ECE 592 – Topics in Data Science

Test 5: Sparsity – Fall 2020

November 16-18, 2020

Please remember to justify your answers carefully.

Last name: ____________________ First name: ____________________

Please recall the course academic integrity policy for tests:
No cooperation or "collaboration" between students is allowed. Especially during an online course experience, it could be tempting to text or email a friend. This is not allowed. You will be allowed to use your notes, books, a browser, and software such as Matlab and/or Python.[1] However, while working on the test you should not text, email, or communicate with other people (certainly not other students) in any way, unless you are consulting with the course staff. **By submitting the test, you will be acknowledging that you completed the work on your own without the help of others in any capacity.** Any such aid would be unauthorized and a violation of the academic integrity policy.

[1] You can use the browser to access Moodle, the course webpage, and look up technical topics. Similar to a normal test, you must not communicate with other people.

Question 1 (Denoiser.)
Consider a communication system that modulates a binary vector, $x \in \{-2, +3\}^N$. The possible values in x are -2 and +3, and they appear with probabilities 0.3 and 0.7, respectively. In various applications, for example within iterations of *approximate message passing* (AMP), x is measured with additive white Gaussian noise,

$$v = x + q,$$

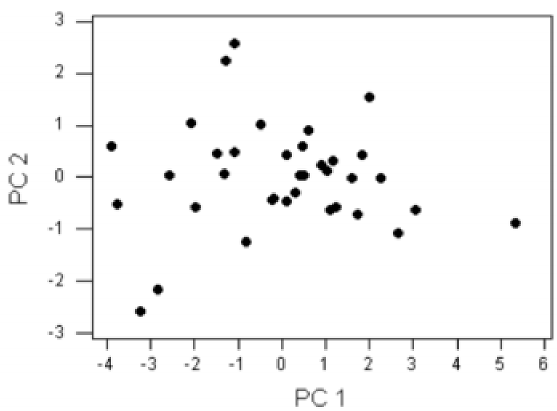
where the random variable Q_n, $n \in \{1, \ldots, N\}$, is *independent and identically distributed* (i.i.d.) with zero mean and variance σ^2. Please develop a denoising function for x, $\widehat{x} = E[X|V = v]$. Recall that the lower case refers to numerical values taken by v or x, and capital letters denote random variables, and that the probability density function for each entry Q_n of Q satisfies

$$f_{Q_n}(Q_n = q) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{-\frac{q^2}{2\sigma^2}\}.$$

Question 2 (Principal component analysis.)
In a sample of 36 rabbits, 6 different physical characteristics are measured for each rabbit: width of the head, body length, length of front legs, length of hind legs, weight, and body temperature. The first four are measured in centimeters, the fifth in Kilograms, and the sixth in degrees Celsius. The outcome of a *principal component analysis* (PCA) is given below. Based on these results, please answer the following questions.

Eigenvalue	3.056	2.042	1.571	0.991	0.054	0.002
Proportion	0.600	0.350	0.045	0.004	0.001	$< 10^{-4}$
Cumulative	0.600	0.950	0.995	0.999	1.000	1.000

Variable	PC1	PC2	PC3
head	0.423	0.01	-0.071
body	0.386	-0.055	0.051
front legs	0.100	0.081	0.101
hind legs	0.90	0.091	0.090
weight	-0.05	0.891	0.021
temp	0.03	-0.009	0.541



(a) How much of the variation of the dataset is explained by the first principal component? How much is explained by the first three together?

(b) From the data above how many dimensions can be reduced when PCA is applied to this data? Please justify your answer.

(c) A plot of the principal component scores for the 36 rabbits is shown in the figure above. Imagine that the horizontal and vertical axes are drawn on the plot, dividing it into 4 quadrants: UR (upper right), UL (upper left), LR (Lower Right), and LL (lower left). Suppose that two new rabbits, Flopsy and Peter, are measured. Flopsy is much smaller and lighter than any of the other rabbits, and owing to her small surface area is also warmer than any of the other rabbits. On the other hand, Peter is larger in size but is almost of the same weight as Flopsy. Peter has larger surface area, which allows him to be cooler as compared to all the other rabbits. For both Flopsy and Peter, which quadrant of the above graph would each lie in? Make sure to justify your answer.

