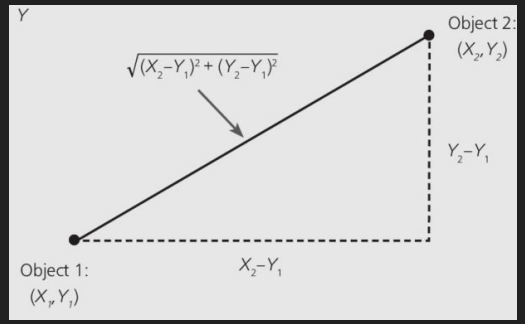
Distance Measures

And their use cases

Euclidean distance

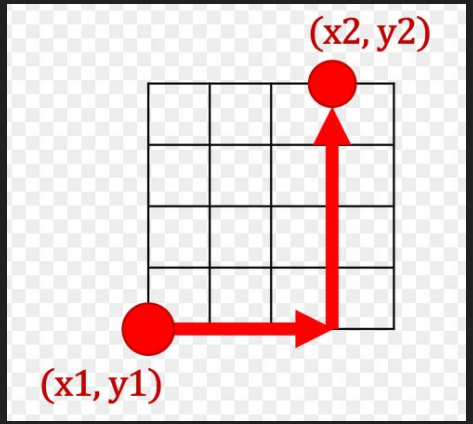
- Distance between real number vectors
- Calculate distance between rows containing numerical values
- The most common with continuous data



$$dist = \sqrt{\sum_{k=1}^{n}(p_k - q_k)^2}$$

Manhatten Distance

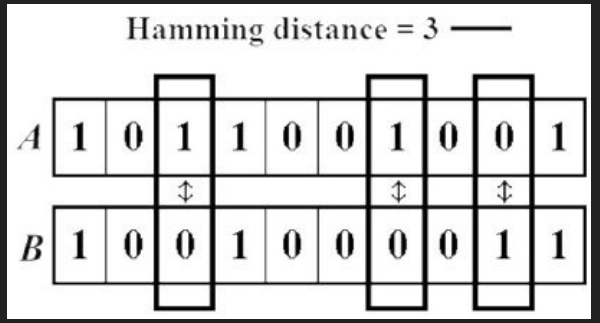
- Distance between real-value vectors
- Useful on data points on a uniform grid



$$\text{Mdist} = |x2 - x1| + |y2 - y1|$$

Hamming Distance

- Distance between vectors containing binary digits
- Useful in datasets with one-hot encoded data



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

Source

- https://machinelearningmastery.com/distance-measures-for-machine-learning/