

MSBD5004 Mathematical Methods for Data Analysis  
Homework 1

RONG Shuo - 21126613

September 2024

**Question 1:**

1. Consider the vector space  $\mathbb{R}^n$ .
- (a) Check that  $\|\mathbf{x}\|_\infty = \max_{1 \leq i \leq n} |x_i|$  is indeed a norm on  $\mathbb{R}^n$ .
- (b) Omitted
- (c) Omitted

**Answer:**

(a)

*Proof.* To show that  $\|\mathbf{x}\|_\infty$  is a norm, we check the following properties:

1.  $\forall \mathbf{x} \in \mathbb{R}^n$ , we have

$$\|\mathbf{x}\|_\infty = \max_{1 \leq i \leq n} |x_i| \geq 0.$$

2. Assume  $\|\mathbf{x}\|_\infty = 0$ . Then,

$$\max_{1 \leq i \leq n} |x_i| = 0 \implies x_i = 0, \forall i \implies \mathbf{x} = \mathbf{0}.$$

3.  $\forall \alpha \in \mathbb{R}$ ,

$$|\alpha \cdot \mathbf{x}|_\infty = \max_{1 \leq i \leq n} |\alpha \cdot x_i| = |\alpha| \cdot \max_{1 \leq i \leq n} |x_i| = |\alpha| \cdot \|\mathbf{x}\|_\infty.$$

4.  $\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ ,

$$\|\mathbf{x} + \mathbf{y}\|_\infty = \max_{1 \leq i \leq n} |x_i + y_i| \leq \max_{1 \leq i \leq n} (|x_i| + |y_i|) \leq \max_{1 \leq i \leq n} |x_i| + \max_{1 \leq i \leq n} |y_i| = \|\mathbf{x}\|_\infty + \|\mathbf{y}\|_\infty.$$

Thus,  $\|\mathbf{x}\|_\infty$  satisfies all properties of a norm.

□

(b)

*Proof.* we have

$$\max_{1 \leq i \leq n} |x_i| \leq \sum_{i=1}^n |x_i| \leq n \cdot \max_{1 \leq i \leq n} |x_i|$$

s.t.

$$\|\mathbf{x}\|_\infty \leq \|\mathbf{x}\|_1 \leq n \cdot \|\mathbf{x}\|_\infty$$

□

(c)

*Proof.* we have

$$|\mathbf{x}^T \mathbf{y}| = \left| \sum_{i=1}^n x_i \cdot y_i \right| \leq \sum_{i=1}^n |x_i \cdot y_i| \leq \sum_{i=1}^n |x_i| |y_i| \leq \sum_{i=1}^n \left( \max_{1 \leq i \leq n} |y_i| \right) \cdot |x_i| = \max_{1 \leq i \leq n} |y_i| \cdot \sum_{i=1}^n |x_i| = \|\mathbf{x}\|_1 \|\mathbf{y}\|_\infty$$

s.t.

$$\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n, |\mathbf{x}^T \mathbf{y}| \leq \|\mathbf{x}\|_1 \|\mathbf{y}\|_\infty$$

□

**Question 2:**2. For any  $\mathbf{A} \in \mathbb{R}^{m \times n}$ , we have defined

$$\|\mathbf{A}\|_2 = \max_{\mathbf{x} \in \mathbb{R}^n, \|\mathbf{x}\|_2=1} \|\mathbf{Ax}\|_2.$$

(a) Prove that  $\|\cdot\|_2$  is a norm on  $\mathbb{R}^{m \times n}$ .(b) Prove that  $\|\mathbf{Ax}\|_2 \leq \|\mathbf{A}\|_2 \|\mathbf{x}\|_2$  for any  $\mathbf{A} \in \mathbb{R}^{m \times n}$  and  $\mathbf{x} \in \mathbb{R}^n$ .(c) Prove that  $\|\mathbf{AB}\|_2 \leq \|\mathbf{A}\|_2 \|\mathbf{B}\|_2$  for all  $\mathbf{A} \in \mathbb{R}^{m \times n}$  and  $\mathbf{B} \in \mathbb{R}^{n \times p}$ .**Answer:**

