Homework 1. PCA and MDS

Instructor: Yuan Yao   *Due: 1.5 weeks later*

The problem below marked by * is optional with bonus credits. For the experimental problem, include the source codes which are runnable under standard settings. Since there is NO TA assigned for this class, homework will not be graded. But if you would like to submit your exercise, please send your homework to the address (datascience.hw@gmail.com) with a title "CSIC5011: Homework #". I'll read them and give you 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 languages (e.g. ipynb) 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 $\tilde{X} = U S_k V^T$;

   (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 $k = 1$, order the images (x_i) $(i = 1, \ldots, n)$ according to the top first *right* singular vector, v_1, in an ascending order of $v_1(i)$;

   (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 $B_\alpha = -\frac{1}{2} H_\alpha D H_\alpha^T \succeq 0$ for matrix $D = [d_{ij}]$.

    (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. *Distance:* Suppose that $d : \mathbb{R}^p \times \mathbb{R}^p \to \mathbb{R}$ is a distance function.

    (a) Is d^2 a distance function? Prove or give a counter example.

    (b) Is $\sqrt{d}$ a distance function? Prove or give a counter example.

5. **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.

    (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

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

(d) *Schatten p-norms:* A matrix norm $\|\cdot\|$ that satisfies

$$\|QAZ\| = \|A\|,$$

for all Q and Z orthogonal matrices is called a unitarily invariant norm. The Schatten p-norm of a matrix A is given by the ℓ_p norm ($p \geq 1$) of its vector of singular values, namely,

$$\|A\|_p = \left(\sum_i \sigma_i^p \right)^{1/p}.$$

Show that the Schatten p-norm is unitarily invariant. Note that the case $p = 1$ is sometimes called the nuclear norm of the matrix, the case $p = 2$ is the Frobenius norm, and $p = \infty$ is the operator norm.

(e) *Best rank-k approximation for unitarily invariant norms:* Show that the SVD provides the best rank-k approximation for any unitarily invariant norm. See also 7.4.51 and 7.4.52 in:
Matrix Analysis, Horn and Johnson, Cambridge University Press, 1985.

(f) *Closest rotation:* Given a square $n \times n$ matrix A whose SVD is $A = U\Sigma V^T$, show that its closest (in the Frobenius norm) orthogonal matrix R (satisfying $RR^T = R^T R = I$) is given by $R = UV^T$. That is, show that

$$\|A - UV^T\|_F = \min_{RR^T = R^T R = I} \|A - R\|_F,$$

where $A = U\Sigma V^T$. In other words, R is obtained from the SVD of A by dropping the diagonal matrix Σ. Use this observation to conclude what is the optimal rotation that aligns two sets of points $p_1, p_2, \ldots, p_n$ and $q_1, \ldots, q_n$ in $\mathbb{R}^d$, that is, find R that minimizes $\sum_{i=1}^n \|Rp_i - q_i\|^2$. See also (the papers are posted on course website):

- [Arun87] Arun, K. S., Huang, T. S., and Blostein, S. D., "Least-squares fitting of two 3-D point sets", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **9** (5), pp. 698–700, 1987.

- [Keller75] Keller, J. B., "Closest Unitary, Orthogonal and Hermitian Operators to a Given Operator", *Mathematics Magazine*, **48** (4), pp. 192–197, 1975.

- [FanHoffman55] Fan, K. and Hoffman, A. J., "Some Metric Inequalities in the Space of Matrices", *Proceedings of the American Mathematical Society*, **6** (1), pp. 111–116, 1955.