Topic 3 – Metric in Vector Spaces

We have seen that any linear combination of vectors is also in that Vector Space. But this property is not enough in many applications. For example in machine learning tasks like Clustering. Here we need to group different input vectors into several groups. Then we able to compare the closeness of the two vectors. The vector space self don't tell us about the closeness of two vectors. So we need a <u>metric</u> on Vector Spaces.

Let V be a Vector Space.
Let $x, y \in V$. Then

$$\text{dist}(x, y) = \text{dist}(x-y,\ x-y) \quad \text{(Shift invariant}$$
$$= \text{dist}(x-y,\ 0) \ (= \text{length of } x-y)$$

To define distance, we only need to define the length of vectors.

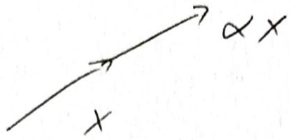
To define the length:

Let $x \in V$. Denote $\|x\|$ be the length of x.
Properties of $\|x\|$:

1. $\|x\| \geq 0$ (the length should be non-negative)

   $\|x\| = 0 \iff x = 0$. (Only $\vec{0}$ has 0 length)

2.



$$\|\alpha x\| = |\alpha| \cdot \|x\| \quad , \forall x \in V.$$

3. $$\|x+y\| \leq \|x\| + \|y\| \qquad (\Delta\text{-Inequality})$$

<u>Definition.</u>

Let V be a Vector Space. A norm on V is a function $\|.\| : V \to \mathbb{R}$ such that

1. $\|x\| \geq 0 \ \forall x \in V$ and $\|x\| = 0 \iff x = 0 \ \forall x \in V$
2. $\|\alpha x\| = |\alpha| \, \|x\| \quad , \forall \alpha \in \mathbb{R}, \ x \in V$
3. $\|x+y\| \leq \|x\| + \|y\| \ , \ \forall x, y \in V$

Remark: • $\|x\|$ is a length of x

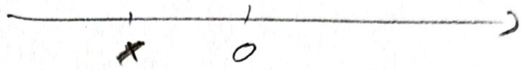
• $dist(x, y) = \|x - y\|$

Example 1:

$\mathbb{R}$ is a Vector Space over $\mathbb{R}$.

- $\|x\| := |x|$ defines a norm on $\mathbb{R}$.

(Norm is a generalization of absolute value function)



- $\|x\| := \frac{1}{2}|x|$ is also a norm on $\mathbb{R}$

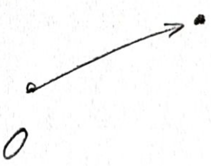
$\vdots$

- $\|x\| := c|x|, \quad c>0$ is a norm on $\mathbb{R}$.

There are many norms on the same Vector Space. (infinetely)

Example 2: $\mathbb{R}^n$ is a Vector Space.

- Euclidean norm (2-norm)



$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

$$\|x\|_2 = \left(\sum_{i=1}^{n} x_i^2 \right)^{1/2}$$

- check $\|.\|_2$ is indeed a Norm.

Proof:

1. $\|x\|_2 \geq 0$

and

$$\|x\|_2 = 0 \iff \sum_{i=1}^{n} x_i^2 = 0 \iff x_i^2 = 0 \ \forall i$$

$$\iff x_i = 0 \ \forall i = 1, \dots, n \iff x = 0.$$

2. $$\|\alpha x\|_2 = \left(\sum_{i=1}^{n} (\alpha x_i)^2\right)^{1/2} = \left(\alpha^2 \sum_{i=1}^{n} x_i^2\right)^{1/2}$$

$$= (\alpha^2)^{1/2} \cdot \left(\sum_{i=1}^{n} x_i^2\right)^{1/2}$$

$$= |\alpha| \cdot \|x\|_2 .$$

3. Do it later!

$\square$

- 1-Norm:

$$\|x\|_1 := \sum_{i=1}^{n} |x_i| \quad \forall x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

- p-Norm:

$$\|x\|_p := \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p}, \quad p \geq 1$$