(a)

*Proof.* 1.  $\forall \mathbf{A} \in \mathbb{R}^{m \times n}$ 

$$\|\mathbf{A}\|_2 = \max_{\mathbf{x} \in \mathbb{R}^n, \|\mathbf{x}\|_2=1} \|\mathbf{Ax}\|_2 \geq 0,$$

since  $\|\mathbf{Ax}\|_2$  is non-negative for all  $\mathbf{x}$ .2. If  $\|\mathbf{A}\|_2 = 0$ , then:

$$\max_{\mathbf{x} \in \mathbb{R}^n, \|\mathbf{x}\|_2=1} \|\mathbf{Ax}\|_2 = 0 \implies \|\mathbf{Ax}\|_2 = 0, \forall \mathbf{x},$$

which means  $\mathbf{Ax} = \mathbf{0}, \forall \mathbf{x}$ .Choosing  $\mathbf{x}$  as the standard basis vectors (e.g.  $e_1 = [1, 0, 0, \dots], e_i = [0, \dots, 1, 0, \dots]$ ),we conclude that  $\mathbf{A} = \mathbf{0}$ .Conversely, if  $\mathbf{A} = \mathbf{0}$ , then clearly  $\|\mathbf{A}\|_2 = 0$ 3. For any scalar  $\alpha$ ,

$$\|\alpha \mathbf{A}\|_2 = \max_{\mathbf{x} \in \mathbb{R}^n, \|\mathbf{x}\|_2=1} \|\alpha \mathbf{Ax}\|_2 = \max_{\mathbf{x} \in \mathbb{R}^n, \|\mathbf{x}\|_2=1} |\alpha| \|\mathbf{Ax}\|_2 = |\alpha| \max_{\mathbf{x} \in \mathbb{R}^n, \|\mathbf{x}\|_2=1} \|\mathbf{Ax}\|_2 = |\alpha| \|\mathbf{A}\|_2.$$

4. For any matrices  $\mathbf{A}, \mathbf{B}$ :

$$\|\mathbf{A} + \mathbf{B}\|_2 = \max_{\mathbf{x} \in \mathbb{R}^n, \|\mathbf{x}\|_2=1} \|\mathbf{Ax} + \mathbf{Bx}\|_2$$

By the triangle inequality for the  $\|\mathbf{x}\|_2$  norm:

$$\|\mathbf{Ax} + \mathbf{Bx}\|_2 \leq \|\mathbf{Ax}\|_2 + \|\mathbf{Bx}\|_2.$$

Thus,

$$\|\mathbf{A} + \mathbf{B}\|_2 \leq \max_{\mathbf{x} \in \mathbb{R}^n, \|\mathbf{x}\|_2=1} \|\mathbf{Ax}\|_2 + \max_{\mathbf{x} \in \mathbb{R}^n, \|\mathbf{x}\|_2=1} \|\mathbf{Bx}\|_2 = \|\mathbf{A}\|_2 + \|\mathbf{B}\|_2$$

5. Conclusion Since all four properties of norms are satisfied, we conclude that  $\|\cdot\|_2$  is a norm on  $\mathbb{R}^{m \times n}$ . □

(b)

*Proof.*  $\|\mathbf{x}\|_2 \neq 0$ :

for any  $\mathbf{x} \in \mathbb{R}^n$ , we know:

$$\|\mathbf{x}\| \neq 0 \implies \frac{\mathbf{x}}{\|\mathbf{x}\|_2} \text{ is a unit vector: } \mathbf{u}$$

Since  $\|\mathbf{u}\| = 1$ , we have:

$$\|\mathbf{A}(\frac{\mathbf{x}}{\|\mathbf{x}\|_2})\|_2 = \frac{1}{\|\mathbf{x}\|_2} \|\mathbf{Ax}\|_2 = \|\mathbf{Au}\|_2 \leq \|\mathbf{A}\|_2 \implies \|\mathbf{Ax}\|_2 \leq \|\mathbf{A}\|_2 \|\mathbf{x}\|_2$$

