Conferenza Italiana sui Sistemi Intelligenti

Prototype Selection for Dissimilarity-based Classifiers

Robert P.W.Duin, Elżbieta Pękalska, Pavel Paclik

Faculty of Electrical Engineering, Mathematics and Computer Science

Delft University of Technology, The Netherlands

Perugia, 16 September 2004

P.O. Box 5031, 2600GA Delft, The Netherlands.
Phone: +(31) 15 2786143, FAX: +(31) 15 2781843,
E-mail: r.p.w.duin@ewi.tudelft.nl

Contents

A Problem

Bring me an apple!!

??

Training

These are apples ...

... and these are pears

!!

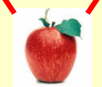
How to learn? Model based?

- Store all relevant properties of an apple

- Generalise over apple examples to obtain an 'apple class' <u>model</u>

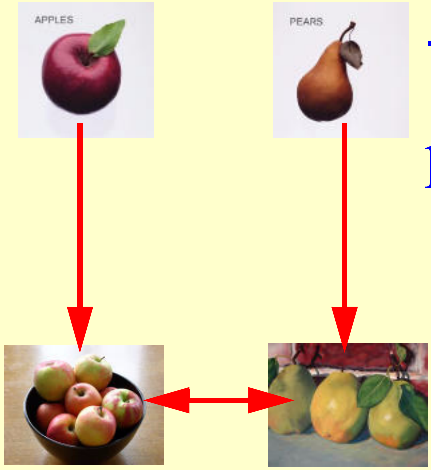
- Repeat for pears

- Find a way to compare 'apple-ness' with 'pear-ness'



We only look at class differences at classification time!!

How to learn? Feature based?



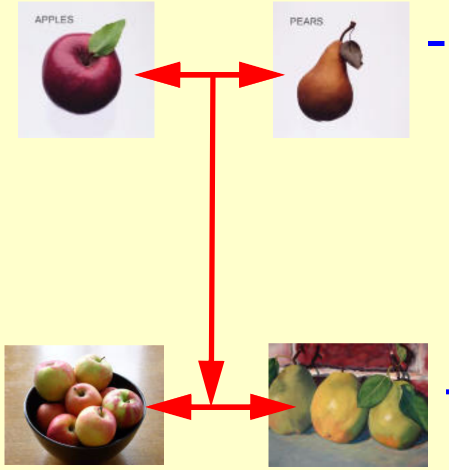
- Choose possible <u>features</u> to represent individual apples and pears

- Select relevant <u>features</u> for the difference of apples and pears

- Generalise over the examples

We only look at class differences at classification time!!

How to learn? Dissimilarity based?



- Use differences between apples and pears for representation

- Generalise over the examples

We only look at class differences during representation!!

The Pattern Recognition System

Model based Objects Measurements Models Class Differences

The Pattern Recognition System

Model based	Objects	Measurements	Models	Class Differences
Feature based	Objects	Features	Class Differences	Classifier

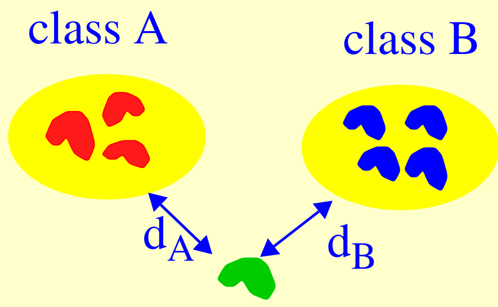
The Pattern Recognition System

Model based	Objects	Measurements	Models	Class Differences
Feature based	Objects	Features	Class Differences	Classifier
Dissimilarity based	Objects	Class Differences	Class Differences	Classifier

Representation Principles

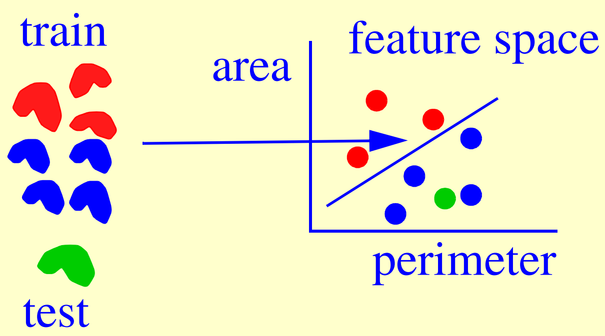
Model based, **dissimilarities of objects and classes (Conceptual)**

