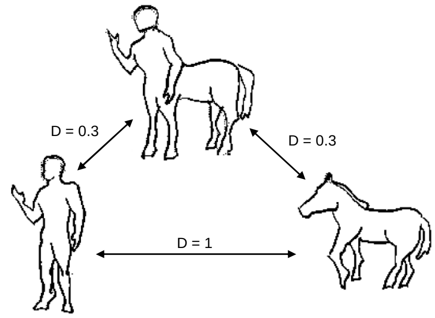
Pattern Recognition
as a
Human Centered non-Euclidean Problem

ICEIS, Funchal, Madeira, 9 June 2010

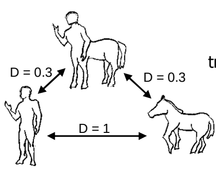
Robert P.W. Duin, Delft University of Technology
(In cooperation with Elżbieta Pękalska, Univ. of Manchester)

Pattern Recognition Lab
Delft University of Technology, The Netherlands
prlab.tudelft.nl

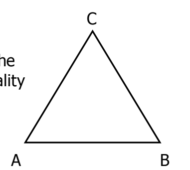


David W. Jacobs, Daphna Weinshall and Yoram Gdalyahu, Classification with Nonmetric Distances: Image Retrieval and Class Representation, *IEEE Trans. Pattern Anal. Mach. Intell*, 22(6), pp. 583-600, 2000.

Human Judgments …



violation of the
triangle inequality



D(A,B) < D(A,C) + D(C,B)

… often don't fit in an Euclidean world.

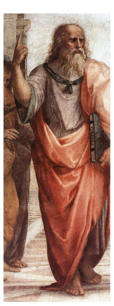
Learning about the world

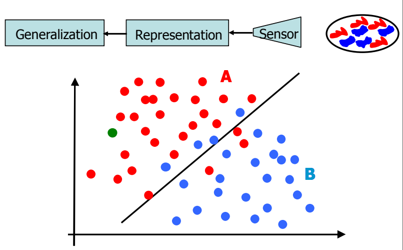
Human knowledge grows
in the debate between

-those who see the patterns, and

-those who know the universal laws

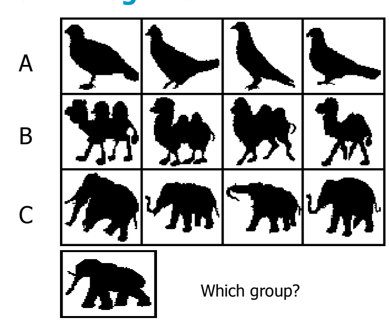
Automatic Pattern Recognition





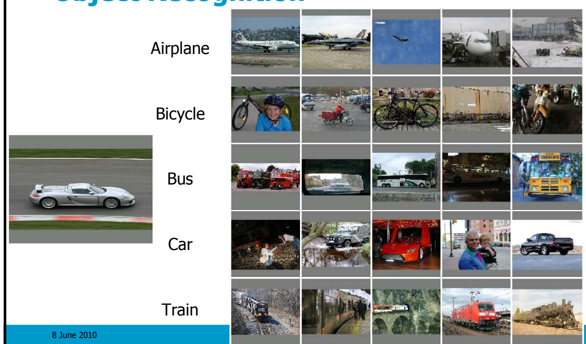
Can we replace our recognition of real world objects by a formal system, also when our mental system is non-Euclidean? How to train? How to apply?

Blob Recognition

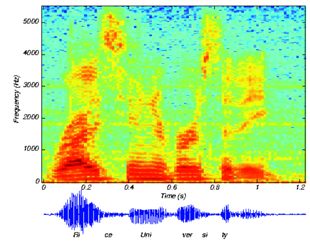


Which group?

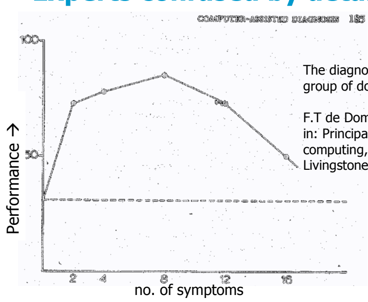
Object Recognition



Pattern Recognition: Speech



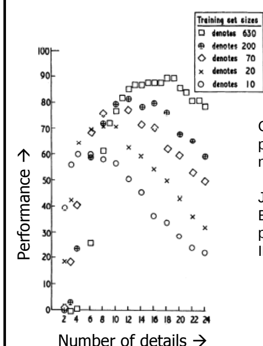
Experts confused by details



The diagnostic classification accuracy of a group of doctors. Sample size is 100.