$\|\mathbf{x}\|_2 = 0$ : we have:

$$\|\mathbf{A}\mathbf{0}\|_2 = 0 = \|\mathbf{A}\|_2 \cdot 0$$

**Conclusion:**

$$\forall \mathbf{A} \in \mathbb{R}^{m \times n} \text{ and } \mathbf{x} \in \mathbb{R}^n : \|\mathbf{Ax}\|_2 \leq \|\mathbf{A}\|_2 \|\mathbf{x}\|_2$$

□

(c)

*Proof.* For all matrix  $\mathbf{M} \in \mathbb{R}^{n \times m}$ , we have:

$$\|\mathbf{Mx}\|_2 \leq \|\mathbf{M}\|_2 \|\mathbf{x}\|_2, \forall \mathbf{x} \in \mathbb{R}^m \quad (1)$$

and

$$\|\mathbf{Mx}\|_2 \leq \|\mathbf{M}\|_2 \|\mathbf{x}\|_2 = \|\mathbf{M}\|_2, \forall \mathbf{x} \in \mathbb{R}^m \text{ and } \|\mathbf{x}\|_2 = 1 \quad (2)$$

Consider  $\mathbf{B} \in \mathbb{R}^{n \times p}$ ,  $\mathbf{x} \in \mathbb{R}^p$  and  $\|\mathbf{x}\|_2 = 1$ , as  $\mathbf{M}, \mathbf{x}$  according formula (2), s.t.

$$\|\mathbf{Bx}\|_2 \leq \|\mathbf{B}\|_2$$

Consider  $\mathbf{A} \in \mathbb{R}^{m \times n}$ ,  $\mathbf{Bx} \in \mathbb{R}^n$  as  $\mathbf{M}, \mathbf{x}$  according formula (1), s.t.

$$\|\mathbf{ABx}\|_2 \leq \|\mathbf{A}\|_2 \|\mathbf{Bx}\|_2 \leq \|\mathbf{A}\|_2 \|\mathbf{B}\|_2$$

So we can conclude:

$$\|\mathbf{AB}\|_2 = \max_{\mathbf{x} \in \mathbb{R}^p, \|\mathbf{x}\|_2=1} \|\mathbf{ABx}\|_2 \leq \|\mathbf{A}\|_2 \|\mathbf{B}\|_2, \forall \mathbf{A} \in \mathbb{R}^{m \times n} \text{ and } \mathbf{B} \in \mathbb{R}^{n \times p}$$

□

**Question 3:**

3. For any  $\mathbf{A} \in \mathbb{R}^{m \times n}$ , we define the Frobenius norm  $\|\mathbf{A}\|_F = \left( \sum_{i=1}^m \sum_{j=1}^n a_{i,j}^2 \right)^{1/2}$ .  
Prove that

$$\|\mathbf{A}\|_2 \leq \|\mathbf{A}\|_F \leq \sqrt{n} \|\mathbf{A}\|_2$$

**Answer:**

*Proof.* \*\*To prove  $\|\mathbf{A}\|_2 \leq \|\mathbf{A}\|_F$  \*\*

We have,

$$\|\mathbf{A}\|_2 = \max_{\mathbf{x} \in \mathbb{R}^n, \|\mathbf{x}\|_2=1} \|\mathbf{A}\mathbf{x}\|_2$$

