Norm Space

CE282: Linear Algebra

Computer Engineering Department

Sharif University of Technology

Hamid R. Rabiee

Maryam Ramezani

Vector Norms

- p-norm:

$$\|x\|_p = (|x_1|^p + |x_2|^p + \ \dots + |x_n|^p)^{\frac{1}{p}}$$

$$subject\ to\ p \geq 1$$

- What is the shape of $\|x\|_p = 1$?
- Properties?

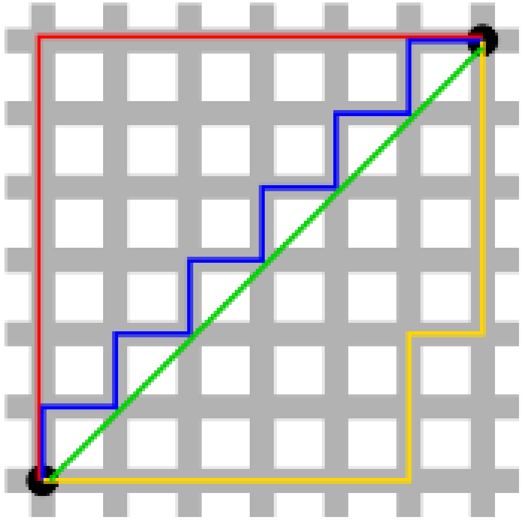
- 1-norm(l_1):

$$\|x\|_1 = (|x_1| + |x_2| + \ldots + |x_n|)$$

- What is the shape of $\|x\|_1 = 1$?
- The distance between two vectors under the l_1 norm is also referred to as the Manhattan Distance.
- Properties?

Example

l_1 distance between $(0, 1)$ and $(1, 0)$?



Norm Derivations

- Square of l_2

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \\ \dots \\ \\ x_n \end{bmatrix}$$

$$\|x\|_2^2 = x_1^2 + x_2^2 + \cdots + x_n^2$$

$$\Rightarrow \begin{cases} \dfrac{d\|x\|_2^2}{dx_1} = 2x_1 \\ \dfrac{d\|x\|_2^2}{dx_2} = 2x_2 \\ \\ \dots \\ \\ \dfrac{d\|x\|_2^2}{dx_n} = 2x_n \end{cases}$$

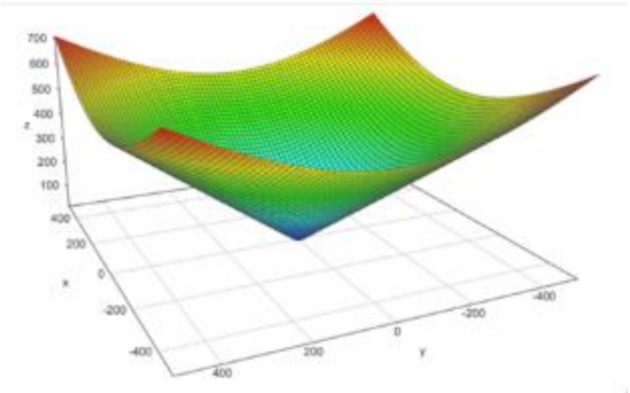
- l_2

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \\ \dots \\ \\ x_n \end{bmatrix} \qquad \|x\|_2 = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2} = \left(x_1^2 + x_2^2 + \cdots + x_n^2\right)^{\frac{1}{2}}$$

