Norm of vector

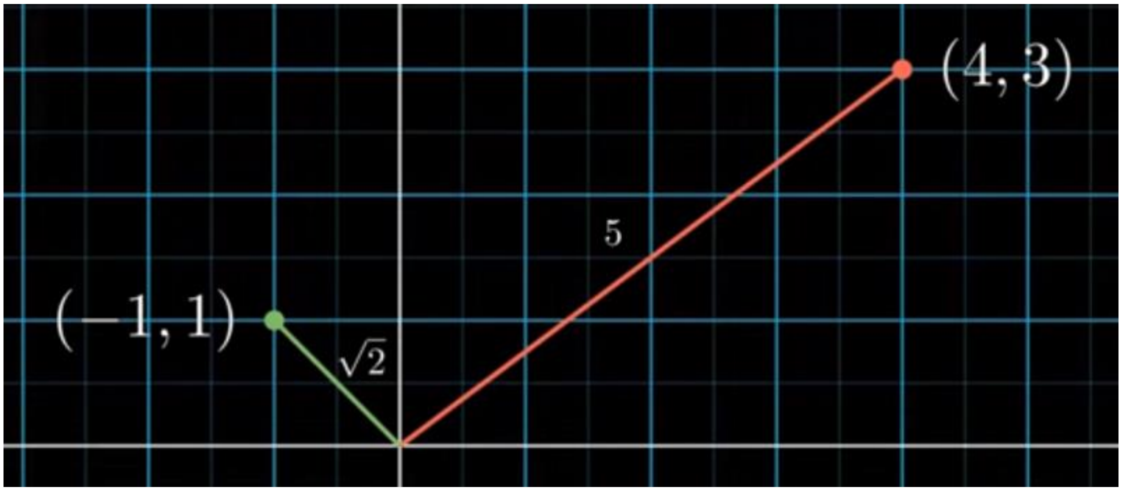
Definition and purpose

In mathematics, a **norm** is a function from a real or complex vector space to the nonnegative real numbers that behaves in certain ways like the distance from the origin:

- it commutes with scaling
- obeys a form of the triangle inequality $p(x + y) \leq p(x) + p(y)$
- and is zero only at the origin

For example, if we want to measure two vectors/points from the 2-D coordinate system, we may want to use the Euclidean Distance (also called l_2 norm), which is written as

$$l_2\ norm\ /\ \text{Euclidean Distance} = \left(\sum_{i=1}^{k} |X_i|^2\right)^{\frac{1}{2}}$$



To generalize, we change 2 into a variable n, so the equation becomes

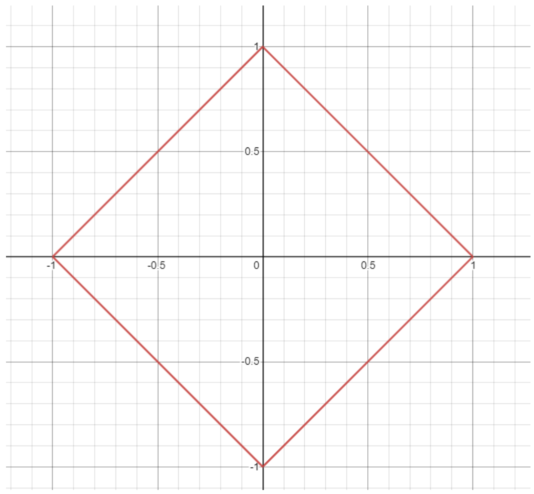
$$l_n\ norm\ /\ Generalized\ Norm = \left(\sum_{i=1}^{k} |X_i|^n\right)^{\frac{1}{n}}$$

