Advanced Topics in Machine Learning (600.692)
Homework 4: PCA with Incomplete Data

Instructor: René Vidal

Due Date: 03/14/2014, 11.59PM Eastern

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

1. **(Orthogonal Power Iteration Method).** Let $A \in \mathbb{R}^{N \times N}$ be a symmetric positive semidefinite matrix with eigenvectors $\{\boldsymbol{u}_i\}_{i=1}^N$ and eigenvalues $\{\lambda_i\}_{i=1}^N$ sorted in descending order. Assume that that $\lambda_1 > \lambda_2$ and let $\boldsymbol{u}^0$ be an arbitrary vector not orthogonal to $\boldsymbol{u}_1$, i.e., $\boldsymbol{u}_1^\top \boldsymbol{u}^0 \neq 0$. Consider the sequence of vectors

$$\boldsymbol{u}_{k+1} = \frac{A\boldsymbol{u}_k}{\|A\boldsymbol{u}_k\|}. \tag{1}$$

   (a) Show that the exist $\{\alpha_i\}_{i=1}^N$ with $\alpha_1 \neq 0$ such that

$$\boldsymbol{u}^k = A^k \boldsymbol{u}^0 = \sum_{i=1}^N \alpha_i \lambda_i^k \boldsymbol{u}_i. \tag{2}$$

   (b) Use this expression to show that $\boldsymbol{u}^k$ converges to $\frac{\alpha_1}{|\alpha_1|}\boldsymbol{u}_1$ with rate $\frac{\lambda_2}{\lambda_1}$. That is, show that there exists a constant $C > 0$ such that for all $k \geq 0$

$$\left\|\boldsymbol{u}^k - \frac{\alpha_1}{|\alpha_1|}\boldsymbol{u}_1\right\| \leq C\Big(\frac{\lambda_2}{\lambda_1}\Big)^k. \tag{3}$$

   (c) Assume that $\lambda_d > \lambda_{d+1}$ and let $U^0 \in \mathbb{R}^{N \times d}$ be an arbitrary matrix whose column space is not orthogonal to the subspace spanned by the top d eigenvectors of A, $\{\boldsymbol{u}_i\}_{i=1}^d$. Consider the sequence of matrices

$$U^{k+1} = AU^k (R^k)^{-1}, \tag{4}$$

   where $Q^k R^k = AU^k$ is the QR decomposition of AU^k. Show that U^k converges to a matrix U whose columns are the top d eigenvectors of A. Moreover, show that the rate of convergence is $\frac{\lambda_{d+1}}{\lambda_d}$.

2. **(Properties of the ℓ_1 Norm).** Let X be a matrix.

   (a) Show that the ℓ_1 norm of X, $f(X) = \|X\|_1 = \sum_{ij} |X_{ij}|$, is a convex function of X.

   (b) Show that the sub-gradient of the ℓ_1 norm is given by

$$(\partial \|X\|_1)_{ij} = \begin{cases} \text{sign}(X_{ij}) & X_{ij} \neq 0 \\ W_{ij} & X_{ij} = 0, \end{cases} \tag{5}$$

   where $W_{ij} \in [0, 1]$.

   (c) Show that the optimal solution of

$$\min_A \quad \frac{1}{2}\|X - A\|_F^2 + \tau \|A\|_1 \tag{6}$$

   is given by $A = \mathcal{S}_\tau(X)$, where $S_\tau(\cdot)$ is the soft-thresholding operator applied to each entry of X as $\mathcal{S}_\tau(x) = \text{sign}(x)\max(|x| - \tau, 0)$.

3. **(Properties of the Nuclear Norm).** Let X be a matrix of rank r.
   (a) Show that the nuclear norm of X, $f(X) = \|X\|_* = \sum_{i=1}^{r} \sigma_i(X)$, is a convex function of X.
   (b) Show that the sub-gradient of the nuclear norm is given by
   $$\partial \|X\|_* = UV^\top + W \tag{7}$$
   where $X = U\Sigma V^\top$ is the compact (rank r) SVD of X and W is a matrix such that $U^\top W = 0$, $WV = 0$ and $\|W\|_2 \leq 1$.
   (c) Show that the optimal solution of
   $$\min_A \quad \frac{1}{2}\|X - A\|_F^2 + \tau \|A\|_* \tag{8}$$
   is given by $A = \mathcal{D}_\tau(X) = U\mathcal{S}_\tau(\Sigma)V^\top$, where $\mathcal{D}_\tau$ is called the singular-value thresholding operator.

4. **Implementation of Power Factorization (PF), Expectation Maximization (EM) and Low Rank Matrix Completion (LRMC).** 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 missing entries, subspace dimension d/D)?

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

Parameters

`X`	$D \times N$ data matrix.
`d`	Number of principal components.
`W`	$D \times N$ binary matrix denoting known (1) or missing (0) entries

Returned values

`mu`	Mean of the data.
`Ud`	Orthonormal basis for the subspace.
`Y`	Low-dimensional representation (or principal components).

Description

Finds the d principal components of a set of points from the data X with incomplete entries as specified in W using the Power Factorization algorithm.

Function `[mu,Ud,sigma]=mppca(X,d,W)`

Parameters

`X`	$D \times N$ data matrix.
`d`	Number of principal components.
`W`	$D \times N$ binary matrix denoting known (1) or missing (0) entries

Returned values

`mu`	Mean of the data.
`Ud`	Basis for the subspace (does not need to be orthonormal).
`sigma`	Standard deviation of the noise.

Description

Finds the parameters of the PPCA model $\boldsymbol{\mu}$ and $\Sigma = U_d U_d^\top + \sigma^2 I$ from the data X with incomplete entries as specified in W using the Expectation Maximization algorithm.

Function `A=lrmc(X,tau,W)`

Parameters

`X`	$D \times N$ data matrix.
τ	Parameter of the augmented Lagrangian.
`W`	$D \times N$ binary matrix denoting known (1) or missing (0) entries

Returned values

`A`	Low-rank completion of the matrix X.

Description

Finds the low-rank approximation of a matrix X with incomplete entries as specified in W using the Low Rank Matrix Completion Algorithm based on the Augmented Lagrangian Method.

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