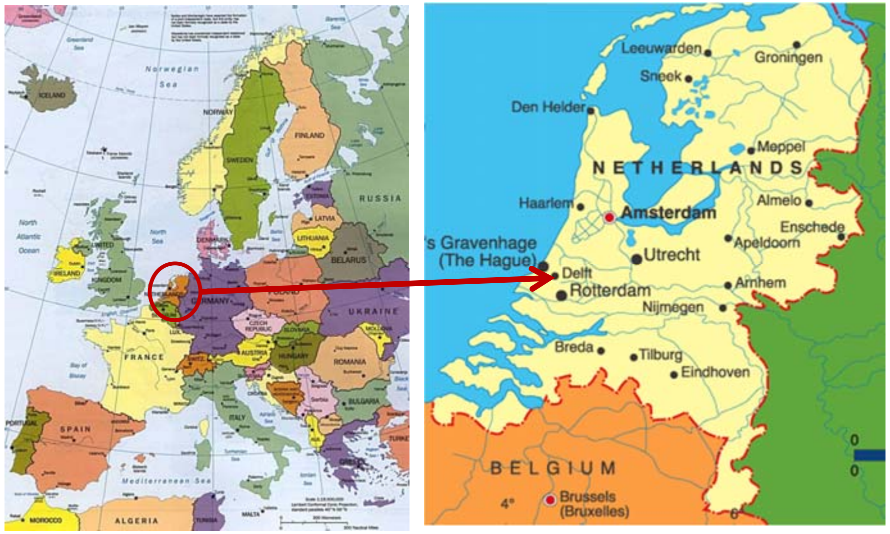



Library

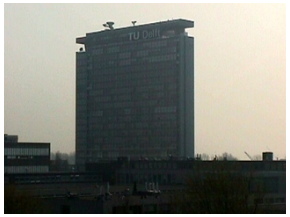

Auditorium


Electr Eng, Math
& Comp Science

Delft according
to Vermeer


Leeuwenhoek


His microscope






Non-Euclidean Problems in Pattern Recognition

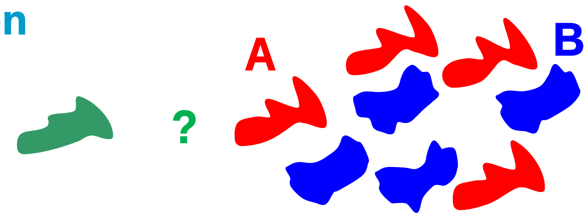
ACIT 2013, Khartoum, Sudan, 17-19 Dec 2013

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

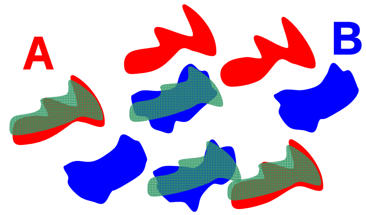
Pattern Recognition Lab
Delft University of Technology, The Netherlands
http://rduin.nl

Introduction

Question



How to represent real world objects,
(with a size and a shape)
given a set of examples
such that we can generalize?



Real world objects and events

Images
Spectra
Time signals
Gestures

→ shapes

How to build a representation?
Features ←→ Structure

Blob Recognition

BACK
BREAST
DRUMSTICK
THIGH-AND-BACK
WING

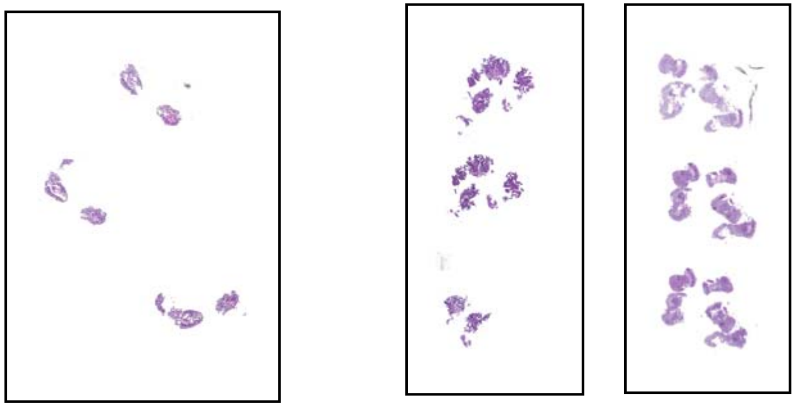
446 binary images, varying size, e.g.: 100 x 130

Andreu, G., Crespo, A., Valiente, J.M.: Selecting the toroidal self-organizing feature maps (TSOFM) best organized to object recogn. In: ICNN. (1997) 1341–1346.

Shape classification by weighted-edit distances (Bunke)

Bunke, H., Buhler, U.: Applications of approximate string matching to 2D shape recognition. Pattern recognition ***26*** *(1993) 1797–1812*

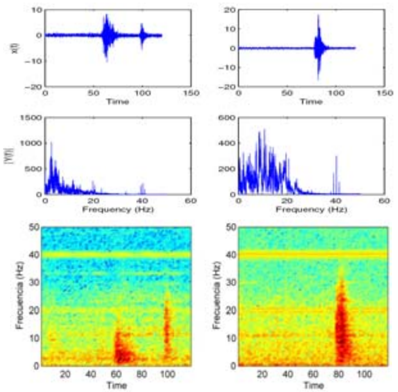
Colon Tissue Recognition



??? normal pathological

Volcano / Seismic Signal Classification

Volcano-Tectonic Long Period



150 000 events (1994 – 2008)
5 volcanos
40 stations
15 classes

J. Makario,
INGEOMINAS, Manizales, Colombia

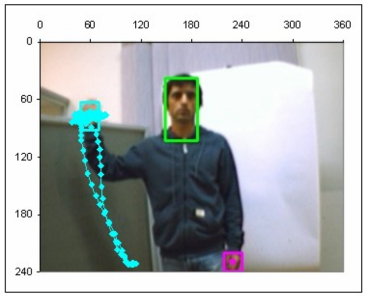
M. Orozco-Alzate,
Nat. Univ. Colombia, Manizales

