Advanced Topics in Machine Learning (600.692)
Homework 5: Robust PCA

Instructor: René Vidal

Due Date: 03/28/2014, 11.59PM Eastern

READING MATERIAL: Chapter 3 and Appendix A of GPCA book.

1. **(Properties of the $\ell_{2,1}$ Norm).**

   (a) Let $\boldsymbol{x}$ be a vector. Show that the sub-gradient of the ℓ_2 norm is given by

   $$\partial\|\boldsymbol{x}\|_2 = \begin{cases} \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|_2} & \text{if } \boldsymbol{x} \neq \mathbf{0} \\ \boldsymbol{w} : \|\boldsymbol{w}\|_2 \leq 1\} & \text{if } \boldsymbol{x} = \mathbf{0} \end{cases} \tag{1}$$

   (b) Let X be a matrix. Show that the $\ell_{2,1}$ norm of X, $f(X) = \|X\|_{2,1} = \sum_j \|X_{\cdot,j}\|_2 = \sum_j \sqrt{\sum_i X_{ij}^2}$, is a convex function of X.

   (c) Show that the sub-gradient of the $\ell_{2,1}$ norm is given by

   $$(\partial\|X\|_{2,1})_{ij} = \begin{cases} \frac{X_{ij}}{\|X_{\cdot,j}\|_2} & X_{\cdot,j} \neq \mathbf{0} \\ W_{ij} : \|W_{\cdot,j}\|_2 \leq 1 & X_{\cdot,j} = \mathbf{0}. \end{cases} \tag{2}$$

   (d) Show that the optimal solution of

   $$\min_A \quad \frac{1}{2}\|X - A\|_F^2 + \tau\|A\|_{2,1} \tag{3}$$

   is given by $A = X\mathcal{S}_\tau(\text{diag}(\boldsymbol{x}))\text{diag}(\boldsymbol{x})^{-1}$, where $\boldsymbol{x}$ be a vector whose j-th entry is given by $x_j = \|X_{\cdot,j}\|_2$, and $\text{diag}(\boldsymbol{x})$ is a diagonal matrix with the entries of $\boldsymbol{x}$ in its diagonal. By convention, if $x_j = 0$, then the j-th entry of $\text{diag}(\boldsymbol{x})^{-1}$ is also zero.

2. Let $X = L_0 + E_0$ be a matrix formed as the sum of a low rank matrix L_0 and a matrix of corruptions E_0, where the corruptions can be either outlying entries (gross errors) or outlying data points (outliers).

   (a) **(PCA with robustness to outliers).** Assuming that the matrix X is fully observed and and that the matrix E_0 is a matrix of outliers, propose an algorithm for solving the following optimization problem

   $$\min_{L,E} \quad \|L\|_* + \lambda\|E\|_{2,1} \quad \text{s.t.} \quad X = L + E. \tag{4}$$

   (b) **(PCA with robustness to missing entries and gross errors).** Assuming that you observe only a fraction of the entries of X as indicated by a set Ω and that the matrix E_0 is a matrix of gross errors, propose an algorithm for solving the following optimization problem

   $$\min_{L,E} \quad \|L\|_* + \lambda\|E\|_1 \quad \text{s.t.} \quad P_\Omega(X) = P_\Omega(L + E). \tag{5}$$

3. **Implementation of Reweighted Least Squares and Robust PCA.** Implement the functions below using as few lines of MATLAB code as possible. Compare the performance of these methods by modifying the sample code test_matrixcompletion.m. Which method works better and which regime (e.g., depending on percentage of corrupted entries (or corrupted data points), subspace dimension d/D)?

Function `[mu,Ud,Y]=rpca_rls(X,d)`

Parameters

`X`	$D \times N$ data matrix.
`d`	Number of principal components.

Returned values

`mu`	Mean of the data.
`Ud`	Basis for the subspace.

Description

Finds the parameters of the PCA model $\boldsymbol{\mu}$ and U_d and the low-dimensional representation using re-weighted least squares with weights $w(e) = \frac{\sigma^2}{e^2+\sigma^2}$.

Function `[L,E]=rpca(X,tau,'method')`

Parameters

`X`	$D \times N$ data matrix.
τ	Parameter of the augmented Lagrangian.
`method`	'L1' for gross errors or 'L21' for outliers

Returned values

`L`	Low-rank completion of the matrix X.
`E`	Matrix of errors.

Description

Solves the optimization problem $\min_{L,E} \|L\|_* + \lambda\|E\|_\ell$ subject to $X = L + E$ where $\ell = \ell_1$ or $\ell = \ell_{2,1}$ using the ADMM algorithm.

Submission instructions. Please follow the same instructions as in HW1.