Indian Institute of Technology Kharagpur
Centre of Excellence in Artificial Intelligence

AI61003 Linear Algebra for AI and ML
Assignment 1, Due on: October 10, 2022

ANSWER ALL THE QUESTIONS

1. Let $\mathbf{x} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}^\top \in \mathbb{R}^n$ and $\mathbf{1}_n \in \mathbb{R}^n$ be the n-vector with all entries equal to one. Let $\text{avg}(\mathbf{x})$ and $\text{std}(\mathbf{x})$ be as defined as follows.

$$\begin{aligned} \text{avg}(\mathbf{x}) &= \frac{1}{n}\mathbf{1}_n^\top \mathbf{x} \\ \text{std}(\mathbf{x}) &= \frac{\|\mathbf{x} - \text{avg}(\mathbf{x})\mathbf{1}_n\|_2}{\sqrt{n}} \end{aligned}$$

Then for any $\alpha, \beta \in \mathbb{R}$ prove the following.

   (a) $\text{avg}(\alpha\mathbf{x} + \beta\mathbf{1}_n) = \alpha\text{avg}(\mathbf{x}) + \beta$

   (b) $\text{std}(\alpha\mathbf{x} + \beta\mathbf{1}_n) = |\alpha|\text{std}(\mathbf{x})$

   (c) If k entries in $\mathbf{x} \in \mathbb{R}^n$ are such that $|x_i - \text{avg}(\mathbf{x})| \geqslant a$ for some $a > 0$, then prove

$$\frac{k}{n} \leqslant \left(\frac{\text{std}(\mathbf{x})}{a}\right)^2$$

2. Let $\mathbf{w} \in \mathbb{R}^n$ be a given vector with $w_i > 0$ for $i = 1, 2, \ldots, n$. Then for any $\mathbf{x} \in \mathbb{R}^n$, define the function

$$\|\mathbf{x}\|_{\mathbf{w}} = \sqrt{\sum_{i=1}^{n} w_i x_i^2}$$

Show that the function $\|\cdot\|_w$ defines a norm called as weighted norm.

3. Let $\mathbf{z} = (\mathbf{A} + \mathbf{B})(\mathbf{x} + \mathbf{y})$ where $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{m\times n}$ and $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. Consider that the multiplication is done using two approaches.

   (a) Approach 1: Add $\mathbf{A}$ and $\mathbf{B}$, add $\mathbf{x}$ and $\mathbf{y}$ and then multiply $(\mathbf{A}+\mathbf{B})(\mathbf{x}+\mathbf{y})$ to get $\mathbf{z}$.

   (b) Approach 2: Compute $\mathbf{Ax}$, $\mathbf{Ay}$, $\mathbf{Bx}$ and $\mathbf{By}$ and then add these vectors to get $\mathbf{z}$.

What are conditions on m and n such that the second approach is computationally efficient than the first one?

4. **2-D Convolution**: Let $\mathbf{A} \in \mathbb{R}^{m\times n}$ and $\mathbf{B} \in \mathbb{R}^{p\times q}$ be two matrices. Then the convolution of $\mathbf{A}$ with $\mathbf{B}$, denoted as $\mathbf{C} = \mathbf{A} \star \mathbf{B} \in \mathbb{R}^{(m+p-1)\times(n+q-1)}$, is defined as follows.

$$C_{rs} = \sum_{i+k=r+1, j+\ell=s+1} A_{ij}B_{k\ell}, \quad r = 1, 2, \ldots, m+p-1, s = 1, 2, \ldots, n+q-1$$

(a) Generate a matrix $\mathbf{A}$ of size 8×6 randomly with entries either 0 or 1. Define

$$\mathbf{B} = \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix}$$

Compute $\mathbf{A} \star \mathbf{B}$.

(b) Take an image of digit zero from MNIST handwritten digit data set. Let $\mathbf{A} \in \mathbb{R}^{28 \times 28}$ be the matrix representation of this image. Let $\mathbf{D} = \begin{bmatrix} 1 & -1 \end{bmatrix}$. Then compute $\mathbf{A} \star \mathbf{D}$ and plot the image. Repeat the exercise for an image of digit 1 and digit 9 as well. What is the operation $\mathbf{A} \star \mathbf{D}$ doing in the original image?

