Coordinate update methods for LASSO and NMF

(Due on Oct.-18-2019)

Instruction: Each student needs to submit the source code file and a report by Latex. In the report file, you can summarize what you observe or difficulties you have. **You will be evaluated based on correctness of your algorithms, accuracy and efficiency of algorithms, performance on applications, and also the report**. Compress your files into a single .zip file, name it as "MATP6960_Assignment2_YourName", and send it to `optimization.rpi@gmail.com`

1 LASSO

The Lasso problem is formulated as

$$\underset{\mathbf{x}}{\text{minimize}} \ \frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2 + \lambda\|\mathbf{x}\|_1, \tag{1}$$

where $\mathbf{A} \in \mathbb{R}^{m \times n}$ and $\mathbf{b} \in \mathbb{R}^m$ are given, and $\lambda > 0$ is to balance the data fidelity term and the sparsity-promoting term. Under certain conditions on $\mathbf{A}$, a sparse vector can be recovered by solving (1) with an appropriate choice of λ.

Requirements

Include every item below in a single report and attach your code. Use the provided testfile to test your code.

1. Develop a solver for (1) using the cyclic coordinate descent method. The input of the solver should include $\mathbf{A}$, $\mathbf{b}$, λ, and other parameters like the stopping tolerance.

2. Develop a solver for (1) using the randomized coordinate descent method. The input of the solver should include $\mathbf{A}$, $\mathbf{b}$, λ, and other parameters like the stopping tolerance.

3. **[Bonus]:** Develop a solver for (1) using the greedy coordinate descent method. The input of the solver should include $\mathbf{A}$, $\mathbf{b}$, λ, and other parameters like the stopping tolerance.

4. Run the provided test file (add by yourself the greedy CD solver if you code it) and report the recovered vectors.

2 Sum-to-One Nonnegative matrix factorization

The nonnegative matrix factorization (NMF) aims to find two nonnegative matrices $X \in \mathbb{R}^{m\times r}$ and $Y \in \mathbb{R}^{r\times n}$ such that their product XY approximates a given nonnegative matrix $M \in \mathbb{R}^{m\times n}$. Under the assumption of Gaussian noise, it can be modeled as

$$\underset{X,Y}{\text{minimize}}\, \frac{1}{2}\|XY - M\|_F^2 + \phi_1(X) + \phi_2(Y), \text{ s.t. } X \in \mathbb{R}_+^{m\times r}, Y \in \mathbb{R}_+^{r\times n}, \tag{2}$$

where ϕ_1 and ϕ_2 are regularization terms to encourage certain structures of the solution. Because of nonnegativity, NMF can learn local features and has better interpretability than the principle component analysis (PCA), which often learns global features. In this assignment, we let $\phi_1 \equiv 0$ and $\phi_2(Y)$ be the indicator function on the column-sum-to-one set

$$\left\{Y \in \mathbb{R}^{r\times n} : \sum_{i=1}^{r} Y_{ij} = 1, \forall j = 1, \dots, n\right\},$$

i.e., we enforce that every column of Y has unit ℓ_1-norm. Since we also require Y to be nonnegative, each column of Y corresponds to a distribution. This way, we can use Y for clustering.

Requirements

Include every item below in a single report and attach your code. Use the provided datasets to test your code.

1. Develop one solver for (2) by the block coordinate update method. You can choose the alternating minimization method or the alternating proximal gradient method. If you choose the alternating minimization method, you need to develop a subsolver for the subproblem. Directly applying Matlab's nonnegative quadratic program solver is NOT allowed. Your solver should include M and r as the input, and also some other parameters like the stopping tolerance as the input.

2. Run the provided test file with the MNIST data set; report the running time and 10 selected images in each cluster.