



Distance & Similarity Measures



Dr. Konstantinos Karampidis



Comparing Patterns in Feature Space (1/2)

- ❑ How can we determine whether two patterns are similar?
- ❑ How can we quantify how different two patterns are?
- ❑ How does the choice of a distance or similarity measure affect recognition?



Comparing Patterns in Feature Space (1/2)

Example applications:

- ❑ Comparing handwritten digits
- ❑ Comparing images
- ❑ Finding similar objects
- ❑ Nearest-neighbor classification
- ❑ Clustering

Similar patterns should have similar representation
in an appropriate feature space



Comparing Patterns

In Pattern Recognition, we often need to compare two patterns. Each pattern can be represented by a feature vector.

Consider two patterns:

$$X = \begin{bmatrix} 2 \\ 3 \end{bmatrix}, Y = \begin{bmatrix} 5 \\ 7 \end{bmatrix}$$

The two patterns are represented by points in a 2-dimensional feature space.

A natural question is:

How far apart are the two patterns?



Comparing Patterns

Example:

Suppose the two features are height and width of an object.

Pattern x : height = 2, width = 3

Pattern y : height = 5, width = 7

Key idea:

A distance measure converts the difference between two patterns into a numerical value.



Comparing Patterns

Exercise:

Which pair of patterns is more similar

$x_1=(2,3)$, $x_2=(3,4)$

$x_3=(2,3)$, $x_4=(8,9)$

Explain your answer without calculating a distance yet.



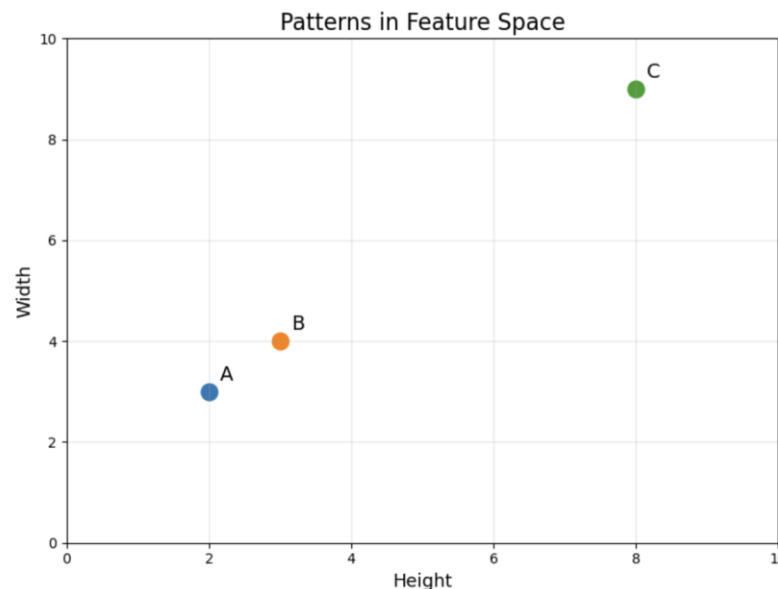
Patterns in Feature Space

- ❑ A feature vector represents a pattern as a point in a feature space.
- ❑ Each dimension corresponds to one feature.
- ❑ Patterns with similar feature values tend to appear close to each other.
- ❑ Patterns with very different feature values tend to appear farther apart.



Patterns in Feature Space

- ❑ Example — Objects
- ❑ Let the features be: x_1 : Height and x_2 : Width
- ❑ Three patterns:
 - ❑ $\mathbf{x}_A = (2, 3)$
 - ❑ $\mathbf{x}_B = (3, 4)$
 - ❑ $\mathbf{x}_C = (8, 9)$
- ❑ In the feature space:
 - ❑ A and B are close.
 - ❑ A and C are far apart.
 - ❑ B and C are also far apart.





Similarity vs. Dissimilarity

□ Similarity

Measures how alike two patterns are.

Higher similarity → more similar patterns.

Similarity is often represented by a numerical score.

□ Dissimilarity

Measures how different two patterns are.

Higher dissimilarity → more different patterns.

A distance measure is a common way to quantify dissimilarity.



Similarity vs. Dissimilarity

□ Example

Consider:

$$A = (2, 3), B = (3, 4)$$

The two patterns are close in feature space.

Therefore:

- High similarity
- Low dissimilarity
- Small distance



Similarity vs. Dissimilarity

Exercise:

Consider three pairs of objects represented by their Height and Width:

$$A=(2,3) , B=(3,4)$$

$$C=(2,3) , D=(2.5,3.5)$$

$$E=(2,3) , F=(8,9)$$

Tasks:

- ❑ Which pair contains the most similar patterns?
- ❑ Which pair contains the least similar patterns?
- ❑ Rank the three pairs from most similar to least similar.
- ❑ Rank the three pairs from least dissimilar to most dissimilar.
- ❑ For each pair, indicate whether you expect the distance between the two patterns to be small, medium, or large.

Do not calculate any distance yet. Use only the position of the patterns in the feature space.



Similarity vs. Dissimilarity

Exercise:

Consider three pairs of objects represented by their Height and Width:

$A=(2,3)$, $B=(3,4)$

$C=(2,3)$, $D=(2.5,3.5)$

$E=(2,3)$, $F=(8,9)$

Answer:

