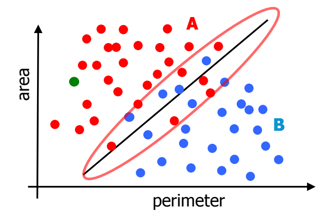
Non-Euclidean Representations

Embedding
Pseudo Euclidean Space
Classifiers in PE Space
Euclidean Corrections
Causes
Informativeness

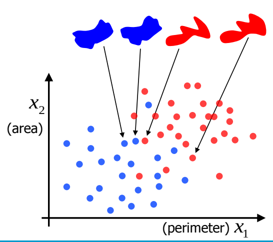
Feature Representation

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



Compactness

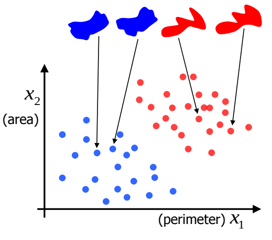
Representations of real world similar objects are close.
There is no ground for any generalization (induction) on representations that do not obey this demand.



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.

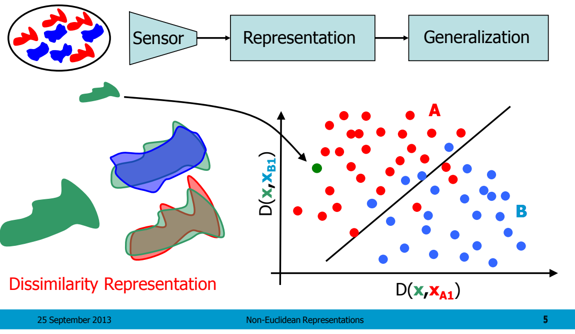
True Representations



Similar objects are close
and
Dissimilar objects are distant.

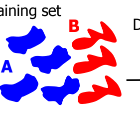
→ no probabilities needed, domains are sufficient!

Dissimilarities → True Representation



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

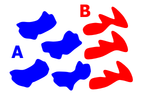
$$d_x = (d_{x1}\ d_{x2}\ d_{x3}\ d_{x4}\ d_{x5}\ d_{x6}\ d_{x7})$$

1. Dissimilarity Space
2. Embedding

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

Embedding of (non-Euclidean) Dissimilarities

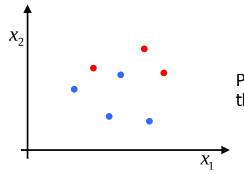
Alternative 2: Embedding



Training set

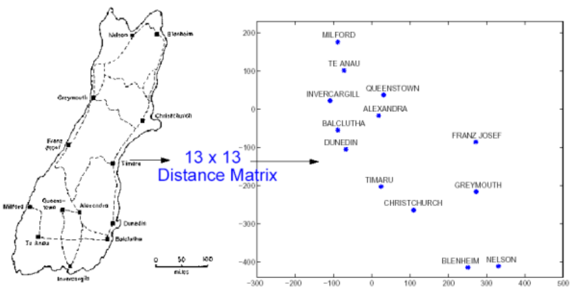
→ Dissimilarity matrix D → X

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

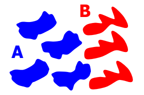


Position points in a vector space such that their Euclidean distances → D

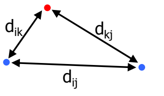
Embedding



Embedding of non-metric measurements

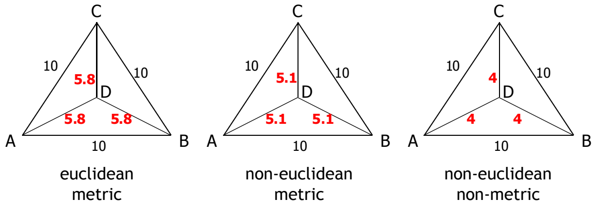


→ Dissimilarity matrix D → X



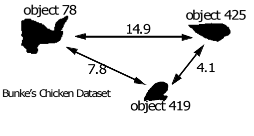
If the dissimilarity matrix cannot be explained from a vector space,
(e.g. for Hausdorff and Hamming distance of images)
or if $d_{ij} > d_{ik} + d_{kj}$ (triangle inequality not satisfied)
embedding in Euclidean space not possible
→ Pseudo-Euclidean embedding

