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 $\|\boldsymbol{x}\|_\infty = \max_{1\le i\le n} |x_i|$ is indeed a norm on $\mathbb{R}^n$.
(b) Omitted
(c) Omitted

Answer:

(a)

Proof. To show that $\|\boldsymbol{x}\|_{i}nfty$ is a norm, we check the following properties:

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

$$\|\boldsymbol{x}\|_\infty = \max_{1\le i\le n} |x_i| \ge 0.$$

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

You also need to show $x = 0 \Rightarrow \|x\|_\infty = 0$

-1

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

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

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

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

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

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

□

(b)

Proof. we have

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

s.t.

$$\|\boldsymbol{x}\|_\infty \le \|\boldsymbol{x}\|_1 \le n \cdot \|\boldsymbol{x}\|_\infty$$

□

(c)

Proof. we have

$$|\boldsymbol{x^T y}| = |\sum_{i=1}^{n} x_i \cdot y_i| \le \sum_{i=1}^{n} |x_i \cdot y_i| \le \sum_{i=1}^{n} |x_i||y_i| \le \sum_{i=1}^{n} (\max_{1\le i\le n} |y_i|) \cdot |x_i| = \max_{1\le i\le n} |y_i| \cdot \sum_{i=1}^{n} |x_i| = \|\boldsymbol{x}\|_1 \|\boldsymbol{y}\|_\infty$$

s.t.

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

□

Question 2:

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

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

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

Answer:

(a)

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

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

since $\|\boldsymbol{Ax}\|_2$ is non-negative for all $\boldsymbol{x}$.

2. If $\|\boldsymbol{A}\|_2 = 0$, then:

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

which means $\boldsymbol{Ax} = \boldsymbol{0}, \forall \boldsymbol{x}$.
Choossing $\boldsymbol{x}$ as the standard basis vectors(e.g. $e_1 = [1, 0, 0, ...], e_i = [0, ..., 1, 0, ...]$),
we conclude that $\boldsymbol{A} = \boldsymbol{0}$.
Conversely, if $\boldsymbol{A} = \boldsymbol{0}$, then clearly $\|\boldsymbol{A}\|_2 = 0$

3. For any scalar α,

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

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

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

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

$$\|\boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{x}\|_2 \leq \|\boldsymbol{A}\boldsymbol{x}\|_2 + \|\boldsymbol{B}\boldsymbol{x}\|_2.$$

Thus,

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

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. **$\|\boldsymbol{x}\|_2 \neq 0$**:
for any $\boldsymbol{x} \in \mathbb{R}^n$, we know:

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

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

$$\|\boldsymbol{A}(\frac{\boldsymbol{x}}{\|\boldsymbol{x}\|_2})\|_2 = \frac{1}{\|\boldsymbol{x}\|_2}\|\boldsymbol{A}\boldsymbol{x}\|_2 = \|\boldsymbol{A}\boldsymbol{u}\|_2 \leq \|\boldsymbol{A}\|_2 \implies \|\boldsymbol{A}\boldsymbol{x}\|_2 \leq \|\boldsymbol{A}\|_2\|\boldsymbol{x}\|_2$$

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

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

Conclusion:

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

□

(c)
Proof. For all matrix $\boldsymbol{M} \in \mathbb{R}^{n\times m}$, we have:

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

and

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

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

$$\|\boldsymbol{B}\boldsymbol{x}\|_2 \leq \|\boldsymbol{B}\|_2$$

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

$$\|\boldsymbol{A}\boldsymbol{B}\boldsymbol{x}\|_2 \leq \|\boldsymbol{A}\|_2\|\boldsymbol{B}\boldsymbol{x}\|_2 \leq \|\boldsymbol{A}\|_2\|\boldsymbol{B}\|_2$$

So we can conclude:

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

□

Question 3:

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

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

Answer:

Proof. **To prove $\|\boldsymbol{A}\|_2 \leq \|\boldsymbol{A}\|_F$**
We have,

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

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

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

We know:

$$\left(\sum_{i=1}^n a_i b_i\right)^2 \leq \sum_{i=1}^n a_i^2 \sum_{i=1}^n b_i^2$$

$$\left(\sum_{j=1}^n a_{i,j} x_j\right)^2 \leq \sum_{j=1}^n a_{i,j}^2 \sum_{j=1}^n x_j^2 = \sum_{j=1}^n a_{i,j}^2$$

s.t.:

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

s.t.:

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

**To prove $\|\boldsymbol{A}\|_F \leq \sqrt{n}\|\boldsymbol{A}\|_2$ **
Let $\boldsymbol{e}_i$ be the standard basis vector in $\mathbb{R}^n$.