domains, structural, connectivity included



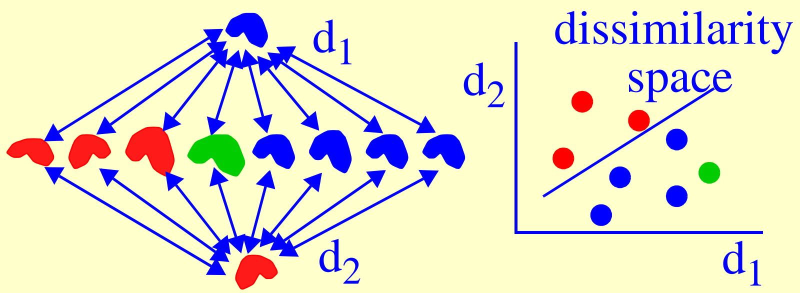
Feature based **(Absolute)**

distributions, connectivity neglected



Dissimilarity based **(Relative)**

distributions or domains, connectivity possibly included

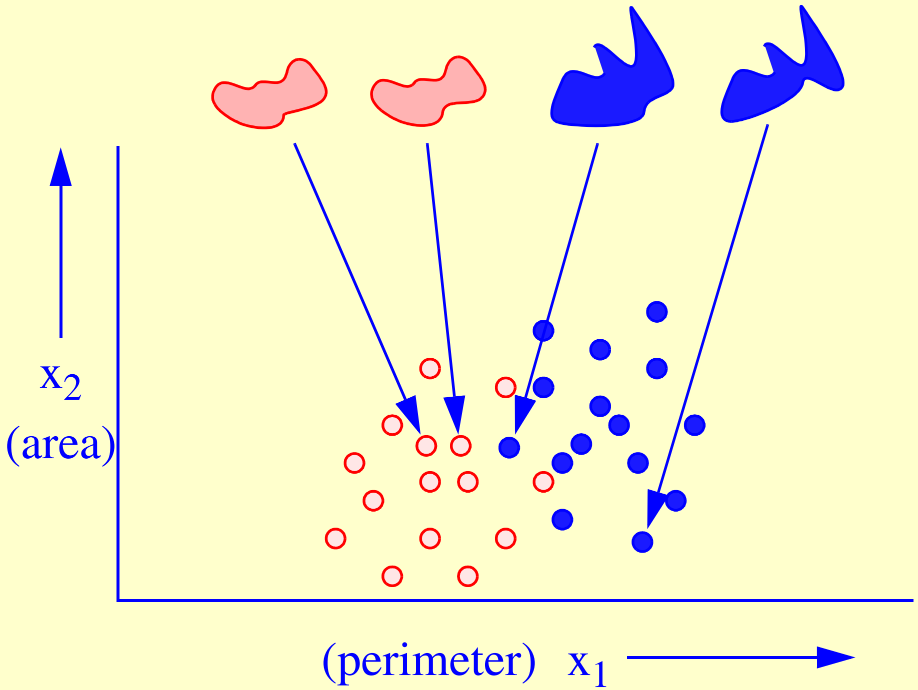


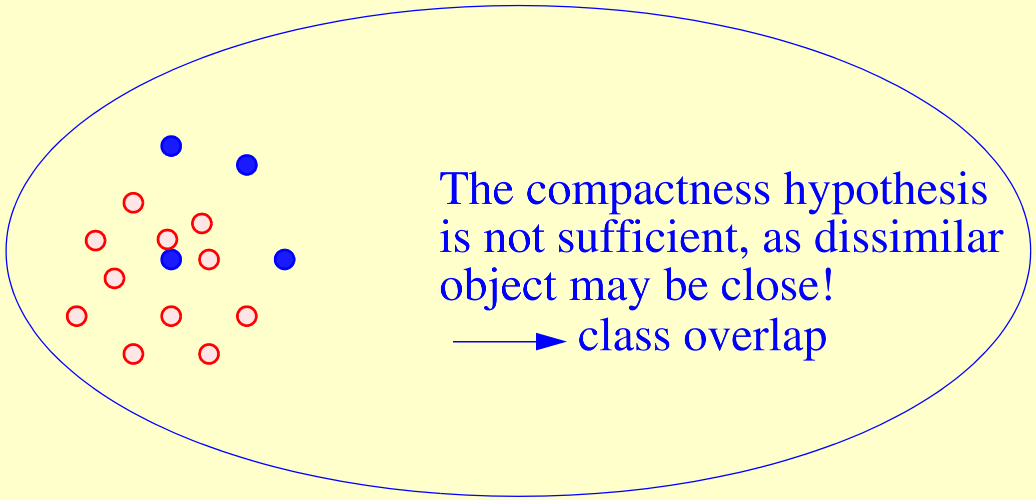
The Compactness Hypothesis

Representations of real world similar objects are close.

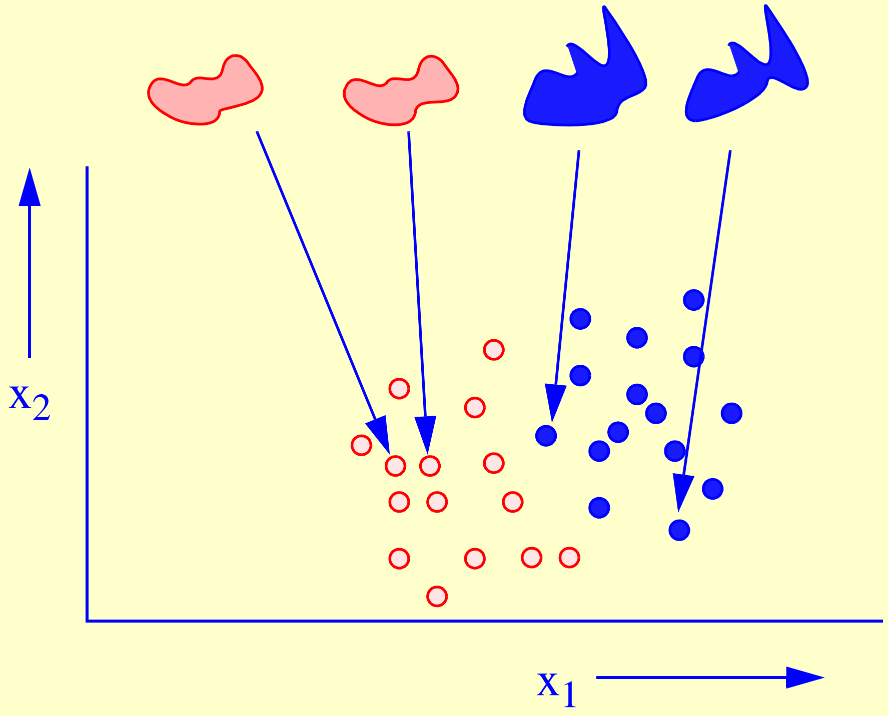
There is no ground for any generalization (induction) on representations
that do not obey this demand.

