Cosine Distance

Measures the angular difference between vectors, ignoring their magnitude

$$d_{\cos}(\mathbf{x}, \mathbf{y}) = 1 - \frac{\mathbf{x} \cdot \mathbf{y}}{\|\mathbf{x}\| \cdot \|\mathbf{y}\|}$$

Terms Explained:

- $\mathbf{x}, \mathbf{y}$: Non-zero vectors in $\mathbb{R}^n$
- $\mathbf{x} \cdot \mathbf{y}$: Dot product
- $\|\mathbf{x}\|, \|\mathbf{y}\|$: Euclidean norms
- Range: 0 (same direction) to 2 (opposite directions)

Use Cases:

Information Retrieval: Document similarity

Recommender Systems: User preference matching

@AIinMinutes

Euclidean Distance

Measures the straight-line distance between two vectors in space; equal to the length of their difference vector

$$d(\boldsymbol{x}, \boldsymbol{y}) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} = \|\boldsymbol{x} - \boldsymbol{y}\|_2$$

Terms Explained:

- $\boldsymbol{x}, \boldsymbol{y}$: Vectors in $\mathbb{R}^n$
- $x_i - y_i$: Difference at dimension i
- $\|\boldsymbol{x} - \boldsymbol{y}\|_2$: L_2 norm of the difference vector

Use Cases:

k-Nearest Neighbors: Finding similar data points
k-Means: Clusters data by minimizing intra-cluster distances.

@AIinMinutes

Mahalanobis Distance

Measures distance while accounting for correlations among features

$$d(\boldsymbol{x}, \boldsymbol{y}) = \sqrt{(\boldsymbol{x} - \boldsymbol{y})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{y})}$$

Terms Explained:

- $\boldsymbol{x}, \boldsymbol{y}$: Vectors in $\mathbb{R}^n$
- $\boldsymbol{\Sigma}^{-1}$: Inverse covariance matrix
- Normalizes by feature covariances

Use Cases:

Outlier Detection: Accounts for feature correlations
Classification: Handles different feature scales and correlations

@AIinMinutes

Hellinger Distance

Measures how different two probability distributions are

$$H(\boldsymbol{P}, \boldsymbol{Q}) = \frac{1}{\sqrt{2}} \sqrt{\sum_{i=1}^{n} (\sqrt{P_i} - \sqrt{Q_i})^2}$$

Terms Explained:

- $\boldsymbol{P}, \boldsymbol{Q}$: Probability distributions
- $\sqrt{P_i}$: Square root of probability at position i
- $H(\boldsymbol{P}, \boldsymbol{Q}) \in [0, 1]$: $0 =$ identical, $1 =$ no overlap

Use Cases:

Anomaly Detection: Identifies statistical deviations
Imbalance-aware Algorithms: Used in Hellinger Distance Decision Trees for handling class imbalance.

@AIinMinutes

Jaccard Distance

Measures how different two sets are by comparing their shared and unique elements

$$d_J(X, Y) = 1 - \frac{|X \cap Y|}{|X \cup Y|}$$

Terms Explained:

- X, Y: Two sets
- $|X \cap Y|$: Size of intersection
- $|X \cup Y|$: Size of union
- Range: 0 (identical) to 1 (disjoint)

Use Cases:

Document Similarity: Comparing text as word sets
Recommender Systems: Finding similar user preferences

@AIinMinutes

Manhattan Distance

Measures distance as the sum of absolute differences along each axis

$$d(\boldsymbol{x}, \boldsymbol{y}) = \sum_{i=1}^{n} |x_i - y_i| = \|\boldsymbol{x} - \boldsymbol{y}\|_1$$

Terms Explained:

- $\boldsymbol{x}, \boldsymbol{y}$: Vectors in $\mathbb{R}^n$
- $|x_i - y_i|$: Absolute difference at dimension i
- $\|\boldsymbol{x} - \boldsymbol{y}\|_1$: L_1 norm (taxicab norm)

Use Cases:

Grid Navigation: Calculating city block distances
Feature Selection: L1 Regularizer

@AIinMinutes

Correlation Distance

Measures dissimilarity based on how variables are statistically related

$$d_{\text{corr}}(\boldsymbol{x}, \boldsymbol{y}) = 1 - \rho(\boldsymbol{x}, \boldsymbol{y}) = 1 - \frac{\text{cov}(\boldsymbol{x}, \boldsymbol{y})}{\sigma_{\boldsymbol{x}} \sigma_{\boldsymbol{y}}}$$

Terms Explained:

- $\boldsymbol{x}, \boldsymbol{y}$: Data vectors of equal length
- $\rho(\boldsymbol{x}, \boldsymbol{y})$: Pearson correlation coefficient
- $\text{cov}(\boldsymbol{x}, \boldsymbol{y})$: Covariance between $\boldsymbol{x}$ and $\boldsymbol{y}$
- $\sigma_{\boldsymbol{x}}, \sigma_{\boldsymbol{y}}$: Standard deviations

Use Case:

Feature Agglomeration: Correlation Clustering

@AIinMinutes

Dice Distance/Loss

Measures set dissimilarity, placing greater emphasis on shared elements than the Jaccard distance

$$d_D(X, Y) = 1 - \frac{2|X \cap Y|}{|X| + |Y|}$$

Terms Explained:

- X, Y: Two sets
- $|X \cap Y|$: Size of intersection
- $|X| + |Y|$: Sum of set sizes
- Range: 0 (identical) to 1 (no overlap)

Use Cases:

Image Segmentation: Evaluates segmentation overlap in image analysis; also as a loss function

@AIinMinutes

Hamming Distance

Counts the number of positions where two sequences differ

$$d_H(\boldsymbol{x}, \boldsymbol{y}) = \sum_{i=1}^{n} \mathbb{I}(x_i \neq y_i)$$

Terms Explained:

- $\boldsymbol{x}, \boldsymbol{y}$: Equal-length sequences
- $\mathbb{I}(x_i \neq y_i)$: Indicator function (1 if $x_i \neq y_i$, 0 otherwise)
- Counts positions where elements differ

Use Cases:

Error Detection: Hamming codes for transmission errors
Bioinformatics: Comparing DNA sequences

@AIinMinutes

Chebyshev Distance

Measures distance between vectors using the largest absolute difference in any dimension

$$d_\infty(\boldsymbol{x}, \boldsymbol{y}) = \max_i |x_i - y_i| = \|\boldsymbol{x} - \boldsymbol{y}\|_\infty$$

Terms Explained:

- $\boldsymbol{x}, \boldsymbol{y}$: Vectors in $\mathbb{R}^n$
- $\max_i |x_i - y_i|$: Maximum absolute difference
- $\|\boldsymbol{x} - \boldsymbol{y}\|_\infty$: L_∞ norm (chessboard distance)

Use Cases:

Anomaly Detection: Flags outliers based on the largest deviation across features
Warehouse Optimization: Finding minimax distances

@AIinMinutes