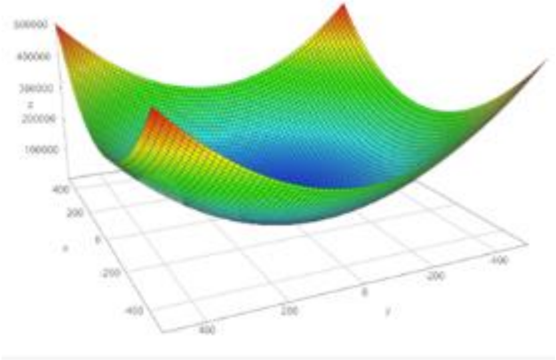
$$\begin{aligned}
\frac{d\|x\|_2}{dx_1} &= \frac{1}{2}\left(x_1^2 + x_2^2 + \ \dots + x_n^2\right)^{\frac{1}{2}-1} . \frac{d}{dx_1}\left(x_1^2 + x_2^2 + \ \dots + x_n^2\right) \\
&= \frac{1}{2}\left(x_1^2 + x_2^2 + \ \dots + x_n^2\right)^{-\frac{1}{2}} . \frac{d}{dx_1}\left(x_1^2 + x_2^2 + \ \dots + x_n^2\right) \\
&= \frac{1}{2} . \frac{1}{\left(x_1^2 + x_2^2 + \ \dots + x_n^2\right)^{\frac{1}{2}}} . \frac{d}{dx_1}\left(x_1^2 + x_2^2 + \ \dots + x_n^2\right) \\
&= \frac{1}{2} . \frac{1}{\left(x_1^2 + x_2^2 + \ \dots + x_n^2\right)^{\frac{1}{2}}} . 2. x_1 \\
&= \frac{x_1}{\left(x_1^2 + x_2^2 + \ \dots + x_n^2\right)^{\frac{1}{2}}}
\end{aligned}$$

$$\Rightarrow \begin{cases}
\dfrac{d\|x\|_2}{dx_1} = \dfrac{x_1}{\left(x_1^2 + x_2^2 + \ \dots + x_n^2\right)^{\frac{1}{2}}} \\
\dfrac{d\|x\|_2}{dx_2} = \dfrac{x_2}{\left(x_1^2 + x_2^2 + \ \dots + x_n^2\right)^{\frac{1}{2}}} \\
\qquad \dots \\
\dfrac{d\|x\|_2}{dx_n} = \dfrac{x_n}{\left(x_1^2 + x_2^2 + \ \dots + x_n^2\right)^{\frac{1}{2}}}
\end{cases}$$