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





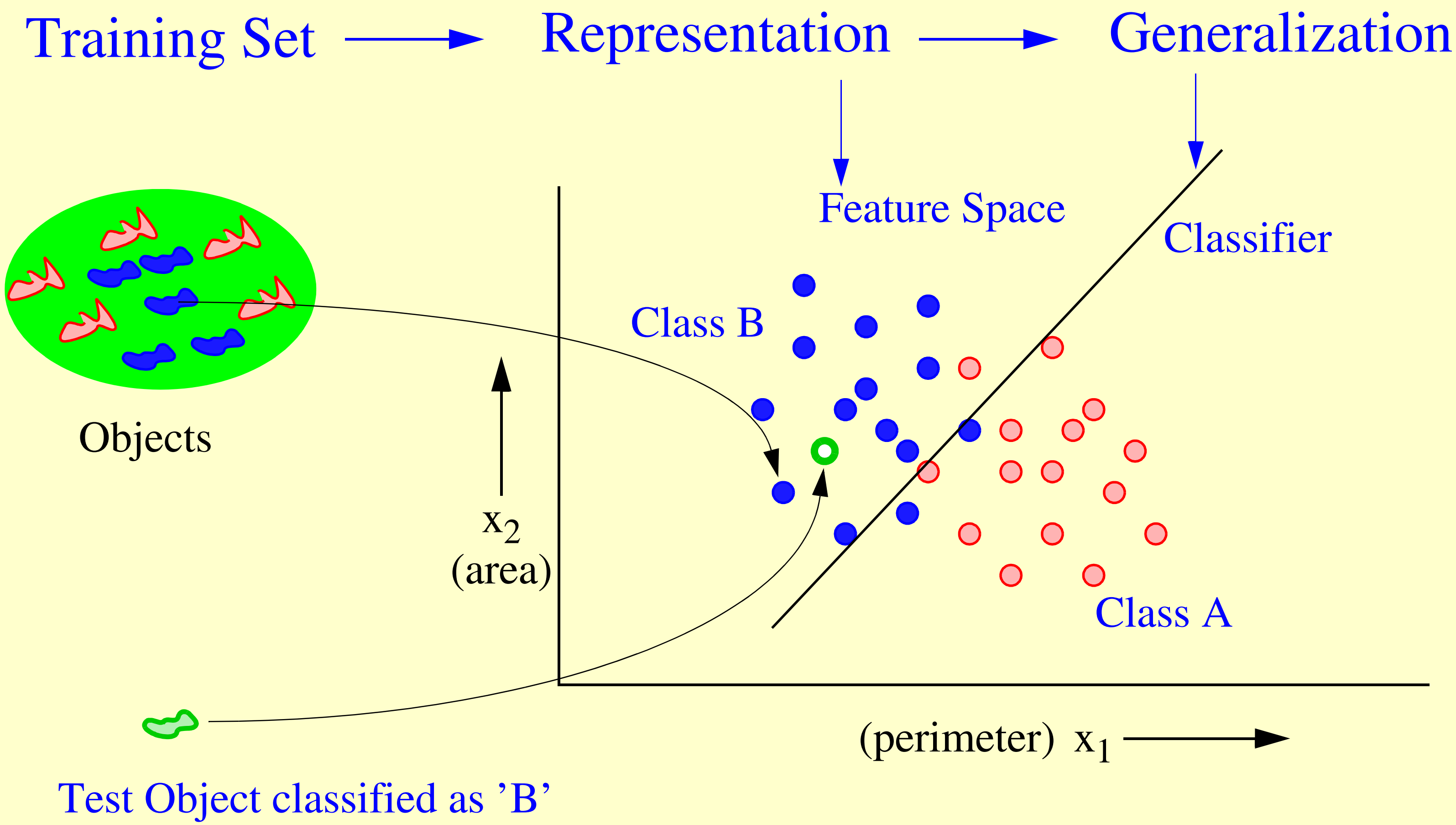
True Representations



Similar object should be close and dissimilar objects should be distant

→ Dissimilarity representations based on measurement signals describing the 'whole' object fulfill this.

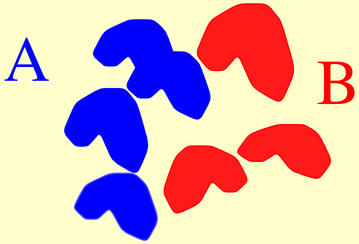
Generalisation from Features



Feature representation → Object reduction → Class overlap → Probabilities

Dissimilarity Representation (DisRep)

Define dissimilarity measure d_{ij} between raw data of objects i and j



Given labeled training set T

Unlabeled object x to be classified

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

$$d_x = (d_1 \; d_2 \; d_3 \; d_4 \; d_5 \; d_6 \; d_7)$$

The traditional Nearest Neighbor rule (template matching) just finds:

label(argmin$_{\text{trainset}}$(di)),

without using DT. Can we do any better?

Dissimilarity representations: motivation

- DRs describe objects by their pair wise dissimilarities.
- A class is a set of similar objects.
- DRs are general:
  - { based on raw measurements
  - { in feature spaces
  - { in structural approach (for strings, grammars, trees
  - { based on human judgments
- They allow to integrate/unify both statistical and structural approaches.
- Shape/image recognition.
- Spectra/image analysis.
- For heterogeneous data, e.g. described by continuous and categorical variables.

Dissimilarity representations: measures (I)

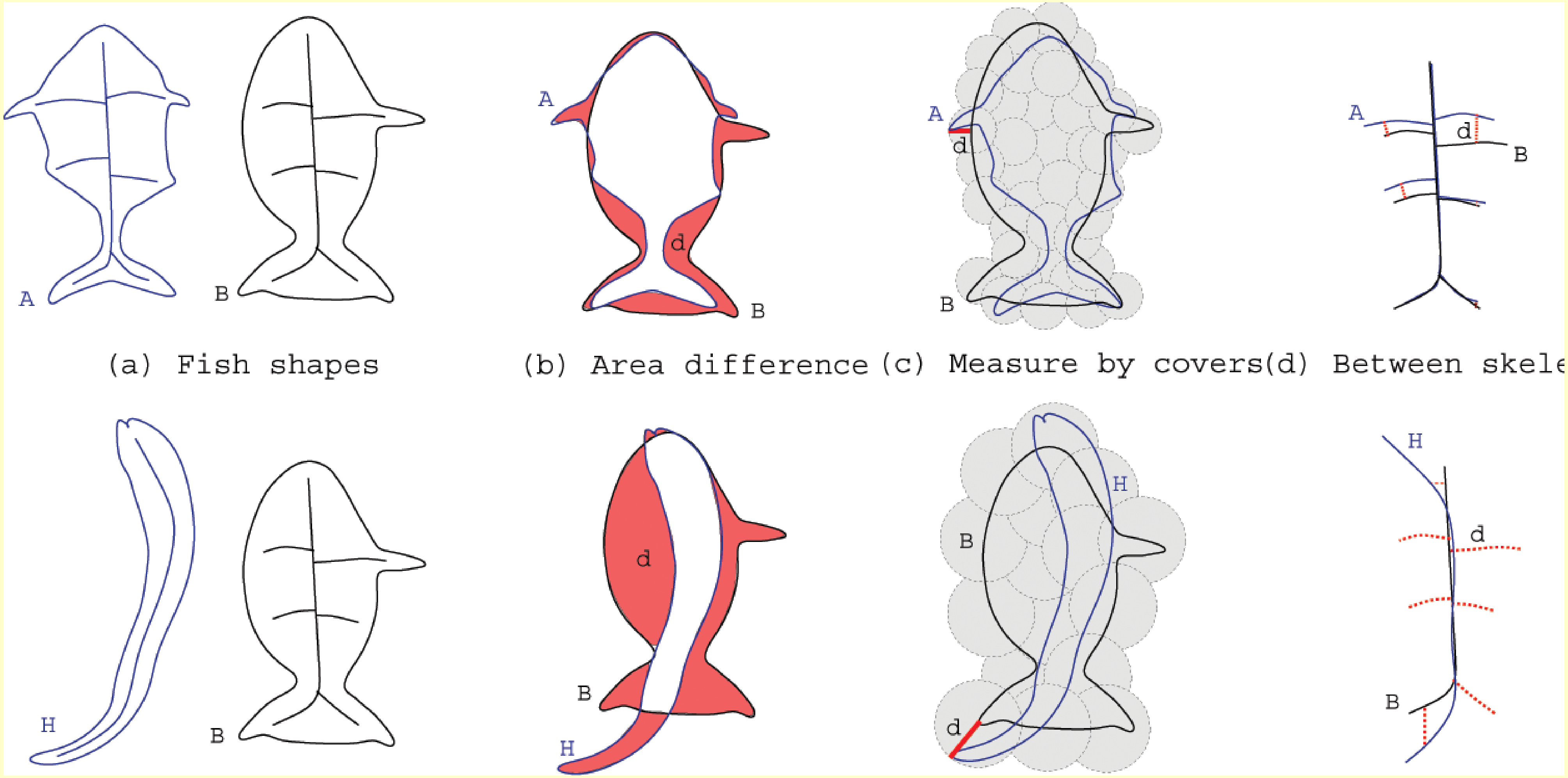
Dissimilarity (distance) measure d describes the degree of difference.