Euclidean - Non Euclidean - Non Metric

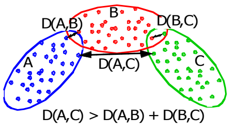


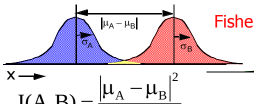
Non-metric distances

Weighted-edit distance for strings



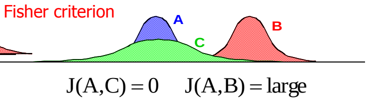
Single-linkage clustering





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

Fisher criterion

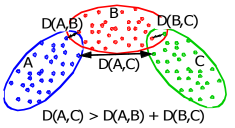


$$J(A,C) = 0 \quad J(A,B) = \text{large}$$
$$J(C,B) = \text{small} \neq J(A,B)$$

Intrinsicly Non-Euclidean Dissimilarity Measures
Single Linkage

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

Single-linkage clustering



Pseudo-Euclidean Space

(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 $\mathfrak{I}_k$ be a k x k diag. matrix, $\mathfrak{I}_k(i,i)$ = 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: $Z = B_z Q_k |\Lambda_k|^{-\frac{1}{2}} \mathfrak{I}_k$

(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 $\mathfrak{I}_k$ be a k x k diag. matrix, $\mathfrak{I}_k(i,i)$ = 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: $Z = B_z Q_k |\Lambda_k|^{-\frac{1}{2}} \mathfrak{I}_k$

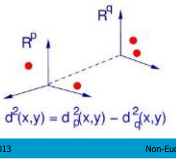
PES: Pseudo-Euclidean Space (Krein Space)

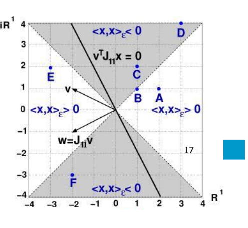
If D is non-Euclidean, B has p positive and q negative eigenvalues.

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:

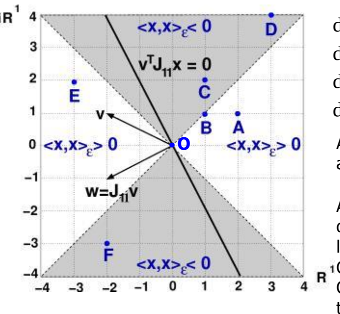
$$\langle x, y \rangle_\varepsilon = x^T \mathfrak{I}_{pq} y = \sum_{i=1}^{p} x_i y_i - \sum_{j=p+1}^{q} x_j y_j \qquad \mathfrak{I}_{pq} = \begin{bmatrix} I_{p \times p} & 0 \\ 0 & -I_{q \times q} \end{bmatrix}$$

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





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

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

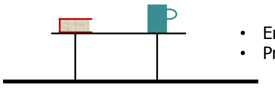
Pseudo Euclidean embedding D → {Xp,Xq}

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

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

Embedding Example

	Table	Book	Cup
Table	0	0	0
Book	0	0	1
Cup	0	1	0
Floor	**0**	**1**	**1**

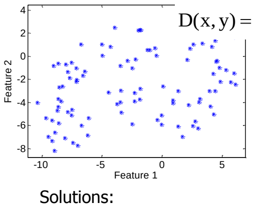


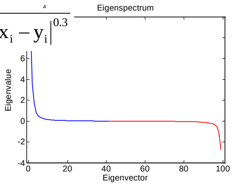
- Embed table, book, cup in a 2D PE space
- Project the floor

Pseudo-Euclidean Embedding

If D is non-Euclidean then B has **p** positive and **q** negative eigenvalues

$$D(x, y) = \sqrt[0.3]{\sum |x_i - y_i|^{0.3}}$$





Solutions:

- Remove all eigenvectors with small and negative eigenvalues
- or, take absolute values of eigenvalues and proceed
- or, construct a pseudo-Euclidean space

Negative Euclidean Fraction

$$\mathrm{NEF} = \frac{\sum_{\lambda_i < 0} |\lambda_i|}{\sum_{\forall \lambda_i} |\lambda_i|}$$

$$0 \le \mathrm{NEF} \le 1$$

PE Space ←→ Kernels

$$K(x, y) = -\tfrac{1}{2} J D(x, y)^{(2)} J \qquad J = I - \tfrac{1}{m}\mathbf{11}$$

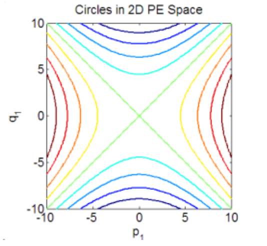
may be considered as a kernel. If

$$K(x, y) = < L(x), L(y) >$$

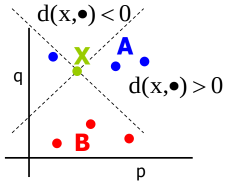
- The **kernel trick** may be used: operations defined on inner products in kernel space can be operated directly on K(x,y) **without embedding**!
- True for **Mercer kernels** (all eigenvalues ≥ 0).
- Difficult for **indefinite kernels**.
- Studying classifiers in **PE space** is studying the indefinite **kernel space**.
- Dissimilarities are more informative than kernels (due to normalization).

