Chapter 7

Norms and Distance Measures

Vector Norms

- Norms are functions which measure the magnitude or length of a vector.
- They are commonly used to determine similarities between observations by measuring the distance between them.
  - Find groups of similar observations/customers/products.
  - Classify new objects into known groups.
- There are many ways to define both distance and similarity between vectors and matrices!

Norms in General

A Norm, or distance metric, is a function that takes a vector as input and returns a scalar quantity ($f : \mathbb{R}^n \to \mathbb{R}$). A vector norm is typically denoted by two vertical bars surrounding the input vector, $\|\mathbf{x}\|$, to signify that it is not just any function, but one that satisfies the following criteria:

1. If c is a scalar, then

$$\|c\mathbf{x}\| = |c|\|x\|$$

2. The triangle inequality:

$$\|\mathbf{x}+\mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$$

3. $\|\mathbf{x}\| = 0$ if and only if $\mathbf{x} = 0$.
4. $\|\mathbf{x}\| \geq 0$ for any vector $\mathbf{x}$

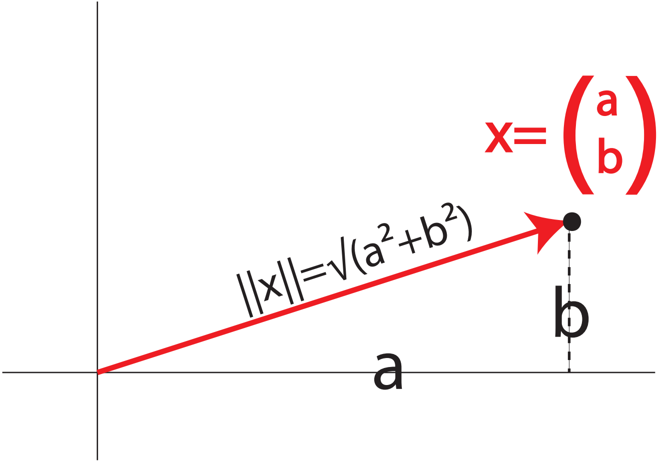
Euclidean Norm (Euclidean Distance)

The **Euclidean Norm**, also known as the 2-**norm** simply measures the Euclidean length of a vector (i.e. a point's distance from the origin). Let $\mathbf{x} = (x_1, x_2, \ldots, x_n)$. Then,

$$\|\mathbf{x}\|_2 = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}$$

- $\|\mathbf{x}\|_2 = \sqrt{\mathbf{x}^T\mathbf{x}}$.
- Often write $\| \star \|$ rather than $\| \star \|_2$ to denote the 2-norm, as it is by far the most commonly used norm.
- This is merely the "distance formula" from undergraduate mathematics, measuring the distance between the point $\mathbf{x}$ and the origin.

Euclidean Norm (Euclidean Distance)

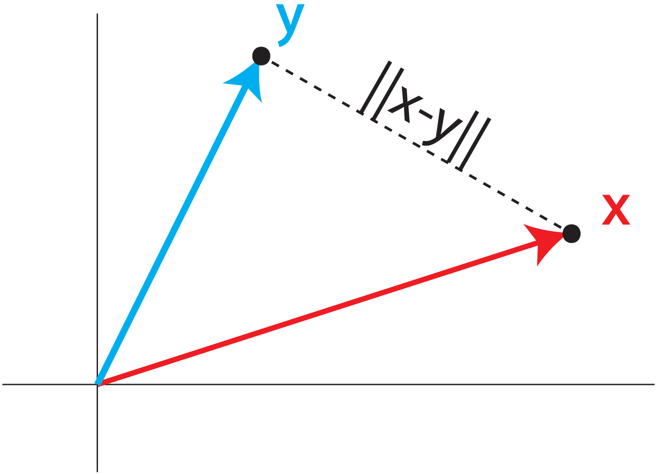


Length vs. Distance

Why do we care about the length of a vector? Two Reasons

- We will often want to make all vectors the same length (A form of standardization).
- The length of the vector $\mathbf{x} - \mathbf{y}$ gives the distance between $\mathbf{x}$ and $\mathbf{y}$.

Euclidean Distance



Euclidean Distance

$$\mathbf{x} - \mathbf{y} = \begin{pmatrix} x_1 - y_1 \\ x_2 - y_2 \\ \vdots \\ x_n - y_n \end{pmatrix}$$

$$\|\mathbf{x} - \mathbf{y}\| = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_n - y_n)^2}$$

Square Root Sum of Squared Differences between the two vectors.