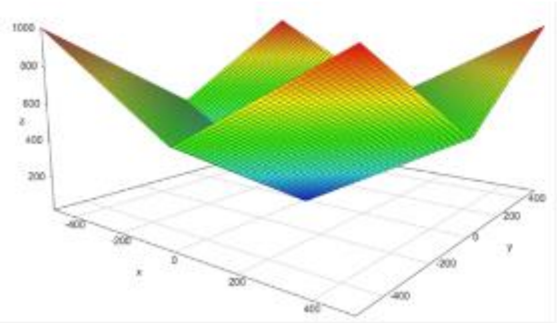
Norm Comparisons



l_2 norm

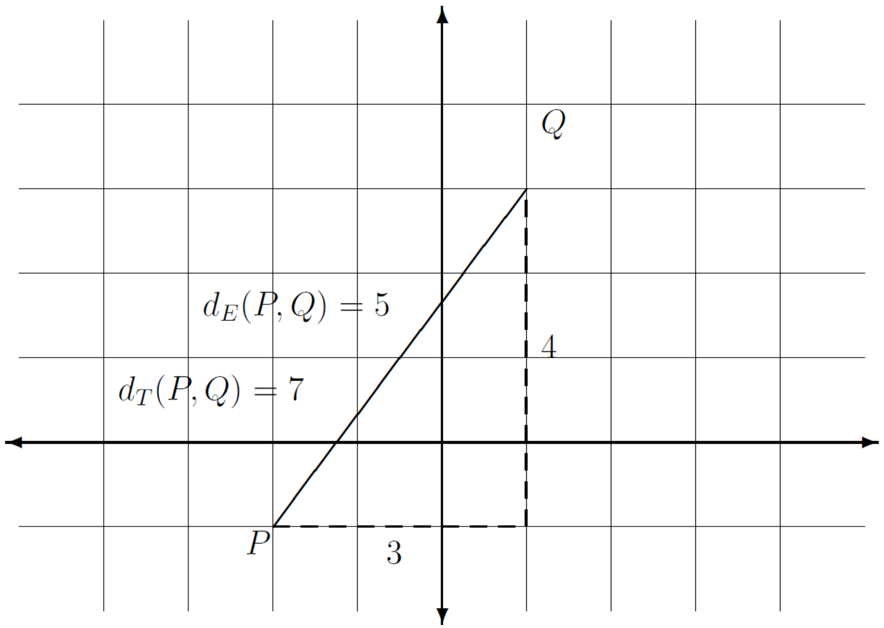


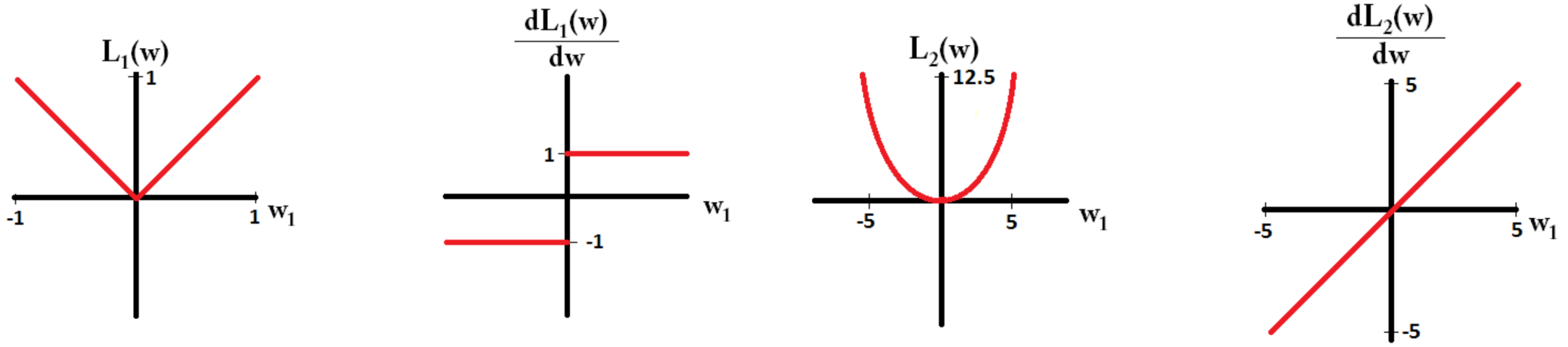
Square l_2 norm



l_1 norm

L1 and L2 norm comparisons

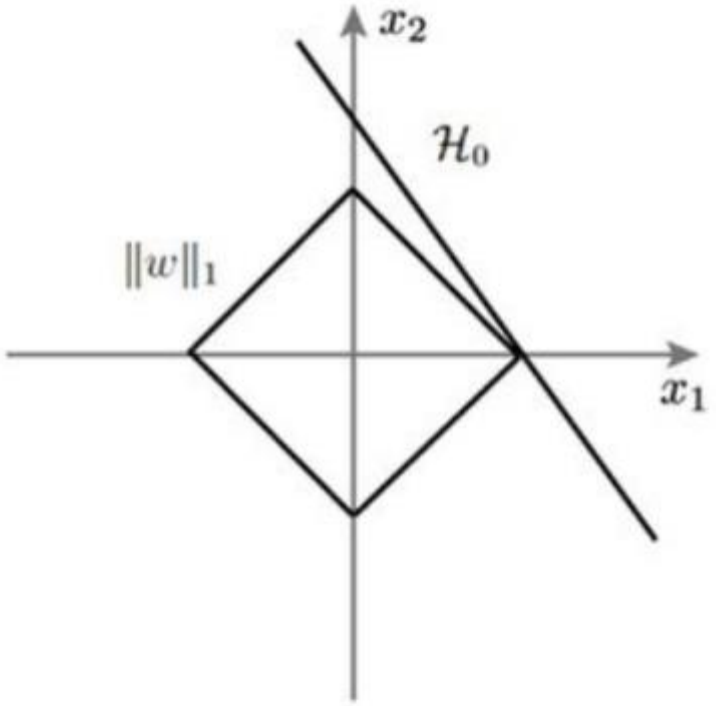




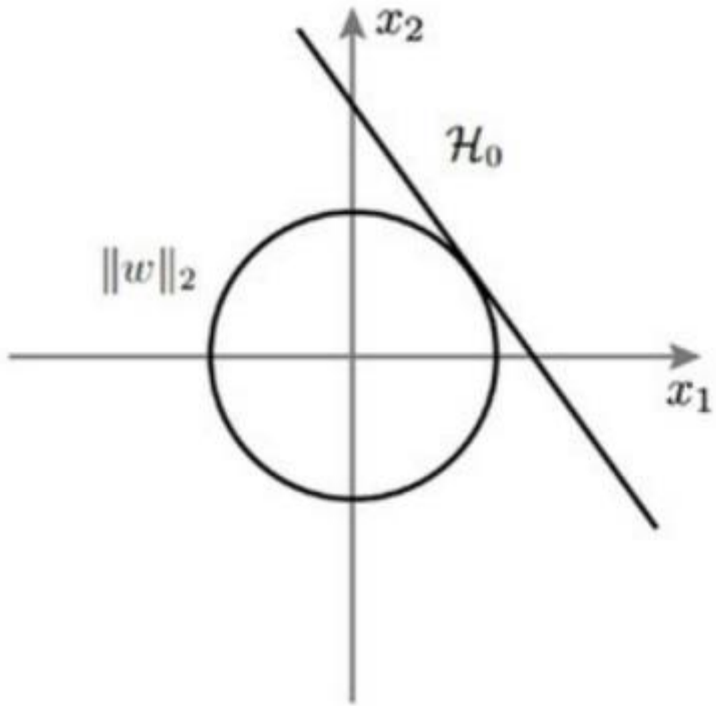
L1 and L2 norm comparisons

- Robustness is defined as resistance to outliers in a dataset. The more able a model is to ignore extreme values in the data, the more robust it is.
- Stability is defined as resistance to horizontal adjustments. This is the perpendicular opposite of robustness.
- Solution numeracy
- Computational difficulty
- Sparsity

Why is l_1 supposed to lead to sparsity than l_2?



l_1 reqularization



l_2 reqularization

Vector Norms

- ∞-norm(l_∞)(max norm):

$$l_\infty = \max(|x_1|, |x_2|, \dots, |x_n|)$$

- What is the shape of $|x|_\infty = 1$?
- Properties?

Vector Norms

- ❑ $\frac{1}{2}$-norm($l_{\frac{1}{2}}$)

