



Measuring distance between data points

Peter, Jonas & Mie

Manhattan Distance

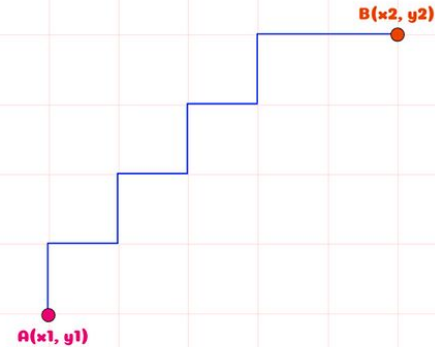
Manhattan distance, also known as taxicab or city block distance, measures the sum of the absolute differences between corresponding coordinates.

Example:

$$d_{\text{Manhattan}} = |3-6| + |4-7| = 3+3=6$$

Manhattan Distance

$$\text{Manhattan}(A,B) = |x_1 - x_2| + |y_1 - y_2|$$



Euclidean Distance

Euclidean distance calculates the straight-line distance between two points in Euclidean space.

Example:

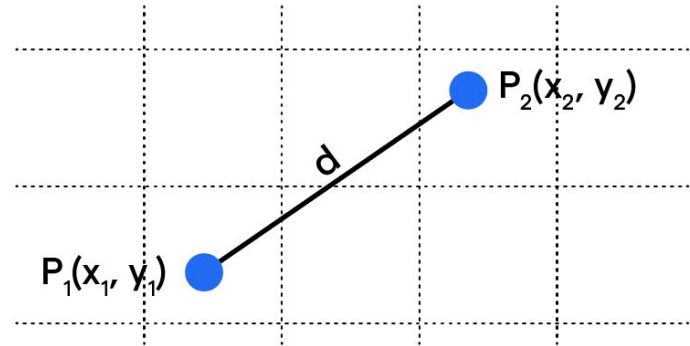
$$d = \sqrt{(3 - 2)^2 + (5 - 3)^2}$$

$$d = \sqrt{(1)^2 + (2)^2}$$

$$d = \sqrt{1 + 4}$$

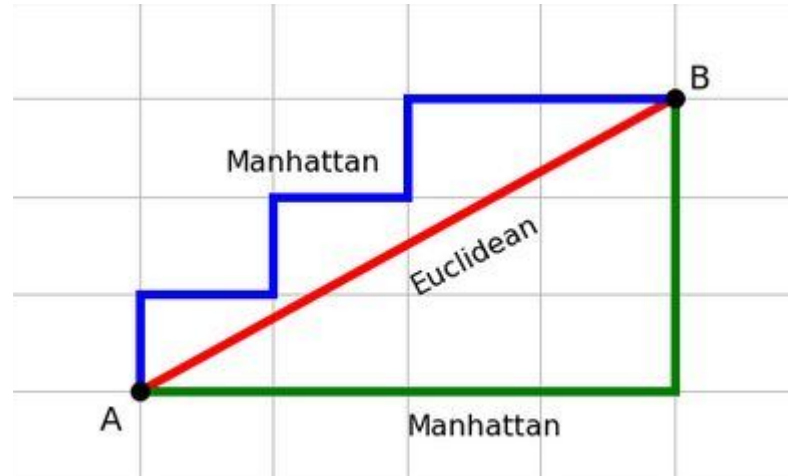
$$d = \sqrt{5}$$

Euclidean Distance



$$\text{Euclidean Distance (d)} = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

Manhattan versus Euclidean

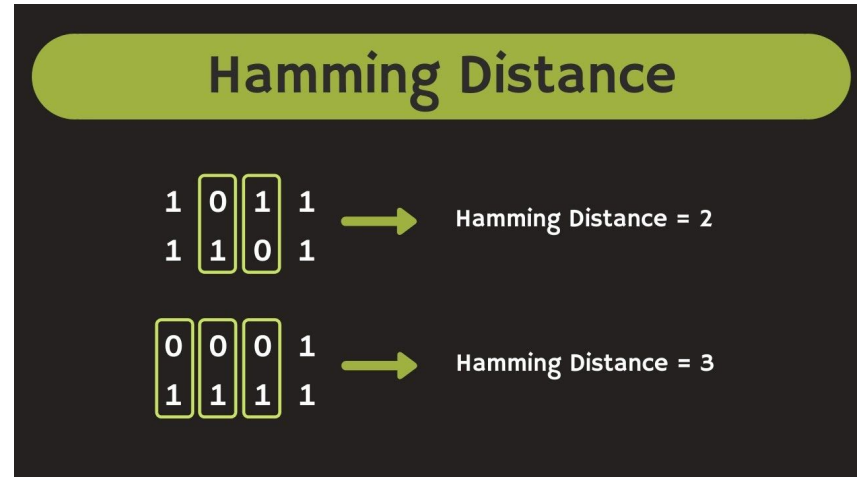


Hamming distance

Hamming distance is used for comparing strings of equal length. It measures the minimum number of substitutions required to change one string into another.

Example:

Strings "0001" and "1111" have a Hamming distance of 3. (Count the differing bits.)





Why are there different measures

- Each distance measure serves specific purposes based on the nature of the data and the problem we try to solve.
- Manhattan distance is useful for grid-based systems and when movement is restricted to horizontal and vertical paths.
- Euclidean distance is suitable for continuous data and when the spatial relationship between points matters.
- Hamming distance is essential in fields like computer science, genetics, and error detection.