

# **Euclidean Distance Calculation Project**

## Project Summary

Euclidean distance measures are central to the k-Nearest Neighbor algorithm, thus this project utilizes 5 algorithms to calculate distance point value differences and each algorithm's detailed numerical complexity analysis. This Python script compares the performance of different implementations of the Euclidean distance calculation for a wine dataset. It also tests the effect of increasing the number of attributes on the calculation time. The script defines five functions that implement the Euclidean distance calculation in different ways.

Euclidean distance is a fundamental concept in mathematics and has many applications in various fields such as physics, statistics, computer science, and machine learning. Some of the reasons why Euclidean distance is important are:

**Measurement of distance:** Euclidean distance is the straight-line distance between two points in a  $n$ -dimensional space. It provides a way to measure the distance between two points and is commonly used in geometry and trigonometry.

**Clustering:** Euclidean distance is often used in clustering algorithms to group similar data points together. In this context, the distance between two data points can be seen as a measure of their similarity.

**Dimensionality reduction:** Euclidean distance can be used in dimensionality reduction techniques such as Principal Component Analysis (PCA) and Multidimensional Scaling (MDS) to reduce the number of dimensions of a dataset while preserving the relationships between data points.

**Machine learning:** Euclidean distance is commonly used in machine learning algorithms such as k-Nearest Neighbors (k-NN) and Support Vector Machines (SVMs) to classify and predict new data points based on their proximity to existing data points in a  $n$ -dimensional space.

**Optimization:** Euclidean distance is also used in optimization algorithms such as gradient descent to find the minimum or maximum of a function. In this context, the distance between the current point and the target point can be seen as a measure of the error or loss to be minimized.

This Project utilizes different functions. It contains the implementation of Pandas for Series/Dataframes, Loaded Operators and Higher order functions. Similarly it also uses NumPy, SciPy and Timeit functions.