R. Duin, TUDelft

M. Bicego, Univ. of Verona, Italy

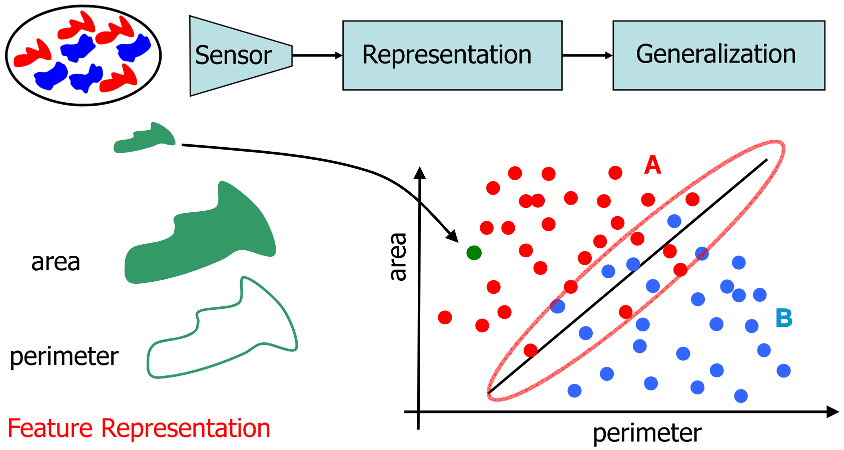
Cenatav, Havana, Cuba

Gesture Recognition



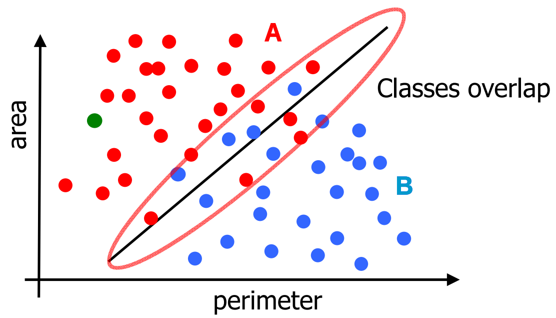
Is this gesture in the database?

Pattern Recognition System



Feature Representation

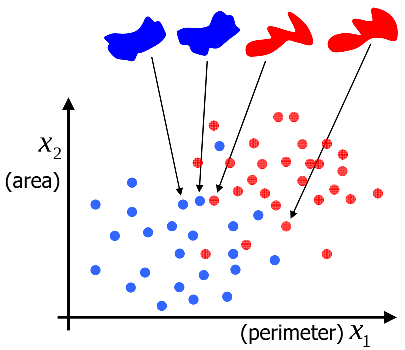
Objects → points in a Euclidean Space
Features reduce → classes overlap
→ to be solved by statistics



Representation

Compactness

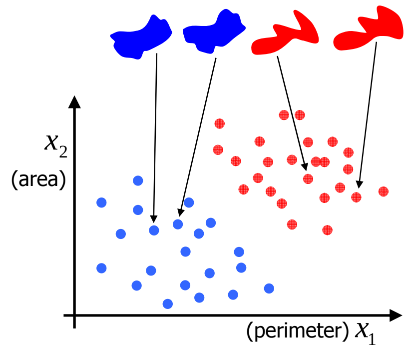
Representations of real world similar objects are close.
There is no ground for any generalization (induction) on representations that do not obey this demand.



The compactness hypothesis is not sufficient for perfect classification as dissimilar objects may be close.
→ class overlap
→ probabilities

A.G. Arkedev and E.M. Braverman, Computers and Pattern Recognition, 1966.

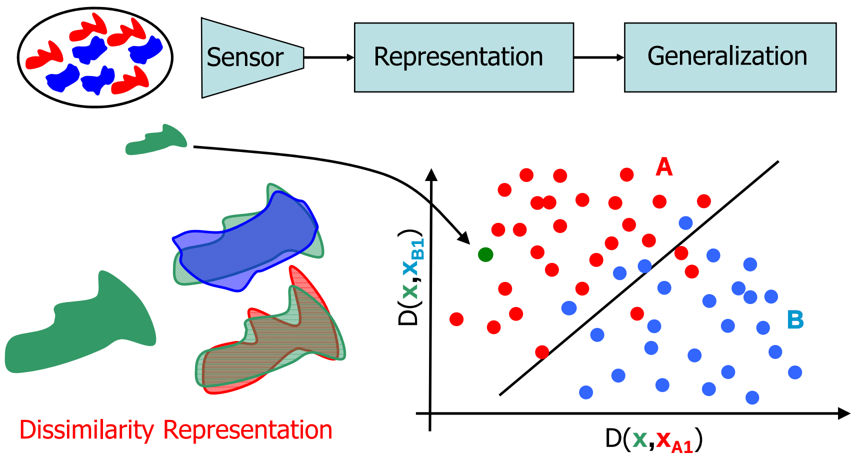
True Representations



Similar objects are close
and
Dissimilar objects are distant.

→ no probabilities needed, domains are sufficient!

