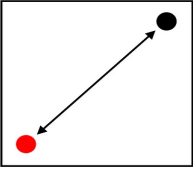
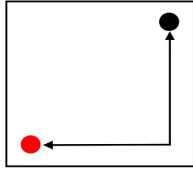
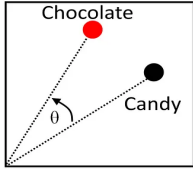
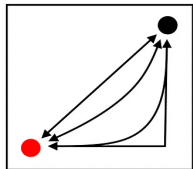
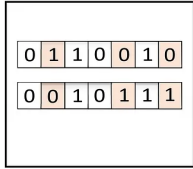
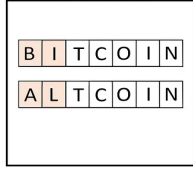
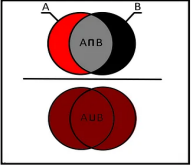
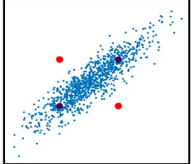
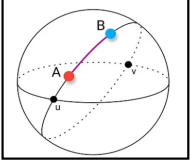
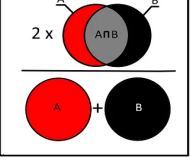
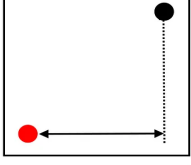
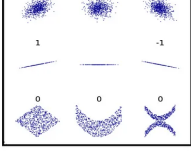
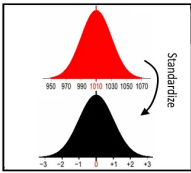
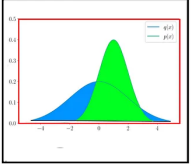
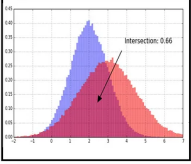
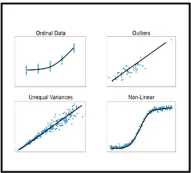
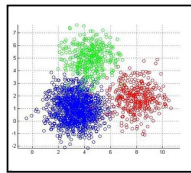


Picture	Distance Measure	Application	Features	Disadvantages	Formula	Big O notation	Processing Speed
Euclidean 	Euclidean Distance	General distance measurement, Clustering, classification, regression	Measures the straight line distance between two points in n-dimensional space.	Sensitive to outliers, Can be affected by scale differences	$\sqrt{\sum((p1 - p2)^2)}$	O(n)	Fast
Manhattan 	Manhattan Distance	Distance on grid networks, Routing algorithms, Image processing	Measures the distance between two points on a grid network, where movement is limited.	Only takes into account differences in individual coordinates, Ignores diagonal movement, May not be as useful for high-dimensional data,	$\sum(\text{abs}(p1 - p2))$	O(n)	Fast
Cosine Chocolate Candy 	Cosine Similarity	Text document clustering, Text analysis, Recommendation systems	Measures the cosine of the angle between two vectors	Ignores magnitude of vectors, May not be as useful for negative values or data with a high degree of correlation	$\text{dot product}(p1, p2) / (\text{norm}(p1) * \text{norm}(p2))$	O(n)	Fast
Minkowski 	Minkowski Distance	General distance measurement	Measures the distance between two points in n-dimensional space, where r determines the metric used.	Sensitive to outliers	$(\sum(\text{abs}(p1 - p2)^r))^{1/r}$	O(n)	Fast
Hamming 	Hamming Distance	Measuring string similarity, Error-correcting codes, DNA sequencing	Measures the number of positions at which the corresponding symbols are different.	Only for same length strings, May not be as useful for continuous data	$\sum(p1 \neq p2)$	O(n)	Fast
Levenshtein 	Levenshtein Distance	Measuring string similarity	Measures the minimum number of single-character edits required to transform one string into another.	More expensive for long strings	Dynamic Programming	O(n^2)	Slow