F.T de Dombal, Computer-assisted diagnosis, in: Principals and practice of medical computing, Whitby and Lutz (eds.), Churchill Livingstone, London, 1971, pp 179 - 199.

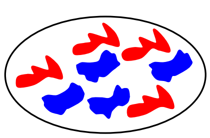
Computers confused by details



Character recognition classification performance as a function of the number of n-tuples used.

J.R. Ullmann,
Experiments with the n-tuple method of pattern recognition,
IEEE Trans. on Computers, 1969, 1135-1136

Pattern Recognition: Shapes

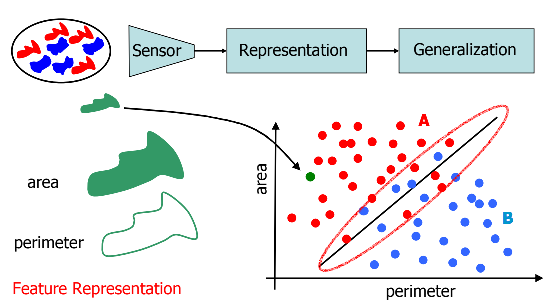


Examples of objects for different classes

Object of unknown class to be classified

A ? B

Pattern Recognition System



Measuring Human Relevant Information

A

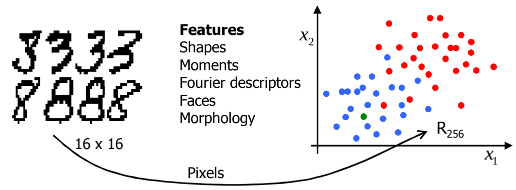
B

Nearest neighbours sorted:

B A A A B A A B A B

TU Delft

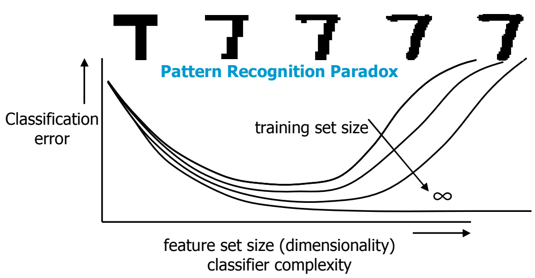
Pixel Representation



Pixels are more general, initially complete representation
Large datasets are available → good results for OCR

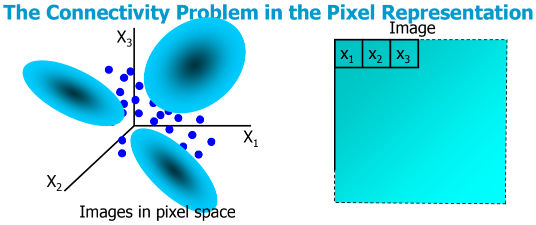
TU Delft

Peaking Phenomenon, Overtraining
Curse of Dimensionality, Rao's Paradox



TU Delft

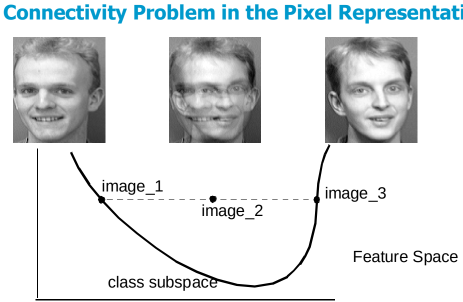
The Connectivity Problem in the Pixel Representation



Dependent (connected) measurements are represented independently.
The dependency has to be refound from the data.

