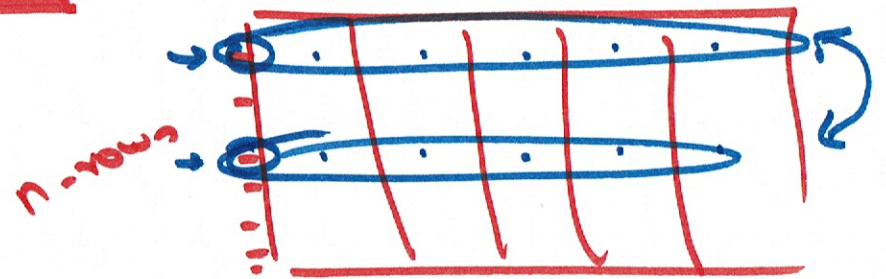


Distance

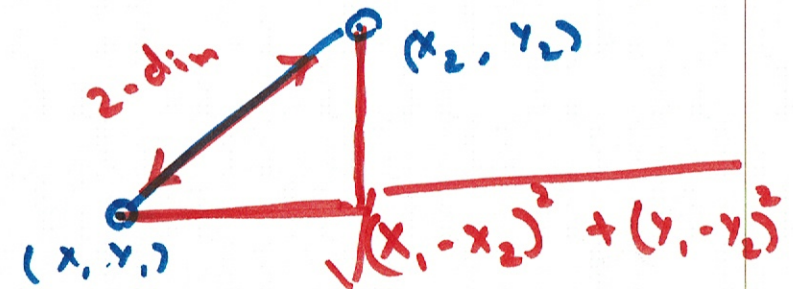


- Do define "similarity" you need a measure of distance
- Examples of common distance measures

- Manhattan Distance

- Eucledian Distance →

- Chebyshev Distance ←

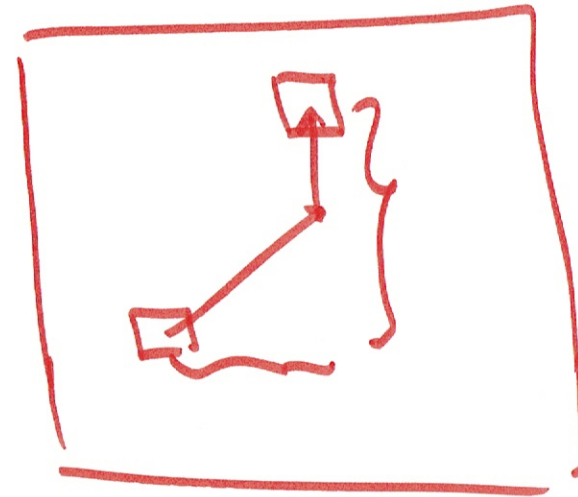


$$|x_1 - x_2| + |y_1 - y_2| + |z_1 - z_2| + \dots$$

$$\sqrt{\frac{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2 + \dots}{n}}$$

m -dim

$$\max(|x_1 - x_2|, |y_1 - y_2|, |z_1 - z_2|, \dots)$$



Minkowski

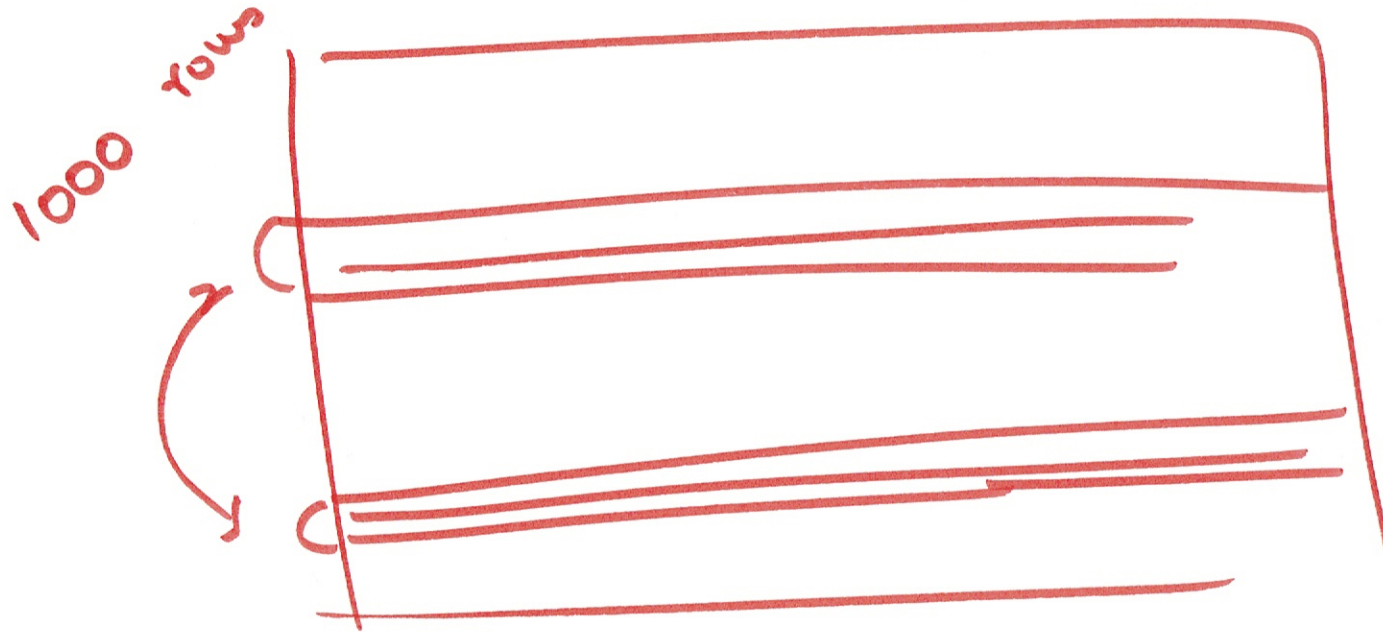
	1	2	m
x	x_1	x_2				x_m
y	y_1	y_2				y_m

$$\left(\sum_{i=1}^m |x_i - y_i|^p \right)^{1/p}$$

if $p = 2 \Rightarrow$ Euclidean dist.

if $p = 1 \Rightarrow$ Manhattan dist.

if $p = \infty \Rightarrow$ Chebyshev dist.



Connectivity based \sim roughly begin by computing
500,000 dist $\frac{n \times (n+1)}{2}$

Centroid based \sim roughly begin by computing }
5-group 5×1000 dist.

$5 \times n$