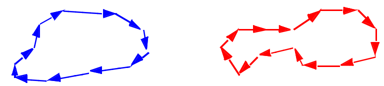
Dissimilarities → True Representation



TUDelft

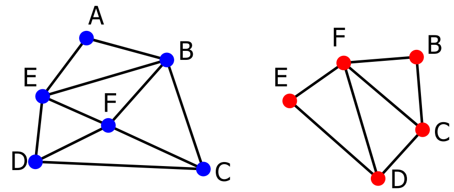
Structural Representation

Strings



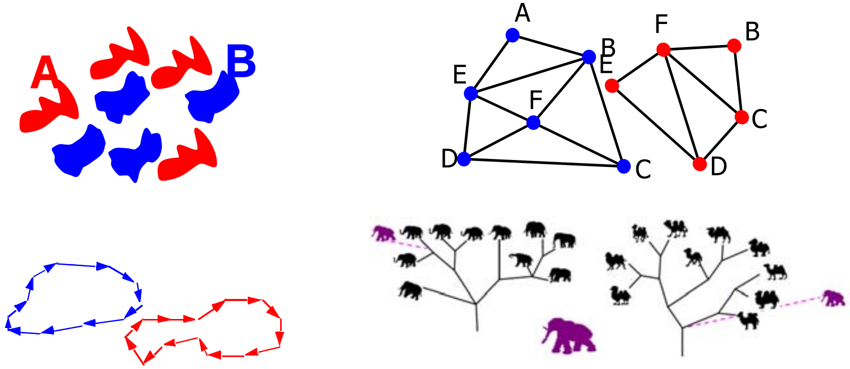
$X = (x_1, x_2, \ldots, x_k)$ $Y = (y_1, y_2, \ldots, y_n)$

Graphs



TUDelft

Structural Representation



How to generalize? Distances!

TUDelft

Dissimilarities

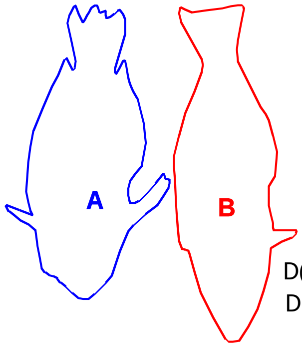
Dissimilarities

Objects → Shape distances

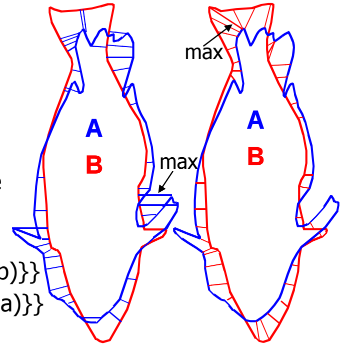
Objects → Features → Euclidean distances

Objects → Graphs → Graph distances

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)}}

D(**A**,**B**) ≠ D(**B**,**A**)

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

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

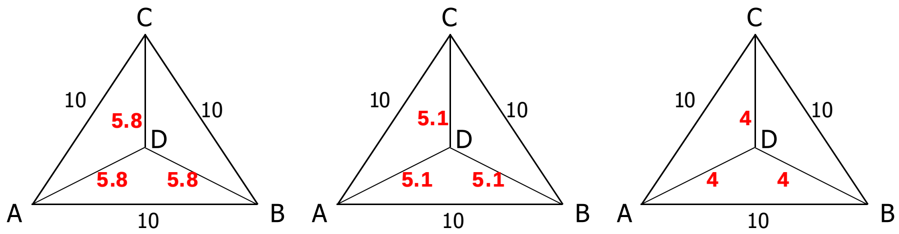
Dubuisoon & Jain, Modified Hausdorff distance for object matching, ICPR12, 2004,, voll 1, 566-568.

Dissimilarities – Possible Assumptions

Metric

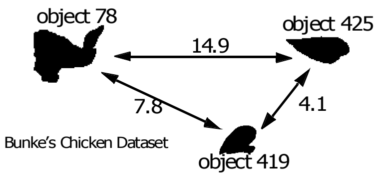
1. Positivity: $d_{ij} \geq 0$
2. Reflexivity: $d_{ii} = 0$
3. Definiteness: $d_{ij} = 0$ iff objects i and j are identical
4. Symmetry: $d_{ij} = d_{ji}$
5. Triangle inequality: $d_{ij} < d_{ik} + d_{kj}$
6. Compactness: if the objects i and j are very similar then $d_{ij} < \delta$.
7. True representation: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

Euclidean - Non Euclidean - Non Metric

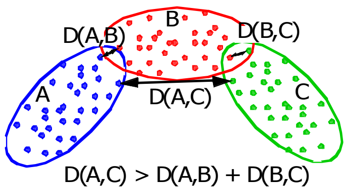


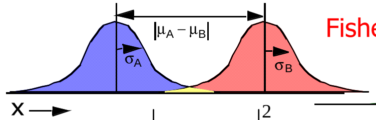
Non-metric distances

Weighted-edit distance for strings

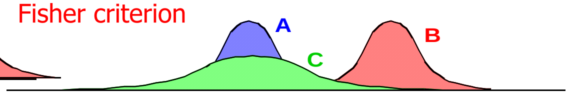


Single-linkage clustering





Fisher criterion



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

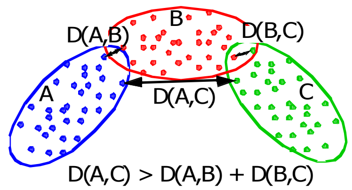
$$J(A,C) = 0 \qquad J(A,B) = \text{large}$$