Classifiers in Pseudo-Euclidean Space

Distance based classifiers in PE Space



Metric in PE Space.
Equidistant points to the origin.



X assigned to B

Nearest Neighbour and
Nearest Mean can be properly defined.
SVM ? What is the distance to a line?

TUDelft

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

SVM in PE Space

- SVM on indefinite kernels may not converge as Mercer's conditions are not fulfilled.
- However, if it converges the solution is proper:

$$|w^T \Im w|$$

is minimized.

- *See also: B. Haasdonk, Feature Space Interpretation of SVMs with Indefinite Kernels, IEEE PAMI, 24, 482-492, 2005.*

TUDelft

Densities in PE Space

- Densities can be defined in a vector space on de basis of volumes, without the need of a metric.
- Density estimates however, often need a metric.
  E.g. the Parzen estimator:

$$\hat{f}(x) = \tfrac{1}{n} \sum_{y_i} c \exp\left(-\frac{d(x,y_i)^2}{2h^2}\right)$$

needs a distance definition $d(x,y)$.
- There is no problem, however, in case for all objects $d(x,y) > 0$.
- How can Gaussian densities be defined?
- Note that QDA in PES is identical to the QDA in AES as the signature cancels. The relation with a Gaussian distribution, however, is lost.

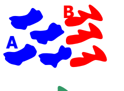
TUDelft

Dissimilarity based classifiers compared

TUDelft

Dissimilarity based classification procedured compared

Training set



→ Dissimilarity matrix D

Test object x

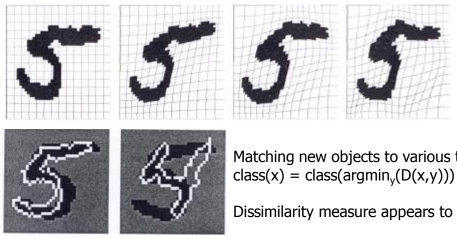
→ Dissimilarities d_x with training set

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

Examples Dissimilarity Measures



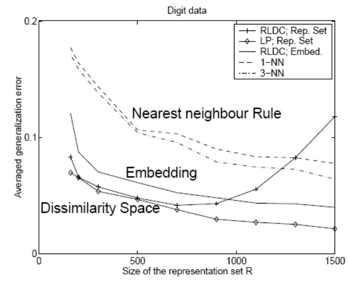
Matching new objects to various templates:
class(x) = class(argmin_y(D(x,y)))

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.

TU Delft

Three Approaches Compared for the Zongker Data



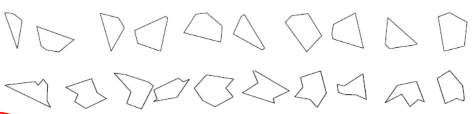
Dissimilarity Space equivalent to Embedding better than Nearest Neighbour Rule

TU Delft

Polygon Data

Convex Pentagons

Heptagons

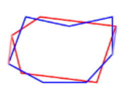


no class overlap zero error

Minimum edge length: 0.1 of maximum edge length

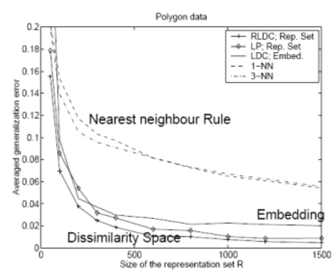
Distance measures: Hausdorff D = max { $\max_i(\min_j(d_{ij}))$, $\max_j(\min_i(d_{ij}))$ }.
Modified Hausdorff D = max { $\mathrm{mean}_i(\min_j(d_{ij}))$, $\mathrm{mean}_j(\min_i(d_{ij}))$ }. (no metric!)
d_{ij} = distance between vertex i of polygon_1 and vertex j of polygon_2.
Polygons are scaled and centered.

Find the largest of the smallest vertex distances



TU Delft

Dissimilarity Based Classification of Polygons



Zero error difficult to reach!