5. Let $\mathbf{x} \in \mathbb{R}^n$ be a given vector. $\mathbf{x}$ is said to be symmetric if $x_k = x_{n-k+1}$ and is antisymmetric if $x_k = -x_{n-k+1}$ for $k = 1, 2, \ldots, n$. Show that every vector $\mathbf{x}$ can be uniquely decomposed as $\mathbf{x} = \mathbf{x}_s + \mathbf{x}_a$ where $\mathbf{x}_s$ is symmetric and $\mathbf{x}_a$ is antisymmetric.

6. Let $\mathbf{A} \in \mathbb{R}^{n \times m}$. Prove that left inverse of $\mathbf{A}$ exists if and only if the columns of $\mathbf{A}$ are linearly independent.

7. Let $\mathbf{A} \in \mathbb{R}^{(n+1) \times (n+1)}$ be a matrix defined as follows.

$$\mathbf{A} = \begin{bmatrix} \mathbf{I}_n & \mathbf{x} \\ \mathbf{x}^\top & 0 \end{bmatrix}$$

where $\mathbf{x} \in \mathbb{R}^n$ is any vector and $\mathbf{I}_n$ is an $n \times n$ identity matrix. Derive a condition on $\mathbf{x}$ such that the matrix $\mathbf{A}$ is invertible. Give an expression of $\mathbf{A}^{-1}$ in terms of $\mathbf{x}$.

8. Let $\mathbf{A} \in \mathbb{R}^{m \times n}$ be a matrix with linearly independent columns and $\mathbf{b} \in \mathbb{R}^m$ be a given vector. Let $\widehat{\mathbf{x}} \in \mathbb{R}^n$ be the least squares solution to the problem $\mathbf{A}\mathbf{x} = \mathbf{b}$. Then prove that for any vector $\mathbf{y} \in \mathbb{R}^n$, $(\mathbf{A}\mathbf{y})^\top \mathbf{b} = (\mathbf{A}\mathbf{y})^\top (\mathbf{A}\widehat{\mathbf{x}})$. Using this further prove that

$$\frac{(\mathbf{A}\widehat{\mathbf{x}})^\top \mathbf{b}}{\|\mathbf{A}\widehat{\mathbf{x}}\|_2 \|\mathbf{b}\|_2} = \frac{\|\mathbf{A}\widehat{\mathbf{x}}\|_2}{\|\mathbf{b}\|_2}$$

9. Let $u_1, u_2, \ldots, u_T$ and $y_1, y_2, \ldots, y_T$ be observed time series with the following relationship.

$$y_t \approx \widehat{y}_t = \sum_{j=1}^{n} h_j u_{t-j+1}, \quad t = 1, 2, \ldots, T,$$

where $u_k = 0$ for $k \leqslant 0$ and $\mathbf{h} \in \mathbb{R}^n$ is unknown vector. Find a matrix $\mathbf{A}$ and vector $\mathbf{b}$ for which

$$\|\mathbf{A}\mathbf{h} - \mathbf{b}\|_2^2 = (y_1 - \widehat{y}_1)^2 + \cdots + (y_T - \widehat{y}_T)^2.$$

10. **k-means clustering**: Consider k-means clustering algorithm as follows with the standard terminology and notation.
<u>Input:</u> $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N \in \mathbb{R}^n$. Initial list of k cluster representatives $\mathbf{z}_1, \ldots, \mathbf{z}_k$.

Output: Cluster assignment $c_1, c_2, \ldots, c_N$
Repeat until convergence

1. Cluster assignment based on cluster representatives.
2. Update cluster representatives.

(a) In Step 1, what is the computational complexity?

(b) In Step 2, what is the computational complexity?

(c) Assuming 10 iterations are performed, how many number of computations are involved to obtain the cluster assignment for the given data points?

11. **Image Clustering**: Consider the MNIST database of handwritten digits. Choose 100 images of each digit from this data set. In the notation of Problem 10, determine values N and n. Fix a reasonable convergence criterion. Perform the following exercises (a),(b) and (c) in two cases:
case (i) random initialization of cluster representatives;
case(ii) choose cluster representatives from the given data set.

(a) For $k = 20$, run the above algorithm to cluster the given images into 20 clusters. Plot the cluster representatives after the algorithm converges. Count the number of iterations.

(b) Choose 50 images (not chosen previously) from the MNIST data set randomly and assign the clusters to these *test* images. What is the accuracy of cluster assignment?

(c) For $k = 5$ to $k = 20$, tabulate the values of J^{clust} and discuss what may be the optimal size of number of clusters.

Does the choice of initial condition have any effect on the performance of k-means clustering algorithm?

****************** THE END ******************