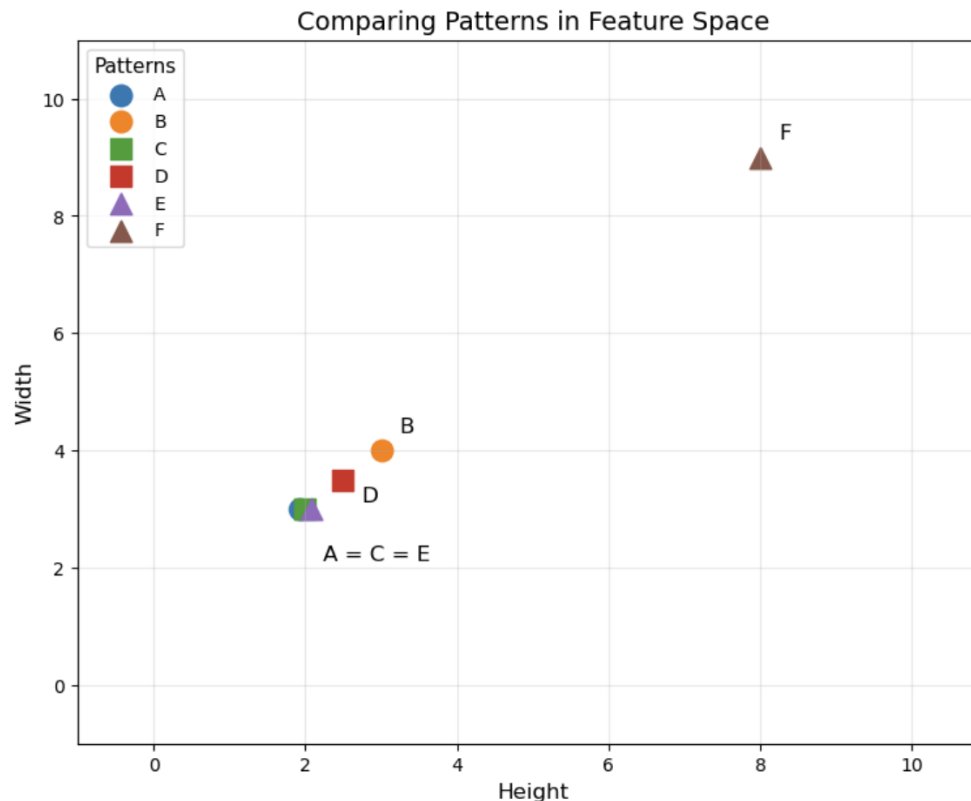
Pair	Patterns	Expected similarity	Expected distance
Pair 1	$A = (2, 3), B = (3, 4)$	High	Small
Pair 2	$C = (2, 3), D = (2.5, 3.5)$	Very High	Very Small
Pair 3	$E = (2, 3), F = (8, 9)$	Low	Large

- Ranking by similarity: Pair 2 > Pair 1 > Pair 3
- Ranking by distance/dissimilarity: Pair 2 < Pair 1 < Pair 3

Similarity vs. Dissimilarity

Important observation:

- The more similar two patterns are, the smaller their distance is expected to be.





What Is a Distance?

Distance

A distance measure quantifies how far apart two patterns are in feature space.

For two patterns represented by feature vectors
 x and y

the distance is written as

$$d(x, y)$$

Interpretation

- ❑ Small distance $\rightarrow$ patterns are similar
- ❑ Large distance $\rightarrow$ patterns are different
- ❑ Distance depends on the chosen distance measure.



What Is a Distance?

Example

Consider:

$$x = (2, 3)$$

$$y = (3, 4)$$

These two points are close in feature space, so we expect:
 $d(x, y)$ to be small

□ Key Idea

Distance transforms geometric difference into a numerical value.



Properties of a Distance Measure

A distance measure should satisfy some basic properties.

1. Non-negativity - A distance cannot be negative.

$$d(x, y) \geq 0$$

2. Identity - Two identical patterns have zero distance.

$$d(x, y) = 0 \iff x = y$$

3. Symmetry - The distance from x to y is the same as from y to x .

$$d(x, y) = d(y, x)$$

4. Triangle Inequality - Going directly from x to z cannot be longer than going through y .

$$d(x, z) \leq d(x, y) + d(y, z)$$



Properties of a Distance Measure

Consider three patterns A , B , and C , and suppose their pairwise distances are:

$$d(A, A) = 0, \quad d(A, B) = 4, \quad d(B, A) = 4, \quad d(B, C) = 3, \quad d(A, C) = 5$$

- ❑ **Non-negativity** $d(A, B) = 4 \geq 0$ ✓ **Satisfied.**
- ❑ **Identity** $d(A, A) = 0$ ✓ **Satisfied.**
- ❑ **Symmetry** $d(A, B) = 4 = d(B, A)$ ✓ **Satisfied.**
- ❑ **Triangle inequality** $d(A, C) = 5$ while
 $d(A, B) + d(B, C) = 4 + 3 = 7$

Therefore: $5 \leq 7$ ✓ Satisfied.

Key Idea

- ❑ **A valid distance measure must satisfy all four properties.**



Euclidean Distance

Definition

The Euclidean distance is the straight-line distance between two patterns in feature space.

For two feature vectors

$$\mathbf{x} = (x_1, x_2, \dots, x_d)$$

and

$$\mathbf{y} = (y_1, y_2, \dots, y_d)$$

the Euclidean distance is:

$$d_E(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{i=1}^d (x_i - y_i)^2}$$



Euclidean Distance

Example

Consider two patterns:

$$A=(2,3) , B=(5,7)$$

Then:

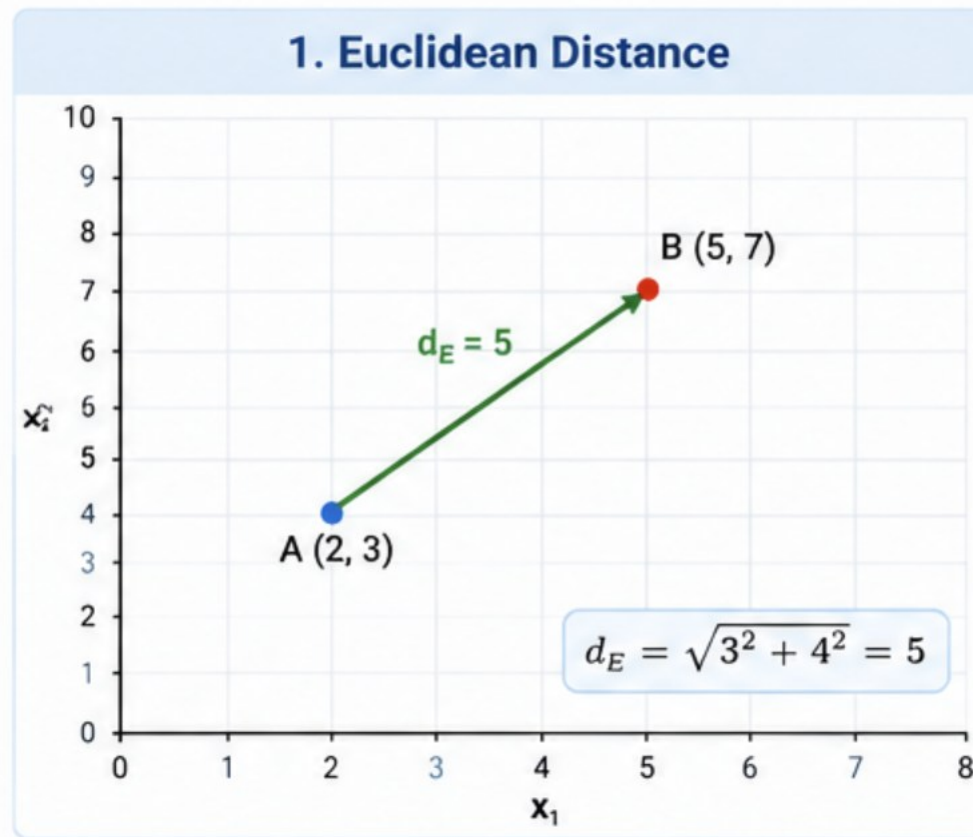
$$\begin{aligned}d_E(A, B) &= \sqrt{(5 - 2)^2 + (7 - 3)^2} \\&= \sqrt{3^2 + 4^2} \\&= \sqrt{25} = 5\end{aligned}$$