TU Delft

Non-Euclidean Representations: Causes

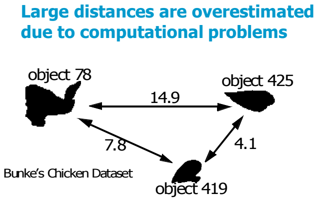
Computational Noise

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

TU Delft

Computational Problems



Weighted edit distance for strings

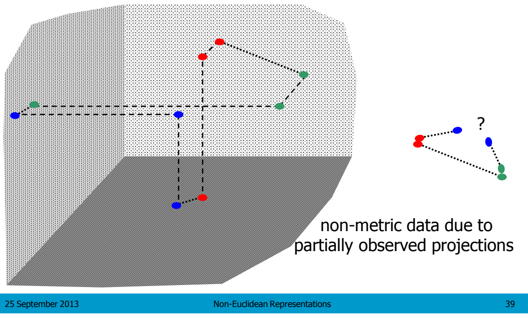
Lack of information

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.

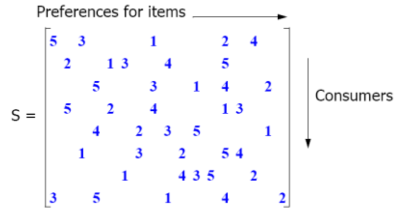
D(V,J) = 5 days
D(V,X) = 4 hours
D(X,J) = 4 hours

Projections - Occlusions

Small distances are underestimated

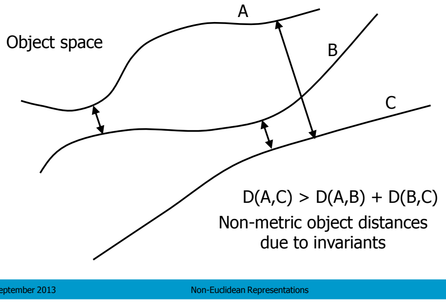


Projections - Occlusions



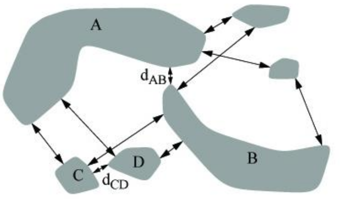
Example: consumer preferences for recommendation systems

Intrinsicly Non-Euclidean Dissimilarity Measures Invariants



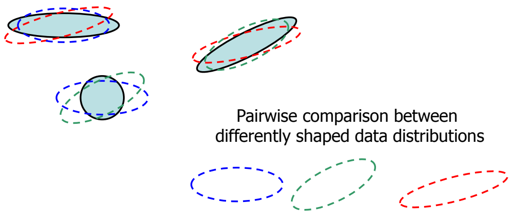
D(A,C) > D(A,B) + D(B,C)
Non-metric object distances
due to invariants

Boundary distances



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

Intrinsicly Non-Euclidean Dissimilarity Measures Mahalanobis



Different pairs → different comparison frameworks
→ non-Euclidean

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

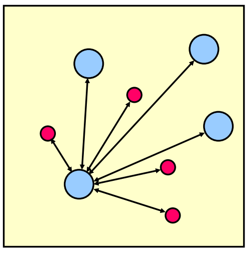
→ Non-Euclidean dissimilarities

Causes of Non-Euclidean Dissimilarities

- **Computational / Observational Limitations**
  - numerical accuracy problems
  - overestimated large distances (too difficult to compute)
  - underestimated small distances (one-sided view of objects)
- **Essential non-Euclidean distance definitions**
  - the human distance concept differs from the mathematical one
  - no global framework
  - invariants

Non-Euclidean Representations: Informativeness

Artificial Example ,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

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	$\tilde{d}_{ij}^2 = d_{ij}^2 + c, i \neq j$	$X = \text{Embedding}(\tilde{D})$
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

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)

TUDelft

	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	0.791	**0.020**	0.067	0.173
Chickenpieces90	446	5	0	0.152	0.791	**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	0.103	**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	0.00[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**	**0.262**	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	**0.34**[illegible]	[illegible]	[illegible]
NewsGroups	600	4	4e-5	0.202	0.733	**0.198**	0.213	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

Representation of non-Euclidean relations

- New objects may not fit into the space
- They should be included in the representation
- → Semi-supervised learning
- → Transductive learning
- → Use out-of-data examples
- → Generalized dissimilarities

TUDelft

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

TUDelft

Final Words

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

TUDelft