EUCLIDEAN DISTANCE AND MANHATTAN DISTANCE

Q) Why we check distance between two points?

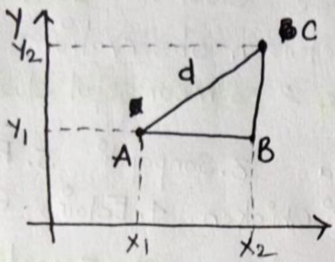
Ans. If two points are near to each other, might be chances they are similar to each other.

So two methods to check **distance between two points** →
1) **Euclidean distance**
2) **Manhatten distance**

EUCLIDEAN DISTANCE

→ It is simply **calculated by Pythagoras theorem**.

Suppose in two dimension :



Suppose we have to find distance d.

$$d,\ AC^2 = AB^2 + BC^2$$

$$AC = \sqrt{AB^2 + BC^2}$$

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

For 3 dimensions,

$$AC = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}$$

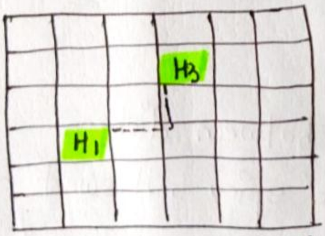
MANHATTAN DISTANCE

→ In this we take the **absolute distance between the point**, instead of underroot and square.

$$AC = |(x_2 - x_1)| + |(y_2 - y_1)|$$

REAL TIME SCENARIO WHERE MANHATTAN DISTANCE USED —

Suppose we see houses and road in the graph.



Suppose **line are the roads**

□ **are different houses**.

---- is the **road follow to reach House H_3 from H_1**

In this case we **use manhatten distance.**

IS EUCLIDEAN DISTANCE meaningful for high dimensional data?

Short answer is **NO**. At high dimension euclidean distance **losses preety much all meaning.**

→ **Main issue** is **"Curse of dimensionality"**

→ **First** → This pattern starts to fall away if we have different dimensions are correlated. If we do PCA or something similar to re-project into lower dimension space with small amount of loss, then metrics are still meaningful.

Second → **Critical problem** is **Sparsity** and **value of any distance metric at high dimension**. If we are **considering narrow "neighbourhood"** of nearest neighbour, though many **approximate nearest neighbours** solutions becomes **very ineffective at high dimensions.**