Name: Sidharth Vishwakarma
Roll No: 20CS10082
Subject: LINEAR ALGEBRA FOR AI & ML (AI61003)

1. $avg(x) = \frac{1}{n} 1_n^T x$

$std(x) = \frac{\| x - avg(x) 1_n \|_2}{\sqrt{n}}$

(a) LHS $= avg(\alpha x + \beta 1_n)$

$= \frac{1}{n} 1_n^T (\alpha x + \beta 1_n)$

$= \frac{1}{n}(1_n^T \alpha x + \beta 1_n^T 1_n)$

$= \frac{1}{n}(1_n^T \alpha x + \beta \| 1_n \|_2^2)$

$= \frac{1}{n}(\alpha 1_n^T x) + \frac{1}{n} \beta \underbrace{(1^2 + 1^2 + \dots + 1^2)}_{n \text{ times}}$

$= \frac{1}{n} \alpha 1_n^T x + \beta$

$= \alpha \frac{1}{n} 1_n^T x + \beta$

$= \alpha\, avg(x) + \beta$

$=$ RHS

(b) ~~std~~ LHS $= std(\alpha x + \beta 1_n)$

$= \frac{\| (\alpha x + \beta 1_n) - avg(\alpha x + \beta 1_n) 1_n \|_2}{\sqrt{n}}$

$= \frac{\| \alpha x + \beta 1_n - (\alpha\, avg(x) + \beta) 1_n \|_2}{\sqrt{n}}$ (from part (a))

$= \frac{\| \alpha x + \beta 1_n - (\alpha\, avg(x) 1_n + \beta 1_n) \|_2}{\sqrt{n}}$

$= \frac{\| \alpha x + \cancel{\beta 1_n} - \alpha\, avg(x) 1_n - \cancel{\beta 1_n} \|_2}{\sqrt{n}}$

$$= \frac{\|\alpha(x - avg(x))\|_2}{\sqrt{n}}$$

$$= \frac{|\alpha| \, \|x - avg(x)\|_2}{\sqrt{n}}$$ (By property of norm)

$$= |\alpha| \frac{\|x - avg(x)\|_2}{\sqrt{n}}$$

$= |\alpha| \, std(x)$

$= RHS$

(c) Now $\|x - avg(x) 1_n\|_2 = \left\| \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} - \begin{pmatrix} avg(x) \\ avg(x) \\ \vdots \\ avg(x) \end{pmatrix} \right\|_2$

$$= \left\| \begin{bmatrix} x_1 - avg(x) \\ x_2 - avg(x) \\ \vdots \\ x_n - avg(x) \end{bmatrix} \right\|_2$$

$$= \sqrt{(x_1 - avg(x))^2 + (x_2 - avg(x))^2 + \dots + (x_n - avg(x))^2}$$

Now 'k' out of these 'n' $|x_i - avg(x)| \geq a$

$\Rightarrow$ for k such x_i's

$|x_i - avg(x)| \geq a$

$\Rightarrow |x_i - avg(x)|^2 \geq a^2$

$\Rightarrow (x_i - avg(x))^2 \geq a^2$

$\Rightarrow$ for the remaining x_i's

$(x_i - avg(x))^2 \geq 0$ holds

$\therefore \|x - avg(x) 1_n\|_2 \geq \sqrt{a^2 + a^2 + \dots (k \text{ times}) + 0 + 0 \dots + 0}$ ~~(n times)~~ (n-k times)

$\Rightarrow \|x - avg(x) 1_n\|_2 \geq \sqrt{ka^2}$

$\Rightarrow \|x - avg(x) 1_n\|_2 \geq a\sqrt{k}$ $(\because a > 0)$ —— ①

Now $\text{std}(x) = \dfrac{\|x - \text{avg}(x) 1_n\|_2}{\sqrt{n}}$

$\|x - \text{avg}(x) 1_n\|_2 = \sqrt{n} \; \text{std}(x)$

from eq^n ① -

$\sqrt{n} \; \text{std}(x) \geq a\sqrt{k}$

$\Rightarrow \sqrt{\dfrac{k}{n}} \leq \dfrac{\text{std}(x)}{a}$ $\quad (\because a>0, \sqrt{k}>0, \sqrt{n}>0)$

$\Rightarrow \dfrac{k}{n} \leq \left(\dfrac{\text{std}(x)}{a}\right)^2$

2. Norm of a vector $x \in \mathbb{R}^n$ is defined by $f : \mathbb{R}^n \to \mathbb{R}$ written as $f(x) = \|x\|$.

The norm should satisfy some properties given by.

