FINM 331: MULTIVARIATE DATA ANALYSIS
FALL 2021
PROBLEM SET 1

The required files for all problems can be found in the subfolder hw1 under ‘Files’ in Canvas or at
the following URL:
http://www.stat.uchicago.edu/~lekheng/courses/331/hw1/
The file name indicates which problem the file is for (p1*.txt for Problem 1, etc). You are welcomed
to use any programming language or software packages you like.

1. Verify our claim in the lecture that the condensed svd of a matrix $A \in \mathbb{R}^{n \times p}$ may be expressed
   as
   $$A = \sigma_1 u_1 v_1^T + \cdots + \sigma_r u_r v_r^T,$$
   where $r = \text{rank}(A)$, $u_1, \ldots, u_r \in \mathbb{R}^n$ and $v_1, \ldots, v_r \in \mathbb{R}^p$ are left and right singular vectors of
   $A$. Find a similar expression for the evd of a symmetric matrix $B \in \mathbb{R}^{p \times p}$.

2. Let $A, B \in \mathbb{R}^{p \times p}$ and $O$ be the zero matrix. For each of the following statement, either give a
   proof or a counterexample:
   (a) If all eigenvalues of $A$ are zero, then $A = O$.
   (b) If all singular values of $A$ are zero, then $A = O$.
   (c) If $A = A^T$, then the evd and svd of $A$ are identical.
   (d) If $A$ and $B$ are similar (i.e., $A = XBX^{-1}$ for some nonsingular $X$), then $A$ and $B$ have the
       same eigenvalues.
   (e) If $A$ and $B$ are similar, then $A$ and $B$ have the same singular values.

3. Let $y_1, \ldots, y_n \in \mathbb{R}^n$. Let $p \leq n$ and $G_p \in \mathbb{R}^{p \times p}$ be the matrix

   $$G_p = \begin{bmatrix}
   y_1^T y_1 & y_1^T y_2 & \cdots & y_1^T y_p \\
   y_2^T y_1 & y_2^T y_2 & \cdots & y_2^T y_p \\
   \vdots & \vdots & \ddots & \vdots \\
   y_p^T y_1 & y_p^T y_2 & \cdots & y_p^T y_p
   \end{bmatrix}.$$

   This is called a Gram matrix or more precisely the Gram matrix of $y_1, \ldots, y_p$.
   (a) Show that $y_1, \ldots, y_p$ are linearly independent $\iff$ $\text{rank}(G_p) = p$.
   (b) Show that $y_1, \ldots, y_p$ are orthonormal $\iff$ $G_p = I_p$.
   Here $I_p \in \mathbb{R}^{p \times p}$ is the $p \times p$ identity matrix.
   (c) Suppose $G_p = I_p$. Let $P_p \in \mathbb{R}^{n \times n}$ be the orthogonal projection matrix for the subspace
       $\text{span}\{y_1, \ldots, y_p\} \subseteq \mathbb{R}^n$.
       What is the relation between $G_p \in \mathbb{R}^{p \times p}$ and $P_p \in \mathbb{R}^{n \times n}$ in terms of the matrix
       $Q_p := [y_1, \ldots, y_p] \in \mathbb{R}^{n \times p}$?
   (d) Show that if $G_p = I_p$, then for any $y \in \mathbb{R}^n$,

       $$\sum_{i=1}^p (y^T y_i)^2 \leq \|y\|_2^2.$$

       Give an example to show that strict inequality can happen.

Date: October 11, 2021 (Version 1.0); due: October 24, 2021.
(e) Show that if $G_n = I_n$, then for any $y \in \mathbb{R}^n$, 
\[
\sum_{i=1}^n (y^T y_i)^2 = \|y\|_2^2.
\]

(f) Show that if $G_n = I_n$, then for any $y \in \mathbb{R}^n$, 
\[
\sum_{i=1}^n (y^T y_i)y_i = y.
\]

4. You should do this problem ‘by hand’, i.e., without relying on any computer program. Answers
must be justified with mathematical arguments — how you get from one step to the next; there will be no credit for just plugging the matrices into MATLAB or Mathematica or R and reproducing the answers.

(a) Find the singular value decomposition of the following matrices
\[
A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} \in \mathbb{R}^{3 \times 2}, \quad B = \begin{bmatrix} 5 & -1 \\ -1 & 5 \\ 2 & 2 \end{bmatrix} \in \mathbb{R}^{3 \times 2}.
\]

(b) Show that the left singular vectors of $A$ and $B$ give orthonormal bases for the subspaces
\[
W_A := \text{span}\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \right\}, \quad W_B := \text{span}\left\{ \begin{bmatrix} 5 \\ -1 \\ 2 \end{bmatrix}, \begin{bmatrix} -1 \\ 5 \\ 2 \end{bmatrix} \right\}
\]
respectively.

