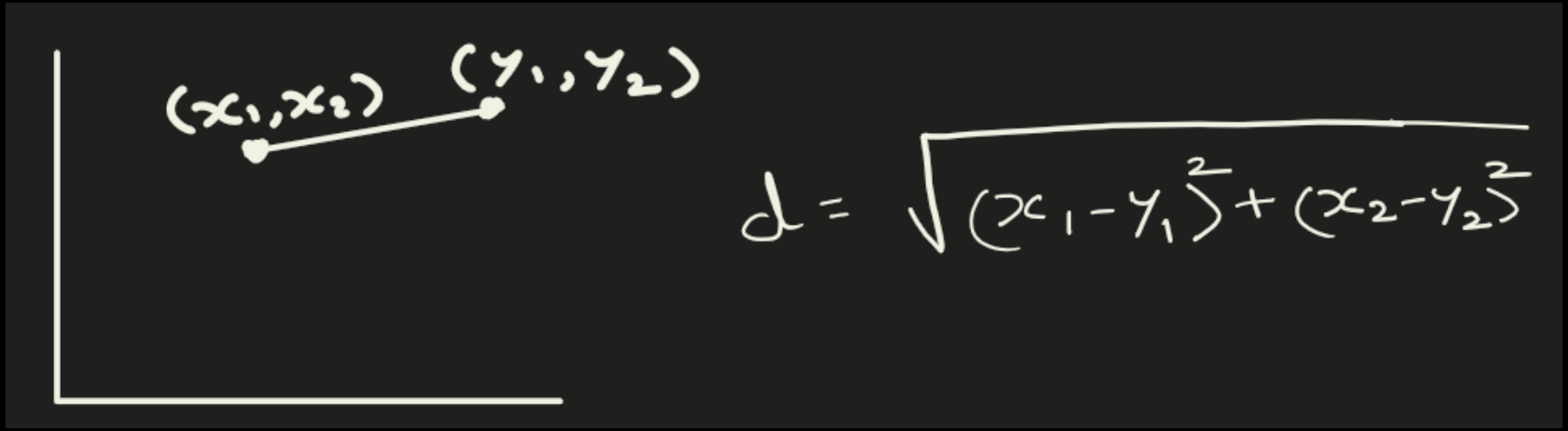
K-Nearest Neighbor (KNN): Distance Metric

1) Euclidean distance as our distance metric since it's the most popular method.