(i) $\|x\| \geq 0$

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

since $w_i \geq 0$ and $x_i^2 \geq 0$

$\therefore \sum_{i=1}^{n} w_i x_i^2 \geq 0$

$\Rightarrow \sqrt{\sum_{i=1}^{n} w_i x_i^2} \geq 0$

$\Rightarrow \|x\|_W \geq 0$

(ii) $\|x\| = 0 \Leftrightarrow x = \vec{0}$

[$\Rightarrow$]

$\|x\|_W = 0$

$\Rightarrow \sqrt{\sum_{i=1}^{n} w_i x_i^2} = 0$

$\Rightarrow \sum_{i=1}^{n} w_i x_i^2 = 0$

now, since $w_i > 0$ and $x_i^2 \geq 0$

$\therefore w_i x_i^2 = 0 \quad \forall i$

$x_i^2 = 0 \quad (\because w_i > 0)$

$x_i = 0 \quad \forall i$

$\therefore \quad x = \vec{0}$

$[\Leftarrow]$

$x = \vec{0}$

$\Rightarrow x_i = 0 \quad \forall i$

Now $\sum_{i=1}^{n} w_i x_i^2 = 0$

$\Rightarrow \sqrt{\sum_{i=1}^{n} w_i x_i^2} = 0$

$\Rightarrow \|x\|_w = 0$

(iii) $\|\alpha x\| = |\alpha| \|x\|$

$$\|\alpha x\|_w = \sqrt{\sum_{i=1}^{n} w_i (\alpha x_i)^2}$$

$$= \sqrt{\sum_{i=1}^{n} w_i \alpha^2 x_i^2}$$

$$= \sqrt{\alpha^2 \sum_{i=1}^{n} w_i x_i^2}$$

$$= |\alpha| \sqrt{\sum_{i=1}^{n} w_i x_i^2}$$

$$= |\alpha| \|x\|_w$$

(iv) $\|x+y\| \leq \|x\| + \|y\|$

$$\|x+y\|_w = \sqrt{\sum_{i=1}^{n} w_i (x_i + y_i)^2}$$

$$= \sqrt{\sum_{i=1}^{n} (\sqrt{w_i}\, x_i + \sqrt{w_i}\, y_i)^2}$$

Construct another vectors X & Y s.t $X, Y \in \mathbb{R}^n$, and

$$X = \begin{bmatrix} \sqrt{w_1}\, x_1 \\ \sqrt{w_2}\, x_2 \\ \vdots \\ \sqrt{w_n}\, x_n \end{bmatrix} \text{ and } Y = \begin{bmatrix} \sqrt{w_1}\, y_1 \\ \sqrt{w_2}\, y_2 \\ \vdots \\ \sqrt{w_n}\, y_n \end{bmatrix}$$

Now $\|x+y\|_w = \|X+Y\|_2$ (By the definition of 2-norm)

Now $\|x\|_w = \sqrt{\sum_{i=1}^{n} w_i x_i^2}$

$= \sqrt{\sum_{i=1}^{n} (w_i x_i)^2}$

$= \|X\|_2$

And, $\|y\|_w = \|Y\|_2$, (Similarly)

Now from the property of 2-norm

$\|X+Y\|_2 \leq \|X\|_2 + \|Y\|_2$

$\Rightarrow \|x+y\|_w \leq \|x\|_w + \|y\|_w$

Hence the function $f: \mathbb{R}^n \to \mathbb{R}$ defined by $\|x\|_w = \sqrt{\sum_{i=1}^{n} w_i x_i^2}$

defines a norm function.

3. Consider a matrix addition of $A, B \in \mathbb{R}^{p \times q}$.

We add each & every element pointwise resulting in pq additions.

Now, consider matrix multiplication ~~of~~ ~~AB~~ $A \in \mathbb{R}^{p \times q}$ & $B \in \mathbb{R}^{q \times r}$

For generating each element of the final resultant matrix $C \in \mathbb{R}^{p \times r}$, we are supposed to do q multiplications followed by $(q-1)$ additions.