Question 3 (Haar multi resolution approximation)
This question deals with the approximation of functions $f(t)$, $t \in \mathbb{R}$, in $L_2(\mathbb{R})$, meaning that they have finite energy, i.e., $\int_{t\in\mathbb{R}} |f(t)|^2 dt < \infty$. Under some technical conditions, for example $f(t)$ does not contain discontinuities, the Fourier transform of $f(t)$ exists. For signals without discontinuities, Fourier analysis often provides a sparse representation.
However, the Fourier transform is not defined for $f(t)$ with discontinuities. We will see how *multi resolution‘ approximation* (MRA) can approximate functions in $L_2(\mathbb{R})$ with discontinuities. Consider the Haar MRA space formed by the following scaling function,

$$\phi(t) = \begin{cases} 1, & t \in [0,1) \\ 0, & \text{otherwise.} \end{cases}$$

The space V_0 is comprised of all piecewise constant functions $f(t) \in L_2(\mathbb{R})$ with discontinuities at integer values, $t \in \{\ldots, -1, 0, +1, +2, \ldots\}$. Similarly, V_1 is comprised of all piecewise constant functions $f(t) \in L_2(\mathbb{R})$ with discontinuities at $t \in \{\ldots, -\frac{1}{2}, 0, +\frac{1}{2}, +1, \ldots\}$. More generally, V_j is comprised of all piecewise constant functions $f(t) \in L_2(\mathbb{R})$ with discontinuities at integer multiples of 2^{-j}.
How do we approximate a function $f(t) \in L_2(\mathbb{R})$ within V_j? Over the range $t \in [\frac{k}{2^j}, \frac{k+1}{2^j})$, the value of the approximation will be the average of $f(t)$ over that range,

$$f_j(k) = \int_{t=\frac{k}{2^j}}^{\frac{k+1}{2^j}} f(t)dt,$$

and the squared error between $f(t)$ and its approximation over this range is

$$e_j(k) = \int_{t=\frac{k}{2^j}}^{\frac{k+1}{2^j}} |f_j(k) - f(t)|^2 dt. \tag{1}$$

(Note that $f_j(k)$, the average over $f(t)$ for this range, minimizes the squared error in (1); any other number $c \neq f_j(k)$ would yield a larger squared error.) Below, we will explore how the approximation error, $e_j(k)$, decays as j increases.

(a) The function $f(t) \in L_2(\mathbb{R})$ can be smooth within range k, i.e., $t \in [\frac{k}{2^j}, \frac{k+1}{2^j})$, or may contain a discontinuity. In this part, suppose that $f(t)$ is smooth. Therefore, there exists a first order Taylor approximation,

$$f(t) = f(t_0) + \alpha(t - t_0). \tag{2}$$

Using the Taylor approximation (2), compute $f_j(k)$ and $e_j(k)$ as a function of α and j.

(b) In this part, suppose that $f(t)$ contains a discontinuity. To keep things simple, we will assume that there exists a single discontinuity at t_k such that

$$f(t_k^-) = \lim_{\epsilon\to 0, \epsilon<0} f(t) \neq \lim_{\epsilon\to 0, \epsilon>0} f(t) = f(t_k^+).$$

Bound $e_j(k)$ as a function of $\frac{k}{2^j}$, $\frac{k+1}{2^j}$, t_k, $f(t_k^-)$, and $f(t_k^+)$. (Hint: the error will be largest if t_k is in the middle of the range. Therefore, you can assume that $t_k = \frac{k+0.5}{2^j}$.)

Discussion: Recall that the Fourier transform of $f(t)$ with discontinuities does not exist; any such discontinuity would create Fourier coefficients whose magnitude is inversely proportional to the frequency, hence there would be many significant Fourier coefficients. In contrast, when the Haar MRA is extended to Haar wavelets, the discontinuity is localized, and as we move from coarse scales to finer scales, the wavelet approximation for the area of $f(t)$ surrounding the discontinuity will only need $\Theta(1)$ Haar coefficients per scale j. The approximation quality of part (b) can be achieved with only $\Theta(j)$ wavelet coefficients.
The discussion is intended to connect the details of this question to the bigger picture of sparse approximation using wavelets. There is no need to answer anything.