Geometric Interpretation

The Euclidean distance is the straight-line distance between the two points. For: $A = (2, 3)$, $B = (5, 7)$ the horizontal difference is 3, the vertical difference is 4, and the distance is 5.

Euclidean Distance

Example





Manhattan Distance

Definition

The Manhattan distance measures the total distance travelled along each feature dimension.

For two feature vectors:

$\mathbf{x} = (x_1, x_2, \dots, x_d)$ and $\mathbf{y} = (y_1, y_2, \dots, y_d)$

the Manhattan distance is:

$$d_M(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^d |x_i - y_i|$$



Manhattan Distance

Example

Consider:

$$A = (2, 3) , B = (5, 7)$$

$$\text{Then: } d_M(A, B) = |5-2| + |7-3| = 3+4=7$$

Geometric Interpretation

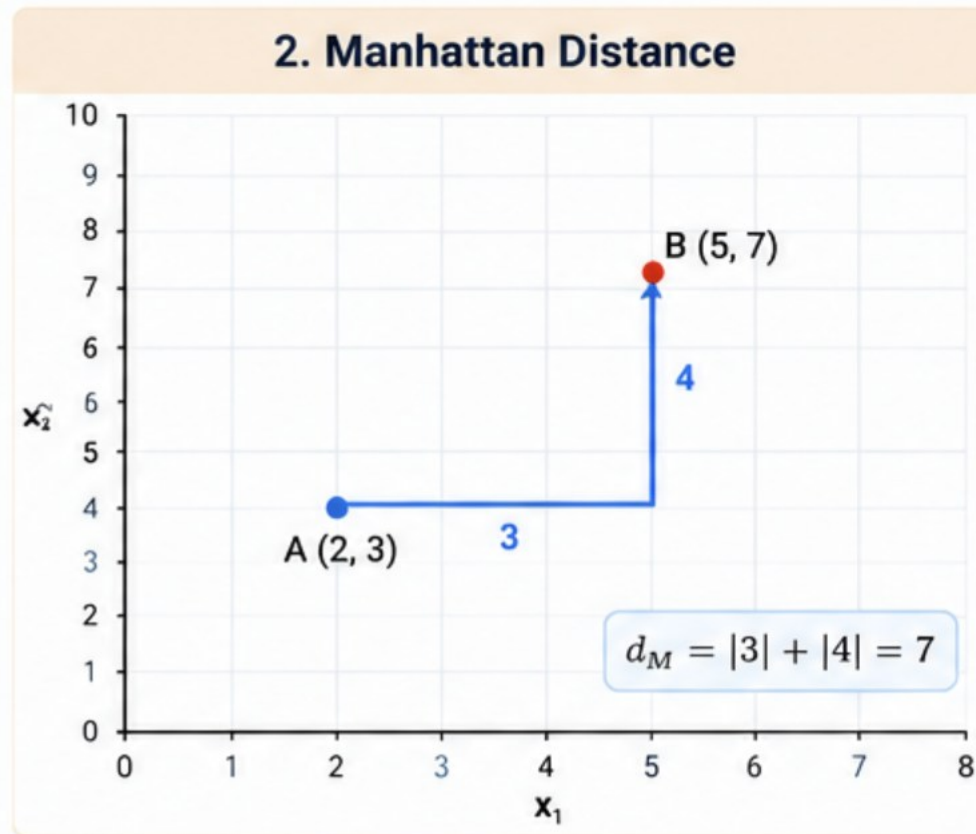
Instead of measuring the direct straight-line path, Manhattan distance measures movement along the feature axes. For the same two points:

- ❑ Horizontal difference: 3
- ❑ Vertical difference: 4
- ❑ Manhattan distance: $3+4=7$

Euclidean distance measures straight-line separation, while Manhattan distance measures the total axis-aligned separation.

Manhattan Distance

Example





Minkowski Distance

Definition

The Minkowski distance is a generalized distance measure that includes several commonly used distance measures as special cases.

For two feature vectors:

$$\mathbf{x} = (x_1, x_2, \dots, x_d) \text{ and } \mathbf{y} = (y_1, y_2, \dots, y_d)$$

the Minkowski distance is:

$$d_p(\mathbf{x}, \mathbf{y}) = \left(\sum_{i=1}^d |x_i - y_i|^p \right)^{1/p}$$

where $p \geq 1$



Minkowski Distance

Example

Consider $A = (2, 3), B = (5, 7)$

For $p = 1$:

$$d_1(A, B) = |5 - 2| + |7 - 3| = 3 + 4 = \boxed{7}$$

This is the Manhattan distance.

For $p = 2$:

$$d_2(A, B) = \sqrt{(5 - 2)^2 + (7 - 3)^2} = \sqrt{25} = \boxed{5}$$

This is the Euclidean distance.