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

Intrinsicly Non-Euclidean Dissimilarity Measures Single Linkage

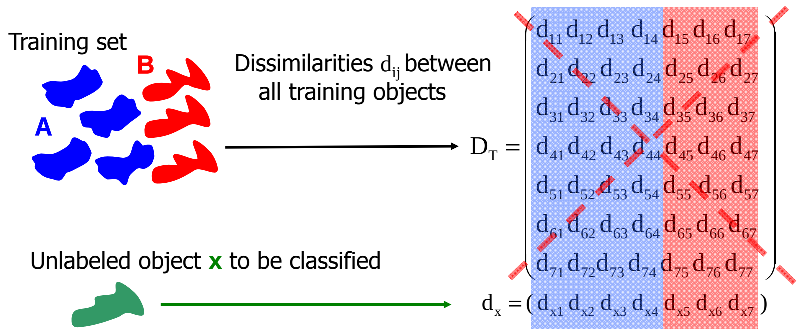
Distance(Table,Book) = 0
Distance(Table,Cup) = 0
Distance(Book,Cup) = 1

Single-linkage clustering



Dissimilarity Representation

Alternatives for the Nearest Neighbor Rule



1. Dissimilarity Space
2. Embedding

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

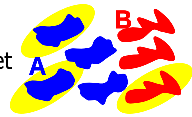
Alternative 1: Dissimilarity Space

Dissimilarities

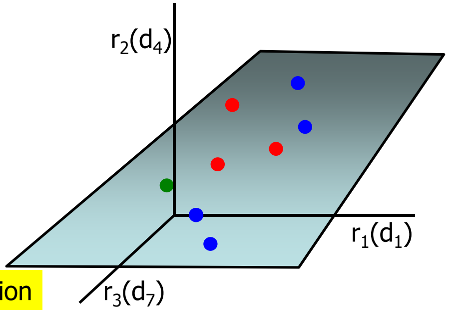
$$
\begin{array}{cc}
 & \begin{array}{ccccccc} r_1 & & & r_2 & & & r_3 \end{array} \\
D_T = & \left(\begin{array}{ccccccc}
d_{11} & d_{12} & d_{13} & d_{14} & d_{15} & d_{16} & d_{17} \\
d_{21} & d_{22} & d_{23} & d_{24} & d_{25} & d_{26} & d_{27} \\
d_{31} & d_{32} & d_{33} & d_{34} & d_{35} & d_{36} & d_{37} \\
d_{41} & d_{42} & d_{43} & d_{44} & d_{45} & d_{46} & d_{47} \\
d_{51} & d_{52} & d_{53} & d_{54} & d_{55} & d_{56} & d_{57} \\
d_{61} & d_{62} & d_{63} & d_{64} & d_{65} & d_{66} & d_{67} \\
d_{71} & d_{72} & d_{73} & d_{74} & d_{75} & d_{76} & d_{77}
\end{array} \right)
\end{array}
$$

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

Given labeled training set

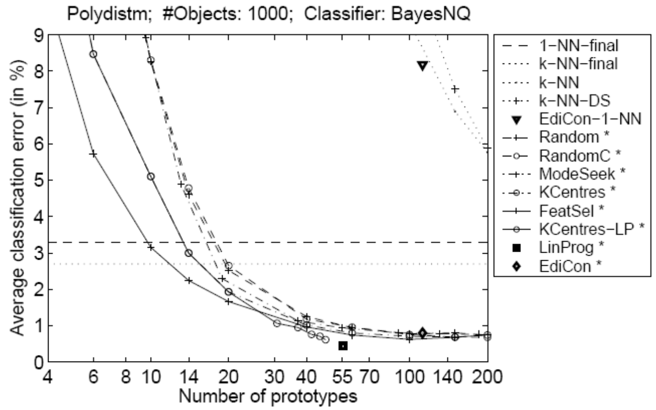


Unlabeled object to be classified



Selection of 3 objects for representation

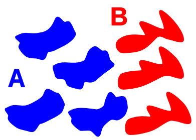
Prototype Selection: Polygon Dataset



The classification error as a function of the number of selected prototypes. For 10-20 prototypes results are already better than by using 1000 objects in the NN rules.

Pekalska et al., Prototype selection for dissimilarity-based classification, Pattern Recognition, 2006, 189-208.

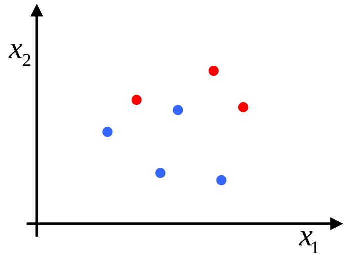
Alternative 2: Embedding



Training set

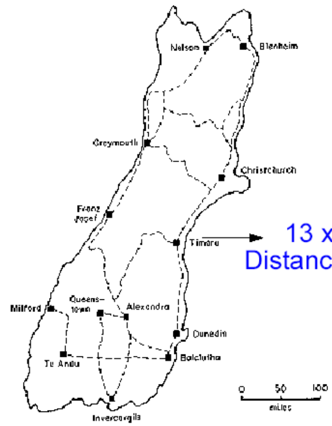
→ Dissimilarity matrix D → X

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

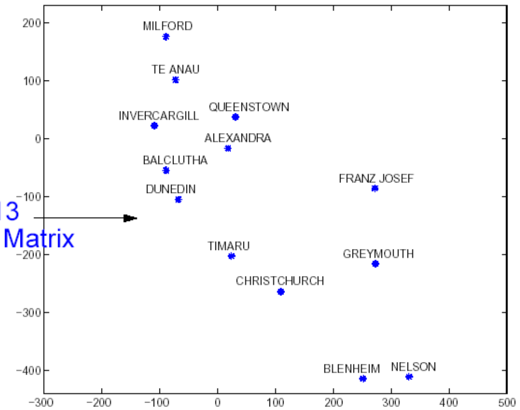


Position points in a vector space such that their Euclidean distances → D

Embedding



13 x 13
Distance Matrix



(Pseudo-)Euclidean Embedding

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

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

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

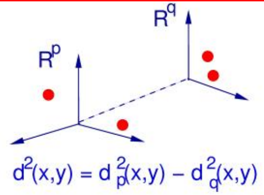
PES: Pseudo-Euclidean Space (Krein Space)

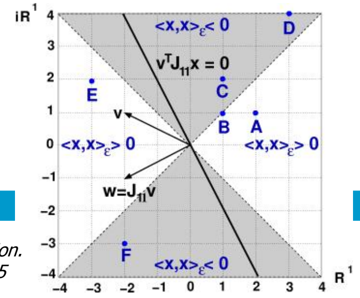
If D is non-Euclidean, B has p positive and q negative eigenvalues.

