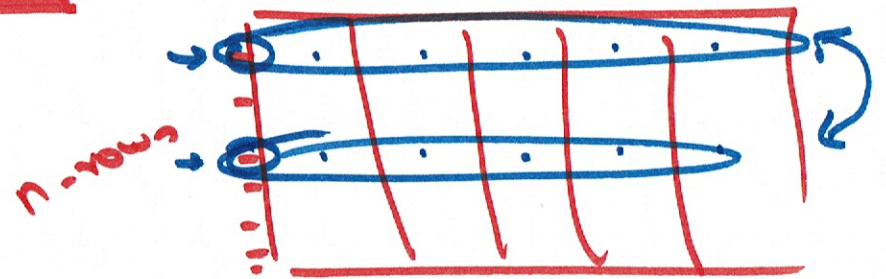


## Distance

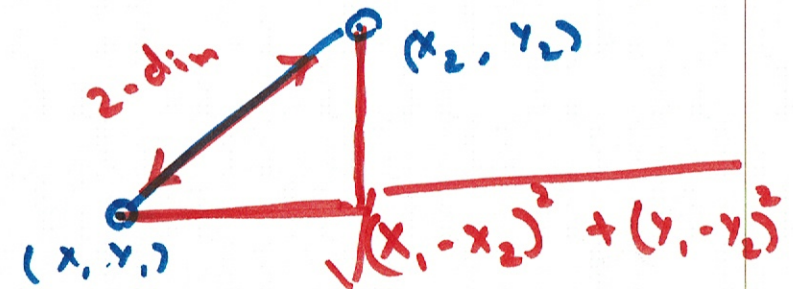


- Do define "similarity" you need a measure of distance
- Examples of common distance measures

- Manhattan Distance

- Eucledian Distance →

- Chebyshev Distance ←

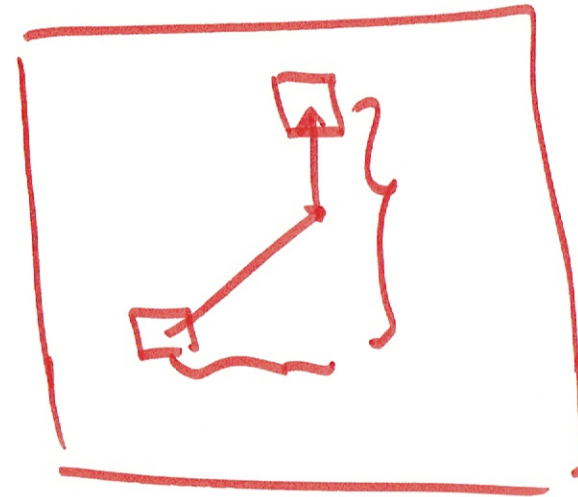


$$|x_1 - x_2| + |y_1 - y_2| + |z_1 - z_2| + \dots$$

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

$m$ -dim

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



# Minkowski

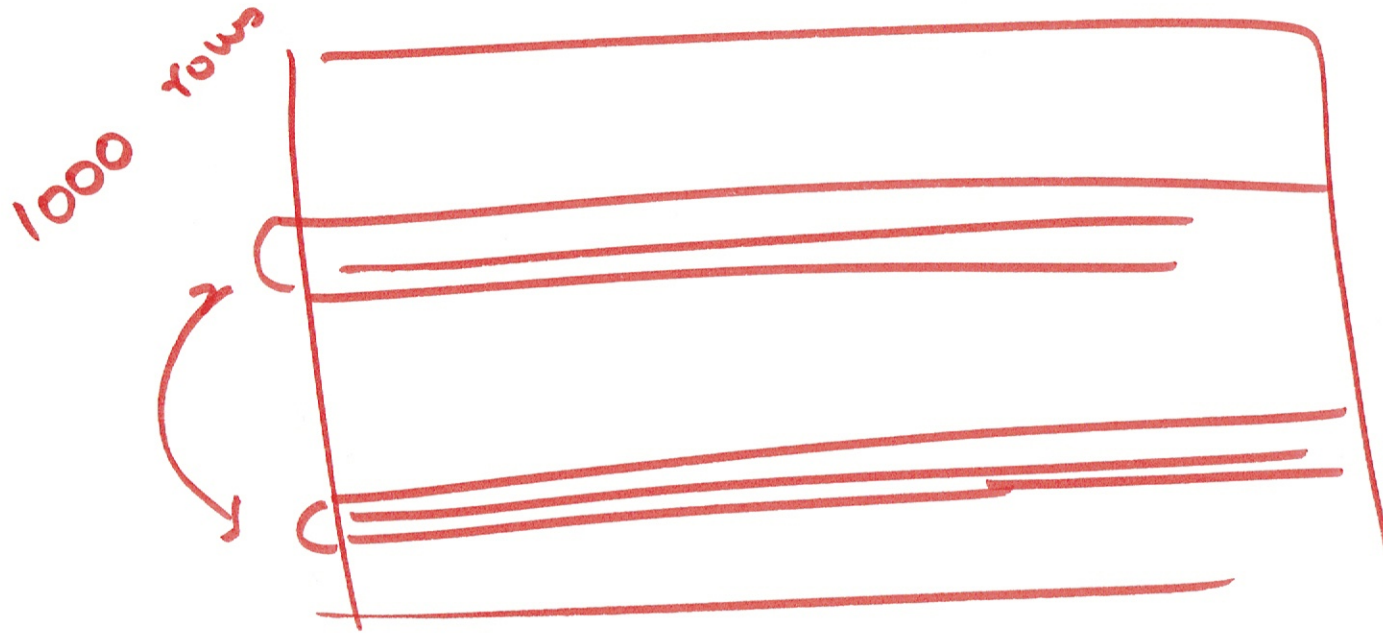
	1	2	...	...	...	m
x	x <sub>1</sub>	x <sub>2</sub>				x <sub>m</sub>
y	y <sub>1</sub>	y <sub>2</sub>				y <sub>m</sub>

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

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  $5 \times 1000$  dist.

$5 \times n$