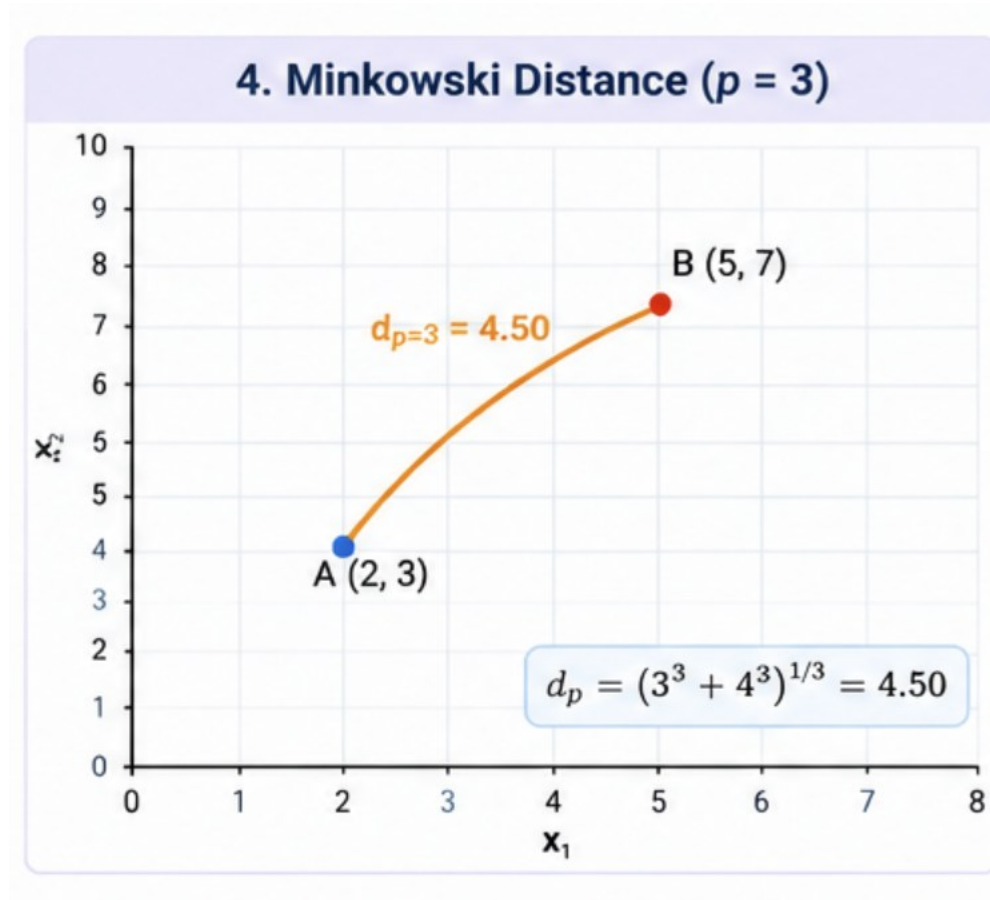
Important Relationship

- $p = 1 \rightarrow \text{Manhattan}$
- $p = 2 \rightarrow \text{Euclidean}$

Minkowski provides a family of distance measures controlled by the parameter p .

Minkowski Distance

Example





Chebysev Distance

Definition

The Chebyshev distance is the maximum absolute difference along any feature dimension.

For two feature vectors:

$$\mathbf{x} = (x_1, x_2, \dots, x_d) \text{ and } \mathbf{y} = (y_1, y_2, \dots, y_d)$$

the Chebyshev distance is:

$$d_C(\mathbf{x}, \mathbf{y}) = \max_i |x_i - y_i|$$



Chebysev Distance

Example

Consider:

$$A = (2, 3), \quad B = (5, 7)$$

The differences are:

$$|5 - 2| = 3$$

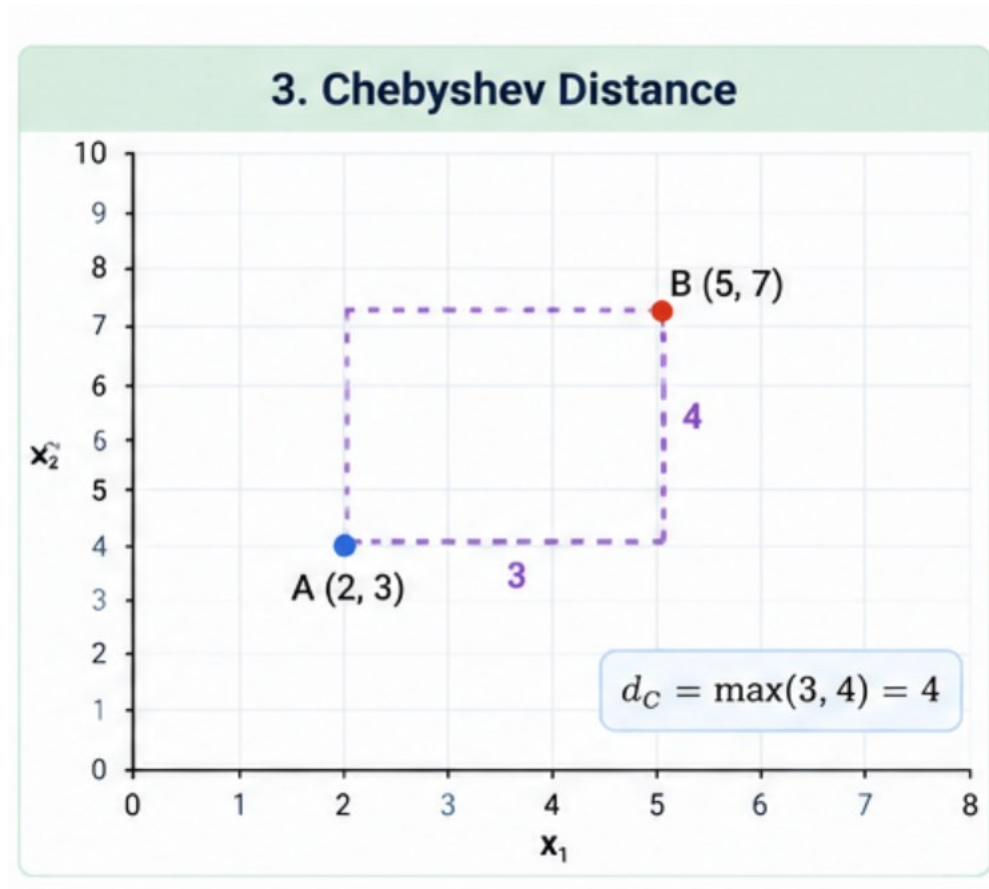
$$|7 - 3| = 4$$

Therefore:

$$d_c(A, B) = \max(3, 4) = 4$$

Chebyshev Distance

Example





Distances

Example

For the same two patterns:

$$A=(2,3) , B=(5,7)$$

Calculate and compare the distances

<i>Distance</i>	<i>Value</i>
Manhattan	7
Euclidean	5
Chebyshev	4

The same two patterns can have different distances depending on the distance measure.