TU Delft

The Connectivity Problem in the Pixel Representation



Interpolation does not yield valid objects

TU Delft

Representations

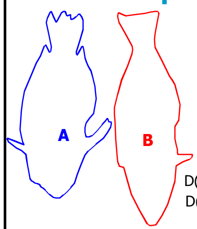
Features – details lost

Pixels – no connectivity

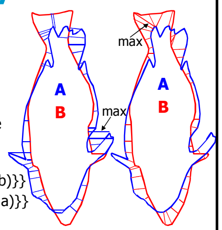
Dissimilarities – shape dependent

TU Delft

Examples Dissimilarity Measures



Dist(**A**,**B**):
a ∈ **A**, points of **A**
b ∈ **B**, points of **B**
d(a,b): Euclidean distance



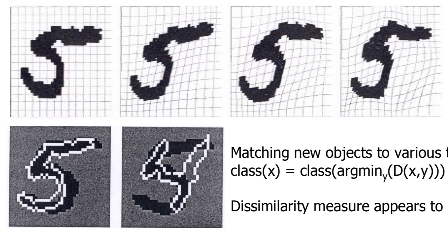
D(**A**,**B**) = max_a{min_b{d(a,b)}}
D(**B**,**A**) = max_b{min_a{d(b,a)}}

Hausdorff Distance (metric):
DH = max{max_a{min_b{d(a,b)}} , max_b{min_a{d(b,a)}}}

D(**A**,**B**) ≠ D(**B**,**A**)

Modified Hausdorff Distance (non-metric):
DM = max{mean_a{min_b{d(a,b)}},mean_b{min_a{d(b,a)}}}

Examples Dissimilarity Measures



Matching new objects to various templates:
class(x) = class(argmin$_y$(D(x,y)))

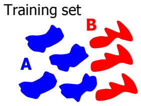
Dissimilarity measure appears to be non-metric.

A.K. Jain, D. Zongker, Representation and recognition of handwritten digit using deformable templates, IEEE-PAMI, vol. 19, no. 12, 1997, 1386-1391.

Dissimilarity Representation

Not used by NN Rule

Training set



Dissimilarities d_{ij} between all training objects

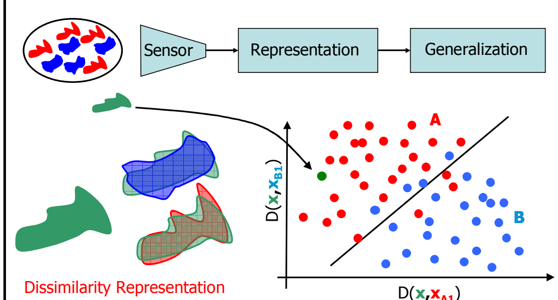
$$D_T = \begin{pmatrix} d_{11} & d_{12} & d_{13} & d_{14} & d_{15} & d_{16} & d_{17} \\ d_{21} & d_{22} & d_{23} & d_{24} & d_{25} & d_{26} & d_{27} \\ d_{31} & d_{32} & d_{33} & d_{34} & d_{35} & d_{36} & d_{37} \\ d_{41} & d_{42} & d_{43} & d_{44} & d_{45} & d_{46} & d_{47} \\ d_{51} & d_{52} & d_{53} & d_{54} & d_{55} & d_{56} & d_{57} \\ d_{61} & d_{62} & d_{63} & d_{64} & d_{65} & d_{66} & d_{67} \\ d_{71} & d_{72} & d_{73} & d_{74} & d_{75} & d_{76} & d_{77} \end{pmatrix}$$

Unlabeled object **x** to be classified

$$d_x = (d_{x1}\ d_{x2}\ d_{x3}\ d_{x4}\ d_{x5}\ d_{x6}\ d_{x7})$$

The traditional Nearest Neighbor rule (template matching) finds:
label(argmin$_{trainset}$\{d_{xi}\}) ,
without using D_T. Can we do any better?

Pattern Recognition System



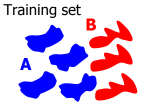
Dissimilarities – Possible Assumptions

Metric

1. Positivity: $d_{ij} \geq 0$
2. Reflexivity: $d_{ii} = 0$
3. Definiteness: $d_{ij} = 0$ objects i and j are identical
4. Symmetry: $d_{ij} = d_{ji}$
5. Triangle inequality: $d_{ij} < d_{ik} + d_{kj}$
6. Compactness: if the objects i and j are very similar then $d_{ij} < \delta$.
7. True representation: if $d_{ij} < \delta$ then the objects i and j are very similar.
8. Continuity of d.

