Distance Measures

- Distance measures, also known as similarity measures, are mathematical techniques used to quantify the similarity or dissimilarity between objects or data points
- As mentioned, in K-means clustering we use distance measures to determine which cluster a data point belongs to
- Some commonly used distance measures include:
  - Euclidean Distance: The straight-line distance between two points (p and q) in Euclidean space. In n dimensional space this is calculated as:

$$\sqrt{\sum_{i=1}^{n}(p_i - q_i)^2}$$

Distance Measures

- Distance measures, also known as similarity measures, are mathematical techniques used to quantify the similarity or dissimilarity between objects or data points
- As mentioned, in K-means clustering we use distance measures to determine which cluster a data point belongs to
- Some commonly used distance measures include:
  - Manhattan Distance: The sum of the absolute differences between the coordinates of two points (p and q). In n dimensional space this is calculated as:

$$\sum_{i=1}^{n} |(p_i - q_i)|$$

Distance Measures

- Distance measures, also known as similarity measures, are mathematical techniques used to quantify the similarity or dissimilarity between objects or data points
- As mentioned, in K-means clustering we use distance measures to determine which cluster a data point belongs to
- Typically, a small value of a distance measure indicates a larger similarity between two points
- In K-means, we will use a Euclidean Distance measure for clustering, however different distance measures can be adapted to adhere to the characteristics of a particular dataset