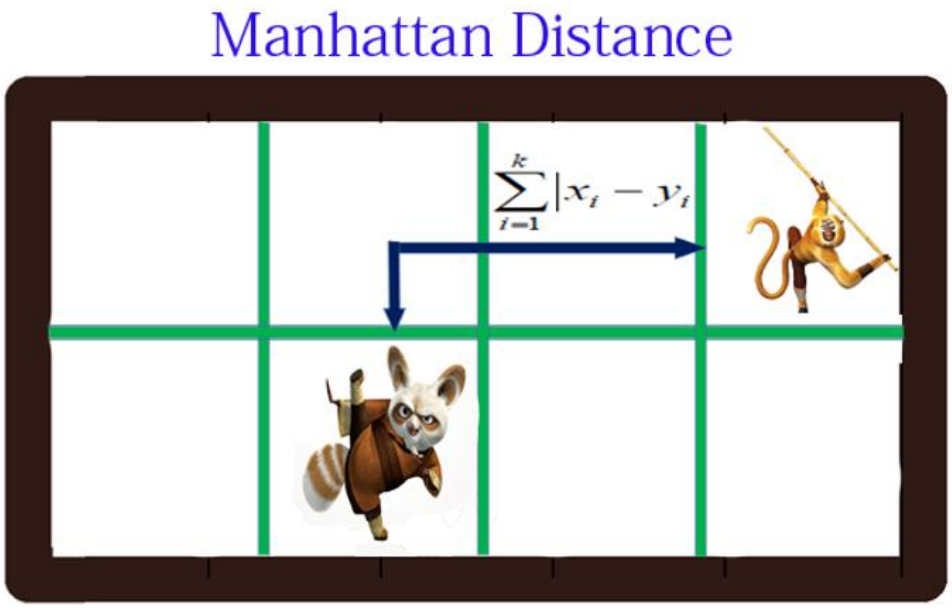
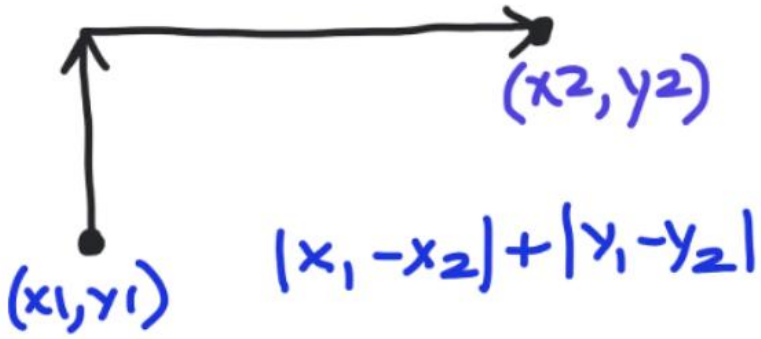
What is Manhattan Distance?

Manhattan distance, also known as L1 distance or city block distance, measures the distance between two points by summing the absolute differences of their coordinates. Imagine navigating a city where you can only travel along predefined streets (horizontally or vertically), like walking blocks in a grid. The Manhattan distance is the shortest path you'd take in such a scenario.



Formula:



Example: Customer Preferences for Breakfast Items

Let's say we have two customers, Alice and Bob, and we've asked them to rate their preference for two breakfast items on a scale of 1 to 5: Pancakes and Waffles.

Alice's Preferences:

- Pancakes: 4
- Waffles: 2
- So, Alice's preference "point" is (4,2)

Bob's Preferences:

- Pancakes: 2
- Waffles: 5
- So, Bob's preference "point" is (2,5)

Now, let's calculate the Manhattan distance between Alice's and Bob's preferences:

- Difference in Pancakes preference (absolute): $|2-4|=|-2|=2$
- Difference in Waffles preference (absolute): $|5-2|=|3|=3$
- Sum the absolute differences: $2+3=5$

So, the Manhattan distance between Alice's and Bob's preferences is 5.

Why "City Block"?

Imagine plotting these points on a grid:

- Alice is at (4,2)
- Bob is at (2,5)

To get from Alice to Bob by only moving horizontally or vertically (like walking blocks in a city):

- You'd move 2 units to the left (from 4 to 2 on the X-axis).
- You'd move 3 units up (from 2 to 5 on the Y-axis).
- Total "blocks" moved = 2 + 3 = 5.

This distance of 5 quantifies the total difference in their preferences when considering each item independently and summing up those differences. It's often preferred over Euclidean distance when individual feature differences are more important than the overall "as-the-crow-flies" distance.

Common Real-Life Applications

1. **Urban Planning and Logistics**

   - **Route Planning:** In cities with a grid system, Manhattan distance is the most accurate way to calculate the shortest travel distance between two points. It's used by delivery services and ride-sharing apps to estimate travel time and cost.
   - **Warehouse Layout:** It's used to optimize the placement of items in a warehouse. Items that are frequently accessed together are placed close to each other, based on their Manhattan distance, to minimize the travel time for workers or robots.

2. **Computer Science and Engineering**

   - **VLSI Design:** In Very Large Scale Integration (VLSI) design, the Manhattan distance is used to estimate the length of a wire connecting two components on a circuit board. This helps in minimizing the wire length to improve signal speed and reduce power consumption.
   - **Computer Graphics:** It's used to calculate the distance between pixels, especially in image processing and computer vision tasks.

3. **Machine Learning and Data Science**

   - **K-Nearest Neighbors (K-NN):** When dealing with high-dimensional data or features that are not independent, Manhattan distance is often a better choice than Euclidean distance. It can be more robust to outliers and is less sensitive to the "curse of dimensionality."
   - **Clustering:** In some clustering algorithms, Manhattan distance is used to form clusters of data points, particularly when the data is not evenly distributed or contains a significant number of outliers.