$\therefore$ Total calculations for generating the product matrix C is $pr(2q-1)$

(a) For approach (1): $(A+B)$ would take mn additions

$(x+y)$ " " n " "

$(A+B)(x+y)$ would take $m(2n-1)$ computations

Total computations in approach (1)

$= mn + n + m(2n-1)$

(b) For approach (2) : Ax would take $m(2n-1)$ computations

Ay " " " "

Bx " " " "

By " " " "

$Ax + Ay + Bx + By$ would take $3 \times n$ additions

Total computations in approach (2)

$$= 4m(2n-1) + 3n$$

For approach (2) to be computationally efficient

$$4m(2n-1) + 3n < mn + n + m(2n-1)$$

$$8mn - 4m + 3n < 3mn + n - m$$

$$5mn + 4n < 5m$$

$$n(5m+4) < 5m$$

$$n < \frac{5m}{5m+4}$$

Since n is an integer & $n > 0$

But $\frac{5m}{5m+4} < 1$ $\quad (\because m \in \mathbb{Z}$ and $m > 0)$

$\therefore$ $n < 1$ is not possible

$\therefore$ Approach (2) can never be computationally efficient than Approach (1)

4. 2-D convolution is defined as follows:

Let $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{p \times q}$ then $C = A * B$

$C \in \mathbb{R}^{(m+p-1) \times (n+q-1)}$ is defined as -

$$C_{rs} = \sum_{\substack{i+k=r+1 \\ j+l=s+1}} A_{ij} B_{kl}, \quad \begin{array}{l} r = 1, 2, \ldots, m+p-1 \\ s = 1, 2, \ldots, n+q-1 \end{array}$$

(a) Let us make the matrix A as follows.

$$A = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \end{bmatrix} \qquad B = \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix}$$

$$C = A * B = \begin{bmatrix} -1 & 1 & -1 & 1 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & -1 & 1 & -1 & 1 & -1 & 0 \end{bmatrix}$$

(b) The convolution operation in the case of images from the MNIST Database transforms the image into textured look and it ~~sems~~ seems as if the images have numbers engraved on it.
For the actual images refer to the jupyter notebook.
<Question 4(b).ipynb>

5. To uniquely decompose a given vector $x \in \mathbb{R}^n$ into symmetric and antisymmetric vectors. Let us firstly denote x^r as the reverse of vector $x \in \mathbb{R}^n$
x^r is defined as follows -

Let $x \in \mathbb{R}^n$, s.t. $x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$, then $x^r = \begin{bmatrix} x_n \\ x_{n-1} \\ \vdots \\ x_1 \end{bmatrix}$

Now in our case where $x = x_s + x_a$, s.t x_s is symmetric and x_a is antisymmmetric.

we define, $x_s = \dfrac{x + x^r}{2}$, element-wise, we can say

$$x_s = \begin{bmatrix} \dfrac{x_1 + x_n}{2} \\ \dfrac{x_2 + x_{n-1}}{2} \\ \vdots \\ \dfrac{x_n + x_1}{2} \end{bmatrix}$$

and, $x_a = \dfrac{x - x^r}{2}$, element-wise, we can say that

$$x_a = \begin{bmatrix} \dfrac{x_1 - x_n}{2} \\ \dfrac{x_2 - x_{n-1}}{2} \\ \vdots \\ \dfrac{x_n - x_1}{2} \end{bmatrix}$$

Clearly $x = x_s + x_a$, and x_s is symmetric and x_a is antisymmmetric.

6. Given that $A \in \mathbb{R}^{n \times m}$

To prove: left inverse of A exists if and only if the columns of A are linearly independent

[if] Let the left inverse of A be C

Therefore, $CA = I$

Now consider the ~~eqⁿ~~ system of eq^ns $Ax = 0$
In order to prove that the columns of A are linearly independent we need to prove that $x = 0$ is the only solution to this system of eq^n.

Now $\quad 0 = C.0 = C.(Ax) = (CA)x = Ix = x$

$\Rightarrow \quad x = 0$

[only if]

If matrix is $m \times n$ then since the columns are linearly independent this implies the matrix has rank n. Thus the m rows span an n-dimentional subspace of $\mathbb{R}^n$, which is $\mathbb{R}^n$ itself.

$\therefore$ The ~~lear~~ linear combination of rows can make up each basis vectors.

We construct ~~a~~ a left inverse in such a way s.t. the k^{th} row of any left inverse will be the coefficients of the linear combination of the k^{th} basis vector, any matrix having this form will act as a left inverse for the given matrix $A \in \mathbb{R}^{m \times n}$

Generally in matrix product $AB = C$ the k^{th} row in C is a linear combination of the rows in B by the coeff in the k^{th} row of A.