Assume  $\operatorname{argmax}(\|\mathbf{A}\|_2) = \mathbf{x}$ ,  $\mathbf{x} = [x_1 \cdots x_j \cdots x_n]$

$$\|\mathbf{A}\|_2^2 = \sum_{i=1}^m \left( \sum_{j=1}^n a_{i,j} x_j \right)^2$$

We know:

$$\begin{aligned} \left( \sum_{i=1}^n a_i b_i \right)^2 &\leq \sum_{i=1}^n a_i^2 \sum_{i=1}^n b_i^2 \\ \left( \sum_{j=1}^n a_{i,j} x_j \right)^2 &\leq \sum_{j=1}^n a_{i,j}^2 \sum_{j=1}^n x_j^2 = \sum_{j=1}^n a_{i,j}^2 \end{aligned}$$

s.t.:

$$\|\mathbf{A}\|_2^2 \leq \sum_{i=1}^m \sum_{j=1}^n a_{i,j}^2 = \|\mathbf{A}\|_F^2$$

s.t.:

$$\|\mathbf{A}\|_2 \leq \|\mathbf{A}\|_F$$

\*\*To prove  $\|\mathbf{A}\|_F \leq \sqrt{n} \|\mathbf{A}\|_2$  \*\*

Let  $\mathbf{e}_i$  be the standard basis vector in  $\mathbb{R}^n$ .

$$\|\mathbf{A}\|_F^2 = \sum_{i=1}^n \|\mathbf{A}\mathbf{e}_i\|_2^2$$

We know:

$$\begin{aligned} \|\mathbf{A}\mathbf{e}_i\|_2^2 &\leq \|\mathbf{A}\|_2^2 \\ \sum_{i=1}^n \|\mathbf{A}\mathbf{e}_i\|_2^2 &\leq n \|\mathbf{A}\|_2^2 \end{aligned}$$

s.t.

$$\begin{aligned} \|\mathbf{A}\|_F^2 &\leq n \|\mathbf{A}\|_2^2 \\ \|\mathbf{A}\|_F &\leq \sqrt{n} \|\mathbf{A}\|_2 \end{aligned}$$

□

**Question 4:**

4. A magic square  $M_n$  is a  $n \times n$  matrix containing the integers from 1 to  $n^2$  whose row and column sums are all the same. For example:

$$\begin{bmatrix} 16 & 2 & 3 & 13 \\ 5 & 11 & 10 & 8 \\ 9 & 7 & 6 & 12 \\ 4 & 14 & 15 & 1 \end{bmatrix}$$

This magic square appears in the Renaissance engraving *Melencolia I* by the German painter, engraver, and amateur mathematician Albrecht Durer (1471-1528).

Let  $a_n$  denote the magic constant of  $M_n$ , so that  $a_n = n(n^2 + 1)/2$ . Let  $\mathbf{d}$  denote a vector in  $\mathbb{R}^n$  with each element equal to 1.

- (a) Determine  $M_n \mathbf{d}$  and  $\mathbf{d}^T M_n$ . Conclude that  $a_n$  is an eigenvalue of  $M_n$ .
- (b) Show that the row and column sums of  $M_n^2$  are all the same.
- (c) Determine  $\|M_n\|_2$ .

**Answer:**

(a)

*Proof.* Assume that  $M_n = [\mathbf{c}_1 \cdots \mathbf{c}_i \cdots \mathbf{c}_n]$ , each  $\mathbf{c}$  indicates the column of  $M_n$ .

We know:

$$M_n \mathbf{d} = [\mathbf{c}_1 \cdots \mathbf{c}_n] \mathbf{d} = \sum_{i=1}^n \mathbf{c}_i = a_n \mathbf{d}$$

Assume that  $M_n = [\mathbf{r}_1 \cdots \mathbf{r}_i \cdots \mathbf{r}_n]^T$ , each  $\mathbf{r}$  indicates the row of  $M_n$ .

$$\mathbf{d}^T M_n = \mathbf{d}^T [\mathbf{r}_1 \cdots \mathbf{r}_n]^T = a_n \mathbf{d}^T$$

Since  $M_n \mathbf{d} = a_n \mathbf{d}$ , s.t.  $a_n$  is an eigenvalue of  $M_n$  with an eigenvector  $\mathbf{d}$ .

Since:

$$\mathbf{d}^T M_n = a_n \mathbf{d}^T$$

We can also conclude that  $a_n$  is an eigenvalue with a left eigenvector.