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:

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

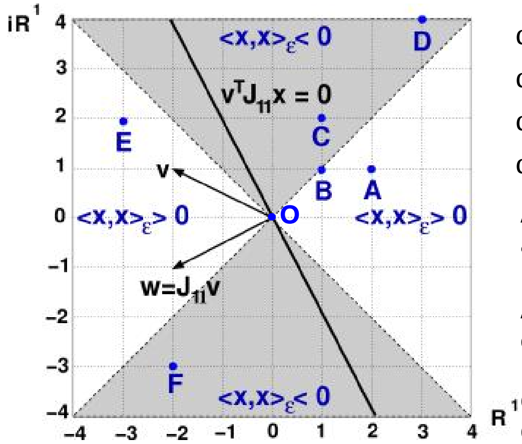
$$\boxed{d_\varepsilon^2(x,y) = \langle x-y, x-y \rangle_\varepsilon = d_p^2(x,y) - d_q^2(x,y)}$$





Pekalska, The Dissimilarity Representation for Pattern Recognition. Foundations and Applications. World Scientific, Singapore , 2005

Distances in PES



$d^2(O,A) > 0$
$d^2(O,E) > 0$
$d^2(O,B) = 0$
$d^2(O,D) < 0$

All points in the grey area are closer to O than O itself !?

Any point has a negative square distance to some points on the line $v^T Jx=0$.
Can it be used as a classifier?
Can we define a margin as in the SVM?

Pseudo Euclidean Space

Euclidean embedding D → X

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

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

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

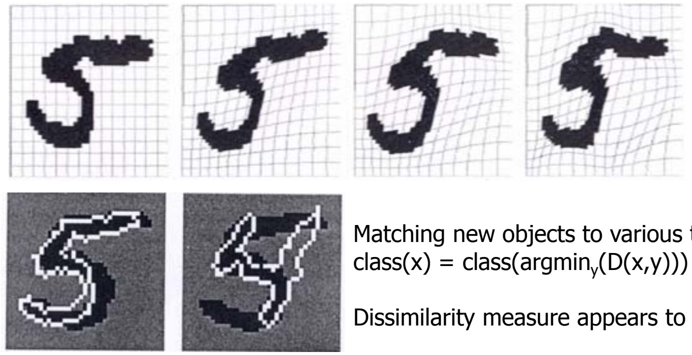
'Positive' and 'negative' space,
Compare Minkowsky space in relativity theory

PE-Space classifiers

- kNN, Parzen, Nearest Mean
  As object distances can be computed (are known)
- LDA, QDA
  As PE inner possibly product definitions cancel they can be computed, interpretation … ?
- SVM
  May get a result (indefinite kernel), possibly not optimal
- Others ??

TUDelft

Examples Dissimilarity Measures



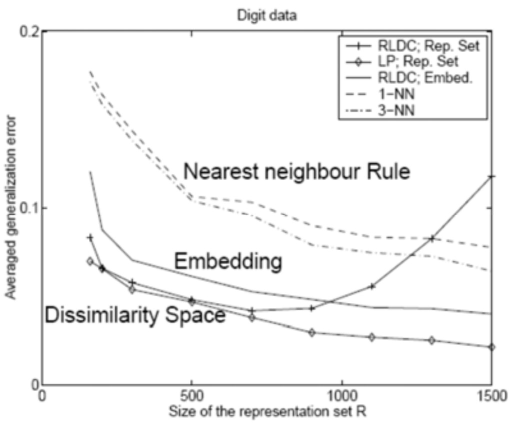
Matching new objects to various templates:
class(x) = class(argmin$_y$(D(x,y)))

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.

TUDelft

Three Approaches Compared for the Zongker Data



Dissimilarity Space equivalent to Embedding better than Nearest Neighbour Rule

Pekalska, The Dissimilarity Representation for Pattern Recognition. Foundations and Applications. World Scientific, Singapore , 2005

TUDelft

Representation Strategies

Avoiding the PE space

Dissimilarity Space: $X = D$

Correcting

Associated space	$X = \{[Xp, Xq], \varnothing\}$	$\tilde{d}_{ij}^2 = d_p^2(x_i, x_j) + d_q^2(x_i, x_j)$
Positive space	$X = X_p$	$\tilde{d}_{ij}^2 = d_p^2(x_i, x_j)$
Negative space	$X = X_q$	$\tilde{d}_{ij}^2 = d_q^2(x_i, x_j)$
Additive Correction	$\tilde{d}_{ij}^2 = d_{ij}^2 + c, i \neq j$	$X = \mathrm{Embedding}(\tilde{D})$

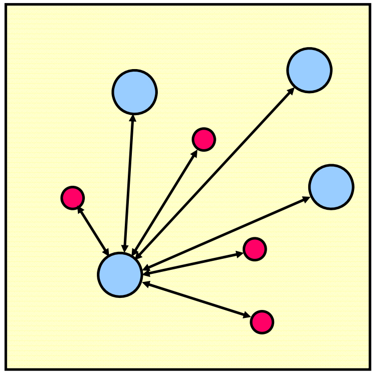
As it is

Pseudo Euclidean Space $X = \{Xp, Xq\}$ $d_{ij}^2 = d_p^2(x_i, x_j) - d_q^2(x_i, x_j)$

Classifiers to be developed further

TUDelft

Ball Distances



- Generate sets of balls (classes) uniformly, in a (hyper)cube; not intersecting.
- Balls of the same class have the same size.
- Compute all distances between the ball surfaces.

