Objectives

- Introduction to Vectors
- Distance Metrics
- Distance Metric Examples
- "Homework" Problems

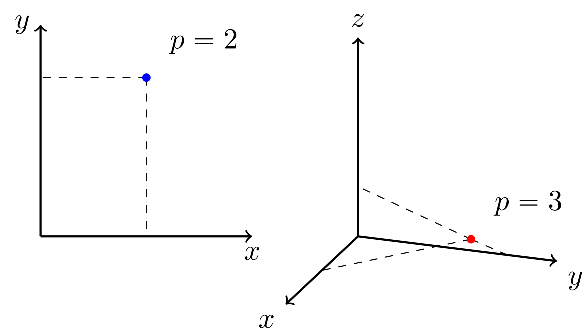
---

1 Introduction to Vectors

All vectors are part of a real space that has p dimensions. We can show them using points in a Euclidean space.

$$\vec{x} \in \mathbb{R}^p$$

p is always the number of **components** (number of elements) in a vector. For vectors with one component, p would be equal to one. p represents the dimension of space that vector is in.



Vectors are notated like so:

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_p \end{bmatrix}$$

In the programming language you choose, make sure to know whether it is **Column Major** or **Row Major**. Almost all vectors will be shown in the **Column major** form.

Transposing a matrix will switch the matrix from column-major to row-major form or vice versa.

$$x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_p \end{bmatrix}$$

$$x^T = [\ x_1,\ x_2,\ x_3,\ \cdots\ x_p\ ]$$

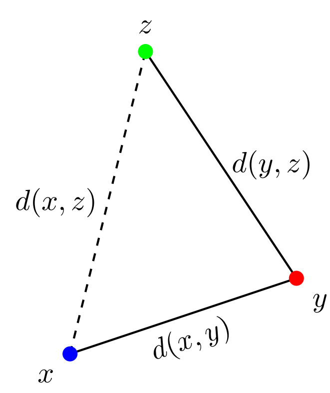
2 Distance Metrics

A **Distance Metric** is a general and abstract concept for the distances between vectors. They describe a function that takes two vectors and maps them to a real number greater than or equal to zero.

$$d(x, y) := \vec{x} \times \vec{y} \to \mathbb{R}^+$$

There are four conditions that a distance metric has to satisfy. To prove that a function on two vectors is a distance metric, show that these four statements are true for the function:

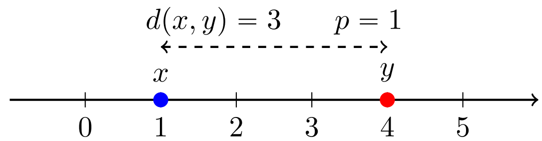
- $d(x, y) \geqslant 0$. No negative values for distance
- $d(x, y) = 0$ if $x = y$. Two of the same vectors should have 0 distance
- $d(x, y) = d(y, x)$. The function must be symmetric and work with either order.
- $d(x, z) \leqslant d(x, y) + d(y, z)$. *Triangle Inequality* shown below:



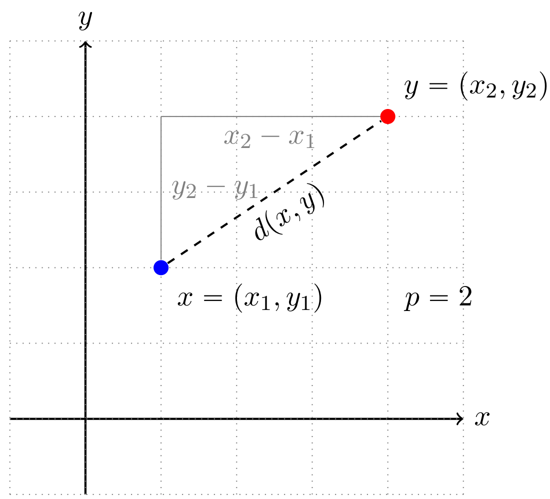
Triangle Inequality: $d(x, z) \leqslant d(x, y) + d(y, z)$

3 Distance Metric Examples

One of the distance metrics we all know of is **Euclidean Distance**, which is generally what we think of when the word *distance* is used.



Where $d(x, y) = |x - y|$ and $d(x, y)$ is the distance metric.



Where $d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}$

The Euclidean distance is one example of a metric used to define distance; however, there are many metrics used to define distances between vectors.

Another example of a distance metric is the **Hamming Distance**. The value of the Hamming Distance is found by counting the number of components that are not the same in two vectors. The Hamming distance is denoted by $d_H(x, y)$. For example:

$$d_H\left(\mathbf{x} = \begin{bmatrix}1\\3\\4\\7\end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix}1\\6\\4\\9\end{bmatrix}\right) = 2 \qquad d_H\left(\mathbf{x} = \begin{bmatrix}2\\3\\8\\7\end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix}1\\6\\4\\9\end{bmatrix}\right) = 4$$

The Hamming distance uses the **Binary Indicator Function** on each component in two vectors to determine the total distance. The binary indicator function takes two components and outputs 1 if they are different and 0 if they are equal.

$$\mathcal{I}(x, y) = \begin{cases} 1 & if\ x \neq y \\ 0 & if\ x = y \end{cases}$$

Using the **Binary Indicator Function** we can describe the **Hamming Distance**:

$$d_H(x, y) := \sum_{i=1}^{p} \mathcal{I}(x_i\ ,\ y_i)$$

4 "Homework" Problems

XPL 1) Is $d(x, y) = \sqrt{|x| + |y|}$ a valid distance metric?

We can show this by confirming the 4 conditions of a distance metric:

1) $d(x, y) \geqslant 0$: This is **true** as the range of the square root function is $[0, \infty)$.

2) $d(x, y) = 0$ if $x = y$. This is **not true**. If $x = y$, then we have $\sqrt{2|x|} \neq 0$.

3) $d(x, y) = d(y, x)$. This is **true**, as $|x| + |y| = |y| + |x|$.

4) $d(x, z) \leqslant d(x, y) + d(y, z)$. Lets not bother, we already proved this false with number 2.

XPL 2) Is $d(x, y) = (x - y)^2$ a valid distance metric?

XPL 3) Is $d(x, x) = 0 \wedge d(x, y) = 1,\ \forall x \neq y$ a valid distance metric?

XPL 4) Code out the Binary Indicator Function using a functional programming paradigm

XPL 5) Prove $d_H(x, y)$ satisfies the Triangle Inequality