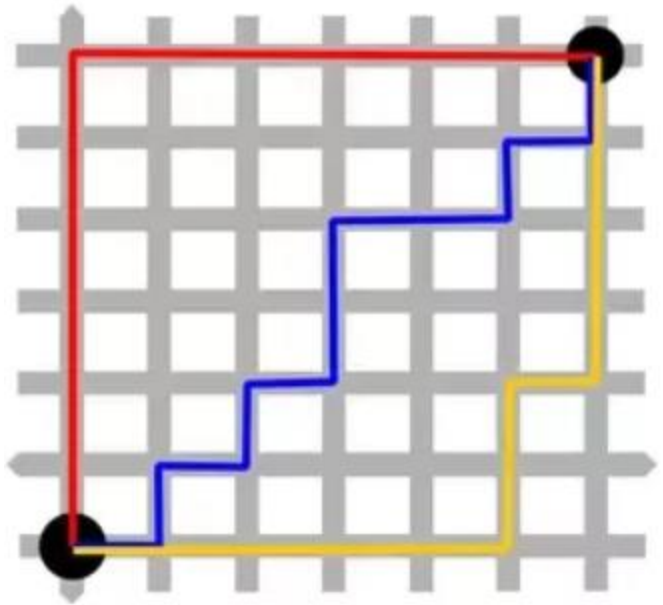
Vectors and Classification

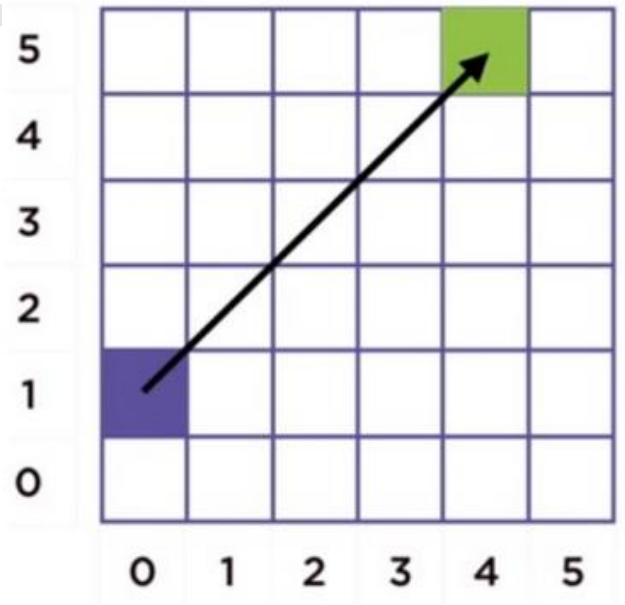
Manhattan Distance

- Vectors in ML: use to represent numeric characteristics (features)
- sum of all the real distances between source and destination



Euclidean Distance

- Shortest distance between two points

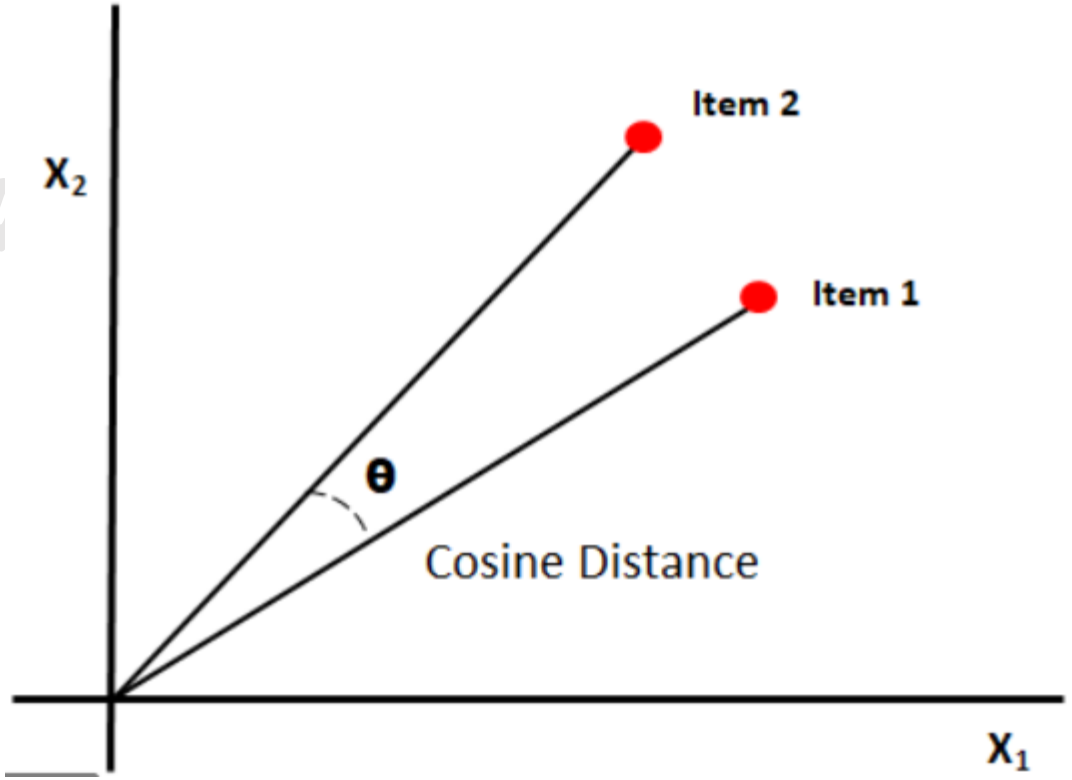


Cosine Distance

- Cosine_distance =1- cos_similarity

- Cosine_similarity

$$\frac{x \bullet y}{\sqrt{x \bullet x}\sqrt{y \bullet y}}$$



Classification

- A classification problem is when the output variable is a category, such as “red” or “blue” or “disease” and “no disease”

K – Nearest Neighbours

- It assumes that similar things exist in close proximity.

Algorithm:

* Step 1: Choose the no. K of neighbours
* Step 2: Take the K nearest neighbours of the new data points by Euclidean distance
* Step 3: Among K Neighbours, count the no. of data points in each category
* Step 4: Assign new data point to the category where you counted most neighbour

Case 1: K = 3, Solid Circle - Assigned to Red Triangle Group

Case 2: K = 5, Dashed Circle - Assigned to Blue Square Group