Alternatives for the Nearest Neighbor Rule

Training set



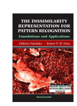
Dissimilarities d_{ij} between all training objects

$$D_T = \begin{pmatrix} d_{11} & d_{12} & d_{13} & d_{14} & d_{15} & d_{16} & d_{17} \\ d_{21} & d_{22} & d_{23} & d_{24} & d_{25} & d_{26} & d_{27} \\ d_{31} & d_{32} & d_{33} & d_{34} & d_{35} & d_{36} & d_{37} \\ d_{41} & d_{42} & d_{43} & d_{44} & d_{45} & d_{46} & d_{47} \\ d_{51} & d_{52} & d_{53} & d_{54} & d_{55} & d_{56} & d_{57} \\ d_{61} & d_{62} & d_{63} & d_{64} & d_{65} & d_{66} & d_{67} \\ d_{71} & d_{72} & d_{73} & d_{74} & d_{75} & d_{76} & d_{77} \end{pmatrix}$$

Unlabeled object **x** to be classified

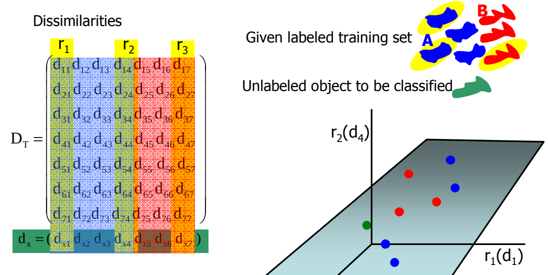
$$d_x = (d_{x1}\ d_{x2}\ d_{x3}\ d_{x4}\ d_{x5}\ d_{x6}\ d_{x7})$$

1. Dissimilarity Space
2. Embedding



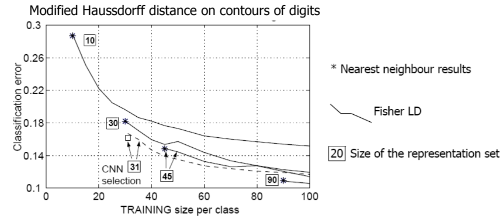
Pekalska, The dissimilarity representation for PR. World Scientific, 2005.

Alternative 1: Dissimilarity Space



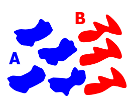
Selection of 3 objects for representation

Dissimilarity Space Classification ←→ Nearest Neighbor Rule



Dissimilarity based classification outperforms the nearest neighbor rule.

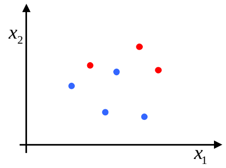
Embedding



Training set

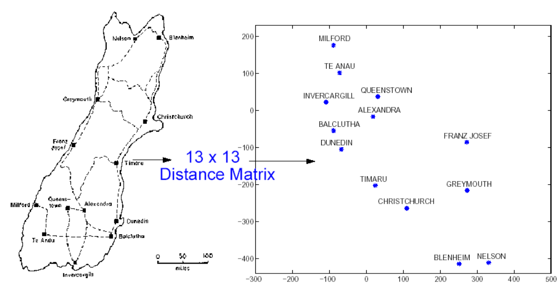
→ Dissimilarity matrix D → X

Is there a feature space for which Dist(X,X) = D ?

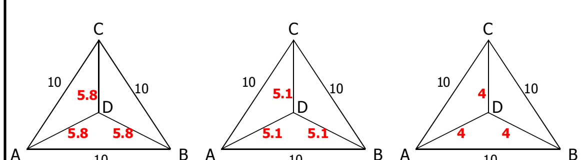


Position points in a vector space such that their Euclidean distances → D

Embedding

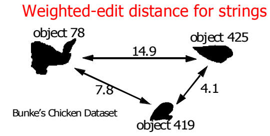


Euclidean - Non Euclidean - Non Metric

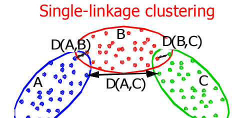


Non-metric distances

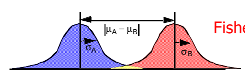
Weighted-edit distance for strings



Single-linkage clustering

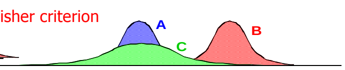


D(A,C) > D(A,B) + D(B,C)