Mahalanobis Distance

Euclidean distance treats all feature dimensions as if:

- they have comparable scales
- they are independent
- the variability is similar in all directions

In real datasets, this may not be true.

Definition

- For two patterns $\mathbf{x}$ and $\mathbf{y}$:

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

where:

S = covariance matrix of the features

S^{-1} = inverse covariance matrix



Mahalanobis Distance

Main Idea

Mahalanobis distance takes into account:

1. Feature variability

- A difference along a feature with high variability is considered less significant.

2. Correlation between features

- If two features are strongly correlated, their joint variation is taken into account.



Mahalanobis Distance

Intuitive Example

Suppose we have two features:

- Height
- Width

If height and width are strongly correlated, points tend to lie along a particular direction.

- Two points may have a large Euclidean distance, but if their difference follows the normal pattern of the data, their Mahalanobis distance may be relatively small.
- Conversely, a relatively small Euclidean difference in an unusual direction may result in a larger Mahalanobis distance.

Key Idea

- Mahalanobis distance measures distance relative to the statistical distribution of the data.

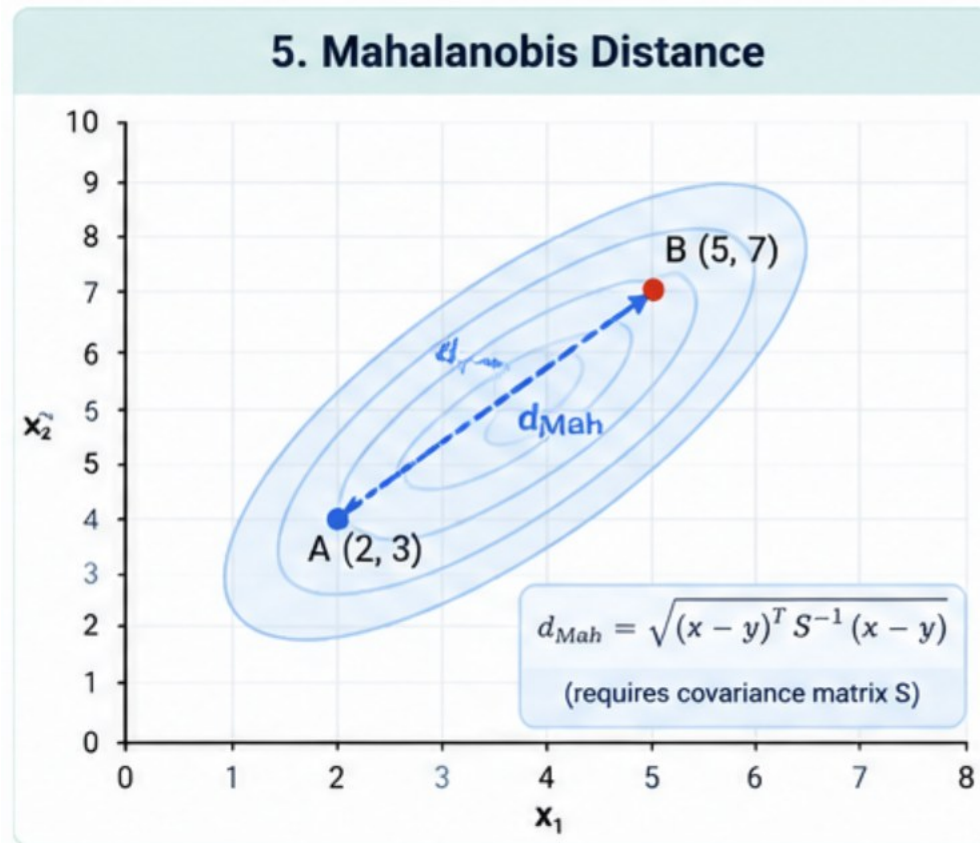


Mahalanobis Distance

Important Comparison

Distance	Takes variance into account?	Takes correlation into account?
Euclidean	No	No
Manhattan	No	No
Chebyshev	No	No
Mahalanobis	Yes	Yes

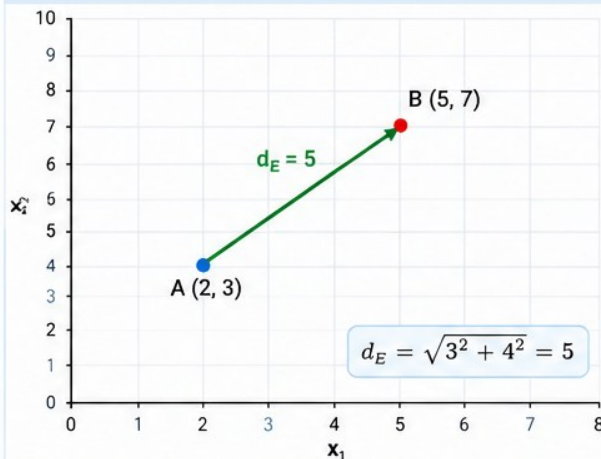
Mahalanobis Distance



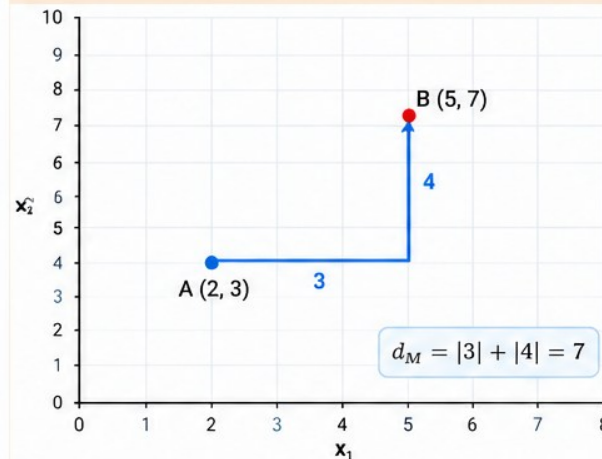
How Different Distance Measures Interpret the Same Two Points

