ELEC 378 – Spring 2023
Homework 3

Due: Friday February 3, 5PM

1 Cauchy-Schwarz Decoding III

Suppose you are eavesdropping on a noisy communication channel and observe the signals contained in `eavesdropping.mat`, which RichB has segmented into packets using his impressive DSP prowess (i.e., each row is a carrier waveform as before). However, you don't know the communication scheme, i.e., the mapping from waveforms to message symbols.

(a) Use PCA to show visually in 2D that the intercepted transmission is likely binary communication, i.e., that each received packet is just one of two distinct carrier waveforms. Include an illustrative scatterplot in your writeup.

(b) Classify each waveform as a 0 or 1 bit and color your scatterplot accordingly to confirm correctness.

(c) Suppose the intercepted transmission actually encodes a 7-bit ASCII message. Decode the bits under this assumption and include the intercepted message in your writeup.

2 Denoising with PCA

When using PCA to reduce dimensionality, only the principal directions that capture large amounts of energy in the data are retained. Directions that capture very small amounts of energy can be thought of as unnecessary in representing the data, or, "noisy". Denoising with PCA can be accomplished by reducing the dimensionality of the data, followed by reconstructing the data from its dimensionality reduced representation. Demonstrate this property on the MNIST dataset, contained in `digits.npy`.

3 Dr. Data Science

Download `cancer.mat`, which contain 6,830 gene expression measurements, stored in the data matrix `X`, for 64 patients, each of whom has 1 of 14 different cancer types, stored in the label vector `Y`.

(a) Use PCA to show visually in 2D that patients with melanoma are similar in some sense. Which of the first two principal directions is most informative of a patient's melanoma diagnosis?

(b) Recall that the coordinate of the i^{th} data point projected onto the j^{th} principal direction is given by

$$\langle \mathbf{x}_i, \mathbf{v}_j \rangle = \sum_{k=1}^{p} x_i[k] v_j[k],$$

where $\mathbf{v}_j$ is the j^{th} eigenvector of $\mathbf{X}^\intercal \mathbf{X}$. (e.g., $\langle \mathbf{x}_i, \mathbf{v}_1 \rangle$ gives the location of $\mathbf{x}_i$'s projection onto the first principal direction.) With this in mind, if $\mathbf{v}_j$ is the most informative principal direction regarding some diagnosis, what can you say about the genes $x_i[k]$ for which $v_j[k]$ is large in magnitude?

(c) Using the results from the previous two parts, suggest a method for determining which genes are responsible for a given diagnosis. Support your suggestion with visual evidence for the diagnosis of melanoma by displaying a heatmap of the data matrix after sorting the columns by magnitude of $v_j[k]$, where $\mathbf{v}_j$ is the most informative principal direction of a melanoma diagnosis.

4 Next Week Sneak Peek

Consider the following systems of linear equations:

a)

$$\begin{aligned} -x_1 + 2x_2 + 3x_3 &= 16 \\ 4x_1 - 2x_2 - x_3 &= -2 \\ x_2 + x_3 &= 7 \end{aligned}$$

b)

$$\begin{aligned} x_1 - x_2 - 2x_3 &= -2 \\ 2x_1 + 2x_2 + x_3 &= 5 \\ 3x_1 - 3x_2 - 6x_3 &= -6 \end{aligned}$$

c)

$$\begin{aligned} 4x_1 + 3x_2 + 2x_3 &= 17 \\ x_1 + 2x_2 + 3x_3 &= 8 \end{aligned}$$

d)

$$\begin{aligned} 2x_1 + x_2 &= 4 \\ x_1 + 3x_2 &= 4 \\ -x_1 + 2x_2 &= 3 \end{aligned}$$

Solve each system by hand or numerically, i.e., without using a symbolic solver, and provide an expression for all possible solutions when they exist.

Submission Instructions

Every student must submit their work in PDF format, providing intermediate and final results as well as any necessary code. Submit your homework on Gradescope.

Collaboration Policy

Collaboration both inside and outside class is encouraged. You may talk to other students for general ideas and concepts, but individual write-ups must be done independently.

Plagiarism

Plagiarism of any form will not be tolerated. You are expected to credit all sources explicitly.