A Mathematical Introduction to Data Science **March 15, 2016**

Homework 3. PCA and MDS

Instructor: Yuan Yao *Due: Tuesday March 22, 2016*

The problem below marked by * is optional with bonus credits.

1. *PCA experiments:* Take any digit data ('0',...,'9'), or all of them, from website

   `http://www-stat.stanford.edu/~tibs/ElemStatLearn/datasets/zip.digits/`

   and perform PCA experiments with Matlab or other language you are familiar:

   (a) Set up data matrix $X = (x_1, \ldots, x_n) \in \mathbb{R}^{p \times n}$;

   (b) Compute the sample mean $\hat{\mu}_n$ and form $\tilde{X} = X - e\hat{\mu}_n^T$;

   (c) Compute top k SVD of $\tilde{X} = US_kV^T$;

   (d) Plot eigenvalue curve, i.e. i vs. $\lambda_i(\hat{\Sigma}_n)/tr(\hat{\Sigma}_n)$ $(i = 1, \ldots, k)$, with top-k eigenvalue λ_i for sample covariance matrix $\hat{\Sigma}_n = \frac{1}{n}\tilde{X} * \tilde{X}^T$, which gives you explained variation of data by principal components;

   (e) Use `imshow` to visualize the mean and top-k principle components as *left* singular vectors $U = [u_1, \ldots, u_k]$;

   (f) For $k = 1$, sort the image data (x_i) $(i = 1, \ldots, n)$ according to the top *right* singular vectors, v_1, in an ascending order;

   (g) For $k = 2$, scatter plot (v_1, v_2) and select a grid on such a plane to show those images on the grid (e.g. Figure 14.23 in book [ESL]: Elements of Statistical Learning).

   (h)* You may try the parallel analysis with permutation test to see how many significant principle components you will obtain.

2. *MDS of cities:* Go to the following website

   `http://www.geobytes.com/citydistancetool.htm`

   Perform the following experiment.

   (a) Input a few cities (no less than 7) in your favorite, and collect the pairwise *air traveling* distances shown on the website in to a matrix D;

   (b) Make your own codes of Multidimensional Scaling algorithm for D;

   (c) Plot the normalized eigenvalues $\lambda_i/(\sum_i \lambda_i)$ in a descending order of magnitudes, analyze your observations (did you see any negative eigenvalues? if yes, why?);

   (d) Make a scatter plot of those cities using top 2 or 3 eigenvectors, and analyze your observations.

3. *Positive Semi-definiteness:* Recall that a n-by-n real symmetric matrix K is called positive semi-definite (*p.s.d.* or $K \succeq 0$) iff for every $x \in \mathbb{R}^n$, $x^T K x \geq 0$.

   (a) Show that $K \succeq 0$ if and only if its eigenvalues are all nonnegative.

   (b) Show that $d_{ij} = K_{ii} + K_{jj} - 2K_{ij}$ is a squared distance function, *i.e.* there exists vectors $u_i, v_j \in \mathbb{R}^n$ ($1 \leq i, j \leq n$) such that $d_{ij} = \|u_i - u_j\|^2$.

   (c) Let $\alpha \in \mathbb{R}^n$ be a signed measure s.t. $\sum_i \alpha_i = 1$ (or $e^T \alpha = 1$) and $H_\alpha = I - e\alpha^T$ be the Householder centering matrix. Show that $B_\alpha = -\frac{1}{2} H_\alpha D H_\alpha^T \succeq 0$.

   (d) If $A \succeq 0$ and $B \succeq 0$ ($A, B \in \mathbb{R}^{n \times n}$), show that $A + B = [A_{ij} + B_{ij}]_{ij} \succeq 0$ (elementwise sum), and $A \circ B = [A_{ij} B_{ij}]_{ij} \succeq 0$ (Hadamard product or elementwise product).

4. **Singular Value Decomposition:* The goal of this exercise is to refresh your memory about the singular value decomposition and matrix norms. A good reference to the singular value decomposition is Chapter 2 in this book:
   Matrix Computations, Golub and Van Loan, 3rd edition.
   Parts of the book are available online here:
   `http://www.math.pku.edu.cn/teachers/yaoy/reference/golub.pdf`

   (a) *Existence:* Prove the existence of the singular value decomposition. That is, show that if A is an $m \times n$ real valued matrix, then $A = U \Sigma V^T$, where U is $m \times m$ orthogonal matrix, V is $n \times n$ orthogonal matrix, and $\Sigma = \text{diag}(\sigma_1, \sigma_2, \ldots, \sigma_p)$ (where $p = \min\{m, n\}$) is an $m \times n$ diagonal matrix. It is customary to order the singular values in decreasing order: $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_p \geq 0$. Determine to what extent the SVD is unique. (See Theorem 2.5.2, page 70 in Golub and Van Loan).

   (b) *Best rank-k approximation - operator norm:* Prove that the "best" rank-k approximation of a matrix in the operator norm sense is given by its SVD. That is, if $A = U \Sigma V^T$ is the SVD of A, then $A_k = U \Sigma_k V^T$ (where $\Sigma_k = \text{diag}(\sigma_1, \sigma_2, \ldots, \sigma_k, 0, \ldots, 0)$ is a diagonal matrix containing the largest k singular values) is a rank-k matrix that satisfies

   $$\|A - A_k\| = \min_{\text{rank}(B)=k} \|A - B\|.$$

   (Recall that the operator norm of A is $\|A\| = \max_{\|x\|=1} \|Ax\|$. See Theorem 2.5.3 (page 72) in Golub and Van Loan).

   (c) *Best rank-k approximation - Frobenius norm:* Show that the SVD also provides the best rank-k approximation for the Frobenius norm, that is, $A_k = U \Sigma_k V^T$ satisfies

   $$\|A - A_k\|_F = \min_{\text{rank}(B)=k} \|A - B\|_F.$$