Measuring distance between data points

By Isak, Helena & Jamie

Manhattan Distance

- Taxicab
- Grid based
- Used in: Taxi's, Video games & Robotics

For two points:

$D_{Manhattan} = |x_2 - x_1| + |y_2 - y_1|$

For n-dimensional space

$D_{Manhattan} = \sum |x_i - y_i|$ (for i = 1 to n)

Example of two points

$D_{Manhattan} = |9 - 3| + |2 - 7|$

$= |6| + |-5|$

$= 6 + 5$

$= 11$

Euclidean Distance

- L2 distance
- Pythagorean theorem
- Used for: physical distance

For two points:

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

For n-dimensional space

$D_{Euclidean} = \sqrt{(\sum (x_i - y_i)^2)}$ (for i = 1 to n)

Example of two points

$D_{Euclidean} = \sqrt{((9 - 3)^2 + (2 - 7)^2)}$

$= \sqrt{(6^2 + (-5)^2)}$

$= \sqrt{(36 + 25)}$

$= \sqrt{61}$

≈ 7.81

Hamming Distance

- Binary data
- Used for: Data transmission & DNA comparison

Equation:

$D_{Hamming} = \sum$ (Number of differing bits)

Example of binary strings:

1011101

1001001

Differences at: (3rd, 5th bits)

$D_{Hamming} = 2$