Properties of d on X:

1. **Nonnegativity**: $d(x, x) \geq 0$.
2. **Reflexivity**: $d(x, x) = 0$.
3. Symmetry: $d(x, y) = d(y, x)$.
4. Definiteness: $d(x, y) = 0 \rightarrow x = y$.
5. Triangle inequality: $d(x, y) + d(y, z) \geq d(x, z)$.
6. **Compactness Hypothesis**: if x and y are similar, then $d(x, y) < \delta$.
7. True representation: $d(x, y) < \delta$, then x and y are similar;

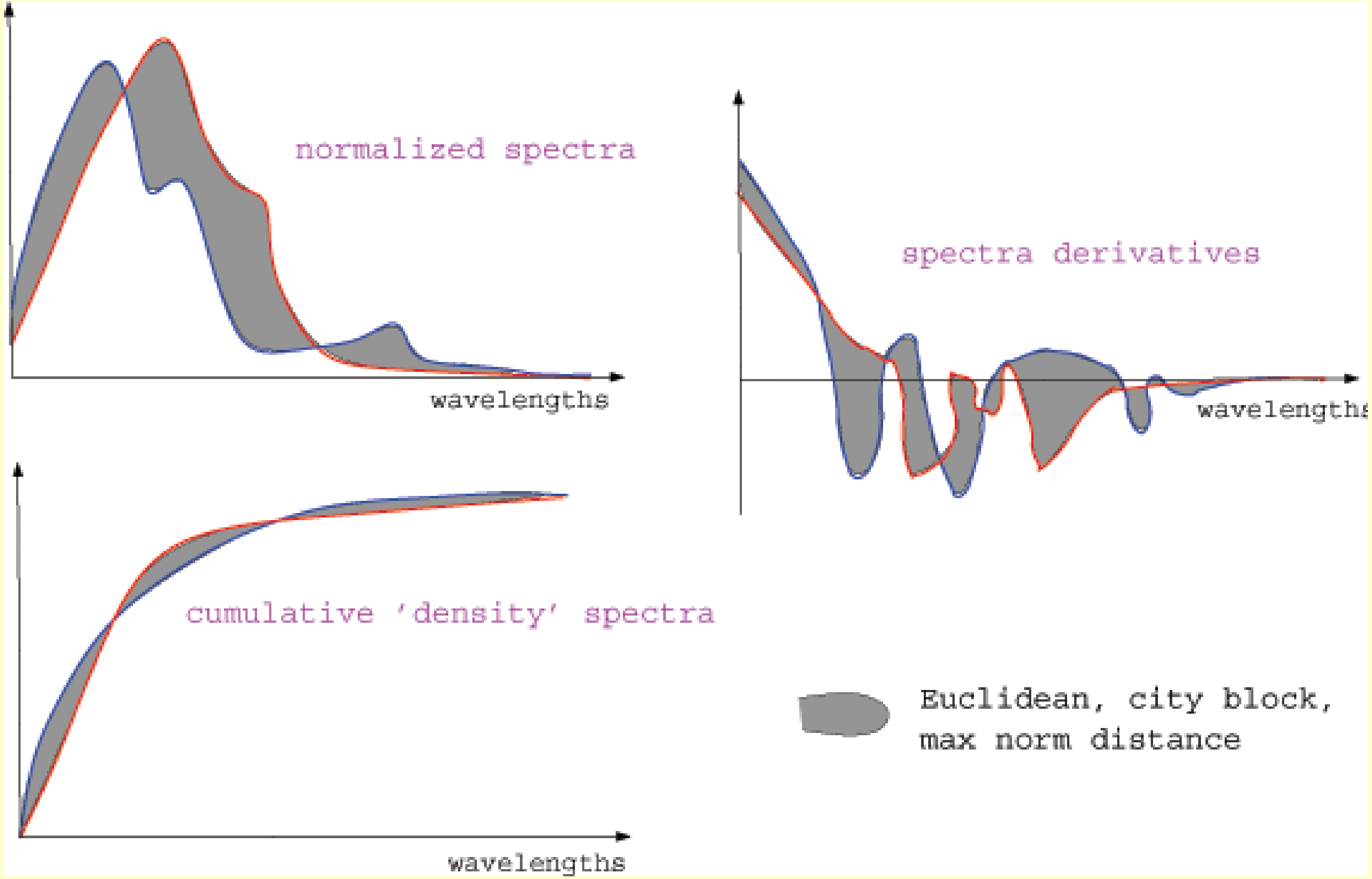
   $d(x, y) >> \delta$, then x and y are dissimilar.

Dissimilarity representations: measures (II)



It is more important that the dissimilarity measure is **descriptive** and **discriminating** for the problem than its metric properties.

Dissimilarity representations: measures (III)



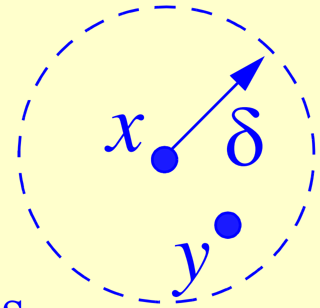
P. Paclik and R.P.W. Duin, Dissimilarity-based classification of spectra: computational issues, Real Time Imaging, vol.9, no. 4, 237-244, 2003.

Dissimilarity representations: measures (IV)

In real applications, the dissimilarity measure should be robust to noise and small aberrations in the (raw) measurements.

Here we study how to learn from dissimilarity representations.

Dissimilarity representations: prospects



Let us assume that

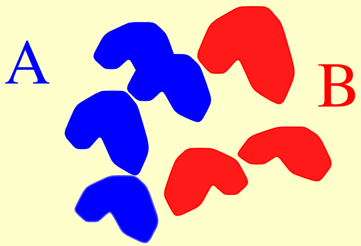
$d(x, y) < \delta$ if and only if the objects x and y are very similar.

If δ is sufficiently small, then x and y belong to the same class, as y is a small distortion of x. True representation is assumed.