$$A = (2, 3), B = (5, 7) \rightarrow \Delta x = 3, \Delta y = 4$$

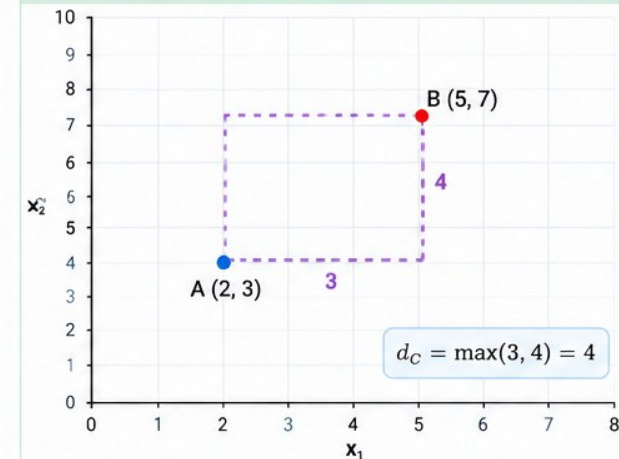
1. Euclidean Distance



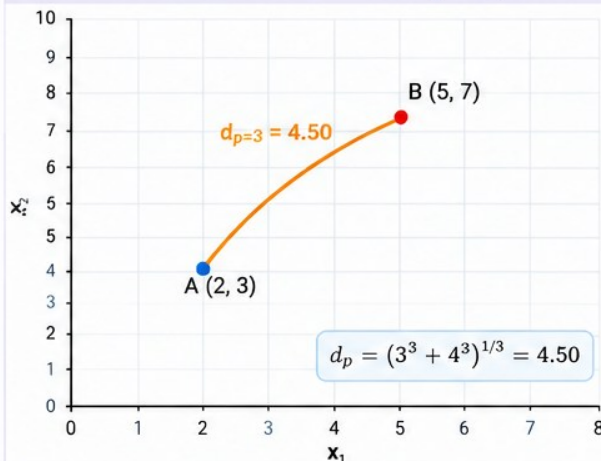
2. Manhattan Distance



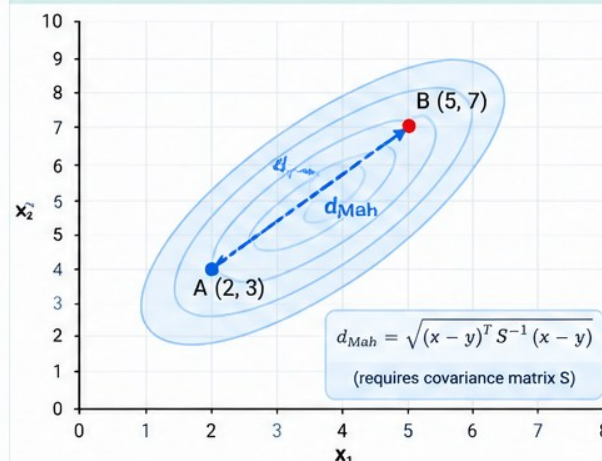
3. Chebyshev Distance



4. Minkowski Distance ($p = 3$)



5. Mahalanobis Distance



Key Takeaways

- 1 **Euclidean:** straight-line distance (5)
- 2 **Manhattan:** horizontal + vertical (7)
- 3 **Chebyshev:** largest coordinate difference (4)
- 4 **Minkowski:** generalization ($p = 3$) (4.50)
- 5 **Mahalanobis:** considers variance and correlation (value depends on data)

The same two points can have different distances depending on the measure, because each measure captures a different notion of "distance".



Hamming Distance

Definition:

- ❑ Measures the number of positions at which two vectors differ.
- ❑ Mainly used for binary, categorical, or discrete representations.
- ❑ It counts mismatches, rather than measuring geometric distance.

For two vectors of equal length:

$$d_H(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^d I(x_i \neq y_i)$$

where $I(\cdot)$ is 1 when the two values differ and 0 otherwise.



Hamming Distance

Example 1 — Binary vectors

$$\mathbf{x} = (1, 0, 1, 1, 0)$$

$$\mathbf{y} = (1, 1, 1, 0, 0)$$

Compare position by position:

Position	x	y	Different?
1	1	1	No
2	0	1	Yes
3	1	1	No
4	1	0	Yes
5	0	0	No

Therefore:

$$d_H(\mathbf{x}, \mathbf{y}) = 2$$

Interpretation: The two patterns differ in 2 out of 5 positions.



Hamming Distance

Example 2 — Categorical features

Suppose a pattern is represented by:

$x = (\text{Red}, \text{SUV}, \text{Diesel}, \text{Yes})$

$y = (\text{Red}, \text{SUV}, \text{Petrol}, \text{Yes})$

Only one feature differs:

$$d_H(x, y) = 1$$

Key idea

Hamming distance counts how many features disagree.



Similarity Measures

Definition:

A similarity measure quantifies how alike two patterns are.

- ❑ Higher similarity $\rightarrow$ more alike patterns.
- ❑ Lower similarity $\rightarrow$ less alike patterns.

Unlike distance, similarity is typically larger for more similar patterns.



Similarity Measures

Example

Consider:

$$A = (2, 3), B = (2.1, 3.1)$$

and

$$C = (8, 9)$$

We expect:

- A and $B \rightarrow$ high similarity
- A and $C \rightarrow$ low similarity

because A and B have very similar feature values.



Distance vs Similarity

	Distance	Similarity
Measures	Difference	Likeness
More similar	Smaller value	Larger value
Less similar	Larger value	Smaller value

Key idea:

- Distance tells us how different two patterns are; similarity tells us how alike they are.



