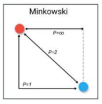
Data Mining 4315: Lecture 01 (Page 2)

J Basravi - Jan 26, 2023 6:30pm

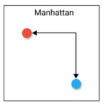
The **Minkowski Distance** is a generalized metric used to measure the **distance between two points**. It's general form is:

$$d(i,j) = \sqrt[h]{|x_{i1} - x_{j1}|^h + |x_{i2} - x_{j2}|^h + \cdots + |x_{ip} - x_{jp}|^h}$$



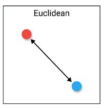
If you change the *h value* in the general form to a **1**, it becomes equivalent to the **Manhattan distance**:

$$d(i,j) = |x_{i1} - x_{j1}| + |x_{i2} - x_{j2}| + \ldots + |x_{ip} - x_{jp}|$$



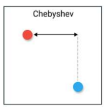
If you change the *h value* in the general form to a **2**, it becomes equivalent to the **Euclidean distance**:

$$d(i,j) = \sqrt{(|x_{i1} - x_{j1}|^2 + |x_{i2} - x_{j2}|^2 + \ldots + |x_{ip} - x_{jp}|^2)}$$



If you change the *h value* in the general form to ∞, it becomes equivalent to the **Chebyshev distance** (also called '**Supremum**'):

$$d(i,j) = \lim_{h \to \infty} \left(\sum_{f=1}^{p} |x_{if} - x_{jf}|^h \right)^{\frac{1}{h}} = \max_{f}^{p} |x_{if} - x_{jf}|$$



GOING FROM POINT A TO B!

MANHATTAN: DISTANCE TRAVELED ON THE ROADS

EUCLIDEAN: DISTANCE TRAVELED IN A STRAIGHT LINE

CHEBYSHEV: THE MAX() OF ALL DIMENSIONS' CHANGES