7. A matrix is invertible if & only if A has linearly independent columns.

This implies $\quad Ax = 0 \Rightarrow x = 0$

Now if ~~$A \in \mathbb{R}^{m \times n}$~~ $A \in \mathbb{R}^{(n+1)\times(n+1)} \Rightarrow x = \mathbb{R}^{n+1}$

or, simply, say, $\quad x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \\ x_{n+1} \end{bmatrix}$

now let $y = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^n$ and $z = x_{n+1} \in \mathbb{R}^1$

Now $Ax = \begin{bmatrix} I_n & X \\ X^T & 0 \end{bmatrix} \begin{bmatrix} y \\ z \end{bmatrix} = \begin{bmatrix} y + Xz \\ X^T y \end{bmatrix}$

Now $Ax = 0$

$\therefore \; y + Xz = 0 \quad \& \quad X^T y = 0$

if $X = 0$, then $y = 0$, $z = k \neq 0$ can be the sol^n of $Ax = 0$, $\therefore$ if $X = 0$ then A is non-invertible

Now, if $X \neq 0$, $y = -Xz$ —— ①

$$\Rightarrow X^T y = 0$$

$$-X^T X z = 0 \qquad \text{(from ①)}$$

$$\|X\|_2 z = 0$$

$$z = 0 \qquad (\because X \neq 0)$$

$$y = 0 \qquad \text{(from ①)}$$

$\therefore$ if $X \neq 0$ we get $x = 0$ $\Rightarrow$ A is invertible

$\therefore$ A is invertible if & only if $X \neq 0$

To find the inverse, Let us solve the system of eq^{n_s} $Ax = b$ and equate it to the result of $A^{-1}b$. Similar to partitioning of $x = \begin{bmatrix} y \\ z \end{bmatrix}$ we partition

$b = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$ where $b_1 \in \mathbb{R}^n$, $b_2 \in \mathbb{R}$

$$Ax = b \Rightarrow \begin{bmatrix} I_n & X \\ X^T & 0 \end{bmatrix} \begin{bmatrix} y \\ z \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} y + Xz \\ X^T y \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

$\Rightarrow \quad y + Xz = b_1 \quad$ and $\quad X^T y = b_2$

From the first eq^n, we get,

$$y = b_1 - Xz \quad \text{——} \quad (1)$$

Putting this in the second eq^n,

$$X^T (b_1 - Xz) = b_2$$

$$\Rightarrow b_2 = X^T b_1 - \|X\|_2^2 z$$

$$\Rightarrow z = \frac{X^T b_1 - b_2}{\|X\|_2^2}$$

Put this in eq^n (1) –

$\cancel{y = b_1 - \left(\frac{X^T b_1 - b_2}{\|X\|_2^2}\right) X}$, ~~we can interchange X & z because z is a scaler~~

$\cancel{y = \|X\|_2^2 I - XX^T}$

$$y = b_1 - X\left(\frac{X^T b_1 - b_2}{\|X\|_2^2}\right)$$

$$y = \left(\frac{\|X\|_2^2 I - XX^T}{\|X\|_2^2}\right) b_1 - \left(\frac{X}{\|X\|_2^2}\right) b_2$$

and $$z = \left(\frac{X^T}{\|X\|_2^2}\right) b_1 - \frac{1}{\|X\|_2^2} b_2$$

$$x = \begin{bmatrix} y \\ z \end{bmatrix} = \frac{1}{\|X\|_2^2} \begin{bmatrix} \|X\|_2^2 I - XX^T & X \\ X^T & -1 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

$$\Rightarrow A^{-1} = \frac{1}{\|X\|_2^2} \begin{bmatrix} \|X\|_2^2 I - XX^T & X \\ X^T & -1 \end{bmatrix}$$

8. Given matrix $A \in \mathbb{R}^{m \times n}$ with linearly independent columns and $b \in \mathbb{R}^m$

Now the solution to the least squares problem ~~is given~~ $Ax = b$ is given by $\hat{x} = A^+ b$ where A^+ is the pseudo-inverse of A given by $A^+ = (A^T A)^{-1} A^T$