(c) Find the formulas for projecting a vector
\[
x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \in \mathbb{R}^3
\]
on to the subspaces $W_A$ and $W_B$.

(d) Find the orthogonal projection matrices $P_A, P_B \in \mathbb{R}^{3 \times 3}$ corresponding to the two projections above.

5. Let $W$ be a $k$-dimensional subspace of $\mathbb{R}^p$. Suppose $a_1, \ldots, a_k \in \mathbb{R}^p$ form a basis for $W$. Find an orthogonal projection matrix for each of the following subspaces:

(a) $W_a = \text{span}\{[1, \ldots, 1]^T\}$. Discuss how this is related to the mean
\[
\bar{x} = \frac{1}{p} \sum_{i=1}^p x_i.
\]

(b) $W_b = \{[x_1, \ldots, x_p]^T \in \mathbb{R}^p : x_1 + \cdots + x_p = 0 \}$. Discuss how this is related to the deviations
\[
x_i - \bar{x}, \quad i = 1, \ldots, p.
\]

(c) $W_c = \text{span}\{w\}$ for a nonzero vector $w \in \mathbb{R}^p$.

(d) $W_d = W_1 \perp W_2$ where $W_1 \perp W_2$, i.e., $x^T y = 0$ for all $x \in W_1, y \in W_2$.

(e) $W_e = E_\lambda = \{x : Bx = \lambda x\}$, the $\lambda$-eigenspace of a symmetric matrix $B \in \mathbb{R}^{p \times p}$.

(f) $W_f = \text{span}\{[1, 1, 1, 1]^T, [1, 1, 0, 0]^T, [1, 1, 1, 0]^T\}$. Here $p = 4$.

6. Let $W$ be a subspace of $\mathbb{R}^p$ and $W^\perp := \{x \in \mathbb{R}^p : x^T y = 0 \text{ for all } y \in W\}$ be its orthogonal complement. Since $\mathbb{R}^p = W \oplus W^\perp$, we may define $P : \mathbb{R}^p \to \mathbb{R}^p$ by $Pv = w$ where $v = w + w'$ with $w \in W$ and $w' \in W^\perp$. We will see that this gives another way to define projection onto $W$.

(a) Show that $P$ is an orthogonal projection matrix and $\text{im}(P) = W$.

(b) Show that such a $P$ is uniquely determined by $W$.  

(c) Show that for every \( \mathbf{v} \in \mathbb{R}^p \), \( \mathbf{v}^T \mathbf{P} \mathbf{v} \geq 0 \).

(d) Show that for every \( \mathbf{v} \in \mathbb{R}^p \), \( \| \mathbf{P} \mathbf{v} \|_2 \leq \| \mathbf{v} \|_2 \).

(e) Show that \( \mathbf{I} - \mathbf{P} \) is the orthogonal projection onto \( \mathbb{W}^\perp \).

(f) Show that for every \( \mathbf{v} \in \mathbb{R}^p \),
\[
\| \mathbf{v} \|_2^2 = \| \mathbf{P} \mathbf{v} \|_2^2 + \| (\mathbf{I} - \mathbf{P}) \mathbf{v} \|_2^2.
\]

(g) Show that \( \mathbf{P} \) is similar to a diagonal matrix of the form \( \text{diag}(1, \ldots, 1, 0, \ldots, 0) \) where the number of 1’s equals \( \dim \mathbb{W} \).

7. Let \( \mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{R}^p \), \( n \geq p \), and set
\[
\mathbf{X} := \begin{bmatrix}
\mathbf{x}_1^T \\
\mathbf{x}_2^T \\
\vdots \\
\mathbf{x}_n^T
\end{bmatrix} = \begin{bmatrix}
x_{11} & x_{12} & \cdots & x_{1p} \\
x_{21} & x_{22} & \cdots & x_{2p} \\
\vdots & \vdots & \ddots & \vdots \\
x_{n1} & x_{n2} & \cdots & x_{np}
\end{bmatrix} = [\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_p] \in \mathbb{R}^{n \times p},
\]
i.e., the row vectors of \( \mathbf{X} \) are \( \mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{R}^p \) and the column vectors of \( \mathbf{X} \) are \( \mathbf{y}_1, \ldots, \mathbf{y}_p \in \mathbb{R}^n \). We will assume throughout this problem that
\[
\mathbf{X}^T \mathbf{1} = \mathbf{0}.
\]

(a) What is the relation between the sample covariance matrix \( \mathbf{S} \) and the Gram matrix \( \mathbf{G}_p \) as defined in Problem 3?