If we set $n = 1$, we get the Manhattan distance

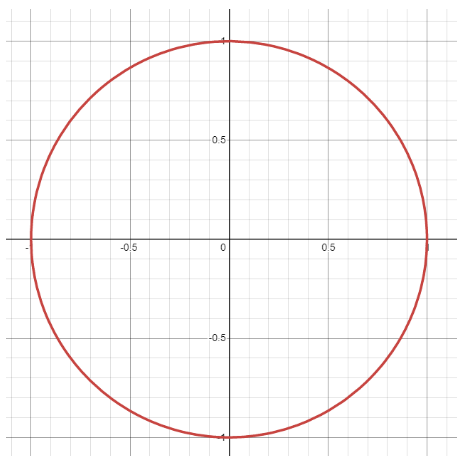
$$l_1\ norm\ /\ \text{Manhattan distance} = \left(\sum_{i=1}^{k} |X_i|^1\right)^{\frac{1}{1}}$$

Usage

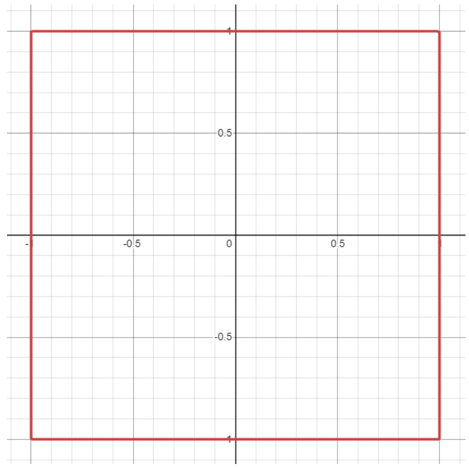
It can be utilized as a constraint. For example, if we consider the 2-D space, with $l_1\ norm \leq 1$, it looks like



For $l_2\ norm \leq 1$, it looks like



For $l_\infty\ norm \leq 1, it\ looks\ like$

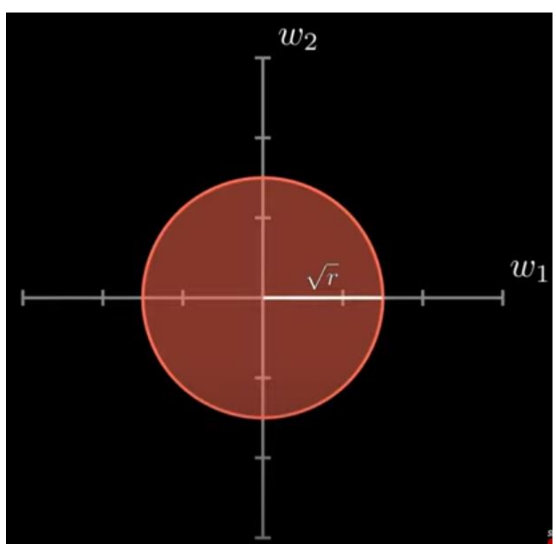


This property has been widely used in multiple theorems. Recall the definition of the Mean Squared Error

$$Mean\ Squared\ Error = \frac{1}{m}\sum_{i=1}^{m}\left(Y_i - \widehat{Y_\iota}\right)^2 = \frac{1}{m}\left[l_2\ norm\left(Y - \hat{Y}\right)\right]^2$$

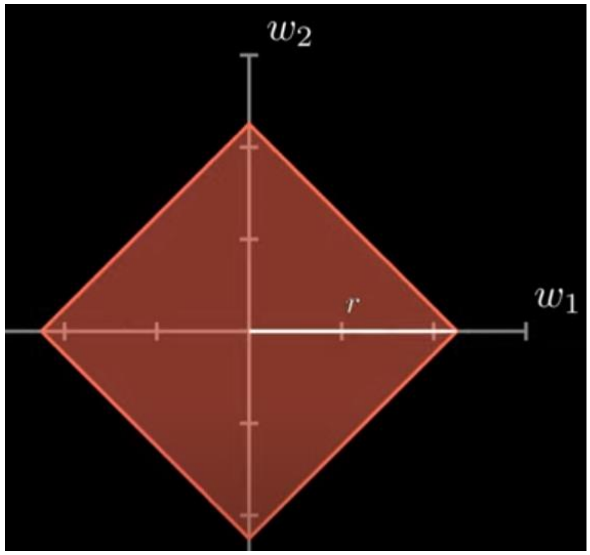
For Ridge Regression Constraint

$$\sum_{i=1}^{k} w_i^2 \le r\ , equal\ to\ l_2\ norm(w) \le \sqrt{r}$$



For Lasso Regression Constraint

$$\sum_{i=1}^{k} |w_i| \leq r \text{, } equal\ to\ l_1\ norm(w) \leq r$$



Norm of Matrix

Definition and purpose

The norm of a matrix is **a real number** which is a **measure of the magnitude of the matrix**.

