Report on assignement 3 of Data Capture and Processinig course

Hamza Omari[1*]

[1*]faculty of science and technology, University Paris est Créteil, Créteil, 94100, France.

Corresponding author(s). E-mail(s): hamza.omari@etu.u-pec.fr;

Abstract

In this report we will explore different norm and metrics used for signals. These norms and distances are widely used in the the fields of data science and signal processing. We will be discussing their results

Keywords: Norms, Distances, Signal processing

1 Generating and defining the functions

We will start first by defining two signals (arrays) x and y randomly and normalizing the both signals to have $mean = 0$ and $std = 1$

Plotting the x and y signals will give us the plot in the figure 1

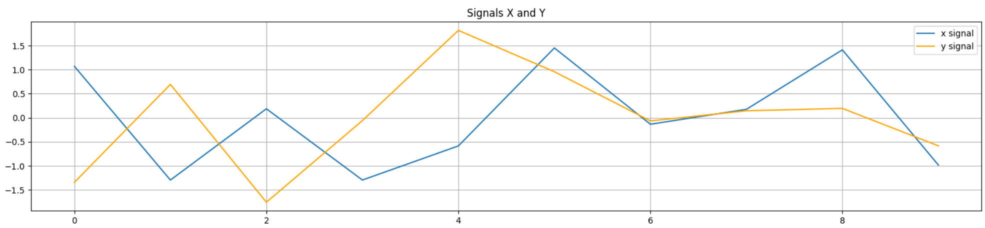

Fig. 1: Random x and y signals

2 Norms and distance calculation

Using numpy library, we will implement allof the below distances/norms :

$$d_{\text{Manhattan}}(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{n} |x_i - y_i|$$

$$d_{\text{Euclidean}}(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$

$$d_{\text{Chebyshev}}(\mathbf{x}, \mathbf{y}) = \max_i |x_i - y_i|$$

$$\text{cosine\_similarity}(\mathbf{x}, \mathbf{y}) = \frac{\mathbf{x} \cdot \mathbf{y}}{\|\mathbf{x}\|_2 \|\mathbf{y}\|_2} = \frac{\sum_{i=1}^{n} x_i y_i}{\sqrt{\sum_{i=1}^{n} x_i^2} \sqrt{\sum_{i=1}^{n} y_i^2}}$$

$$d_{\text{Minkowski}}(\mathbf{x}, \mathbf{y}, p) = \left(\sum_{i=1}^{n} |x_i - y_i|^p \right)^{\frac{1}{p}}$$

3 Values and interpretation

after calculating the norms of the vector x we got the value in the table

Table 1: Different norms value for x

manhatten	8.6
euclidian	3.2
chebyshev	1.45

we can see that manhatten norm > euclidian norm. this is due to the nauture of the formulas themselves. and a certain value for the chebyshev that tells us about the biggest absolute value that exists in the array x

Next we will calculate the distance between y and x vectors and will plot the results in figure 2

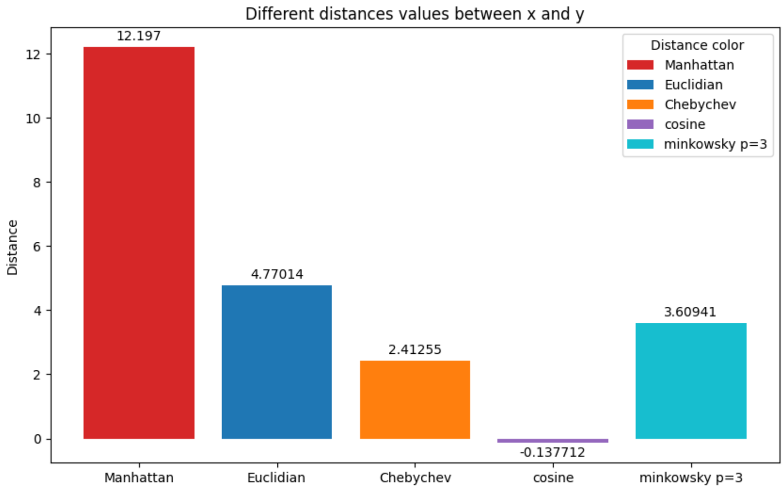

Fig. 2: Barchart of the distances between x and y signals

We can see that at first the **cosine similarity**, has a very small value, this due to the fact that this metric measure the cosinus between the two vectors in their dimensional spaces. So this one will be interesting to just see how much the two vectors are correlated to each other depending on the value. And the direction.

The **manhatten** distance got the biggest value, to the nature of the function itself, this formula is interesting especially for measuring distances in a gird like space, like 2D games.

Euclidian and **chebychev** look alike, for the fact that the first formula is square to the power of 2 and the other is to the power of p. so the first one is a special case from a generalization formula. Choosing one from another will depend on the nature of the problem, and how much we want to normalize the distance.

The **minowsky distance** got a special value, it would be interesting if we want to see the biggest difference between two arrays of a certain features. Where we are inerested to see the biggest difference between all the differences of the different characteristics of our vectors.

4 Code and implementation

All of the methods presented in this report, are implemented in a jupyter notebook using python and libraries like numpy and matplotlib.

Notebook is accessible in this github repository: LINK