(b) Let the EVP of \( \mathbf{S} \) be
\[
\mathbf{S} = \mathbf{V} \Lambda \mathbf{V}^T,
\]
where \( \mathbf{V}^T \mathbf{V} = \mathbf{I} \) and \( \Lambda = \text{diag}(\lambda_1, \ldots, \lambda_p) \) with \( \lambda_1 \geq \cdots \geq \lambda_p \). Let \( \mathbf{G}_p \) be the Gram matrix of \( \mathbf{y}_1, \ldots, \mathbf{y}_p \) as in Problem 3. Show that the eigenvectors of \( \mathbf{S} \), the eigenvectors of \( \mathbf{G}_p \), and the right singular vectors of \( \mathbf{X} \) are all the same. How are the eigenvalues of \( \mathbf{S} \), the eigenvalues of \( \mathbf{G}_p \), and the singular values of \( \mathbf{X} \) related?

(c) Write down an expression for \( \mathbf{P}_W \in \mathbb{R}^{p \times p} \), the orthogonal projection onto the 2-dimensional subspace
\[
W := \text{span}\{\mathbf{v}_j, \mathbf{v}_k\} \subseteq \mathbb{R}^p.
\]
Simplify your expression as much as possible.

(d) Show that to plot the projections of \( \mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{R}^p \) onto \( \mathbb{W} \cong \mathbb{R}^2 \), we may simply plot the \( n \) points
\[
\{ (\mathbf{u}_{ij}, \mathbf{u}_{ik}) \in \mathbb{R}^2 : i = 1, \ldots, n \}
\]
where \( \mathbf{U} = [\mathbf{u}_{ij}] \in \mathbb{R}^{n \times n} \) and \( \Sigma = \text{diag}(\sigma_1, \ldots, \sigma_p) \in \mathbb{R}^{n \times p} \) are the matrix of left singular vectors and matrix of singular values respectively.

(e) Explain what we meant by \( \mathbb{W} \cong \mathbb{R}^2 \) and why the plot in (d) may be interpreted as on a graph whose \( x \)-axis is \( \mathbf{v}_j \) and \( y \)-axis is \( \mathbf{v}_k \).

8. The file p8.txt contains an image with 359 pixels by 371 pixels of gray-scale values 1, 2, \ldots, 64 stored in the form of comma separated values (csv). Read the file and store its values as a matrix \( \mathbf{X} \in \mathbb{R}^{359 \times 371} \).

(a) Compute the singular value decomposition of \( \mathbf{X} \) and plot its singular value profile on a semilog scale, i.e., plot the graph
\[
\{ (i, \log \sigma_i) \in \mathbb{R}^2 : i = 1, \ldots, 359 \}.
\]
Why did we use the log scale on the vertical axis? What if we had instead plotted
\[
\{ (i, \sigma_i) \in \mathbb{R}^2 : i = 1, \ldots, 359 \}?
\]
(b) Find \( X_r \in \mathbb{R}^{359 \times 371} \), the best rank-\( r \) approximation of \( X \), for \( r = 1, 20, 50, 100 \). Your solution should show \( X_1, X_{20}, X_{50}, X_{100} \) in the form of images (do not submit them as matrices of numerical values) alongside with the image of \( X \). Comment on the quality of \( X_1, X_{20}, X_{50}, X_{100} \) relative to the original \( X \).

9. The files \texttt{p9X.csv} and \texttt{p9Y.csv} contain entries of two matrices, \( X, Y \in \mathbb{R}^{1000 \times 2} \) respectively. Each row of them represents a point in \( \mathbb{R}^2 \).

(a) Visualize \( X \) and \( Y \) in a single plot with different colors for points in \( X \) and \( Y \).

(b) Write a program that does orthogonal Procrustes analysis, i.e., given two matrices \( X, Y \in \mathbb{R}^{n \times p} \), your program should compute the orthogonal matrix \( Q \in \mathbb{R}^{p \times p} \) that solves

\[
\min_{Q^TQ=I} \|X - YQ\|_F.
\]

You are free to use any programming language as well as packages/functions to compute \texttt{svd}. Test your code on the \( 4 \times 2 \) example in the lecture notes.

(c) Use the function you wrote to perform orthogonal Procrustes on \( X \) and \( Y \). That is we would like to rotate \( Y \) to be as close to \( X \) as possible. Visualize your \( X \) and \( YQ \) in a single plot with different colors as in (a).

(d) In your plot for (c), do the two matrices look similar? If so, report your error \( \|X - YQ\|_F \). If not, how can you improve your algorithm? What is the corresponding distance between the transformed \( Y \) and \( X \) in your improved algorithm?