**1. Cosine Similarity**

* **Description**: Measures the cosine of the angle between two vectors in a multi-dimensional space.
* **Use Case**: High-dimensional data like text embeddings or TF-IDF vectors.
* **Formula**: Cosine Similarity=A⋅B∥A∥∥B∥\text{Cosine Similarity} = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \|\mathbf{B}\|}Cosine Similarity=∥A∥∥B∥A⋅B​ Where A\mathbf{A}A and B\mathbf{B}B are vectors.

**2. Euclidean Distance**

* **Description**: Measures the straight-line distance between two points (vectors) in space.
* **Use Case**: Quantifies absolute differences; works well with low-dimensional numerical data.
* **Formula**: Euclidean Distance=∑i=1n(ai−bi)2\text{Euclidean Distance} = \sqrt{\sum\_{i=1}^{n}(a\_i - b\_i)^2}Euclidean Distance=i=1∑n​(ai​−bi​)2​

**3. Jaccard Similarity**

* **Description**: Measures the similarity between two sets based on the size of their intersection divided by the size of their union.
* **Use Case**: Comparing sets, such as unique words in documents.
* **Formula**: Jaccard Similarity=∣A∩B∣∣A∪B∣\text{Jaccard Similarity} = \frac{|A \cap B|}{|A \cup B|}Jaccard Similarity=∣A∪B∣∣A∩B∣​

**4. Manhattan Distance (Taxicab Distance)**

* **Description**: Measures the distance between two points by summing the absolute differences along each dimension.
* **Use Case**: Data where direct paths are restricted, like grid-based systems.
* **Formula**: Manhattan Distance=∑i=1n∣ai−bi∣\text{Manhattan Distance} = \sum\_{i=1}^{n} |a\_i - b\_i|Manhattan Distance=i=1∑n​∣ai​−bi​∣

**5. Pearson Correlation Coefficient**

* **Description**: Measures the linear correlation between two sets of data. It assesses how well one variable predicts the other.
* **Use Case**: Finding relationships between variables in numerical datasets.
* **Formula**: r=∑(xi−xˉ)(yi−yˉ)∑(xi−xˉ)2∑(yi−yˉ)2r = \frac{\sum (x\_i - \bar{x})(y\_i - \bar{y})}{\sqrt{\sum (x\_i - \bar{x})^2 \sum (y\_i - \bar{y})^2}}r=∑(xi​−xˉ)2∑(yi​−yˉ​)2​∑(xi​−xˉ)(yi​−yˉ​)​

**6. Hamming Distance**

* **Description**: Measures the number of positions at which two strings of equal length differ.
* **Use Case**: Comparing binary strings or DNA sequences.
* **Formula**: Hamming Distance=Count of mismatches between corresponding elements\text{Hamming Distance} = \text{Count of mismatches between corresponding elements}Hamming Distance=Count of mismatches between corresponding elements

**7. Dice Coefficient (Sorensen-Dice Similarity)**

* **Description**: Similar to Jaccard but gives double weight to shared elements.
* **Use Case**: Set comparison, especially for text overlaps.
* **Formula**: Dice Similarity=2∣A∩B∣∣A∣+∣B∣\text{Dice Similarity} = \frac{2|A \cap B|}{|A| + |B|}Dice Similarity=∣A∣+∣B∣2∣A∩B∣​

**8. Minkowski Distance**

* **Description**: A generalization of Euclidean and Manhattan distances. Controlled by a parameter ppp.
* **Use Case**: When you want to adjust the sensitivity of the measure.
* **Formula**: Minkowski Distance=(∑i=1n∣ai−bi∣p)1/p\text{Minkowski Distance} = \left( \sum\_{i=1}^{n} |a\_i - b\_i|^p \right)^{1/p}Minkowski Distance=(i=1∑n​∣ai​−bi​∣p)1/p Special Cases:
  + p=2p = 2p=2: Euclidean Distance.
  + p=1p = 1p=1: Manhattan Distance.

**9. Bhattacharyya Distance**

* **Description**: Measures the similarity between two probability distributions.
* **Use Case**: Comparing histograms or probability data.

**10. Cosine Overlap Similarity**

* **Description**: A variant of cosine similarity that accounts only for overlapping dimensions.
* **Use Case**: Sparse text data.