$\therefore \hat{x} = (A^T A)^{-1} A^T b$

Now RHS of the given eq^n $= (Ay)^T (A\hat{x})$

$$= (Ay)^T (A (A^T A)^{-1} A^T b)$$

$$= y^T A^T A (A^T A)^{-1} A^T b$$

$$= y^T ((A^T A)(A^T A)^{-1}) A^T b$$

$$= y^T I A^T b$$

$$= y^T A^T b$$

$$= (Ay)^T b$$

$$= LHS$$

Now since $(Ay)^T b = (Ay)^T (A\hat{x})$

Put $y = \hat{x}$

$$(A\hat{x})^T b = (A\hat{x})^T (A\hat{x})$$

$$\Rightarrow (A\hat{x})^T b = \|A\hat{x}\|_2^2$$

$$\Rightarrow \frac{(A\hat{x})^T b}{\|A\hat{x}\|_2} = \|A\hat{x}\|_2$$

$$\Rightarrow \frac{(A\hat{x})^T b}{\|A\hat{x}\|_2 \|b\|_2} = \frac{\|A\hat{x}\|_2}{\|b\|_2}$$ (Divide both sides by $\|b\|_2$)

9. Given $u_1, u_2, \ldots, u_T$ and $y_1, y_2, \ldots, y_T$ with the following relationship:

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

where $u_k = 0$ for $k \leq 0$ & $h \in \mathbb{R}^n$

Also given that, $\|Ah - b\|_2^2 = (y_1 - \hat{y}_1)^2 + \ldots + (y_T - \hat{y}_T)^2$

$$= (\hat{y}_1 - y_1)^2 + \ldots + (\hat{y}_T - y_T)^2$$

One possible value of the vector $(Ah - b)$ would be

$$Ah - b = \begin{bmatrix} \hat{y}_1 - y_1 \\ \hat{y}_2 - y_2 \\ \vdots \\ \hat{y}_T - y_T \end{bmatrix}$$

Suppose $b = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{bmatrix}$

Therefore $Ah = \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_T \end{bmatrix} = \begin{bmatrix} h_1 u_1 \\ h_1 u_2 + h_2 u_1 \\ h_1 u_3 + h_2 u_2 + h_3 u_1 \\ \vdots \\ h_1 u_T + h_2 u_{T-1} + h_3 u_{T-2} + \ldots + h_n u_{T-n+1} \end{bmatrix}$

$$Ah = \begin{bmatrix} u_1 & 0 & 0 & \cdots & 0 \\ u_2 & u_1 & 0 & \cdots & 0 \\ u_3 & u_2 & u_1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ u_T & u_{T-1} & u_{T-2} & \cdots & u_{T-n+1} \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \\ h_3 \\ \vdots \\ h_n \end{bmatrix}$$

$\therefore \quad A = \begin{bmatrix} u_1 & 0 & 0 & \cdots & 0 \\ u_2 & u_1 & 0 & \cdots & 0 \\ u_3 & u_2 & u_1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ u_T & u_{T-1} & u_{T-2} & & u_{T-n+1} \end{bmatrix}_{T\times n}$ & $b = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_T \end{bmatrix}$

10. Consider the algorithm of k-means clustering
Repeat until convergence:
1. Cluster assignment based on cluster representatives
2. Update cluster representatives

(a) Computational complexity in step 1.

$= \underbrace{O(nk)}_{\text{For calculating distances from k-cluster representatives for each feature vector}} \times \underbrace{O(N)}_{\text{No. of feature vectors}}$

$= O(nNk)$

(b) Computational complexity in step 2

$= \underbrace{O(n)}_{\text{To add the elements of each feature vector into cummulative sum of its cluster assignment}} \times \underbrace{O(N)}_{\text{Number of feature vectors}}$

$= O(nN)$

(c) Assume random initial cluster assignment. This would take $O(nk)$ computational complexity to assign k vectors of size n.

For performing each iteration $O(nNk + nN)$ time would be needed.

For 10 iterations a total of $O(10nNk + 10nN)$ time will be needed

$\therefore$ Total time $= O(10nNk + 10nN + nk)$

11. Considering the MNIST database of handwritten digits. Since we are choosing 100 images of each digit from 0-9, therefore, a total of $100 \times 10 = 1000$ feature vectors will be available to us. This makes $N = 1000$, also since each image is represented by a 28×28 matrix therefore, $n = 28 \times 28 = 784$.

Yes, the choice of initial condition have an effect on the performance of k-means clustering algorithm. In case (ii) when cluster representatives were choosen from the given data set, then the k-means clustering algorithm converges earlier and with less no. of iterations as compared with case (i) when there was random initialization of cluster representatives. Moreover we also saw a slight gain in accuracy in the algorithm for case (ii) compared to the case (i)