Suppose I have two vectors in 3-space:

$$\mathbf{x} = (1, 1, 1) \text{ and } \mathbf{y} = (1, 0, 0)$$

Then the magnitude of **x** (i.e. its length or distance from the origin) is

$$\|\mathbf{x}\|_2 = \sqrt{1^2 + 1^2 + 1^2} = \sqrt{3}$$

and the magnitude of **y** is

$$\|\mathbf{y}\|_2 = \sqrt{1^2 + 0^2 + 0^2} = 1$$

and the distance between point **x** and point **y** is

$$\|\mathbf{x} - \mathbf{y}\|_2 = \sqrt{(1-1)^2 + (1-0)^2 + (1-0)^2} = \sqrt{2}.$$

Unit Vectors

In this course, we will regularly make use of vectors with length/magnitude equal to 1. These vectors are called **unit vectors**. For example,

$$\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad \mathbf{e}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

are all *unit vectors* because

$$\|\mathbf{e}_1\| = \|\mathbf{e}_2\| = \|\mathbf{e}_3\| = 1.$$

Simple enough!

Creating a unit vector

If we have some random vector, $\mathbf{x}$, we can always transform it into a *unit vector* by dividing every element by $\|\mathbf{x}\|$.
For example, take

$$\mathbf{x} = \begin{pmatrix} 3 \\ 4 \end{pmatrix}$$

Then, $\|\mathbf{x}\| = \sqrt{3^2 + 4^2} = \sqrt{25} = 5$. The new vector,

$$\mathbf{u} = \tfrac{1}{5} \begin{pmatrix} 3 \\ 4 \end{pmatrix}$$

is a unit vector:

$$\|\mathbf{u}\| = \sqrt{\left(\frac{3}{5}\right)^2 + \left(\frac{4}{5}\right)^2} = \sqrt{\frac{9}{25} + \frac{16}{25}} = 1$$

Note that this implies $\mathbf{u}^T\mathbf{u} = 1$

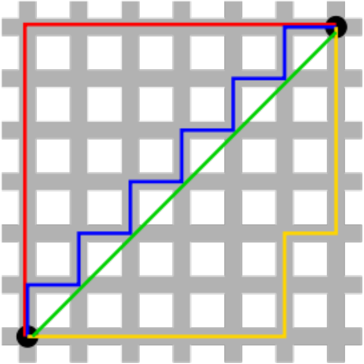
How else can we measure distance?

- $\| \star \|_1$ (1-norm) a.k.a. Taxicab metric, Manhattan Distance, City block distance
- $\| \star \|_\infty$ (∞-norm) a.k.a Max norm, Supremum norm, Uniform Norm
- Mahalanobis Distance (A probabilistic distance that accounts for the variance of variables)

1-norm, $\| \star \|_1$

$$\|\mathbf{x}\|_1 = |x_1| + |x_2| + |x_3| + \cdots + |x_n|$$

This is often called the city block norm because it measures the distance between points along a rectangular grid (as a taxicab must travel on the streets of Manhattan).



1-norm, $\| \star \|_1$

$$\|\mathbf{x}\|_1 = |x_1| + |x_2| + |x_3| + \cdots + |x_n|$$

This is often called the city block norm because it measures the distance between points along a rectangular grid (as a taxicab must travel on the streets of Manhattan).

So the 1 norm distance between two observations/vectors would be

$$\|\mathbf{x} - \mathbf{y}\|_1 = |x_1 - y_1| + |x_2 - y_2| + \cdots + |x_n - y_n|$$

∞-norm, $\| \star \|_\infty$

The infinity norm is sometimes called "max distance":

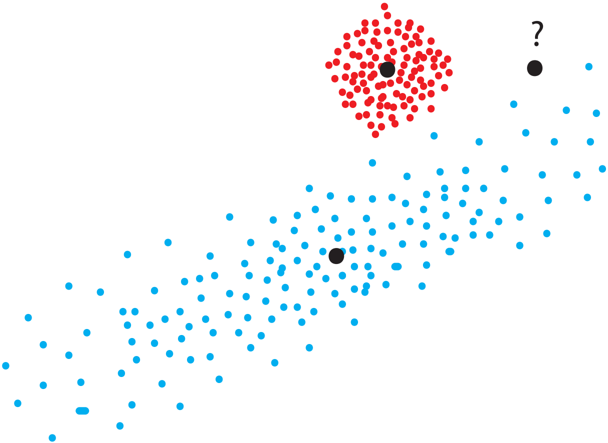
$$\|\mathbf{x}\|_\infty = \max\{|x_1|, |x_2|, |x_3|, \ldots, |x_n|\}$$

So the max distance between points/vectors **x** and **y** would be

$$\max\{|x_1 - y_1|, |x_2 - y_2|, |x_3 - y_3|, \ldots, |x_n - y_n|\}$$

Mahalanobis Distance

Takes into account the distribution of the data, often times comparing distributions of different groups.



YES, this stuff is useful!

Let's take a quick look at an application, which we will probably explore for ourselves later. MovieLens is a website devoted to Non-commercial, personalized movie recommendations:

https://movielens.org

As part of a massive open source project in recommendation system development, this website releases large amounts of it's data to the public to play with.

User-Rating Matrix

User	Movie 1	Movie 2	Movie 3	Movie 4
1	5			1
2		2	5	
3	3			5
4	5		5	

LOTS OF MISSING VALUES!!

What can we do with distance alone?

http://lifeislinear.davidson.edu/movieV1.html