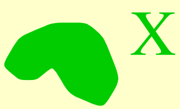
Since for sufficiently large training sets there will be such an object y in the neighborhood of $x \rightarrow$ Zero error classification is possible!

R.P.W. Duin and E. Pekalska, Possibilities of zero-error recognition by dissimilarity representations, Proc. Pattern Recognition in Information Systems, Alicante, ICEIS Press, Portugal, 2002, 20-32.

DisRep Approach: NN Rule, Pre-topological Space



Given labeled training set T



Unlabeled object x to be classified

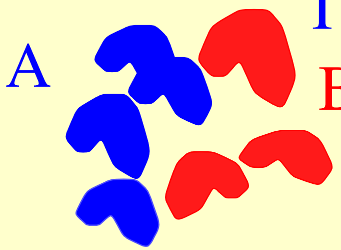
$d_x = (d_1 \;\; d_2 \;\; d_3 \;\; d_4 \;\; d_5 \;\; d_6 \;\; d_7)$

class(x) = label (argmin(d_i))

- Computationally expensive
- Locally sensitive
- Consistent: if size(T) --> ∞ then error --> 0

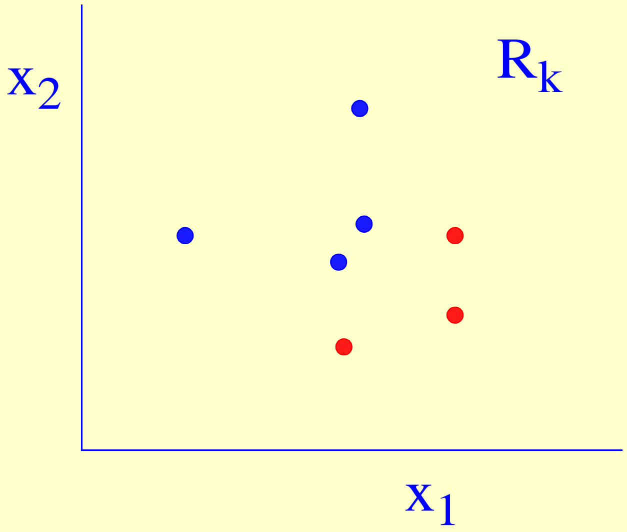
Dissimilarity Representation Approach: Embedding

Training set



→ Dissimilarity matrix D → X?

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



Euclidean distances → D

If D is non-Euclidean, embedding results in a pseudo-Euclidean Space (*Goldfarb, Pekalska*)

Dissimilarity Representation Approach: Dissimilarity Space

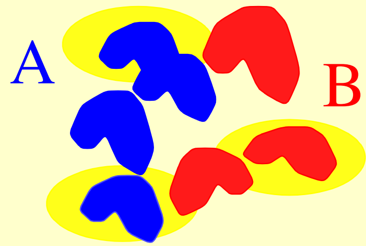
Dissimilarities

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

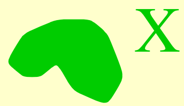
$$d_x = (\, d_1 \;\; d_2 \;\; d_3 \;\; d_4 \;\; d_5 \;\; d_6 \;\; d_7 \,)$$

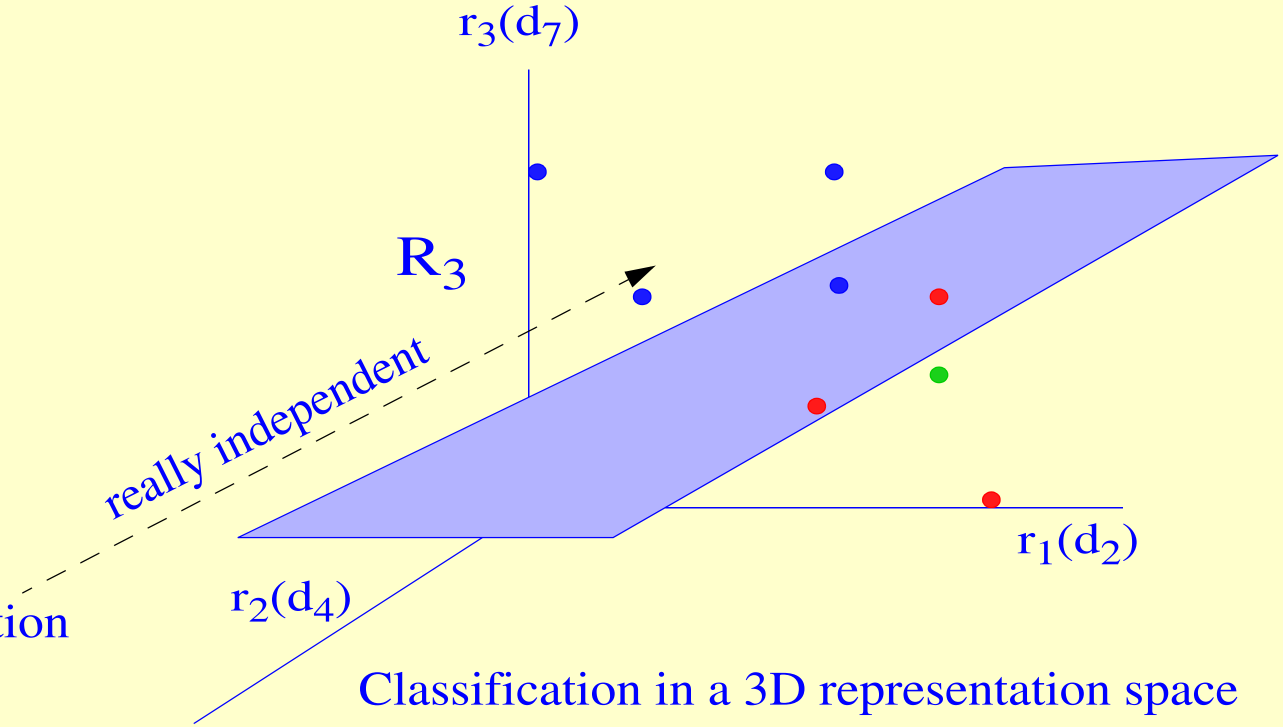
Selection of 3 objects for representation
(at random?)

