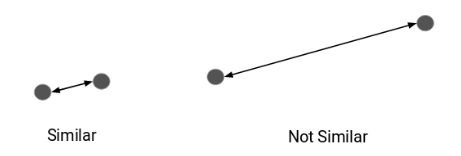
What Are Distance Metrics?

Distance metrics are a key part of several [machine learning algorithms](https://www.analyticsvidhya.com/blog/2017/09/common-machine-learning-algorithms/?utm_source=blog&utm_medium=4-types-of-distance-metrics-in-machine-learning). These distance metrics are used in both supervised and unsupervised learning, generally to calculate the similarity between data points.

An effective distance metric improves the performance of our machine learning model, whether that’s for classification tasks or clustering.

Let’s say you need to create clusters using a clustering algorithm such as [K-Means Clustering](https://www.analyticsvidhya.com/blog/2019/08/comprehensive-guide-k-means-clustering/?utm_source=blog&utm_medium=4-types-of-distance-metrics-in-machine-learning) or [k-nearest neighbor algorithm](https://www.analyticsvidhya.com/blog/2018/03/introduction-k-neighbours-algorithm-clustering/?utm_source=blog&utm_medium=4-types-of-distance-metrics-in-machine-learning) (knn), which uses nearest neighbours to solve a classification or regression problem. How will you define the similarity between different observations? How can we say that two points are similar to each other? This will happen if their features are similar, right? When we plot these points, they will be closer to each other by distance.



Types of Distance Metrics in Machine Learning

1. Euclidean Distance
2. Manhattan Distance
3. Minkowski Distance
4. Hamming Distance

Let’s start with the most commonly used distance metric – Euclidean Distance.

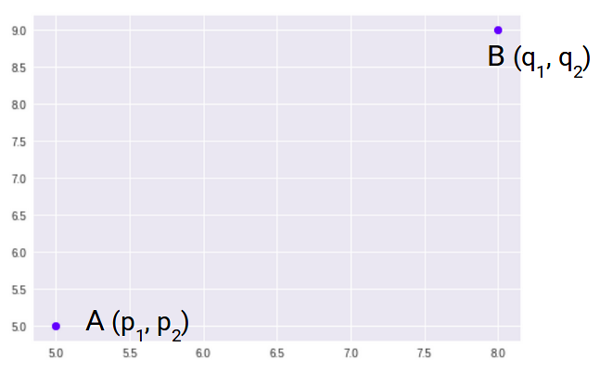
Euclidean Distance

Euclidean Distance represents the shortest distance between two vectors. It is the square root of the sum of squares of differences between corresponding elements.

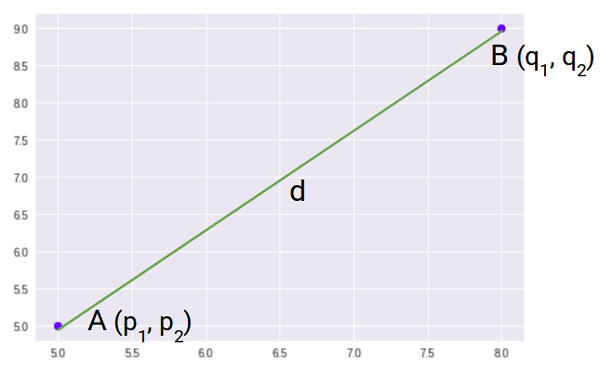
The Euclidean distance metric corresponds to the L2-norm of a difference between vectors and vector spaces. The cosine similarity is proportional to the dot product of two vectors and inversely proportional to the product of their magnitudes.



Most machine learning algorithms, including K-Means use this distance metric to measure the similarity between observations. Let’s say we have two points, as shown below:



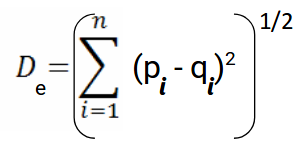
So, the Euclidean Distance between these two points, A and B, will be:



Here’s the formula for Euclidean Distance:

euclidean distance formula | distance metrics

We use this formula when we are dealing with 2 dimensions. We can generalize this for an n-dimensional space as:



Where,

* n = number of dimensions
* pi, qi = data points