Cosine Similarity

Cosine similarity measures the angle between two vectors. It focuses on the direction of the vectors rather than their magnitude.

Vectors pointing in the same direction are highly similar.

Formula:

$$\text{sim}_{\cos}(\mathbf{x}, \mathbf{y}) = \frac{\mathbf{x}^T \mathbf{y}}{\|\mathbf{x}\| \|\mathbf{y}\|}$$

where:

- $\mathbf{x}^T \mathbf{y} \rightarrow$ dot product
- $\|\mathbf{x}\|, \|\mathbf{y}\| \rightarrow$ vector magnitudes



Cosine Similarity

Example

Consider:

$$x = (1, 2)$$

$$y = (2, 4)$$

The two vectors point in exactly the same direction.

$$\text{sim}_{\cos}(x, y) = 1$$

Therefore, they have maximum cosine similarity.

Now consider:

$$z = (-1, 2)$$

x and z point in opposite directions:

$$\text{sim}_{\cos}(x, z) = -1$$



Cosine Similarity

Geometric interpretation

- ❑ Same direction $\rightarrow$ similarity close to 1
- ❑ Perpendicular vectors $\rightarrow$ similarity close to 0
- ❑ Opposite directions $\rightarrow$ similarity close to -1

Key idea:

- ❑ Cosine similarity compares the direction of vectors, not simply how far apart their endpoints are.



Binary Similarity

Core idea:

Some patterns are represented using binary features: 0 or 1.

Example: presence/absence of characteristics.

We can compare two patterns by counting how many binary features they have in common.



Binary Similarity

Suppose a pattern describes whether an object has certain properties:

$A = (1, 1, 0, 1, 0)$ and $B = (1, 0, 0, 1, 1)$

Compare feature by feature:

Feature	A	B	Match?
f_1	1	1	✓
f_2	1	0	✗
f_3	0	0	✓
f_4	1	1	✓
f_5	0	1	✗

They agree in 3 out of 5 features.



Binary Similarity

A simple matching similarity is therefore:

$$S(A, B) = \frac{\text{number of matches}}{\text{total number of features}}$$
$$S(A, B) = \frac{3}{5} = 0.6$$

Important distinction

For binary data, sometimes a shared 1 is more informative than a shared 0.

Example:

- ❑ 1,1 → both have the characteristic
- ❑ 0,0 → both lack the characteristic



Jaccard Similarity

Core idea:

Jaccard similarity is useful for binary or set-like data.

- It focuses on the features that are present in the patterns.
- Shared absences (0,0) are not counted as evidence of similarity.

□ Formula:

For binary patterns:

$$J(A, B) = \frac{M_{11}}{M_{11} + M_{10} + M_{01}}$$

where:

- M_{11} : both patterns have 1
- M_{10} : $A = 1, B = 0$
- M_{01} : $A = 0, B = 1$



Jaccard Similarity

Example

$$A = (1, 1, 0, 1, 0)$$

$$B = (1, 0, 0, 1, 1)$$

Count the combinations:

Feature	A	B	Type
f_1	1	1	M_{11}
f_2	1	0	M_{10}
f_3	0	0	Ignored
f_4	1	1	M_{11}
f_5	0	1	M_{01}

Therefore: $M_{11} = 2, M_{10} = 1, M_{01} = 1$ and $J(A, B) = \frac{2}{2+1+1} = \frac{2}{4} = 0.5$



Jaccard Similarity

Why ignore 0,0 ?

Suppose the features represent characteristics that an object possesses.

- If two objects both *do not* possess a characteristic, that absence may not provide meaningful evidence that the objects are similar.

Key idea:

- Jaccard similarity measures the overlap of shared positive features.



Pearson Correlation as Similarity

Core idea:

Pearson correlation measures how similarly two variables vary together.

It captures the shape and direction of variation, rather than absolute magnitude.

It is useful when two patterns may have different scales but similar trends.

Formula:

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

where $\bar{x}$ and $\bar{y}$ are the means of the two vectors.



Pearson Correlation as Similarity

Example

Consider:

$$x = (1, 2, 3, 4) \text{ and } y = (2, 4, 6, 8)$$

As x increases, y increases proportionally.

Therefore:

$$r_{xy} = 1 \rightarrow \text{perfect positive correlation}$$

Now consider:

$$z = (8, 6, 4, 2)$$

As x increases, z decreases:

$$r_{xz} = -1 \rightarrow \text{perfect negative correlation}$$



Pearson Correlation as Similarity

Interpretation

$r \approx 1 \rightarrow$ similar increasing pattern

$r \approx 0 \rightarrow$ little or no linear relationship

$r \approx -1 \rightarrow$ opposite variation

Important: Two vectors can have different numerical values but still have very high correlation if they follow the same pattern.

Key idea:

Correlation measures similarity in variation, not simply similarity in magnitude.



Comparing Similarity Measures

Different similarity measures capture different notions of "similarity."

Measure	Main idea	Typical data	High value means
Cosine Similarity	Similar direction	Vector data	Same direction
Jaccard Similarity	Shared positive features	Binary / set-like data	Large overlap
Binary Matching	Overall agreement	Binary data	Many matching features
Pearson Correlation	Similar variation / shape	Numerical vectors	Similar pattern of change



Comparing Similarity Measures

Example

Consider two vectors:

$$A = (1, 2, 3, 4)$$
$$B = (10, 20, 30, 40)$$

They have very different magnitudes, but:

- ❑ Cosine similarity $\rightarrow$ very high
- ❑ Pearson correlation $\rightarrow 1$
- ❑ Euclidean distance $\rightarrow$ relatively large

This illustrates an important point:

- ❑ Two patterns can be considered similar under one measure and different under another.



Choosing the measure

The appropriate measure depends on:

- ❑ type of data
- ❑ meaning of the features
- ❑ what "similar" means for the application
- ❑ scale and distribution of the features

Key idea:

- ❑ There is no universally best similarity measure. The measure should match the nature of the data and the recognition task.



Choosing the measure

The choice of distance or similarity measure depends on the type and characteristics of the data.