-> Dissimilarity matrix D

Duin et al., Non-Euclidean dissimilarities: Causes and informativeness, SSSPR 2010, 324-333.

TU Delft

Balls3D

Classifier	PE Sp	Ass Sp	Pos Sp	Neg Sp	Cor Sp
1-NN	47.4 (2.0)	47.4 (2.0)	47.4 (2.0)	44.2 (1.5)	47.4 (2.0)
Parzen	45.7 (1.7)	45.5 (1.6)	45.6 (1.7)	35.5 (1.7)	45.7 (1.7)
NM	47.5 (2.0)	47.7 (2.0)	47.6 (1.9)	49.6 (0.2)	48.1 (1.8)
SVM-1	50.7 (2.2)	50.0 (2.7)	50.0 (2.5)	62.1 (1.7)	50.1 (2.0)

Classifier	PE Dis Sp	Ass Dis Sp	Pos Dis Sp	Neg Dis Sp	Cor Dis Sp
1-NN	49.8 (2.2)	49.8 (2.2)	49.8 (2.2)	5.1 (0.8)	49.7 (2.2)
Parzen	47.9 (2.2)	47.9 (2.2)	47.9 (2.2)	4.6 (0.5)	47.9 (2.2)
NM	49.8 (2.2)	49.8 (2.2)	49.8 (2.2)	5.0 (0.8)	49.9 (2.2)
SVM-1	50.2 (1.6)	50.8 (1.7)	50.7 (1.7)	1.9 (0.5)	49.8 (1.5)

10 x (2-fold crossvalidation of 50 objects per class)

TU Delft

Is the PE Space Informative?

	size	classes	Non-Metric	NEF	Rand Err	+- Original, D	+ Positive, D_p	- Negative, D_q
Chickenpieces45	446	5	0	0.156	0.791	**0.022**	0.132	0.175
Chickenpieces60	446	5	0	0.162	0.79[illegible]	**0.020**	0.067	0.173
Chickenpieces90	446	5	0	0.152	0.79[illegible]	**0.022**	0.052	0.148
Chickenpieces120	446	5	0	0.130	0.791	**0.034**	0.108	0.148
FlowCyto	612	3	1e-5	0.244	0.598	0.105	0.100	0.327
WoodyPlants50	791	14	5e-4	0.229	0.928	**0.075**	0.076	0.442
CatCortex	65	4	2e-3	0.208	0.738	**0.046**	0.077	0.662
Protein	213	4	0	0.001	0.718	0.0[illegible]	[illegible]	[illegible]
Balls3D	200	2	3e-4	0.001	0.500	**0.470**	0.495	0.000
GaussM1	500	2	0	0.262	0.500	**0.202**	0.202	0.228
GaussM02	500	2	5e-4	0.393	0.500	0.204	**0.174**	0.252
CoilYork	288	4	8e-8	0.258	0.750	**0.267**	0.313	0.618
CoilDelftSame	288	4	0	0.027	0.750	**0.413**	0.417	0.597
CoilDelftDiff	288	4	8e-8	0.128	0.750	0.3[illegible]	[illegible]	[illegible]
NewsGroups	600	4	4e-5	0.202	0.733	0.198	[illegible]	0.435
BrainMRI	124	2	5e-5	0.112	0.499	0.226	**0.218**	0.556
Pedestrians	689	3	4e-8	0.111	0.348	**0.010**	0.015	0.030

Informative

Extremely Informative

Not Informative

Examples

Example: Chickenpieces (H. Bunke, Bern)

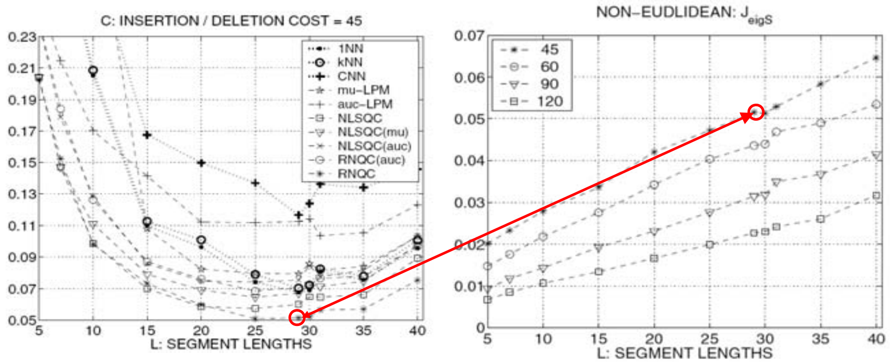
446 binary images, varying size, e.g.: 100 x 130

Andreu, G., Crespo, A., Valiente, J.M.: Selecting the toroidal self-organizing feature maps (TSOFM) best organized to object recogn. In: ICNN. (1997) 1341–1346.