- ❑ What is the shape of $|x|_{\frac{1}{2}} = 1$?

- ❑ Properties?

Vector Norms

- 0-norm(l_0):

$$\|x\|_0 = \lim_{\alpha \to 0^+} \|x\|_\alpha = \left(\sum_{k=1}^{n} |x|^\alpha\right)^{\frac{1}{\alpha}} = \sum_{k=1}^{n} 1_{(0,\infty)}(|x|)$$

- 0-norm, defined as **the number of non-zero elements in a vector**, is an ideal quantity for feature selection. However, minimization of 0-norm is generally regarded as a combinatorially difficult optimization

- $\|x\|_0 = \sum_{x_i \neq 0} 1$

Vector Norms

- Is 0-norm a valid norm?

- What is the shape of $\|x\|_0 = 1$?

Examples

- l_0 distance between $(0, 0)$ and $(0, 5)$?
- l_0 distance between $(1, 1)$ and $(2, 2)$?
- (username, password)

Vector Norms

Class Activity

- l_0 distance between $(0, 0)$ and $(0, 5)$?
- l_0 distance between $(1, 1)$ and $(2, 2)$?
- (username, password)

Or go to the below link
https://forms.gle/xFHSDKJDq1KoL4Kx6

Timer: (2:30 minutes)