$$J(A,B) = \frac{|\mu_A - \mu_B|^2}{\sigma_A^2 + \sigma_B^2}$$

Fisher criterion



$$J(A,C) = 0 \quad J(A,B) = \text{large}$$

$$J(C,B) = \text{small} \neq J(A,B)$$

(Pseudo-)Euclidean Embedding

$m \times m$ D is a given, imperfect dissimilarity matrix of training objects.

Construct inner-product matrix: $B = -\frac{1}{2} J D^{(2)} J \quad J = I - \frac{1}{m}\mathbf{11}$

Eigenvalue Decomposition , $B = Q \Lambda Q^T$

Select k eigenvectors: $X = Q_k \Lambda_k^{\frac{1}{2}}$ (problem: $\Lambda_k < 0$)

Let $\Im_k$ be a $k \times k$ diag. matrix, $\Im_k(i,i)$ = sign($\Lambda_k(i,i)$)

$\Lambda_k(i,i) < 0 \rightarrow$ Pseudo-Euclidean

$n \times m$ D_z is the dissimilarity matrix between new objects and the training set.

The inner-product matrix: $B_z = -\frac{1}{2}(D_z^{(2)} J - \frac{1}{n}\mathbf{11}^T D^{(2)} J)$

The embedded objects: $Z = B_z Q_k |\Lambda_k|^{-\frac{1}{2}} \Im_k$

TUDelft

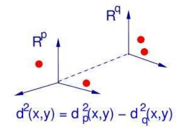
PES: Pseudo-Euclidean Space (Krein Space)

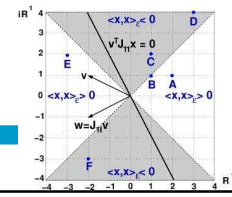
If D is non-Euclidean, B has p positive and q negative eigenvalues.

A pseudo-Euclidean space $\mathcal{E}$ with signature (p,q), $k = p+q$, is a non-degenerate inner product space $\Re_k = \Re_p \oplus \Re_q$ such that:

$$\langle x, y \rangle_\varepsilon = x^T \Im_{pq} y = \sum_{i=1}^{p} x_i y_i - \sum_{j=p+1}^{q} x_j y_j \qquad \Im_{pq} = \begin{bmatrix} I_{p\times p} & 0 \\ 0 & -I_{q\times q} \end{bmatrix}$$

$$d_\varepsilon^2(x,y) = \langle x-y, x-y \rangle_\varepsilon = d_p^2(x,y) - d_q^2(x,y)$$





Pseudo Euclidean Space

Euclidean embedding D → X

$$d_{ij}^2 = \|\mathbf{x}_i - \mathbf{x}_j\|^2$$

Pseudo Euclidean embedding D → {X^p,X^q}

$$d_{ij}^2 = \|\mathbf{x}_i^p - \mathbf{x}_j^p\|^2 - \|\mathbf{x}_i^q - \mathbf{x}_j^q\|^2$$

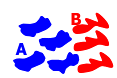
'Positive' and 'negative' space,

Compare Minkowsky space in relativity theory

TUDelft

Dissimilarity based classification procedured compared

Training set



→ Dissimilarity matrix D

Test object x

→ Dissimilarities d_x with training set

1. Nearest Neighbour Rule
2. Reduce training set to representation set

⇒ dissimilarity space

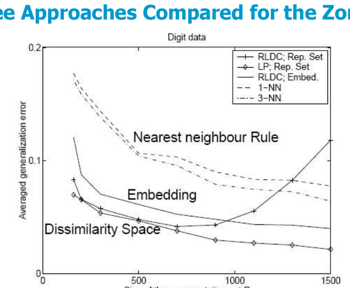
3. Embedding:Select large $\Lambda_{ii} > 0$ ⇒ Euclidean space

