Assignment 1

Due: 2/12/2022 23:59PM Submit to myCourses

- The purpose of this assignment is to test your understanding of the course material and assess your progress. It is a crucial part of your grade and must be completed to the best of your abilities.
- If you are unable to type the mathematical equations in Microsoft Word, you can write it on A4 size white paper, scan it and embed it in a Word file for submission.
- Please follow the instructions carefully and make sure to cite all sources properly. If you have any questions or concerns, do not hesitate to reach out to your instructor for clarification.
- We wish you all the best of luck with your assignment. Please make sure to submit it on time to avoid any issues.

1. Data mining concepts **(20 pts)**

(a) A collection of m items that is broken up into K groups, the i-th group of which is of size m, is supplied to you. What distinguishes the following two sampling methods if the objective is to produce a sample of size n m? (Assume replacement sampling.)

1) We randomly select $n \times m_i / m$ elements from each group.
2) We randomly select n elements from the data set, without regard for the group to which an object belongs.

(b) Classify the following attributes as:

** Numeric: binary, discrete, or continuous, and/or*

** Categorical: nominal or ordinal*

Some cases may have more than one interpretation, so briefly indicate your reasoning if you think there may be some ambiguity.

1. Time in terms of AM or PM.
2. ISBN numbers for books.
3. Angles as measured in degrees between 0° and 360°.
4. Brightness as measured by a light meter.
5. Brightness as measured by people's judgments.
6. Bronze, Silver, and Gold medals as awarded at the Olympics.
7. Height above sea level.
8. Number of patients in a hospital.

2. Matrices Concepts **(30 pts)**

(a) Assume the unknown matrix R is a rotation matrix. Remember that for rotation matrices, the transpose of the matrix has the reverse effect, which means $R^T = R^{-1}$. And this means $R^T R = I$ and $R^T R = I$. Also, assume the unknown matrix A is invertible. Use these facts to solve for Y in the matrix expression

$$RYA = B.$$

Remember, Y and B are not rotation matrices, and you can't assume anything in particular about them.

(b) Let $A = \begin{vmatrix} 2 & 0 \\ 0 & 3 \end{vmatrix}$, $B = \begin{vmatrix} 0 & 1 \\ 0 & 0 \end{vmatrix}$, $C = \begin{vmatrix} 1 & 2 \\ 2 & 1 \end{vmatrix}$, please apply the basic matrix product rule, and check if $(AB)C = A(BC)$.

(c) Assume $A = \begin{bmatrix} 2 & -2 & -4 \\ -1 & 3 & 4 \\ 1 & -2 & -3 \end{bmatrix}$. Prove that $A^2 = A$. Such matrix is called idempotent matrix. Could you find an idempotent matrix whose determinant is NOT zero?

3. SVD application - Matrix Norm **(20 pts)**

Similar to vectors, we may define a norm (a magnitude measurement) for matrices. A matrix norm $\|\cdot\|$ is any function with the following properties:

(i) $\|A\| \geq 0$, with equality iff $A = 0$ (zero matrix).

(ii) $\|cA\| = |c|\|A\|$, for any $c \in \mathbb{R}$.

(iii) $\|A + B\| \leq \|A\| + \|B\|$.

A common matrix norm is the 2-norm:

$$\|A\|_2 = \max_x \sqrt{(Ax)^T(Ax)},$$

where x is a unit vector (i.e. of length 1). Here, the $\max_x(\ )$ notation means "the maximum value of (expression), for any value of x" and the expression $\sqrt{x^T x}$ for a column vector x is another way of writing the length (you should go through the multiplication yourself to make sure you understand this fact). In other words, we can see that $\|A\|_2$ measures the maximum "magnifying power" of A: the max amount A can stretch any unit vector x under multiplication.

(a) At first glance, it looks like it would be hard to calculate $\|A\|_2$. But we can prove that $\|A\|_2 = \sigma_{\max}$, the largest singular value of A. Please prove this using the properties of the Singular Value Decomposition of A. Hint: here is one possible strategy you might follow to prove this:

- Rewrite the formula for $\|A\|_2$ using the knowledge that any matrix may be expressed as its SVD: $A = U\Sigma V^T$.
- Use the properties of transpose and of rotation matrices to eliminate the U's.
- Note that, for any rotation matrix R and unit vector u, Ru produces another unit vector (call it w), and $u^T R^T$ produces w^T. You can use this line of reasoning to eliminate V (as an alternative to this reasoning and the next step, you could skip straight to defining the unit vector which will maximize the expression you have).

(b) Another commonly used norm is the Frobenius norm:

$$\|A\|_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} |a_{ij}|^2}, \quad \text{where } A \text{ is a } m \times n \text{ matrix}$$

If we square it to eliminate the square root, the same operation can be calculated as

$$\|A\|_F^2 = \text{trace}(A^T A)$$

From class, remember that the trace of a matrix is the sum of its diagonal elements. If desired, you can verify that the above expressions match by working through yourself how the diagonal elements of $A^T A$ are calculated. Your task is to prove that:

$$\text{trace}(A^T A) = \sigma_1^2 + \sigma_2^2 + \cdots + \sigma_r^2,$$

the sum of the squares of the singular values of A. Hints:

- Use some of the SVD properties that you used in (a).
- One property of the trace is that it is "basis independent," which means, if we have an invertible matrix B, trace$(B^{-1}AB)$ = trace(A).
- Using the above properties, you should be able to reduce the expression to a matrix whose diagonal elements are $\sigma_1^2 + \sigma_2^2 + \cdots + \sigma_r^2$.

4. SVD for Image Compression (30 pts)

Singular Value Decomposition (SVD) can be effectively used to compress images. Suppose I is the pixel intensity matrix of a large image $n \times n$. The transmission (or storage) of I requires $\mathrm{O}(\mathrm{n}^2)$ numbers. Instead, one could use I_k, that is, the top k singular values $\sigma_1, \sigma_2, \ldots, \sigma_k$ along with the left and right singular vectors $u_1, u_2, \ldots, u_k$ and $v_1, v_2, \ldots, v_k$. This would require using $O(kn)$ real numbers instead of $O(n^2)$ real numbers. If k is much smaller than n, this results in substantial savings.

In this problem, you will explore SVD compression on the **flower.bmp** image we have provided. In addition to your answers to each question, you should also submit your Matlab code and required plots where necessary. Hint: You may find the Matlab SVD command particularly useful for this problem.

(a) Use MATLAB to read in flower.bmp and convert it to grayscale and 'double' format. Apply SVD and give the top 10 singular values. Generate a plot for all singular values versus their rankings (the diag command may be helpful to format the values). What do you notice from this plot?

(b) Verify that you can reconstruct and display the image using the three SVD matrices (note that the svd command returns V, not V^T). Then, perform compression by using only the top k singular values and their corresponding left/right singular vectors. Let k = 10, 50, and 100. Reconstruct and print the compressed images for the three different values of k. Briefly describe what you observe.

(c) Instead of transmitting the original (grayscale) image, you can perform SVD compression on it and transmit only the top k singular values and the corresponding left/right singular vectors. This should be much smaller than the original image for low values of k. With this specific image, will we still save space by compressing when k = 200? Show why or why not.