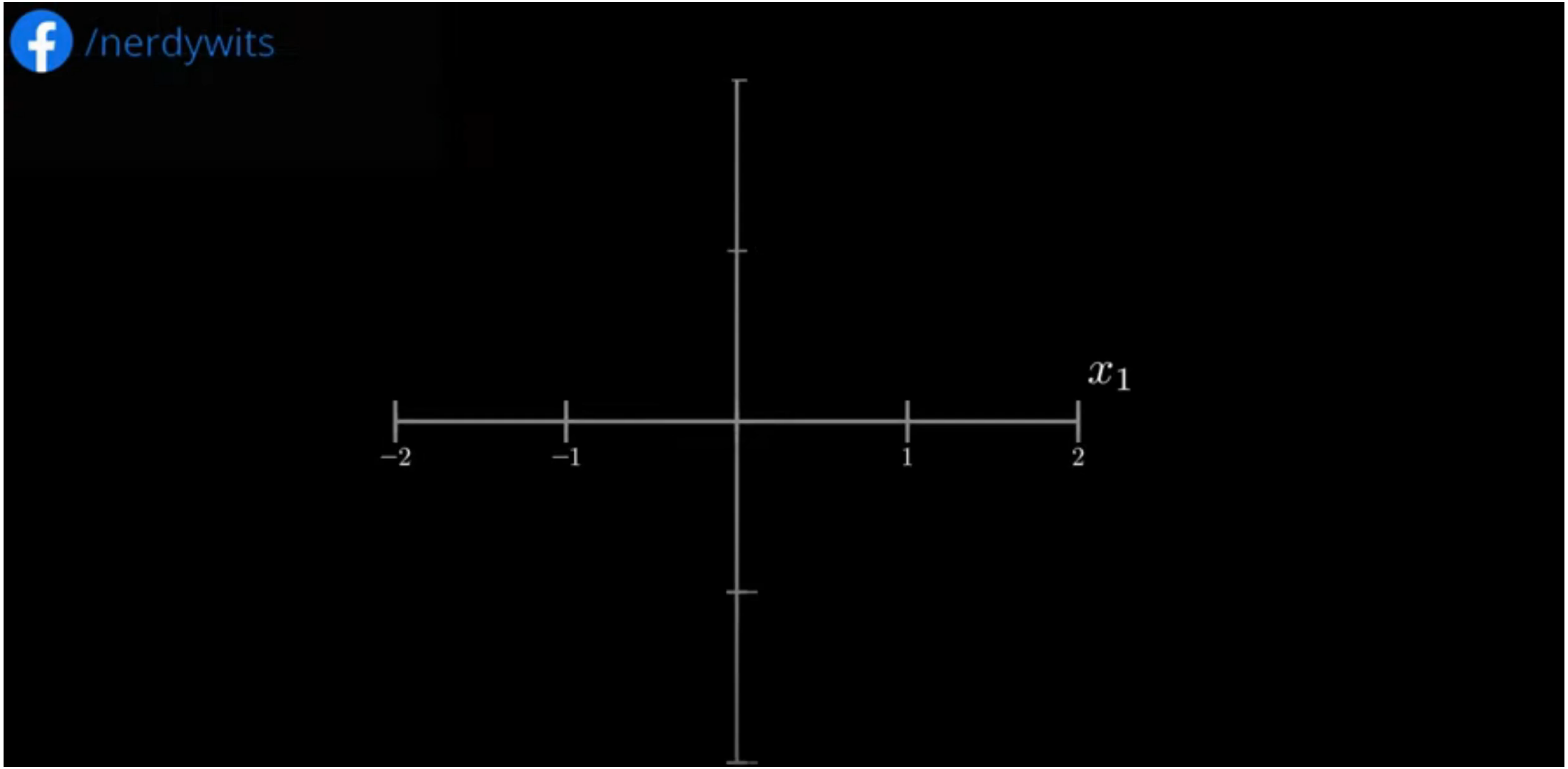
Examples

- l_0 distance between $(0,0)$ and $(0,5)$?
- l_0 distance between $(1,1)$ and $(2,2)$?
- (username, password)

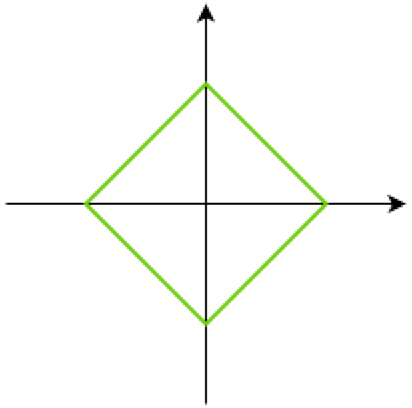
Solution

- 1
- 2
- When l_0 is 0, then we can infer that username and password is a match and we can authenticate the user.

