Math524 Final Exam by Dr. Samuel Shen, FA2023

Due 11:59 PM Thursday/December 14

This is an open-book exam. You can use Google or ChatGPT to search for any helpful information. You can use a calculator, R, computer, or experiment. However, you CANNOT ask natural person to help you. When done, please submit **a single pdf file and a single R file** *online via Canvas.*

Student Name:

Student ID Number:

1. [20 $points$] **SVD for a space-time data matrix**
(a) [10 points] Use R to read the `NOAAGlobalT.csv` data file, and Select the four grid boxes that cover the following four locations: Darwin (12.46S, 130.84E, Australia), Tahiti (17.65S, 149.43W, French Polynesia), Hanga Roa (27.15S, 109.43W, Chile), and Quito (0.18S, 78.47W, Ecuador). Generate a 4×30 space-time data matrix for the December mean surface temperature anomaly data of the four selected grid boxes and the 30 years from 1981 to 2010. Display the first 5 columns of the space-time data matrix. Please note that the longitude in the data file is from 0 to 360. The longitude for a location here is 0 to 180E and 0 to 180W. You need to learn how to convert from one system to another.
Hint: You can download the NOAA Global Surface Temperature (NOAAGlobalT.csv) dataset from this exam site on Canvas. lease note that the longitude in the data file is from 0 to 360. The longitude for a location here is 0 to 180E and 0 to 180W. You need to learn how to convert from one system to another.
(b) [5 points] Make an SVD analysis of the space-time matrix in Part (a). Display the four singular values in the D matrix.
(c) [5 points] Plot the curve of the second principal component, i.e., the second column of the V matrix in the SVD decomposition as a function of time from 1981 to 2010.

2. [20 $points$] **SVD for a time-time matrix**
(a) [6 points] The data file `EarthTemperatureData.csv` contains the global average monthly and annual temperature anomalies from 1880 to 2015. The data file can be downloaded from the exam site on Canvas. Extract the monthly data from this data matrix (i.e., from Column 2 for January to Column 13 for December) for 50 years from 1961 to 2010.
(b) [6 points] Make an SVD analysis for this time-time matrix. The row time is year and the column time indicates seasonal change. Output and display the 12 singular values, i.e., the diagonal elements of the SVD result matrix D.
(c) [4 points] Plot the curve for the first vector of the SVD result matrix U, i.e., the first column of U, as a function of years from 1961 to 2010.
(d) [4 points] Plot the curve for the first vector of the SVD result matrix V, i.e., the first column of V, as a function of months from January to December.

3. [20 $points$] **SVD for a space-space matrix**
(a) [6 points] Use R package `imager` to read Sam's photo file `sam.png`. The figure can be downloaded from the exam site on Canvans.

```
#install.packages('imager')
library(imager)
setwd('/Users/sshen/????')
dat <- load.image('sam.png')
graydat = grayscale(dat)[,,1,1]
dim(graydat)
#[1] 430 460
#So graydat is a 430-by-460 space-space data matrix
```

(b) [6 points] Make an SVD analysis for the above space-space matrix `graydat`. Plot the first 80 singular values against the mode numbers 1 to 80, i.e., plot the SVD singular values $d_1, d_2, \cdots, c_{80}$

against the horizontal axis from 1 to 80.
(c) [4 points] Plot the curve for the first singular vector of the SVD matrix U, i.e., the first column of U, as a function of space points from 1 to 430.
(d) [4 points] Plot the curve for the first singular vector of the SVD matrix V, i.e., the first column of V, as a function of space points from 1 to 460.

4. $[8\ points]$ **Solving linear equations**
Use R to solve the following linear equations $Ax = b$ with A and b given as follows.

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 0 & 9 \end{bmatrix}, \tag{0.1}$$

$$b = \begin{bmatrix} 2 \\ 5 \\ 0 \end{bmatrix}, \tag{0.2}$$

5. $[12\ points]$ Concept problems.
(a) [6 points] Can five 3-dimensional vectors be linearly independent from each other? Justify your answer with a mathematical proof for the general case and an example for a special case.

(b) [6 points] Given that

$$A = \begin{bmatrix} 4 & 1 \\ 1 & 3 \end{bmatrix} \tag{0.3}$$

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \tag{0.4}$$

(i) write down by hand the second-order polynomial of x_1 and x_2, according to the following matrix expression

$$P(x_1, x_2) = \mathbf{x}^t A A^t \mathbf{x}; \tag{0.5}$$

(ii) Use R to compute the eigenvectors and eigenvalues of A; and
(iii) Let (x_1, x_2) be equal to the first eigenvector of A computed in (ii), use R to compute the numerical value of $P(x_1, x_2)$.

6. $[20\ points]$ **Machine learning**
(a) [7 points] Based on the K-means principle of minimal tWCSS described in Section 3.1 of our textbook, use both hand-calculation and R to determine the two clusters from the following three points:

$$P_1(1,1), \quad P_2(2,1), \quad P_3(3,3). \tag{0.6}$$

(b) [3 points] Use R to plot the three points and the two centers similar to Fig. 3.1 in the textbook.
(c) [7 points] Use R to plot the two hyperplanes and the separating plane for the maximum separation between the two clusters in (a). Use Fig. 3.6 in the textbook as a reference.
(d) [3 points] What are the supporting vectors in (c)? What is the distance D_m between the positive hyperplane and the negative hyperplane?