$$d(x, y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

Example in 2 D

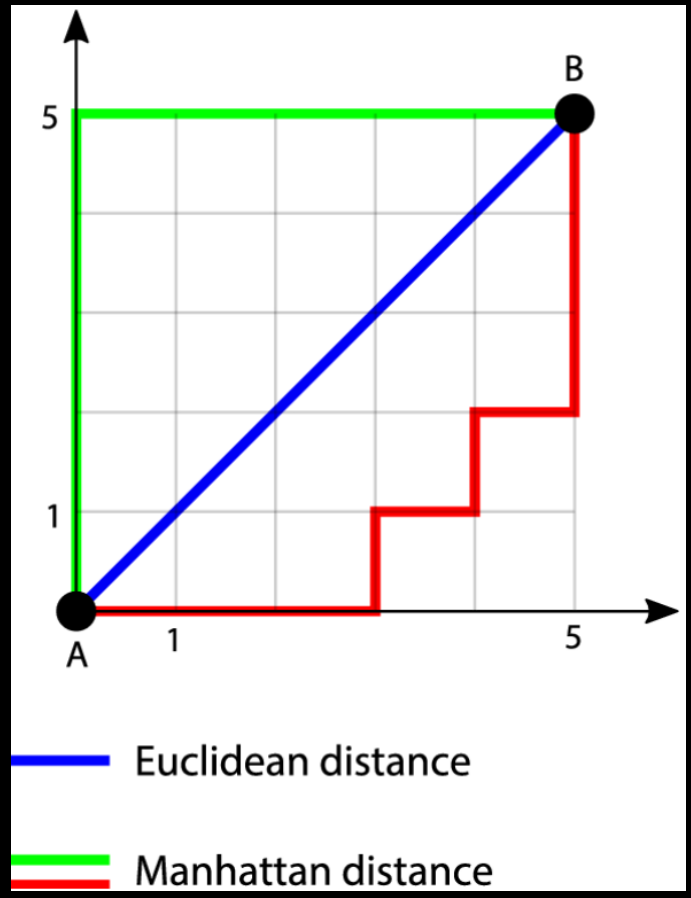


K-Nearest Neighbor (KNN): Distance Metric

2) Manhattan distance: It's also known as the "city block" distance.

$$d(x, y) = \sum_{i=1}^{n} |x_i - y_i|$$

In the example shown:
Manhattan distance for red path = 3 + 1 + 1 + 1 + 1 + 3 = 10
Manhattan distance for green path = 5 + 5 = 10



K-Nearest Neighbor (KNN): Distance Metric

3) Minkowski distance: Generalized form of Euclidean and Manhattan.

$$d(x, y) = \left(\sum_{i=1}^{n} |x_i - y_i|^p \right)^{\frac{1}{p}}$$

If $p = 1$→ Manhattan
If $p = 2$→ Euclidean

K-Nearest Neighbor (KNN): Distance Metric

4) Chebyshev distance: Maximum difference across any dimension.

$$d(x, y) = \max_i |x_i - y_i|$$

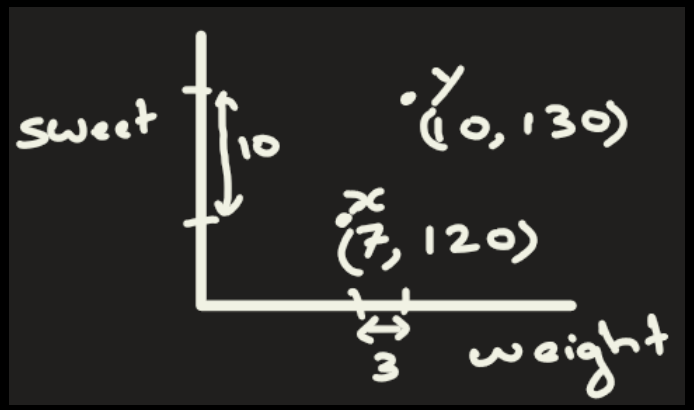
Example:
You have two fruit samples:

$$x = (\text{Sweetness} = 7, \text{Weight} = 120)$$
$$y = (\text{Sweetness} = 10, \text{Weight} = 130)$$

Chebyshev distance = max (|7-10|, |120-130|)

= max (3 , 10)

= 10



K-Nearest Neighbor (KNN): Distance Metric

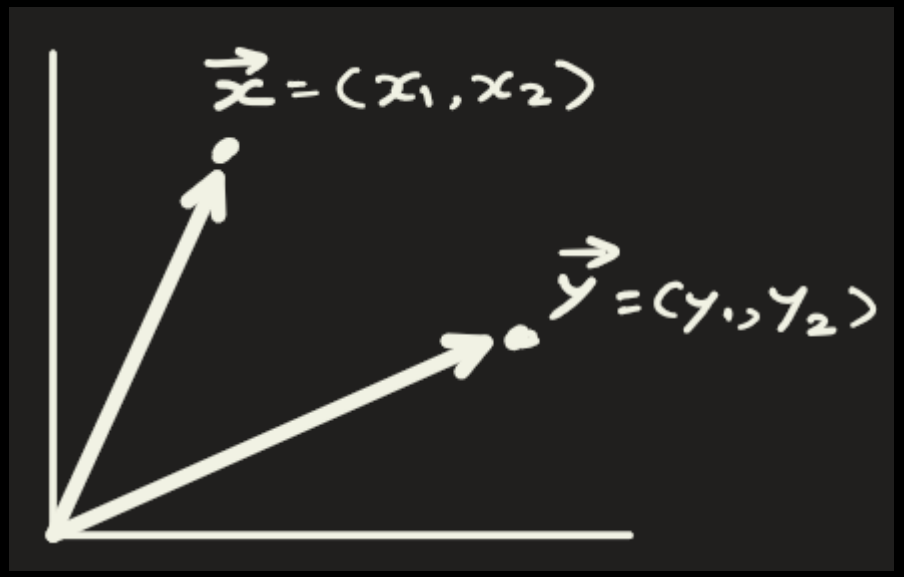
5) Cosine distance: Used for text, high-dimensional data.