Select large $|\Lambda_{ii}| > 0$ → pseudo-Euclidean space

} discriminant function

TUDelft

Three Approaches Compared for the Zongker Data



Dissimilarity Space equivalent to Embedding better than Nearest Neighbour Rule

TUDelft

Non-Euclidean Representations

- Why do we have them?
- Are they essential?
- Can we build classifiers for them? (to some extend)
- Can we transform them into Euclidean representations? (Yes, but at the cost of performance loss)

SIMBAD **Beyond Features**
Similarity-based Pattern Analysis and Recognition

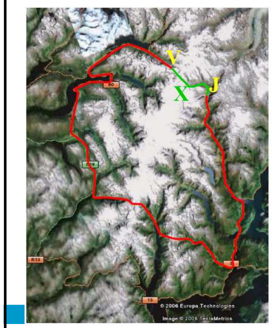
TUDelft

Computational Noise

Even for Euclidean distance matrices zero eigenvalues may show negative, e.g:

- X = N(50,20) : 50 points in 20 dimensions
- D = Dist(X): 50 x 50 distance matrix
- Expected: 49-20 = 29 zero eigenvalues
- Found: 15 negative eigenvalues

Lack of information

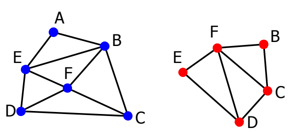


1800:
Crossing the Jostedalsbreen was impossible.
Travelling around (200 km) lasted 5 days.
Untill the shared point X was found.
People could visit each other in 8 hours.

D(V,J) = 5 days
D(V,X) = 4 hours
D(X,J) = 4 hours

Graph Matching → Dissimilarities

Representation by Connected Graphs



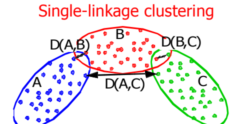
Graph (Nodes, Edges, Attributes)

Distance (Graph_1, Graph_2)

Intrinsicly Non-Euclidean Dissimilarity Measures
Single Linkage

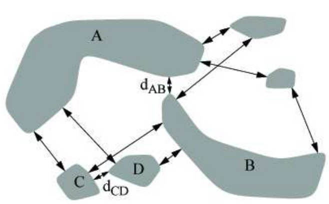
Distance(Table,Book) = 0
Distance(Table,Cup) = 0
Distance(Book,Cup) = 1

Single-linkage clustering



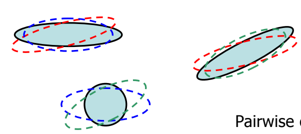
D(A,C) > D(A,B) + D(B,C)

Boundary distances

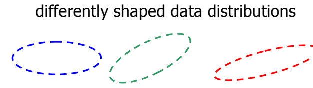


A set of boundary distances may characterize sets of datapoints:
Distances → features

Intrinsicly Non-Euclidean Dissimilarity Measures
Mahalanobis

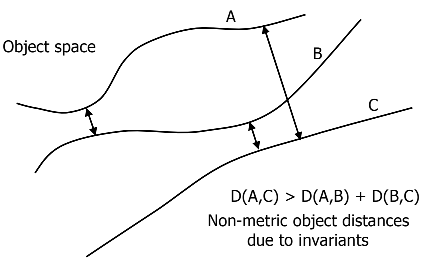


Pairwise comparison between
differently shaped data distributions

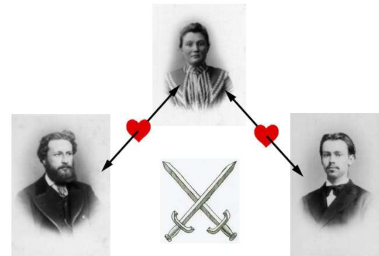


Different pairs → different comparison frameworks
→ non-Euclidean

Intrinsicly Non-Euclidean Dissimilarity Measures
Invariants



Intrinsicly Non-Euclidean Dissimilarity Measures



Non-Euclidean human relations

Objects may have an 'inner life'

In dissimilarity measures the 'inner life' of objects may be taken into account (e.g. invariants).

→ Objects cannot be represented anymore as points

→ Non-Euclidean dissimilarities

Conclusions

- Pseudo Euclidean Space (PES) is sometimes informative (corrections are not helpful).
- The corresponding problems may be intrinsic non-Euclidean
- Classifiers for non-Euclidean data have to be studied further