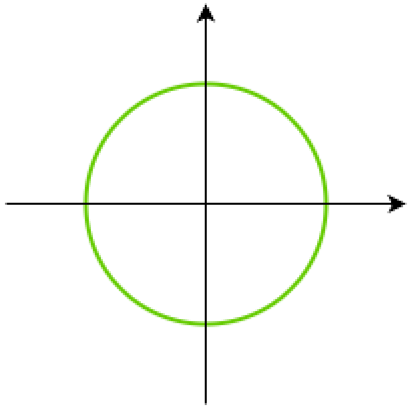
Vector Norms Shapes



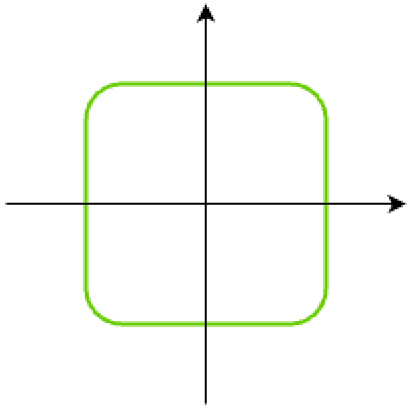
- For $\boldsymbol{p \geq 1}$, $\boldsymbol{l_p}$ norm is convex



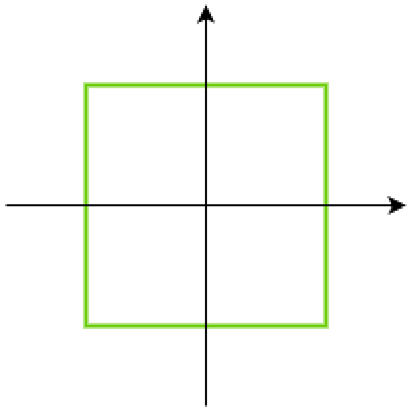
$\|x\|_1 = 1$



$\|x\|_2 = 1$



$\|x\|_p = 1$



$\|x\|_\infty = 1$

L1-L2 norm inequality

Theorem

For all $x \in \mathbb{R}^d$:

$$||x||_2 \leq ||x||_1 \leq \sqrt{d}||x||_2$$

Proof

Max norm inequality

Theorem

For all $x \in \mathbb{R}^d$:

$$||x||_\infty \leq ||x||_1 \leq d||x||_\infty$$

$$||x||_\infty \leq ||x||_2 \leq \sqrt{d}||x||_\infty$$

Proof

Conclusion

- By a normed linear space (briefly normed space) is meant a real or complex vector space E in which every vector x is associated with a real number $|x|$, called its absolute value or norm, in such a manner **that the properties** $(a') - (c')$ holds. That is, for any vectors $x, y \subset E$ and scalar α we have:

i. $|x| \geq 0$

ii. $|x| = 0 \; iif \; x = \vec{0}$

iii. $|\alpha x| = |\alpha||x|$

iv. $|x + y| \leq |x| + |y|$

Inner product and norm

Theorem

Take any inner product $\langle \cdot, \cdot \rangle$ and define $f(x) = \sqrt{\langle x, x \rangle}$. Then f is a norm.

Proof

Note

Every inner product gives rise to a norm, but not every norm comes from an inner product. (Think about norm 2 and norm max)

References

- Linear Algebra and Its Applications, David C. Lay
- Introduction to Applied Linear Algebra Vectors, Matrices, and Least Squares
- https://www.youtube.com/watch?v=76B5cMEZA4Y