Shape classification by weighted-edit distances (Bunke)

*Bunke, H., Buhler, U.: Applications of approximate string matching to 2D shape recognition. Pattern recognition **26** (1993) 1797–1812*

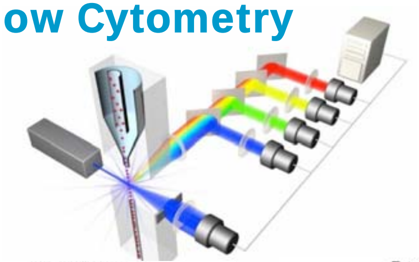
Chickenpieces: Various Dissimilarity Measures

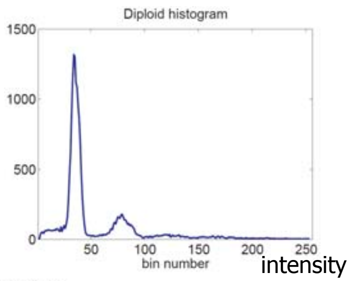


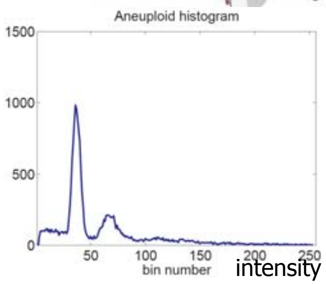
<u>Best classification result is for a very non-Euclidean dissimilarity measure !</u>

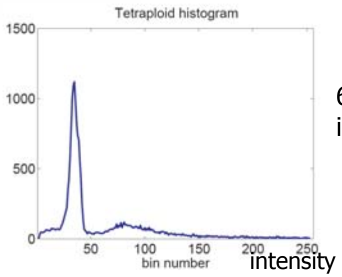
Pekalska et al., On not making dissimilarities Euclidean, SSSPR 2004, 1145-1154.

Flow Cytometry









612 histograms in 3 classes

Nap & van Rodijnen, Atrium Hospital, Heerlen

Flow Cytometry: classification errors

Pairwise, horizontal (intensity calibration):
D(hist1,hist2) = $min_\alpha L_1$(hist1,hist2(α))

⟵ Dissimilarity space ⟶

Data Source	NEF	1-NN	1-NND	SVM-1
Tube 1	0.27	0.38	0.38	0.30
Tube 2	0.27	0.37	0.37	0.29
Tube 3	0.27	0.38	0.40	0.27
Tube 4	0.27	0.42	0.42	0.30
Averaged	0.24	0.27	0.20	0.11

Bio-crystallization

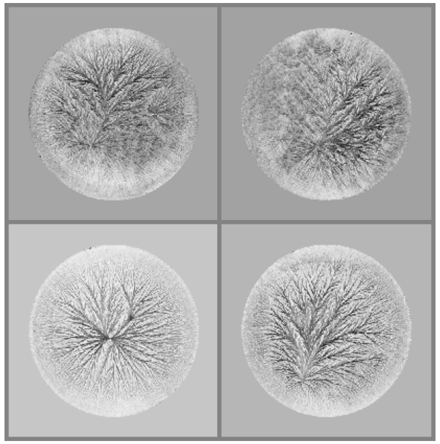
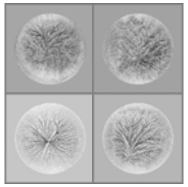


image size: 2114 x 2114
Different food products / quality
2 classes, 54 examples/class

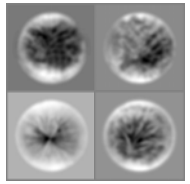
Busscher et al., Standardization of the iocrystallization Method for Carrot Samples, Biological Agriculture and Horticulture, 2010, Vol. 27, pp. 1–23

TUDelft

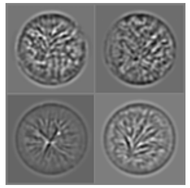
Bio-crystallization: Dissimilarity Measures



Originals



Gauss → L2



Laplace → L2

Laplace → Abs → Histogram → L1

TUDelft

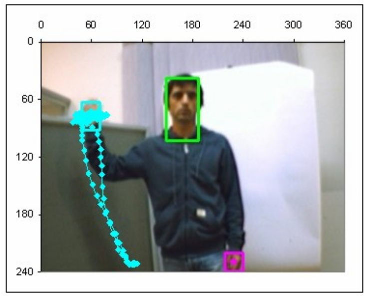
Bio-crystallization: classification errors

Dissimilarity space

Dissimilarity Measure	NEF	1-NN	1-NND	SVM-1
Gauss	0	0.329	0.266	0.106
Laplace	0	0.229	0.313	0.125
Laplace Histogram	0.067	0.107	0.172	0.072
Averaged Dissimilarities	0.004	0.114	0.166	0.057

TUDelft

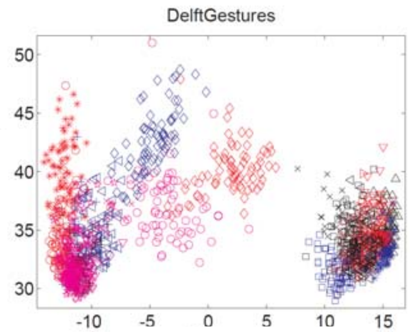
Gesture Recognition



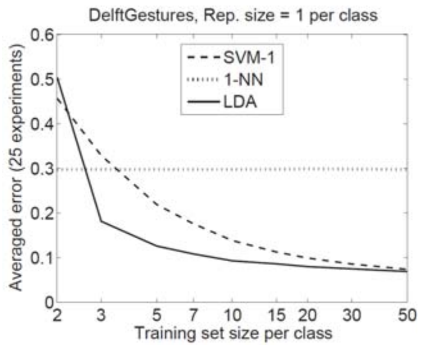
Is this gesture in the database?