□

(b)

*Proof.* We know for each row in  $M_n^2$ , it can be expressed by:

$$\mathbf{r}_i M_n$$

s.t. the sum of the rows can be expressed by:

$$\mathbf{r}_i M_n \mathbf{d} = \mathbf{r}_i a_n \mathbf{d} = a_n \sum_{j=1}^n r_{ij} = a_n^2$$

We know for each column in  $\mathbf{M}_n^2$ , it can be expressed by:

$$\mathbf{M}_n^T \mathbf{c}_i$$

s.t. the sum of the column can be expressed by:

$$\mathbf{d}^T \mathbf{M}_n^T \mathbf{c}_i = a_n^2$$

In Conclusion: the row and the column sums of  $\mathbf{M}_n^2$  are all the same as  $a_n^2$ .  $\square$

(c)

*Proof.* We know, for every element  $c_{ij}$  in  $\mathbf{M}_n$ :

$$\|\mathbf{M}_n\|_2^2 \leq \|\mathbf{M}_n\|_F^2 = \sum_{j=1}^n \sum_{i=1}^n c_{ij}^2$$

We know that  $\mathbf{M}_n$  contains the integers from 1 to  $n^2$ , s.t.:

$$\|\mathbf{M}_n\|_F^2 = \sum_{j=1}^n \sum_{i=1}^n c_{ij}^2 = 1^2 + 2^2 + \dots + (n^2)^2 = \frac{n^2(n^2+1)(2n^2+1)}{6}$$

And we know:

$$a_n^2 = \frac{n^2(n^2+1)^2}{4}$$

And:

$$\frac{a_n^2}{\|\mathbf{M}_n\|_F^2} = \frac{3}{2} \cdot \frac{n^2+1}{2n^2+1} = \frac{3}{2} \cdot \left(1 - \frac{1}{2 + \frac{1}{n^2}}\right) \geq \lim_{n^2 \rightarrow \infty} \frac{3}{2} \cdot \left(1 - \frac{1}{2 + \frac{1}{n^2}}\right) = 1$$

s.t.:

$$a_n^2 \geq \|\mathbf{M}_n\|_F^2 \geq \|\mathbf{M}_n\|_2^2$$

s.t.:

$$\|\mathbf{M}_n\|_2 \leq a_n$$

And:

$$\frac{\|\mathbf{M}_n \mathbf{d}\|_2}{\|\mathbf{d}\|_2} = \frac{\|a_n \mathbf{d}\|_2}{\|\mathbf{d}\|_2} = a_n$$

s.t.:

$$\|\mathbf{M}_n\|_2 = a_n$$

$\square$

### Question 5:

5. Let  $a_1, a_2, \dots, a_m$  be  $m$  given real numbers. Prove that a median of  $a_1, a_2, \dots, a_m$  minimizes

$$\sum_{i=1}^m |a_i - b|$$

over all  $b \in \mathbb{R}$  (As we discussed in the lecture, this result is curcial for deriving the  $K$ -medians algorithm in clustering.)

**Answer:**

*Proof.* Suppose we order this sequence, and have  $a_1 \leq a_2 \leq \dots \leq a_m$ . For  $b \in \mathbb{R}$  and  $a_1 \leq a_k \leq b \leq b+d \leq a_{k+1} \leq a_m$  in the ascending ordered sequence  $a_1, \dots, a_m$ , we have:

$$\begin{aligned} f(b+d) &= \sum_{i=1}^m |a_i - b - d| = \sum_{i=1}^k (b+d - a_i) + \sum_{i=k+1}^m (a_i - b - d) \\ &= \sum_{i=k+1}^m (a_i - b) + \sum_{i=1}^k (b - a_i) + (2k - m)d = f(b) + (2k - m)d \\ f(b+d) - f(b) &= (2k - m)d \end{aligned}$$

We know for  $f(b)$ ,  $2k < m$ ,  $f(b)$  is descending, and  $f(b)$  get the minimum when  $b = a_{k+1}$ . When  $2k > m$ ,  $f(b)$  is ascending, and when  $b = a_k$ ,  $f(b)$  get the minimum.