Given labeled training set T



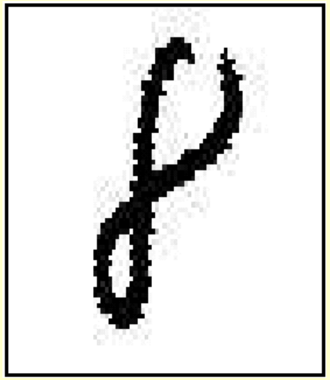
Unlabeled object x to be classified

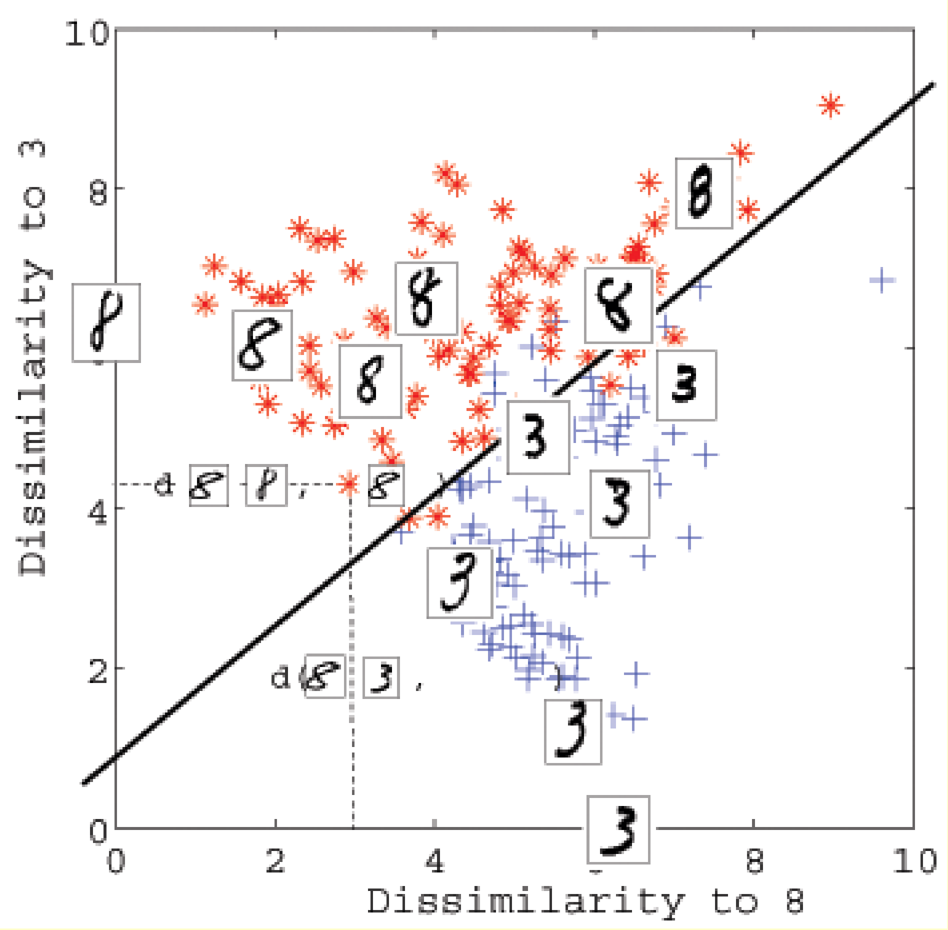




Classification in a 3D representation space

Example In Dissimilarity Spaces





Two randomly selected prototypes

NIST handwritten digits 3 and 8:

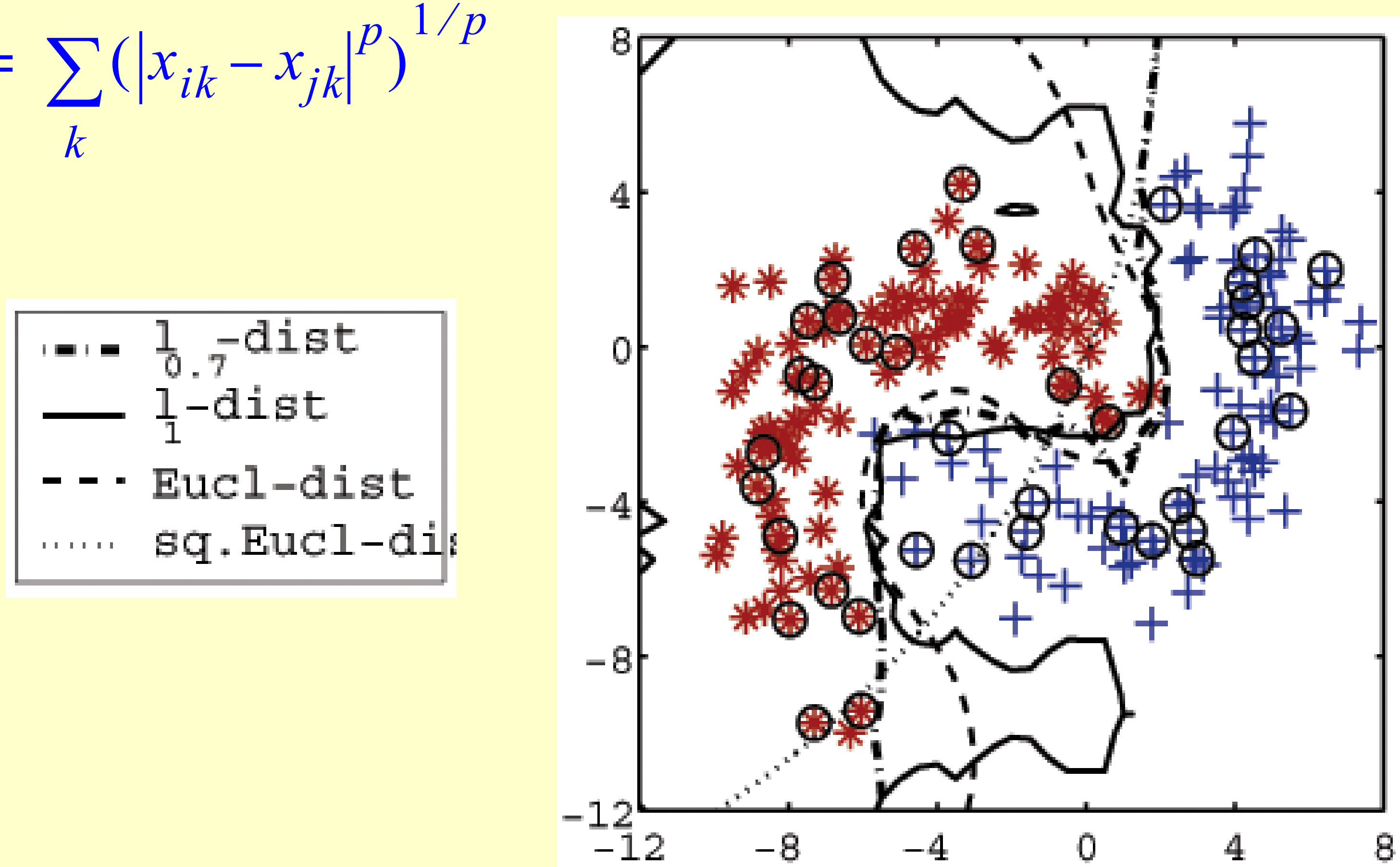
Euclidean distances *D(T;R)* between the Gaussian-smoothed images.