TUDelft

Gesture Recognition



Dissimilarity Space PCA

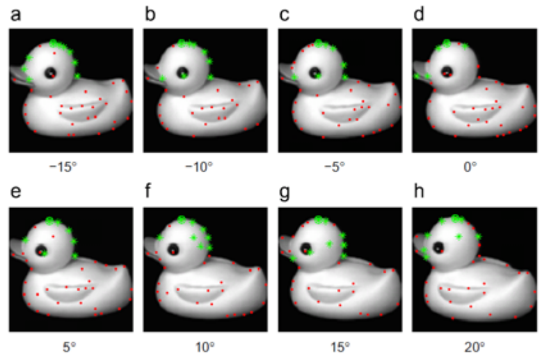


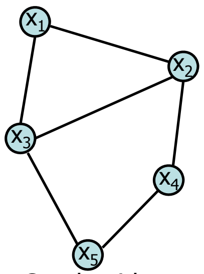
20 signs (classes), 75 examples/sign
Distance measure: DTW

Lichtenauer et al. Sign language recognition by combining statistical DTW and independent classification. PAMI 2008, 2040–2046.

TUDelft

Application: Graphs



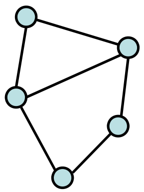


Graph with feature nodes

Taken from: Ren, Aleksic, Wilson, Hancock,
A polynomial characterization of hypergraphs using the Ihara zeta function, Pattern Recognition, 2011, 1941-1957

TUDelft

Interpolating structural and feature space dissimilarities



Structure only
(no features)

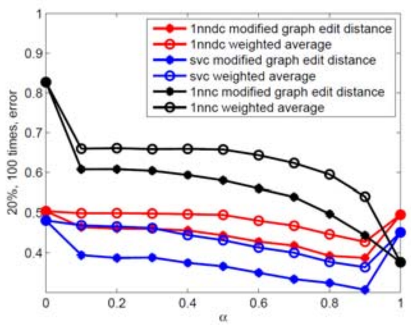


Fig. 5. Results of coil-segment for modified graph edit distance.

$\{x_1\ x_2\ x_3\ x_4\ x_5\}$

Features only
(no structure)

W.R.Lee et al., Bridging Structure and Feature Representations in Graph Matching, IJPRAI,2011, accepted

TUDelft

Conclusion

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)

Beyond Features
Similarity-based Pattern Analysis and Recognition

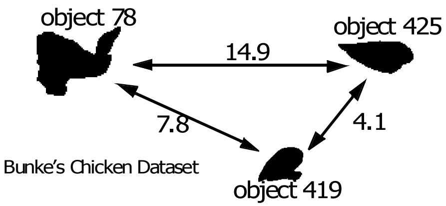
Computational Noise

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

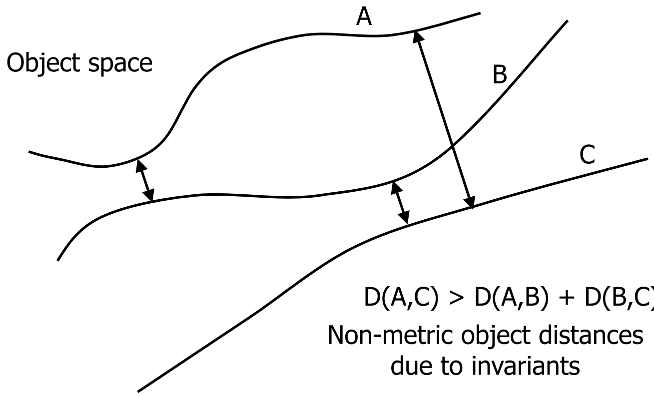
Computational Problems

Large distances are overestimated due to computational problems



Weighted edit distance for strings

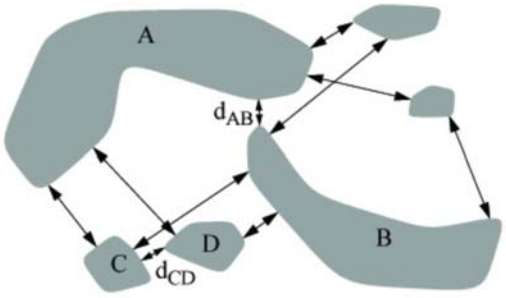
Intrinsicly Non-Euclidean Dissimilarity Measures Invariants



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

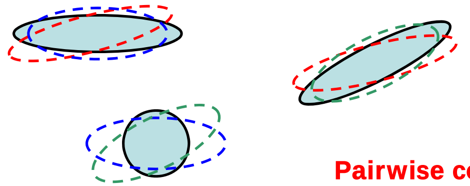
Non-metric object distances due to invariants

Boundary distances

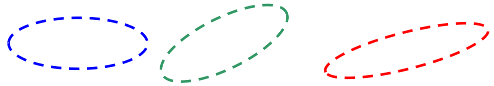


A set of boundary distances may characterize sets of datapoints:
Distances → features

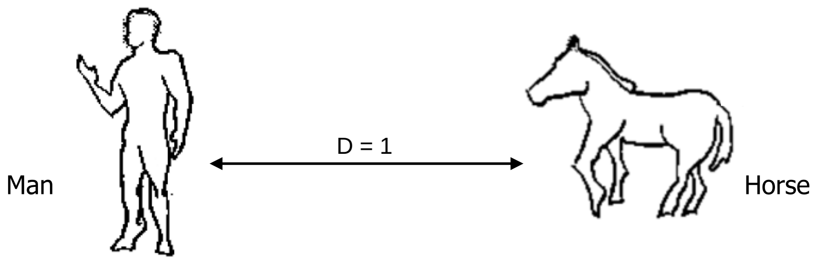
Intrinsicly Non-Euclidean Dissimilarity Measures Mahalanobis



Pairwise comparison between differently shaped data distributions



Different pairs → different comparison frameworks
→ non-Euclidean



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.

Conclusions

- Real world objects are not points
- Objects have a size
- Relations are non-Euclidean
- Non-Euclidean generalization procedures are needed