Measuring Distance Between Data Points

Timothy Mortensen

Manhattan Distance

- Also called **L1 norm** or **city block distance**.
- Formula:

$$D = |x_1 - x_2| + |y_1 - y_2|$$

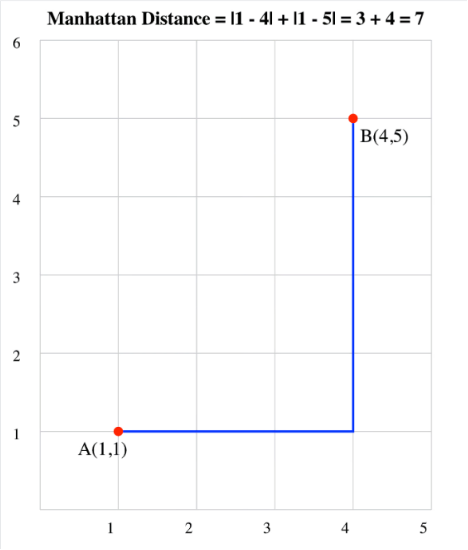
- Example:

$$A = (1, 2), \quad B = (4, 6)$$

$$D = |1 - 4| + |2 - 6| = 3 + 4 = 7$$

- Absolute value = no negative values.

Manhattan Distance Illustration



A visual example of Manhattan distance on a grid layout.

Euclidean Distance

- Also called **L2 norm**.
- Formula:

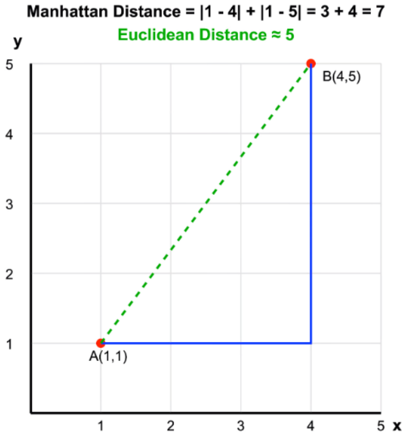
$$D = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

- Example:

$$A = (1, 2), \quad B = (4, 6)$$

$$D = \sqrt{(1 - 4)^2 + (2 - 6)^2} = \sqrt{9 + 16} = 5$$

Euclidean vs manhattan Difference



A visual example of Manhattan distance on a grid layout.

Hamming Distance

- Used for binary or categorical data.
- It counts the number of differing positions.
- Example:

$$A = 10101, \quad B = 10011$$

$$\text{Hamming Distance} = 2 \quad \text{(positions 3 and 5 differ)}$$

Why Different Measures?

- **Manhattan**: robust in high-dimensional space.
- **Euclidean**: works well for continuous data and geometry.
- **Hamming**: best for binary/categorical data.
- Choose based on your data type and model