Classifiers in Dissimilarity Space

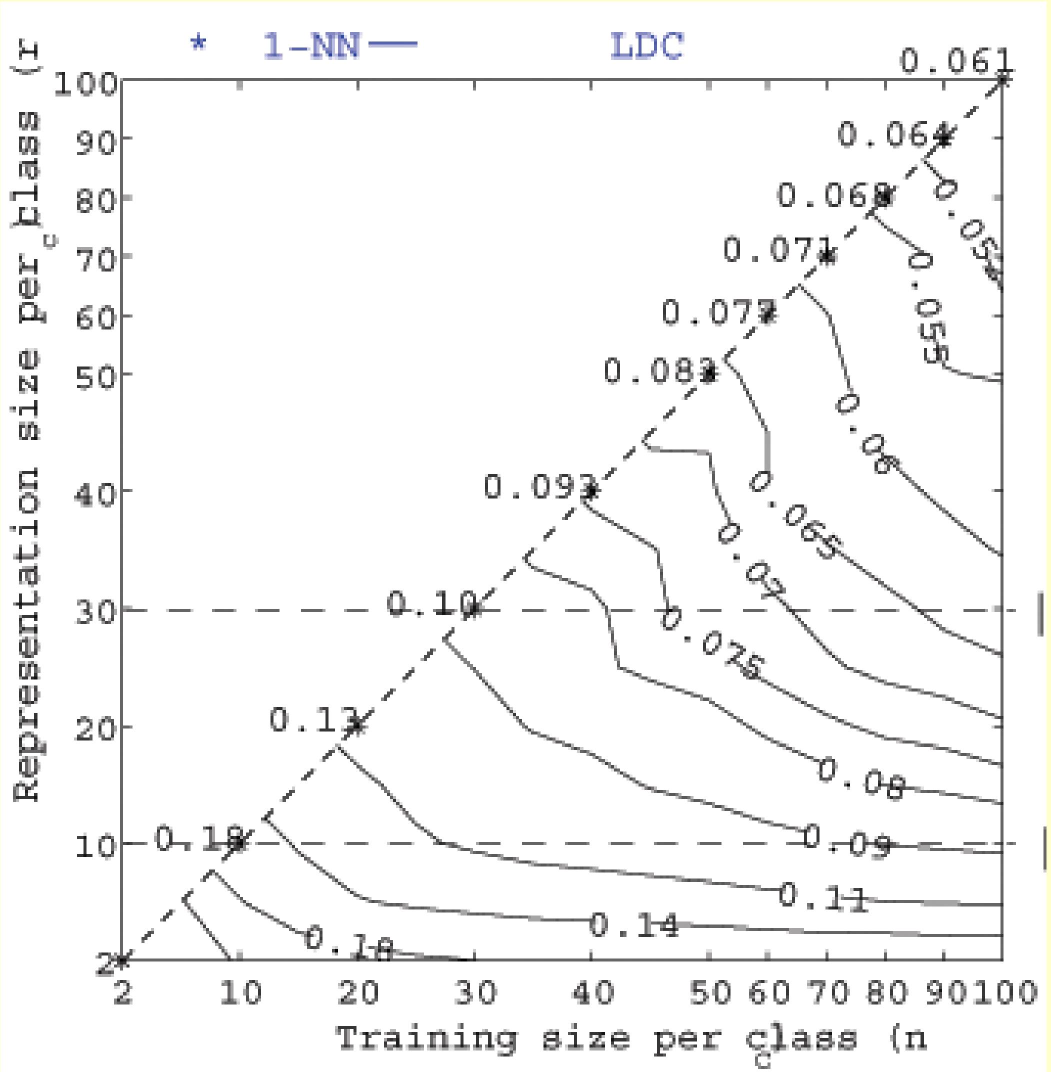
A linear classifier in a dissimilarity space $f(D(x, R)) = w'D(x, R) + w_0$

is a nonlinear classifier in the underlying feature space.

l_p-dist: $d(x_i, x_j) = \sum_k \left(\left| x_{ik} - x_{jk} \right|^p \right)^{1/p}$



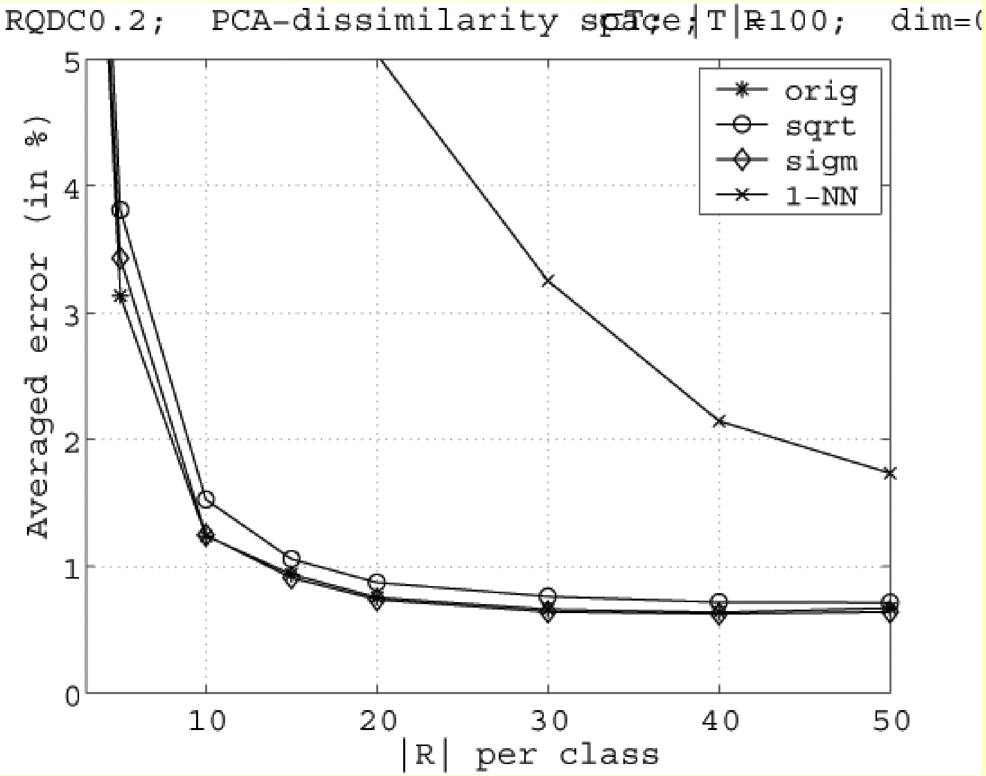
Classification in dissimilarity spaces: proof of principle

