Linear Algebra for Data Science **Name:** ______________________

Fall 2020

Final Exam

10/26/2020

Time Limit: 2 hours (part 1 + part 2)

Instructor: Dr. Anahita Zarei

Grade Table (for instructor use only)

Question	Points	Score
1	5	
2	5	
3	15	
4	20	
Total:	45	

GOOD LUCK!

---

1. (5 points) Matrix A has 3 rows and 2 columns. It has 2 singular values: $\sigma_1 = 4$ and $\sigma_2 = 0$

   (a) (2.5 points) Determine the dimension of U, Σ, and V if $U\Sigma V^T$ denotes the FULL singular value decomposition of A.

   (b) (2.5 points) Determine the dimension of U, Σ, and V if $U\Sigma V^T$ denotes the ECONOMY singular value decomposition of A.

2. (5 points) Consider the system of linear equations $Ax = b$. The singular values of A include $\sigma = \begin{bmatrix} 8200 & 630 & 41 \end{bmatrix}$. Considering that R is accurate to about 16 significant digits, how accurate do you predict your solution x to be? Must show work.

3. (15 points) Suppose matrix X is a 4 by 2 matrix that contains 2 measured features on 4 samples. The singular value decomposition of X returns the following values:
$$u = \begin{bmatrix} -0.6 & 0.4 \\ -0.3 & -0.3 \\ 0.6 & 0.6 \\ 0.3 & -0.7 \end{bmatrix}, \sigma = \begin{bmatrix} 6.7 & 0 \\ 0 & 3.0 \end{bmatrix}, v = \begin{bmatrix} 0.8 & -0.6 \\ 0.6 & 0.8 \end{bmatrix}.$$

   (a) (5 points) What is the direction of the first 2 principal components?

   (b) (5 points) What is the total variance in the data?

   (c) (5 points) Assume that you transformed the data points in the direction of 2 largest variances. What are the coordinates of the data points in the new space?

You need to do the following problem in R and submit your Rmd and html file on Canvas. Please don't write anything on this paper for this problem.

4. (20 points) In this problem, you will use singular value decomposition to compress an image A. Recall that any matrix A with r nonzero singular values can be expressed as $A_{m,n} = U_{m,r} D_{r,r} V^T_{r,n}$. Alternatively, we can express A in its spectral decomposition as $A = \sigma_1 u_1 v_1^T + ... + \sigma_r u_r v_r^T$. We can compress A by truncating its SVD and considering only the first k terms in the spectral decomposition expression. We will denote the compressed image as A_k.

   (a) (5 points) The pixel values for image A can be found in "imageSVD.csv" posted on Canvas under Final Exam. (Note that you don't need to reshape A.) Determine minimum number of singular values/vectors, k_{min}, that are required for the compressed image to preserve 90% of total variance of the original image.

   (b) (5 points) Plot cumulative variance vs. k.

   (c) (5 points) Compress the image A to A_k using SVD and utilizing the value you obtained in part (a). Report the values in A_k for the following pixels: $A_k(1, 1)$, $A_k(256, 256)$, $A_k(512, 512)$.

   (d) (5 points) Using singular values, determine the error (in L_2 norm sense) you have committed by approximating A by A_k.

(e) (2 points) **Bonus**: Display the original and compressed images. Note that you need to add the mean of 123.55 to every pixel in order to display the images properly.