And when  $m$  is even,  $k = \frac{m}{2}$ ,  $f(b)$  to be minimum, when  $a_{\frac{m}{2}} \leq b \leq a_{\frac{m}{2}+1}$ . And when  $m$  is odd,  $k = \frac{m+1}{2}$ ,  $f(b)$  to be minimum, when  $b = a_{\frac{m+1}{2}}$ .  $k = \frac{m-1}{2}$  to be the same.

And when  $b$  is the median of this sequence, it satisfies these inequations.

For  $b < a_1$ , we know  $f(b) > f(a_1)$  all the time.

For  $b > a_n$ , we know  $f(b) > f(a_n)$  all the time.

We can conclude that a median of this sequence minimizes  $\sum_{i=1}^m |a_i - b|$

□

**Question 6:**

6. Suppose that the vector  $\mathbf{x}_1, \dots, \mathbf{x}_N$  in  $\mathbb{R}^n$  are clustered using the  $K$ -means algorithm, with group representative  $\mathbf{z}_1, \dots, \mathbf{z}_k$ .

(a) Suppose the original vectors  $\mathbf{x}_i$  are nonnegative, *i.e.*, their entries are nonnegative. Explain why the representative  $\mathbf{z}_j$  output by the  $K$ -means algorithm are also nonnegative.

(b) Suppose the original vectors  $\mathbf{x}_i$  represent proportions, *i.e.*, their entries are nonnegative and sum to one. (This is the case when  $\mathbf{x}_i$  are word count histograms, for example.) Explain why the representatives  $\mathbf{z}_j$  output by the  $K$ -means algorithm are also represent proportions (*i.e.*, their entries are nonnegative and sum to one).

(c) Suppose the original vectors  $\mathbf{x}_i$  are Boolean, *i.e.*, their entries are either 0 or 1. Give an interpretation of  $(\mathbf{z}_j)_i$ , the  $i$ -th entry of the  $j$  group representative.

**Answer:**

(a)

*Proof.* We know:

$$\mathbf{z}_j = \frac{\sum_{\mathbf{x}_i \in G_j} \mathbf{x}_i}{\text{Count}(G_j)} \text{ where } G_j = \{\mathbf{x} | \text{the representative of } \mathbf{x} \text{ is } \mathbf{z}_j, \mathbf{x} \in \{\mathbf{x}_1, \dots, \mathbf{x}_N\}\}$$

According,  $\mathbf{x}_i$  is nonnegative. We know:

$$\sum_{\mathbf{x}_i \in G_j} \mathbf{x}_i \text{ is nonnegative.}$$

s.t.  $\mathbf{z}_j$  is nonnegative. □

(b)

*Proof.* We know:

$$\mathbf{z}_j = \frac{\sum_{\mathbf{x}_i \in G_j} \mathbf{x}_i}{\text{Count}(G_j)}$$

$$\sum_{m=1}^n z_m = \frac{1}{\text{Count}(G_j)} \sum_{\mathbf{x}_i \in G_j} \sum_{m=1}^n x_m = \frac{1}{\text{Count}(G_j)} \cdot \text{Count}(G_j) = 1$$

We know  $\mathbf{z}_j$  is nonnegative, according to (a). And  $\mathbf{z}_j$  represents proportions, too. □

(c)

*Proof.*

$$(\mathbf{z}_j)_i = \frac{1}{\text{Count}(G_j)} \sum_{\mathbf{x}_i \in G_j} (\mathbf{x}_i)_i = \frac{\sum_{\mathbf{x}_i \in G_j} (\mathbf{x}_i)_i}{\text{Count}(G_j)} = \frac{\text{CountOfTrue}}{\text{CountOfTrue} + \text{CountOfFalse}}$$

We can conclude that  $(\mathbf{z}_j)_i$  represents the proportions of true in  $\{(\mathbf{x}_i)_i | \mathbf{x}_i \in G_j\}$  □