Data / Situation	Appropriate measure	Why?
Continuous numerical features	Euclidean	Geometric separation
Numerical features with different variability / correlation	Mahalanobis	Accounts for covariance
Grid-like / coordinate-based movement	Manhattan	Sum of coordinate differences
Binary vectors	Hamming	Counts mismatches
Binary presence/absence data	Jaccard	Measures shared positive features
Vector direction matters	Cosine	Compares direction
Pattern of variation matters	Pearson correlation	Compares shape/trend



Choosing the measure

Example

Suppose we represent documents using binary features indicating whether certain words are present.

$$A = (1, 1, 0, 1, 0)$$

$$B = (1, 0, 0, 1, 1)$$

A measure such as Jaccard similarity may be more meaningful than Euclidean distance because we care about the shared presence of features.

Key idea

- The distance or similarity measure is part of the pattern representation problem—not just a mathematical choice.



Why Does Feature Scaling Matter?

Distance-based methods are sensitive to the scale of the features.

Consider two features:

- Height: 150-200 cm
- Income: 20,000-100,000 €

If we calculate Euclidean distance directly, the income differences can dominate the distance because its numerical scale is much larger.

□ Example

Consider two patterns:

$$A = (170, 30,000)$$

$$B = (180, 40,000)$$

The differences are:

- $\Delta\text{Height} = 10$
- $\Delta\text{Income} = 10,000$



Why Does Feature Scaling Matter?

Example cntd

Without scaling, the second feature contributes much more to the numerical distance.

Therefore: A feature with a larger numerical scale can dominate the distance calculation.

Why is this a problem?

The distance may reflect mainly the units and scale of a feature rather than its actual importance.

This can affect: k-NN, clustering, distance-based retrieval, similarity calculations

- ❑ Before using a distance-based method, we should consider whether the features are on comparable scales.



Min-Max Normalization

Min-Max normalization transforms a feature to a predefined range, usually $[0, 1]$.

- It preserves the relative ordering of the observations.
- It prevents features with large numerical ranges from dominating distance calculations.

Formula:

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}}$$

where:

- x = original value
- $x_{\min}$ = minimum value of the feature
- $x_{\max}$ = maximum value of the feature
- x' = normalized value



Min-Max Normalization

Example

Suppose the Height feature ranges from:

$$150 \leq x \leq 200$$

For a person with:

$$x = 175$$

we obtain:

$$x' = \frac{175 - 150}{200 - 150} = \frac{25}{50} = 0.5$$

Therefore $175 \rightarrow 0.5$



Standardization

Standardization transforms a feature so that it has:

mean = 0

standard deviation = 1

It is also called Z-score normalization.

Unlike Min-Max normalization, it does not force values into a fixed range such as $[0, 1]$.

Formula:

$$z = \frac{x - \mu}{\sigma}$$

where:

x = original value

μ = mean of the feature

σ = standard deviation

z = standardized value



Standardization

Example

Suppose the height of a person is: $x=180$ and for the dataset:
 $\mu=170, \sigma=10$

Then:

$$z=(180-170)/10=10/10=1$$

Therefore: $180'' \text{ cm} \rightarrow z=1$

Interpretation:

$z=1$ means that the observation is one standard deviation above the mean. Similarly:

$z=0 \rightarrow$ exactly at the mean

$z>0 \rightarrow$ above the mean

$z<0 \rightarrow$ below the mean

Standardization puts features on a common scale by measuring how far each value is from the feature mean in standard-deviation units.



L1 and L2 Norms

A norm measures the magnitude, or length, of a vector.
For:

$$\mathbf{x} = (x_1, x_2, \dots, x_d)$$

The L1 norm is:

$$\|\mathbf{x}\|_1 = \sum_{i=1}^d |x_i|$$

The L2 norm is:

$$\|\mathbf{x}\|_2 = \sqrt{\sum_{i=1}^d x_i^2}$$



L1 and L2 Norms

Example

Consider:

$$\mathbf{x} = (3, 4)$$

L1 norm:

$$\|\mathbf{x}\|_1 = |3| + |4| = 7$$

L2 norm:

$$\|\mathbf{x}\|_2 = \sqrt{3^2 + 4^2} = 5$$

Geometric interpretation

For a 2D vector (3, 4):

L2 norm = 5 → straight-line length from the origin

L1 norm = 7 → total movement along the coordinate axes



L1 vs. L2 Distance

The same two patterns can have different distances depending on the norm used.

For two vectors $\mathbf{x}$ and $\mathbf{y}$:

$$d_{L1}(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|_1$$

$$d_{L2}(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|_2$$

Example

Consider:

$$\mathbf{x} = (2, 3), \mathbf{y} = (5, 7)$$

First calculate the difference:

$$\mathbf{x} - \mathbf{y} = (2 - 5, 3 - 7) = (-3, -4)$$

L1 distance:

$$d_{L1} = |-3| + |-4| = 3 + 4 = 7$$

L2 distance:

$$d_{L2} = \sqrt{(-3)^2 + (-4)^2} = \sqrt{9 + 16} = 5$$

Same two patterns $\rightarrow$ different numerical distance



Distance-Based Methods and Feature Scaling

Distance-based methods rely directly on distances or similarities between patterns.

Examples:

k-NN → uses distance to find nearest neighbors

Clustering → uses distances to determine which patterns are close

Similarity-based retrieval → finds patterns similar to a query pattern

Why scaling matters

Before applying a distance-based method:

Features should usually be placed on a comparable scale.

Otherwise, features with large numerical ranges may dominate the distance.



Questions ?



References

- ❑ Duda, Hart: Pattern Classification and Scene Analysis. J. Wiley & Sons, New York, 1982. (2nd edition 2000).
- ❑ Fukunaga: Introduction to Statistical Pattern Recognition. Academic Press, 1990.
- ❑ Bishop: Neural Networks for Pattern Recognition. Claredon Press, Oxford, 1997.
- ❑ Schlesinger, Hlaváč: Ten lectures on statistical and structural pattern recognition. Kluwer Academic Publisher, 2002.
- ❑ Satoshi Watanabe Pattern Recognition: Human and Mechanical, Wiley, 1985
- ❑ E. Gose, R. Johnsonbaught, S. Jost, Pattern recognition and image analysis, Prentice Hall, 1996.
- ❑ Sergios Thodoridis, Kostantinos Koutroumbas, Pattern recognition, Academiv Press, 1998.