Exam 1: Unsupervised Learning (600.692)

Instructor: René Vidal

April 3, 2017

1. **Model Selection for PCA.** Assume you are given a matrix $X \in \mathbb{R}^{D\times N}$ whose columns lie approximately in a low-dimensional subspace of *unknown dimension* $d < D$. Let $X = U_X \Sigma_X V_X^\top$ be the SVD of X. You would like to approximate X by a low-rank matrix $A \in \mathbb{R}^{D\times N}$ and you consider the following optimization problem
$$\min_A \quad \|X - A\|_F^2 + \tau \|A\|_*^2, \tag{1}$$
where $\tau > 0$ is a fixed parameter.
   (a) **(15 points)** Let $A = U_A \Sigma_A V_A^\top$ be the SVD of A. Show that an optimal solution for U_A and V_A with Σ_A held constant is given by $U_A = U_X$ and $V_A = V_X$.
   (b) **(35 points)** Let $\bar{\sigma}_d(X)$ be the average of the top d singular values of X, where d is the largest integer such that $\sigma_d(X) > \frac{\tau d}{1+\tau d}\bar{\sigma}_d(X)$. Show that an optimal solution for A is:
$$A = U_X \mathcal{S}_\mu(\Sigma_X) V_X^\top \qquad \text{where} \qquad \mu = \frac{\tau d}{1+\tau d}\bar{\sigma}_d(X) \tag{2}$$
and $\mathcal{S}_\mu(Y) = \operatorname{argmin}_A \frac{1}{2}\|Y - A\|_F^2 + \mu\|A\|_1$ is the shrinkage thresholding operator.

      Hint: Show that an optimal solution for Σ_A satisfies $(I_d + \tau \mathbf{1}\mathbf{1}^\top)\sigma(A) = \sigma_{1:d}(X)$, where $\sigma(A) \in \mathbb{R}^d$ is the vector of singular values of A (similarly for X). Show also that $(I_d + \tau\mathbf{1}\mathbf{1}^\top)^{-1} = (I_d - \frac{\tau}{1+\tau d}\mathbf{1}\mathbf{1}^\top)$.
   (c) **(5 points)** What is the estimate of the subspace dimension given by the above model selection approach? What is the estimate when $\tau \to \infty$?
   (d) **(5 points)** Discuss the advantages of this model selection approach versus the one discussed in class
$$\min_A \quad \frac{1}{2}\|X - A\|_F^2 + \tau\|A\|_*, \tag{3}$$
where an optimal solution is given by $A = U_X \mathcal{S}_\tau(\Sigma_X) V_X^\top$. **Hint:** consider the case $\tau \to \infty$.

2. **PCA with missing entries and outliers.** Let $L_0 \in \mathbb{R}^{D\times N}$ be a low-rank matrix, i.e., $\operatorname{rank}(L_0) \ll \min\{D, N\}$. Let $E_0 \in \mathbb{R}^{D\times N}$ be a sparse matrix, i.e., its number of nonzero entries is $\|E_0\|_0 \ll ND$. Suppose you are given a subset of the entries of $X = L_0 + E_0 \in \mathbb{R}^{D\times N}$ indexed by a set $\Omega \subseteq \{1,\dots,D\} \times \{1,\dots,N\}$. To recover L_0 and E_0 you consider the optimization problem, where $\tau > 0$ and $\lambda > 0$ are fixed parameters:
$$\min_{L,E} \quad \frac{1}{2}\|L\|_F^2 + \tau\|L\|_* + \frac{\lambda}{2}\|E\|_F^2 + \lambda\tau\|E\|_1 \quad \text{such that} \quad P_\Omega(X) = P_\Omega(L + E). \tag{4}$$
   (a) **(30 points)** Write down the Lagrangian for this problem and use it to give a detailed derivation of the following (dual ascent) algorithm for solving the above problem
$$L^{k+1} = \mathcal{D}_\tau(Z^k) \tag{5}$$
$$E^{k+1} = \mathcal{S}_\tau(Z^k/\lambda) \tag{6}$$
$$Z^{k+1} = Z^k + \delta_k P_\Omega(X - L^{k+1} - E^{k+1}), \tag{7}$$
where $\mathcal{S}_\tau(Y) = \operatorname{argmin}_A \frac{1}{2}\|Y - A\|_F^2 + \tau\|A\|_1$ is the shrinkage thresholding operator, $\mathcal{D}_\tau(Y) = \operatorname{argmin}_A \frac{1}{2}\|Y - A\|_F^2 + \tau\|A\|_*$ is the singular value thresholding operator, $Z \in \mathbb{R}^{D\times N}$ is the matrix of Lagrange multipliers initialized as $Z^0 = \mathbf{0}$, and $\delta_k > 0$ is a sequence of real numbers.
   (b) **(10 points)** What parameter should be increased to make L low rank and make E sparse? Can you guess sufficient conditions on L_0 and E_0 under which $L^* = L_0$ and $E^* = E_0$ with overwhelming probability.