bases".

Problem 4 [12 points]

(a) You are sampling a signal every 0.25 millisecond. What is the maximum frequency you would be able to see in FFT ?

(b) Suppose you take $N = 1000$ for your FFT. At what frequency resolution would you be able to analyze the signal you are sampling ? A frequency resolution of R Hz means you express the signal at frequencies $[0, R, 2R, ...]$ Hz.

(c) True/False : The FFT of any signal is symmetric around frequency zero. Explain your answer in 1 sentence.

(d) Say X_f is the DFT of a signal x_n. Now, consider $Y_f = X_f.e^{j\phi}$, where ϕ is a constant angle. Is the IDFT(Y_f) a shifted version of the signal x_n ? Briefly argue in favor or against.

Problem 5 [10 points]

(a) Consider a signal $x[n] = \cos(2\pi f_1 n t_s) + 2\sin(2\pi f_1 n t_s) + \sin(4\pi f_1 n t_s)$. Draw the magnitude and phase plots of X_f, which is the DFT of $x[n]$. Assume that f_1 is the fundamental frequency in which you are sampling the signal.

(b) Prove that DFT is linear, i.e., $\text{DFT}(a_1 x[n] + a_2 y[n]) = a_1 X_f + a_2 Y_f$, where X_f and Y_f are the DFTs of $x[n]$ and $y[n]$, respectively.

Problem 6 [5 points]

Use your phone to record your own voice, and say "My name is [Your Name]." Use Python to import the saved audio file, and compute its FFT. Submit the plot of the *magnitude* of the FFT.

Hint : You can use `scipy.io.wavfile.read`[1] to read the audio file and get the sampling rate. If your data has two channels, you can extract 1 with `data = data[:, 0]`. You can then compute the FFT with `scipy.fft`[2].

Problem 7 [30 points]

(a) Prove that : $P(A, B, C, ...Z) = P(A|B, C...Z)P(B|C, D, ...Z)...P(Z)$

(b) Given two random variables X_1 and X_2, write an equation to check if they are uncorrelated.

(c) Say data $D_1 = [2, 9, 10, 7, 2, 6, 8, 16, 14]$ and data $D_2 = [4, 12, 14, 8, 5, 7, 11, 16, 10]$.
Are the data uncorrelated ?

(d) Given two random variables X_1 and X_2, write an equation to check if they are independent.

(e) Can two data streams be dependent but uncorrelated ?
If so, give an example. If not, explain why not.

(f) Use any programming language to generate 3 sequences, each containing 10 integers generated uniform randomly between $[-10, 10]$. Let's call these sequences S_1, S_2, S_3 and say a vector $S = [S_1, S_2, S_3]^T$. Calculate (using pen and paper, or using a calculator) the covariance matrix $Cov(S)$ and write it out in matrix form.

(g) Random variables Y_1 and Y_2 are jointly Gaussian. Write their joint distribution in terms of their mean vector μ_Y and the covariance matrix $\sum_Y$.

(h) Roughly sketch the above PDF when :

(I) $\mu_Y = [0, 0]^T$ and $\sum_Y = I$, the Identity matrix.

(II) $\mu_Y = [1, 3]^T$ and $\sum_Y = I$, the Identity matrix.

1. https://docs.scipy.org/doc/scipy/reference/generated/scipy.io.wavfile.read.html
2. https://docs.scipy.org/doc/scipy/reference/tutorial/fft.html#fast-fourier-transforms

(III) $\mu_Y = [1, 3]^T$ and $\sum_Y \neq I$.

(i) You collect some data as a vector $D = [d_1, d_2, \ ... \ d_n]^T$. Say $\sigma^2 = Var(D)$, where Var implies variance. You now multiply each data value with a constant k and the add c to each of them. What is $Var(kD + c)$ in terms of σ^2.

(j) A Bivariate Gaussian distribution of X_1 and X_2 has a covariance matrix which is an Identity.

(I) Do you think X_1 and X_2 are independent ? Why or why not ?

(II) Calculate $P(X_1 < 0.5 | X_2 = 0.9)$, assuming both X_1 and X_2 are mean zero random variables.