Lab 4
Sparsity Aware Learning
Advanced Machine Learning
DATA 442/642

Exercise 1

Consider an unknown 2-sparse vector $\boldsymbol{\theta}_0$, which when measured with the sensing matrix

$$\mathbf{X} = \begin{bmatrix} 0.5 & 2 & 1.5 \\ 2 & 2.3 & 3.5 \end{bmatrix}$$

that is, of $\mathbf{y} = \mathbf{X}\boldsymbol{\theta}_0$, gives $\mathbf{y} = [1.25 3.75]^\top$. Perform the following tasks in Python.

(a) Based on the pseudoinverse of $\mathbf{X}$, compute $\hat{\boldsymbol{\theta}}_2$, which is the ℓ_2 norm minimized solution. Next, check that this solution $\hat{\boldsymbol{\theta}}_2$ leads to zero estimation error (up to machine precision). Is $\hat{\boldsymbol{\theta}}_2$ a 2-space vector such as the true unknown vector $\boldsymbol{\theta}_0$?

(b) Solve the ℓ_0 minimization task based on an exhaustive search for all possible 1- and 2-sparse solutions and get the best one, $\hat{\boldsymbol{\theta}}_0$. Does $\hat{\boldsymbol{\theta}}_0$ lead to zero estimation error (up to machine precision)?

(c) Compute and compare the ℓ_2 norms of $\hat{\boldsymbol{\theta}}_2$ and $\hat{\boldsymbol{\theta}}_0$. Which is the smaller one? Was this result expected?

Exercise 2: Image Denoising

A typical example in sparsity aware learning is the denoising problem. The problem in signal denoising is that instead of the actual signal samples, $\tilde{\mathbf{y}}$, a noisy version of the corresponding observations, $\mathbf{y}$, is available; that is, $\mathbf{y} = \tilde{\mathbf{y}} + \boldsymbol{\eta}$, where $\boldsymbol{\eta}$ is the vector of noise samples. Under the sparse modeling framework, the unknown signal $\tilde{\mathbf{y}}$ is modeled as a sparse representation in terms of a specific known dictionary $\boldsymbol{\Psi}$, that is, $\tilde{\mathbf{y}} = \boldsymbol{\Psi}\boldsymbol{\theta}$. Moreover, the dictionary is allowed to be redundant (overcomplete). Then the denoising procedure is realized in two steps. First, an estimate of the sparse representation vector, $\boldsymbol{\theta}$, is obtained via any LASSO formulation, and second, the estimate of the true signal is computed as $\hat{\mathbf{y}} = \boldsymbol{\Psi}\hat{\boldsymbol{\theta}}$.
For Exercise 2, you have to reproduce the the denoising results of the case study in Section 9.10. First, extract from the image all the possible sliding patches of size 12×12. Confirm that $(256 - 12 + 1)^2 = 60,025$ patches in total are obtained. Next, a dictionary in which all the patches are sparsely represented needs to be designed. Specifically, the dictionary atoms are going to be those corresponding to the two-dimensional redundant DCT transform, and are obtained as follows

(a) Consider vectors $\mathbf{d}_i = [d_{i,1}, d_{i,2}, \ldots, d_{i,12}]^\top$, $i = 0, \ldots, 13$, being the sampled sinusoids of the form

$$d_{i,t+1} = \cos(\frac{t\pi i}{14}),\ t = 0, \ldots, 11.$$

Then make (12×14) matrix $\bar{\mathbf{D}}$, having as columns the vectors $\mathbf{d}_i$ normalized to unit norm; $\mathbf{D}$ resembles a redundant DCT matrix.

(b) Construct the $(12^2 \times 14^2)$ dictionary $\boldsymbol{\Psi}$ according to $\boldsymbol{\Psi} = \mathbf{D} \otimes \mathbf{D}$, where $\otimes$ denoted Kronecker product. Built in this way, the resulting atoms correspond to atoms related to the overcomplete 2D-DCT transform.

As a next step, denoise each image patch separately. In particular, assuming that $\mathbf{y}_i$ is the ith patch reshaped in column vector, estimate a sparse vector $\boldsymbol{\theta}_i \in \mathbb{R}^{196}$ and obtain the corresponding denoised vector as $\hat{\mathbf{y}}_i = \boldsymbol{\Psi}\hat{\boldsymbol{\theta}}_i$. Finally, average the values of the overlapping patches in order to form the full denoised image.