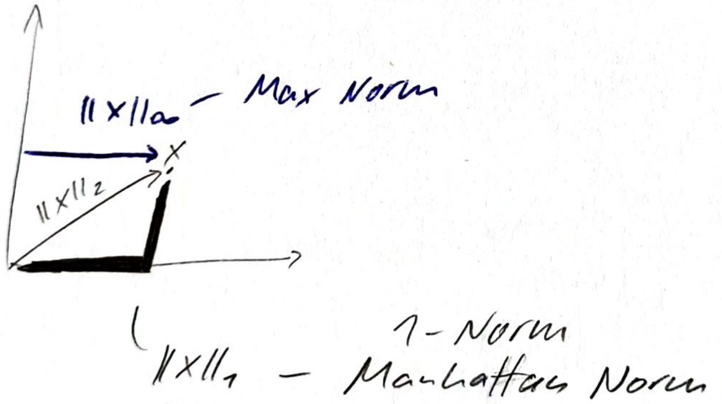
$\|\cdot\|_p$ is a Norm on $\mathbb{R}^n$ if $p \geq 1$.

How about $p \to +\infty$?

We can check $\forall x \in \mathbb{R}^n$: $\lim_{p\to\infty} \|x\|_p = \max_{i=1,\dots,n} |x_i|$

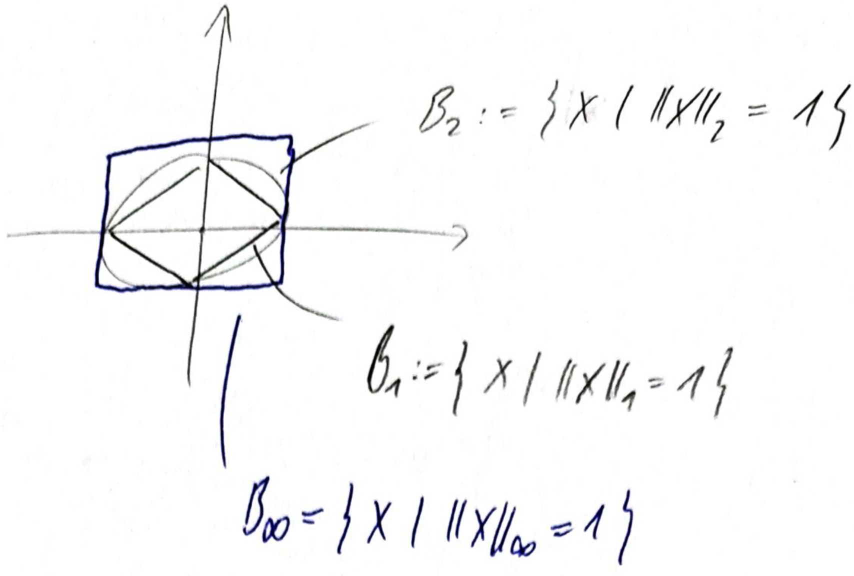
So we define.

$$\|x\|_\infty = \max_{i=1,\dots,n} |x_i|$$ – it is indeed a norm on $\mathbb{R}^n$.



In machine learning:

Comparison of unit balls of p-norms.



- We have

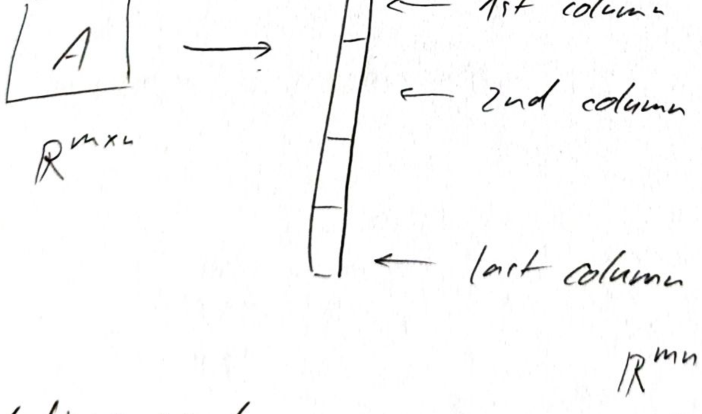
$$\boxed{\|x\|_p \leq \|x\|_q \quad \text{if } p \geq q}$$

- We have other norms on $\mathbb{R}^n$ than p-norms.