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

We know:

$$\|\boldsymbol{A}\boldsymbol{e}_i\|_2^2 \leq \|\boldsymbol{A}\|_2^2$$

$$\sum_{i=1}^n \|\boldsymbol{A}\boldsymbol{e}_j\|_2^2 \leq n\|\boldsymbol{A}\|_2^2$$

s.t.

$$\|\boldsymbol{A}\|_F^2 \leq n\|\boldsymbol{A}\|_2^2$$

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

$\square$

Question 4:

4. A magic square $\boldsymbol{M}_n$ 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 $\boldsymbol{M}_n$, so that $a_n = n(n^2+1)/2$. Let $\boldsymbol{d}$ denote a vector in $\mathbb{R}^n$ with each element equal to 1.
(a) Determine $\boldsymbol{M}_n\boldsymbol{d}$ and $\boldsymbol{d}^T\boldsymbol{M}_n$. Conclude that a_n is an eigenvalue of $\boldsymbol{M}_n$.
(b) Show that the row and column sums of $\boldsymbol{M}_n^2$ are all the same.
(c) Determine $\|\boldsymbol{M}_n\|_2$.

Answer:

(a)

Proof. Assume that $\boldsymbol{M}_n = [\boldsymbol{c}_1 \cdots \boldsymbol{c}_i \cdots \boldsymbol{c}_n]$, each $\boldsymbol{c}$ indicates the column of $\boldsymbol{M}_n$.
We know:

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

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

$$\boldsymbol{d}^T\boldsymbol{M}_n = \boldsymbol{d}^T[\boldsymbol{r}_1 \cdots \boldsymbol{r}_n]^T = a_n\boldsymbol{d}^T$$

Since $\boldsymbol{M}_n\boldsymbol{d} = a_n\boldsymbol{d}$, s.t. a_n is an eigenvalue of $\boldsymbol{M}_n$ with an eigenvector $\boldsymbol{d}$.
Since:

$$\boldsymbol{d}^T\boldsymbol{M}_n = a_n\boldsymbol{d}^T$$

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

(b)

Proof. We know for each row in $\boldsymbol{M}_n^2$, it can be expressed by:

$$\boldsymbol{r}_i\boldsymbol{M}_n$$

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

$$\boldsymbol{r}_i\boldsymbol{M}_n\boldsymbol{d} = \boldsymbol{r_i}a_n\boldsymbol{d} = a_n\sum_{j=1}^{n} r_{ij} = a_n^2$$

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

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

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

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

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

(c)

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

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

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

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

And we know:

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

-3

And:

$$\frac{a_n^2}{\|\boldsymbol{M}_n\|_F^2} = \frac{3}{2} \cdot \frac{n^2+1}{2n^2+1} = \frac{3}{2} \cdot (1 - \frac{1}{2+\frac{1}{n^2}}) \geq \underline{\lim_{n^2 \to \infty} \frac{3}{2}} \cdot (1 - \frac{1}{2+\frac{1}{n^2}}) = 1$$

s.t.:

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

The Proof is only for the case of $n \to \infty$.

s.t.:

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

And:

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

s.t.:

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

□

Question 5:

5. Let $a_1, a_2, \cdots, a_m$ be m given real numbers. Prove that a median of $a_1, a_2, \cdots, a_m$ 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 $a_1 \le a_2 \le \cdots \le a_m$ For $b \in \mathbb{R}$ and $a_1 \le a_k \le b \le b+d \le a_{k+1} \le a_m$ in the ascending ordered sequence $a_1, \cdots, a_m$, we have:

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

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

$$f(b+d) - f(b) = (2k - m)d$$

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 $a_{\frac{m}{2}} \le b \le a_{\frac{m}{2}+1}$
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 $\sum_{i=1}^{m} |a_i - b|$ $\square$

Question 6:

6. Suppose that the vector $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N$ in $\mathbb{R}^n$ are clustered using the K-means algorithm, with group representative $\boldsymbol{z}_1, \cdots \boldsymbol{z}_k$.
(a) Suppose the original vectros $\boldsymbol{x}_i$ are nonnegative, *i.e.*, their entries are nonnegative. Explain why the representative $\boldsymbol{z}_j$ output by the K-means algorithm are also nonnegative.
(b) Suppose the original vectors $\boldsymbol{x}_i$ represent proportions, *i.e.*, their entries are nonnegative and sum to one. (This is the case when $\boldsymbol{x_i}$ are word count hisograms, for example.) Explain why the representatives $\boldsymbol{z}_j$ 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 $\boldsymbol{x}_i$ are Boolean, *i.e.*, their entries are either 0 or 1. Give an interpretation of $(\boldsymbol{z}_j)_i$, the i-th entry of the j group representative.

Answer:

(a)

Proof. We know:

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

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

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

s.t. $\boldsymbol{z}_j$ is nonnegative.

$\square$

(b)

Proof. We know:

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

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

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

$\square$

(c)

Proof.

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

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

$\square$