Picture	Distance Measure	Application	Features	Disadvantages	Formula	Big O notation	Processing Speed
<p>Jaccard</p> 	Jaccard Similarity	Set similarity measurement, Text analysis, recommendation systems	Measures the similarity between two sets by comparing their intersection and union.	Ignores magnitude of sets, May not be as useful for continuous data	$(p1 \cap p2) / (p1 \cup p2)$	$O(n)$	Fast
<p>Mahalanobis</p> 	Mahalanobis Distance	Multivariate statistical analysis, Outlier detection, Clustering	Measures the distance between two points in n-dimensional space, taking into account the correlation between variables.	Requires full covariance matrix, May not be as useful for datasets with a large number of variables	$\sqrt{(p1 - p2)' S^{-1} (p1 - p2)}$	$O(n^3)$	Slow
<p>Haversine</p> 	Haversine Distance	Measuring distance on a sphere, Geographic calculations	Measures the great-circle distance between two points on a sphere, such as the Earth.	May not be as useful for small distances or for applications where a spherical Earth is not an accurate representation	$2r * \arcsin(\sqrt{\sin^2((lat2 - lat1)/2) + \cos(lat1)\cos(lat2)\sin^2((lon2 - lon1)/2)})$	$O(1)$	Slow
<p>Sørensen–Dice</p> 	Sørensen–Dice Index	Measuring similarity of sets, Ecology, biology, genetics	Measures the similarity between two sets	May not be as useful for continuous data and Ignores magnitude of sets	$2 * \text{size}(\text{intersection}(p1, p2)) / (\text{size}(p1) + \text{size}(p2))$	$O(n)$	Fast
<p>Chebychev</p> 	Chebyshev Distance	Measuring maximum difference, Clustering, anomaly detection	Measures the maximum difference between corresponding components of two vectors	Only applicable for continuous data, Sensitive to outliers, may not be as useful for highly correlated data	$\max(\text{abs}(p1 - p2))$	$O(n)$	Fast
<p>Pearson</p> 	Pearson Correlation	Measuring linear correlation	Measures the linear correlation between two variables in a dataset.	Requires linear correlation	$\text{cov}(p1, p2) / (\text{std}(p1) * \text{std}(p2))$	$O(n^2)$	Slow

Picture	Distance Measure	Application	Features	Disadvantages	Formula	Big O notation	Processing Speed
<p>SED</p> 	Squared Euclidean Distance	Clustering algorithms	Measures the square of the straight line distance between two points in n-dimensional space.	More sensitive to outliers	$\text{sum}((p1 - p2)^2)$	$O(n)$	Fast
<p>Jensen-Shannon</p> 	Jensen-Shannon Divergence	Measuring similarity of probability distributions. Commonly used in clustering and recommendation systems	symmetric measure used to compare probability distributions. considers both similarities and differences between the distributions	Only applicable for non-negative vectors	$\text{sqrt}(\text{JS}(p1 p3)/2 + \text{JS}(p2 p3)/2)$	$O(n)$	Slow
<p>Chi-Square</p> 	Chi-Square Distance	Measuring similarity of histograms	Measures the distance between two histograms by comparing their Chi-Square divergence.	Only applicable for non-negative vectors	$\text{sum}((p1 - p2)^2 / (p1 + p2)) / 2$	$O(n)$	Fast
<p>Spearman</p> 	Spearman Correlation	Measuring rank correlation	Measures the correlation between two variables in a dataset based on their rank order.	Only applicable for ordinal data	$\text{cov}(\text{rank}(p1), \text{rank}(p2)) / (\text{std}(\text{rank}(p1)) * \text{std}(\text{rank}(p2)))$	$O(n \log n)$	Fast
<p>Canberra</p> 	Canberra Distance	Measuring distance for sparse data	Measures the distance between two vectors, taking into account their relative magnitudes.	More sensitive to small differences	$\text{sum}(\text{abs}(p1 - p2) / (\text{abs}(p1) + \text{abs}(p2)))$	$O(n \log n)$	Fast