<u>Example 3:</u>

$\mathbb{R}^{m \times n}$ is a vector space.

- $\mathbb{R}^{m \times n}$ can be viewed as $\mathbb{R}^{mn}$



We can define vector norm p-norm for <u>matrices</u>.

- $p=1$:

$$\|A\|_{1,vec} = \sum_{i=1}^{n}\sum_{j=1}^{n} |a_{ij}|$$

- $p=2$:

$$\|A\|_{2,vec} = \left(\sum_{i=1}^{m}\sum_{j=1}^{n} |a_{ij}|^2\right)^{1/2}$$

This norm is widely known as <u>Frobenius Norm</u>

$$\|A\|_F := \|A\|_{2,vec}$$

- $p=\infty$:

$$\|A\|_{\infty,vec} := \max_{j=1,\dots,m}\max_{i=1,\dots,n} |a_{ij}|$$

- <u>$R^{m\times n}$ can be viewed as linear transforms $R^n \to R^m$.</u>

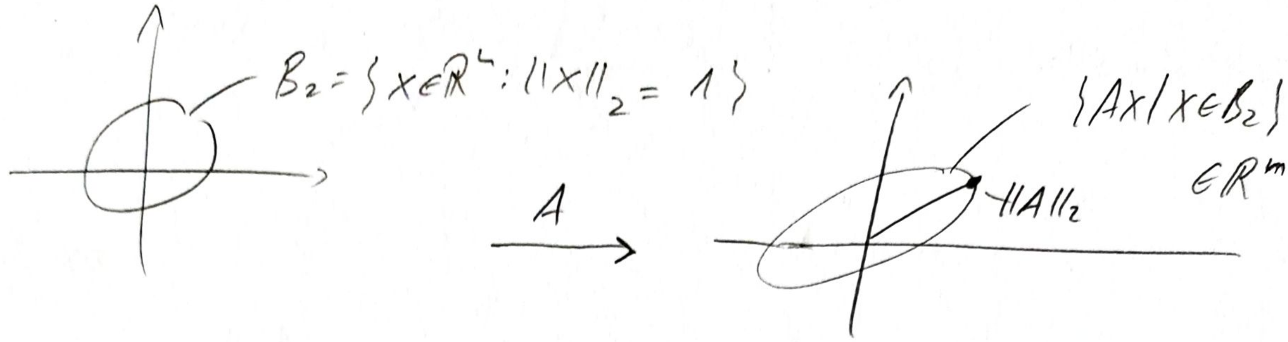
(i.e. $A \in \mathbb{R}^{m\times n} \leftrightarrow A: \mathbb{R}^n \to \mathbb{R}^m$)

We define matrix p-norm by:

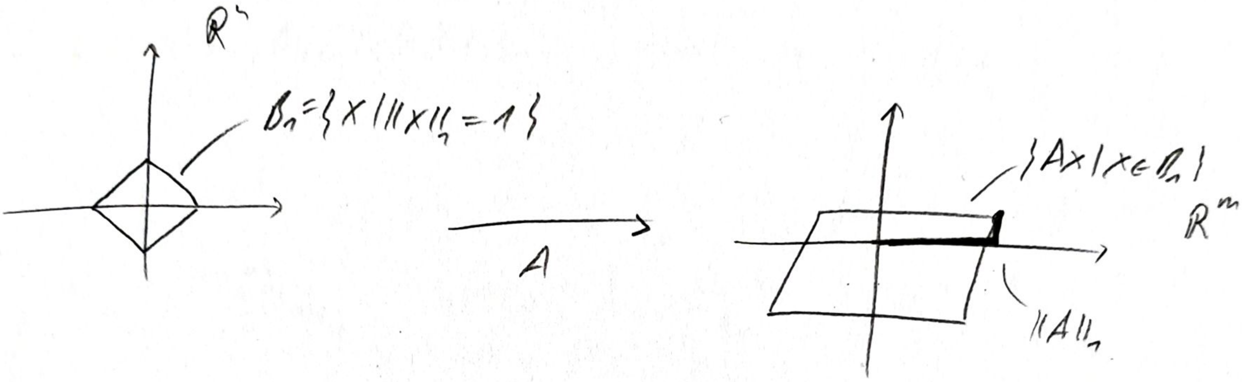
$$\|A\|_p := \max_{\substack{x\neq 0\\ x\in\mathbb{R}^n}} \frac{\|Ax\|_p}{\|x\|_p} \underset{\uparrow}{=} \max_{\|x\|_p=1} \|Ax\|_p$$

