ELEC 378 – Spring 2023
Homework 1

Due: Friday January 20, 5PM

1 ℓ_p Geometry in d Dimensions

The ℓ_p-norm of a vector $\mathbf{x} \in \mathbb{R}^d$ is defined as

$$\|\mathbf{x}\|_p = \left(\sum_{i=1}^{d} |x_i|^p \right)^{\frac{1}{p}},$$

and it can be shown that

$$\lim_{p \to \infty} \|\mathbf{x}\|_p = \max(|x_1|, \ldots, |x_d|).$$

The unit ℓ_p ball is defined as the set $\{\mathbf{x} : \|\mathbf{x}\|_p \leq 1\}$, i.e., the set of all vectors whose ℓ_p-norm is at most one.

a) Sketch the ℓ_1, ℓ_2, and ℓ_∞ balls for $\mathbf{x} \in \mathbb{R}^2$ and $\mathbf{x} \in \mathbb{R}^3$.

b) Using a simulation, estimate the volume between the ℓ_2 ball and ℓ_∞ ball as the dimensionality d grows large. How does this volume in high dimensions compare to your intuition from $\mathbb{R}^2$? What does this result say about the concentration of volume in a hypercube in high dimensions?

2 Curse of Dimensionality, The Movie

Many supervised classification methods predict the label of an unlabeled data point using the known labels of training data points in its "local neighborhood". While this approach seems reasonable in low dimensions (e.g., $\mathbb{R}^2$ or $\mathbb{R}^3$ where we can directly visualize the data), the concept of "local neighborhood" breaks down in high dimensions (in the hundreds, thousands, and beyond) – this phenomenon is called the curse of dimensionality.

To explore this breakdown, consider a collection of training data points uniformly distributed throughout a p-dimensional hypercube $[0,1]^p$, where each side is of length 1. Special cases are the unit interval for $p = 1$, the unit square for $p = 2$, and the unit cube for $p = 3$.

(a) Consider a data point in the unit interval ($p = 1$), and let us define its local neighborhood to be the interval centered at this data point that contains the nearest r percent of the training data. If the training data points are uniformly distributed throughout the unit interval, calculate the expected length of the local interval in terms of r and make a sketch of the situation for $r = 0.01$ (i.e., capturing 1% of the training data).

(b) Consider a data point in the unit square ($p = 2$), and let us define its local neighborhood to be the square centered at this data point that contains the nearest r percent of the training data. If the training data points are uniformly distributed throughout the unit square, calculate the expected side-length of the local square in terms of r and make a sketch of the situation for $r = 0.01$.

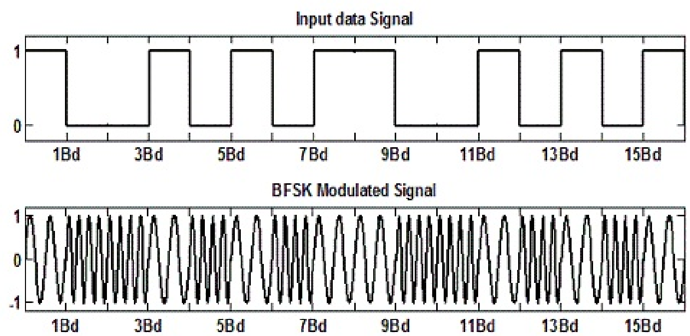
(c) Consider a data point in the unit cube ($p = 3$), and let us define its local neighborhood to be the cube centered at this data point that contains the nearest r percent of the training data. If the training data points are uniformly distributed throughout the unit cube, calculate the expected side-length of the local cube in terms of r and make a sketch of the situation for $r = 0.01$.

(d) Consider a data point in the p-dimensional unit hypercube for arbitrary p, and let us define its local neighborhood to be the hypercube centered at this data point that contains the nearest r percent of the training data. If the training data points are uniformly distributed throughout the unit hypercube, calculate the expected side-length of the local hypercube in terms of r (make sure your general formula agrees with what you found in the previous parts). While you cannot make a sketch for $p > 3$, what is the side-length of the local hypercube when $r = 0.01$?

(e) As p grows very large, would you call the hypercube that captures a ratio of r of the training data points (even for r small) "local"? Why or why not? In particular, how would you describe the region which contains the nearest 1% of the training data when p is very large?

3 Cauchy-Schwarz Decoding

To transmit a digital signal (sequence of bits) via an analog channel (e.g. wireless communication), the digital signal must somehow be embedded in an analog signal that can be physically sent through the channel – this embedding is called *modulation*. One simple modulation scheme is *binary frequency-shift keying* (BFSK), in which the analog signal is constructed by sending a sinusoid $c_1(t)$ with some frequency to represent a 1 and a sinusoid $c_0(t)$ with a different frequency to represent a 0:



`http://www.kics.edu.pk/wdsp/flextrainer/images/dspimages/27.gif`

If the signal were to make it through the channel untouched, one could simply read off the bit sequence from the received waveform. However, every real channel is noisy. A reasonable assumption is that the noise is additive, i.e. if the transmitter sends $x(t)$, the received signal is $y(t) = x(t) + n(t)$ for some noise signal $n(t)$.

(a) Design a system to decode a digital signal from its corrupt BFSK signal. Be sure to justify why the system should work via the Cauchy-Schwarz inequality.

(b) Download `cauchy_schwarz_decoding.mat` from the Resources section of Piazza and implement your system to decode the corrupt signal `y`. You will need the carrier waveforms `c1` and `c0`, also included. Write the decoded bits as a `.jpg` file and include the resulting image in your write-up. Does your system decode the corrupt signal correctly?

(c) Listen to the corrupt signal $y(t)$. What does it sound like? Reconstruct the modulated signal $x(t)$ from the decoded bits and listen to it. How does it compare to $y(t)$? Are you satisfied with your system's performance?

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. Homework, tests, and solutions from previous offerings of this course are off limits, under the honor code.