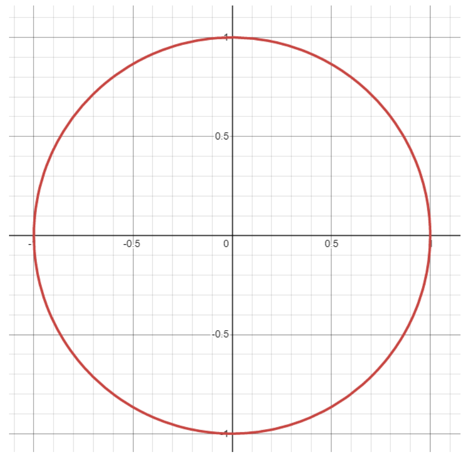
Recall that we use the norm of vectors to compare the “length” of vectors. We also want to come out with a metric to measure the “size” of matrices. E.g.

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, B = \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix}$$

Because matrices are considered as linear transformations, so we define the norm of matrices as, with any given vector $\vec{u}$, it satisfies

$$||A||_2 = Max(A\vec{u})_2 \text{ with constraint } ||\vec{u}||_2 = 1$$

Recall that, $l_2\ norm\ of\ vector = 1\ means\ all\ vectors\ lies\ on\ the\ circle$



So, the $||A||_2$ means, after the transformation, the largest change of $l_2\ norm$ of vectors which lies on this circle

The norm of matrices shares the following properties:

1. $||A|| \geq 0\ for\ any\ square\ matrix\ A.$
2. $||A|| = 0\ if\ any\ only\ if\ the\ matrix\ A = 0$
3. $||kA|| = |k|\ ||A||\ for\ any\ scalar\ k.$
4. $||A + B|| \leq ||A|| + ||B||.$
5. $||AB|| \leq ||A||\ ||B||$

Calculation

The l_1 norm

The $l_1\ norm$ of a square matrix is the maximum of the absolute column sums

$$||A||_1 = \max_{1 \leq j \leq n}(\sum_{i=1}^{n} |a_{ij}|)$$

Put simply, we **sum the absolute values down each column** and then take the **biggest answer**.

e.g.

$$A = \begin{bmatrix} 5 & -4 & 2 \\ -1 & 2 & 3 \\ -2 & 1 & 0 \end{bmatrix}$$

Then

$$\begin{aligned} ||A||_1 &= \max(5 + 1 + 2, 4 + 2 + 1, 2 + 3 + 0) \\ &= \max(8,7,5) \\ &= 8 \end{aligned}$$

The infinity norm

The infinity-norm of a square matrix is the **maximum of the absolute row sums**.

$$||A||_{\infty} = \max_{1\le i\le n}\left(\sum_{i=1}^{n}|a_{ij}|\right)$$

e.g.

$$A = \begin{bmatrix} 5 & -4 & 2 \\ -1 & 2 & 3 \\ -2 & 1 & 0 \end{bmatrix}$$

Then

$$\begin{aligned} ||A||_{\infty} &= \max(5+4+2,1+2+3,2+1+0) \\ &= \max(11,6,3) \\ &= 11 \end{aligned}$$

The l_2 $(Euclidean)$ norm

The Euclidean norm of a square matrix is the **square root of the sum of all the squares of the elements.**

$$||A||_{E} = \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n}(a_{ij})^2}$$

e.g.

$$A = \begin{bmatrix} 5 & -4 & 2 \\ -1 & 2 & 3 \\ -2 & 1 & 0 \end{bmatrix}$$

Then

$$\begin{aligned} ||A||_{E} &= \sqrt{25+16+4+1+4+9+4+1+0} \\ &= \sqrt{64} \\ &= 8 \end{aligned}$$