Can prove

- $p=2$:



- $p=1$:



- $p=\infty$: similar.

- $p=2$:

$$\|A\|_2 = \max_{\|x\|_2=1} \|Ax\|_2$$

$$\Leftrightarrow \|A\|_2^2 = \max_{\|x\|_2=1} \|Ax\|_2^2$$

$$= \max_{\|x\|_2=1} x^T A^T A x$$

$$= \text{max eigenvalue of } A^T A.$$

$$\Leftrightarrow \|A\|_2 = (\text{max eig of } A^T A)^{1/2} = \text{max singular value of } A.$$

- $\|.\|_2$ is also called the operator norm of A.

- $p=1$:

$$\|A\|_1 = \max_{j=1,\dots,n}\left(\sum_{i=1}^{m}|a_{ij}|\right)$$ (max column 1-norms)

- $p=\infty$:

$$\|A\|_\infty = \max_{i=1,\dots,m}\left(\sum_{j=1}^{n}|a_{ij}|\right)$$ (max row 1-norms)

Also we can use different norms in $\mathbb{R}^n$ and $\mathbb{R}^m$.

$\mathbb{R}^n$: p-norm, $\mathbb{R}^m$: q-norm

$$\Rightarrow \quad \|A\|_{p\to q} = \max_{\|x\|_p=1} \|Ax\|_q$$

In ML sometimes we need to find a <u>low-rank matrix</u>. For low-rank-matrix, if the factor has some special structure, we may use this kind of mixed p-q-Norms to solve our ML task.

Example 4:

$C[a,b] = \{ f \mid f$ is continuous function on $[a,b] \}$

is a Vector Space.

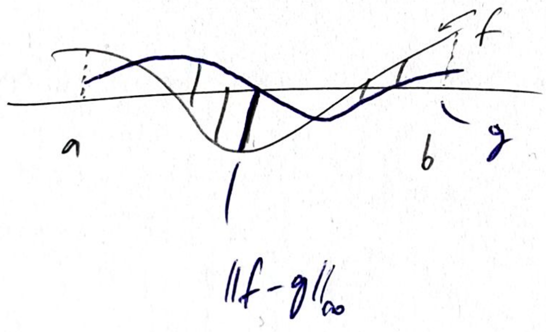
$\forall f \in C[a,b]$, define

$$\|f\|_\infty = \max_{t\in[a,b]} |f(t)|.$$

We can check $\|\cdot\|_\infty$ is a norm on $C[a,b]$.

Then $\forall f, g \in C[a,b]$ their distance can be defined by

$$\|f-g\|_\infty = \max_{t\in[a,b]} |f(t) - g(t)|$$



Some other norms on $C[a,b]$.

- $\|f\|_1 = \int_a^b |f(t)|\, dt$ ... $\|f\|_p = \left(\int_a^b |f(t)|^p dt\right)^{1/p}$

$p \geq 1$!

- $\|f\|_2 = \left(\int_a^b |f|^2 dt\right)^{1/2}$

Example 5:

$l_\infty = \{ a \mid a$ is an infinite sequence and $\exists C > 0$ s.t. $|a_i| \le C \ \forall i \}$

is a Vector Space.

· $\forall a \in l_\infty$, define

$$\|a\|_\infty := \sup_i |a_i|$$

Remarks:

1. For the same Vector Space, we can define infinitely many norms on it.

2. A common technique in Machine Learning is to optimize some norm of unknown vector. Different norms leed to different results. For example: Sparse Vector we use $\|.\|_1$.