CS6015: Linear Algebra and Random Processes

Programming Assignment 1 **Deadline: 7^{th} October 2019, 23:55 hrs**

Instructions:

1. You have to turn in the well-documented code along with a detailed report of the results of the experiments.
2. **Any sort of plagiarism/cheating will be dealt very strictly. Acknowledge any source used for performing the experiments.**
3. Plot your data and analyze before proceeding.
4. Be precise with your explanations. Avoid verbosity. Place relevant results that bolster your conclusions. Report should be **within 10 pages** in a single column format, and with 11pt font. Results/conclusions beyond the page limit will be ignored during evaluation.
5. You can use any programming language for this assignment. However, we recommend Python or MATLAB. You can use the inbuilt function to find eigenvalues and eigenvectors. **Use of any other inbuilt functions like svd, pca, etc. are not allowed.**
6. Create a folder named "TeamNumber_TeamMember1RollNo_TeamMember2RollNo' (for e.g. "1_CS17S011_CS17S016). In this folder, you should have your report and a subfolder "codes" which should have all your codes. Upload this folder(.zip) on Moodle. Please follow the naming convention strictly.
7. Please make only one submission for the team. No emailed reports will be accepted.

Datasets for each team can be found here. Use your smail account to access the link. The shared folder also has sample plots for your reference.

1 Dimensionality reduction

Perform dimension reduction on two datasets provided in the link mentioned above.

1. For the 2-dimensional dataset: (2+2+2+5 marks)
   (a) Plot the dataset and the singular vectors.
   (b) Project the data points onto the top singular vector and plot the original as well as projected data.

(c) What is the percentage of information lost by reducing dimension? Calculate the error between projected and original data points, i.e., $\sum_{i=1}^{N} \|x_i - \hat{x_i}\|_2^2/N$, where x_i is the original data point, $\hat{x_i}$ is the projected data point and N is the total number of data points.
(*Hint: The information lost by reducing dimension is the sum of singular values of all the singular vectors you are neglecting.*)

(d) Consider the given dataset as $\{(x_i, y_i)\}_{i=1}^{N}$ and find the least squares solution. Plot the dataset, the least squares approximation and the top singular vector in one figure. Is the least squares line same as the top singular vector? Compare the least squares solution error with the dimensionality reduction error (see part(c)). Provide detailed inferences using the results.

2. For the 10-dimensional dataset, do the following: (2+2 marks)

   (a) Project the data points onto the top two singular vectors, then plot the data in this reduced dimensional space.

   (b) Is the information lost by reducing to 2-dimension less than 10%? If not, how many singular vectors are required to capture 90% of the information?

2 Image compression

(i) Perform singular value decomposition on the square image by converting it to grayscale. (2 marks)

(ii) Perform eigen value decomposition on the rectangular image by converting it to grayscale. (*Hint: Use $A^T A$, where A is a rectangular matrix*) (2 marks)

The report should include the following for both experiments: (6+3+2 marks)

(a) Reconstruct the image using top N (N = 10%, 25%, 50%) singular/eigen vectors. Plot the reconstructed images along with their corresponding error image. How good is the quality of the reconstructed image?
(*Hint: Use Frobenius norm to measure the quality of the reconstructed image.*)

(b) Try random N (N = 10%) singular vectors instead of top N. Plot the reconstructed image along with its error image. Explain the observed trend in the result.

(c) Plot a comparative graph of the reconstruction error vs N.

3 Save the water

Consider N^2 taps arranged in a $N \times N$ square, i.e., N rows of taps with N taps in each row. Every tap has two states: open, closed. Whenever we toggle the state of a tap, due to malfunction in the water pipeline, the state of taps in the adjacent also toggles, i.e., the tap to the right, to the left, the tap above and the tap below. Initially, all of the taps are open. The goal is to close all the taps.
(*Hint: Since there are only two states, use module 2 arithmetic.*)

Example: For N = 2, the solution sequence is: 1,2,3,4.

$$\begin{array}{|c|c|}\hline 1 & 1 \\ \hline 1 & 1 \\ \hline\end{array} \rightarrow \begin{array}{|c|c|}\hline 0 & 0 \\ \hline 0 & 1 \\ \hline\end{array} \rightarrow \begin{array}{|c|c|}\hline 1 & 1 \\ \hline 0 & 0 \\ \hline\end{array} \rightarrow \begin{array}{|c|c|}\hline 0 & 1 \\ \hline 1 & 1 \\ \hline\end{array} \rightarrow \begin{array}{|c|c|}\hline 0 & 0 \\ \hline 0 & 0 \\ \hline\end{array}$$

Consider $N = 4$ and answer the following with proper explanation: (5+4+1 marks)

1. Formulate the problem in $Ax = b$ form.
2. Find the solution sequence.
3. Is the solution unique? If not, how many solutions are there?

4 Ciphertext

Consider encryption of a message M using the following two matrices:

$$X_1 = \begin{bmatrix} 1 & 2 & 6 & 3 \\ 12 & 1 & 2 & 4 \\ 1 & 10 & 11 & 30 \\ 13 & 24 & 2 & 13 \end{bmatrix}, \qquad X_2 = \begin{bmatrix} 1 & 2 & 6 & 3 \\ 2 & 4 & 12 & 4 \\ 1 & 2 & 6 & 30 \\ 7 & 26 & 42 & 13 \end{bmatrix}$$

To encode a message M, first convert it to a vector $m_{ascii} = [m_1, m_2, \ldots, m_l]^T$, where l is the length of message and m_i contains ascii code of the i^{th} character in the message, then the encoding of a message M is $X * m_{ascii}$.

Example: Encoding of a message $M =$"LARP" obtained using X_1 is "938 1461 4028 3752" and using X_2 is "938 1716 3098 6706".

Answer the following: (2+3 marks)

1. What is encoding of the message "LINEAR" using both X_1 and X_2?
2. Decode "927 1345 4006 3913" and "927 1445 3811 3665 708 1081 1778" using both X_1 and X_2?