KNN

1. Role of Distance in KNN

K-Nearest Neighbors (KNN) is a **non-parametric, instance-based learning algorithm**.
It does not learn an explicit model during training. Instead, it stores all training samples and makes predictions by comparing a query point to stored samples using a **distance metric**.

Given:

- A dataset of points $X = \{x_1, x_2, \dots, x_n\}$
- A query point q
- A distance function $d(x, q)$
- A chosen number of neighbors k

KNN:

1. Computes $d(x_i, q)$for all i
2. Selects the $k$smallest distances
3. Aggregates the labels (classification) or values (regression)

Thus, the **distance function directly determines the behavior, accuracy, and bias of KNN**.

---

2. Euclidean Distance (L2 Norm)

Formula

For two points $x = (x_1, x_2, \dots, x_n)$and
$y = (y_1, y_2, \dots, y_n)$:

$$d_{\text{Euclidean}}(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$

Explanation

- Measures straight-line distance in n-dimensional space
- Derived from the Pythagorean theorem
- Assumes all features are continuous and equally scaled

Properties

- Sensitive to feature scaling
- Dominated by features with large numeric ranges
- Commonly used when features are normalized

Use Cases

- Continuous numerical data
- Low to moderate dimensional spaces

---

3. Manhattan Distance (L1 Norm)

Formula

$$d_{\text{Manhattan}}(x, y) = \sum_{i=1}^{n} | x_i - y_i |$$

Explanation

- Measures distance along axis-aligned paths
- Equivalent to the total absolute coordinate differences
- Represents “city block” distance

Properties

- Less sensitive to outliers than Euclidean distance
- No square or square root operations
- Works well when dimensions are independent

Use Cases

- Sparse data
- High-dimensional spaces
- When feature differences should add linearly

---

4. Minkowski Distance (Generalized Norm)

Formula

$$d_{\text{Minkowski}}(x,y) = \left(\sum_{i=1}^{n} |x_i - y_i|^p\right)^{\frac{1}{p}}$$

Explanation

- Generalizes Euclidean and Manhattan distances
- Parameter $p$controls distance behavior

Special Cases

- $p = 1$: Manhattan distance
- $p = 2$: Euclidean distance
- $p \to \infty$: Chebyshev distance

Properties

- Flexible and tunable
- Computational cost increases with p

---

5. Chebyshev Distance (L∞ Norm)

Formula

$$d_{\text{Chebyshev}}(x,y) = \max_i |x_i - y_i|$$

Explanation

- Measures the maximum difference across any single dimension
- Assumes the worst-case deviation defines similarity

Properties

- Extremely sensitive to single-feature differences
- Ignores cumulative effects of other features

Use Cases

- Quality control
- Chessboard distance

- Threshold-based similarity systems

---

6. Cosine Distance

Formula (Cosine Similarity)

$$\text{cosine similarity}(x, y) = \frac{x \cdot y}{\| x \| \| y \|}$$

Conversion to Distance

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

Explanation

- Measures angle between vectors, not magnitude
- Focuses on orientation rather than absolute values

Properties

- Scale-invariant
- Effective in high-dimensional spaces

Use Cases

- Text classification
- Document similarity
- Recommendation systems

---

7. Hamming Distance

Formula

For categorical or binary vectors:

$$d_{\text{Hamming}}(x, y) = \sum_{i=1}^{n} \mathbb{1}(x_i \neq y_i)$$

Explanation

- Counts the number of mismatched positions
- Suitable only for discrete or binary features

Properties

- Simple and interpretable
- Not applicable to continuous data

Use Cases

- Binary classification
- Error detection
- Genetic sequence comparison

---

8. Mahalanobis Distance

Formula

$$d_{\text{Mahalanobis}}(x, y) = \sqrt{(x - y)^T S^{-1} (x - y)}$$

Where:

- S is the covariance matrix of the data

Explanation

- Accounts for feature correlations
- Normalizes distances based on variance structure

Properties

- Scale-invariant
- Computationally expensive
- Requires invertible covariance matrix

Use Cases

- Correlated features
- Statistical anomaly detection

---

9. Importance of Feature Scaling

Distance-based methods are highly sensitive to scale.

Common scaling techniques:

- Min-Max normalization
- Standardization (Z-score)

Without scaling, distance calculations become biased toward high-variance features.

---

10. Summary Table

DISTANCE METRIC	SENSITIVE TO SCALE	HANDLES CORRELATION	TYPICAL USE
EUCLIDEAN	Yes	No	General numeric data
MANHATTAN	Yes	No	High-dimensional data
MINKOWSKI	Yes	No	Tunable metric
CHEBYSHEV	Yes	No	Max deviation
COSINE	No	No	Text, embeddings
HAMMING	No	No	Categorical data
MAHALANOBIS	No	Yes	Statistical data