$$\text{Cosine Similarity} = \frac{x \cdot y}{\|x\| \, \|y\|}$$

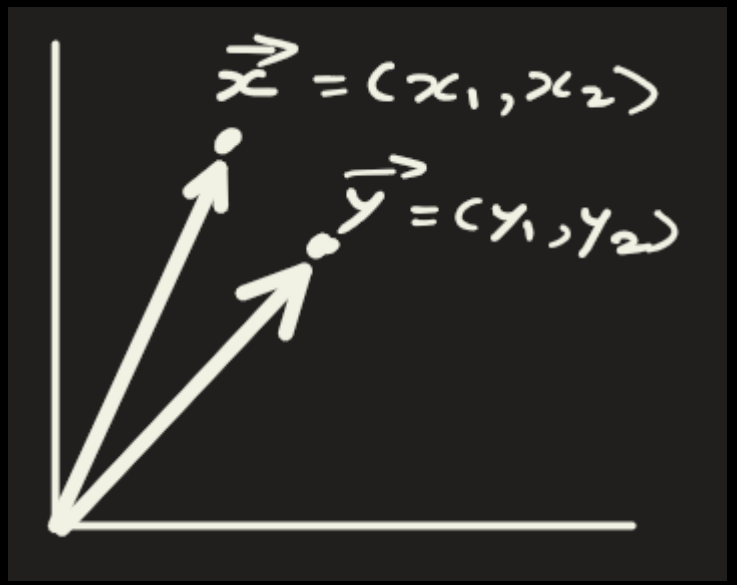
Cosine distance = 1 – cosine similarity

Example: Suppose x and y represents some documents.

x and y are less similar



x and y are more similar



K-Nearest Neighbor (KNN): Distance Metric

6) Hamming distance: Used for **categorical or binary** features.

$$d(x, y) = \sum_{i=1}^{n} \mathbf{1}(x_i \neq y_i)$$

(Counts positions where values differ)

Example:

Feature	Fruit A	Fruit B	Same?	Contribution
Color	Red	Green	No	1
Size	Medium	Medium	Yes	0
Taste	Sweet	Sour	No	1
Skin Texture	Smooth	Smooth	Yes	0

Add the mismatches:

$$\text{Hamming } \textit{Distance} = 1 + 1 = 2$$

EXTRA

Mathematics of KNN

1)Given a positive integer K and a test observation x_0, the KNN classifer first identifes the K points in the training data that are closest to x_0, represented by N_0. The distance could be Euclidean Distance, Manhattan Distance, Minkowski Distance, etc

$$\text{distance}(x, X_i) = \sqrt{\sum_{j=1}^{d}(x_j - X_{i_j})^2]}$$

$$d(x, y) = \sum_{i=1}^{n} |x_i - y_i|$$

$$d(x, y) = \left(\sum_{i=1}^{n} (x_i - y_i)^p\right)^{\frac{1}{p}}$$

2) It then estimates the conditional probability for class j as the fraction of points in N_0 whose response values equal j:

$$\Pr(Y = j | X = x_0) = \frac{1}{K} \sum_{i \in \mathcal{N}_0} I(y_i = j).$$

3) Classify the test observation x0 to the class with the largest probability calculated from above eq.