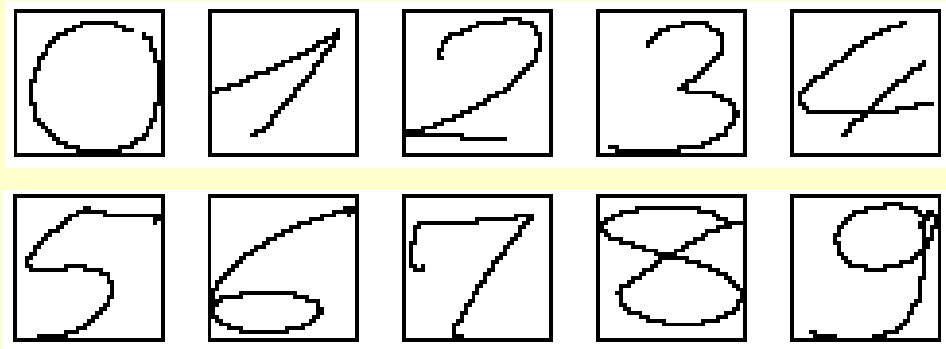


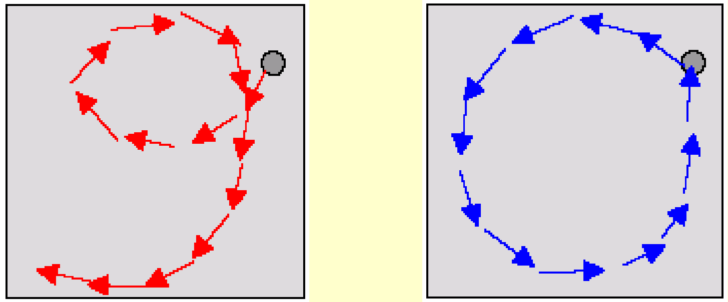
Nist Data, 16 x 16, normalized Euclidean distances between the Gaussian-smoothed images.

TUDelft

Example: Edit Distance between Pen-written Digits





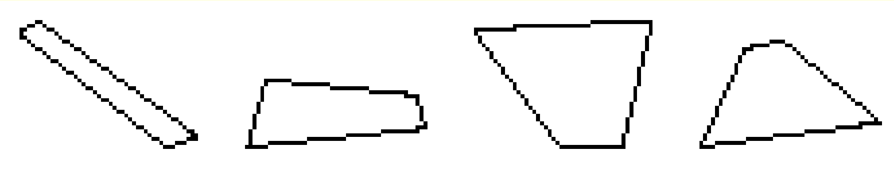


Edit distance with fixed costs for insertion, deletion and substitution.

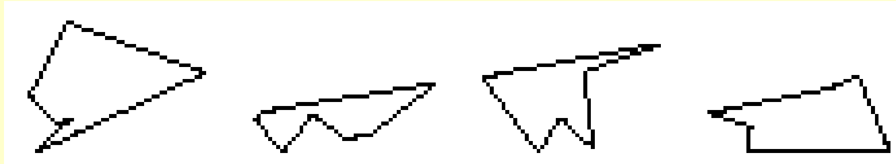
The distance measure is moderately non-metric.

Data: S. Günter, H. Bunke

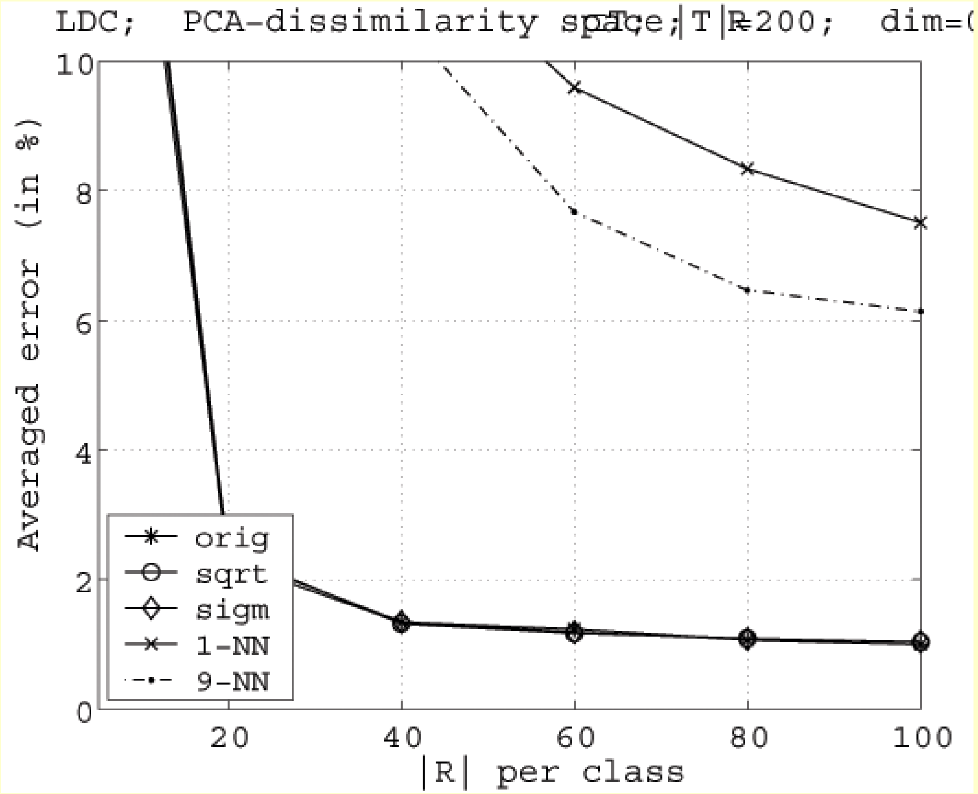
Example: Modified Hausdorff Distance between Randomly Generated Polygons

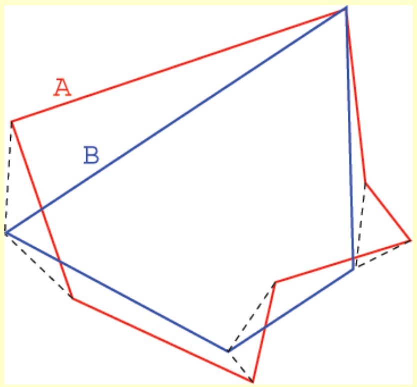


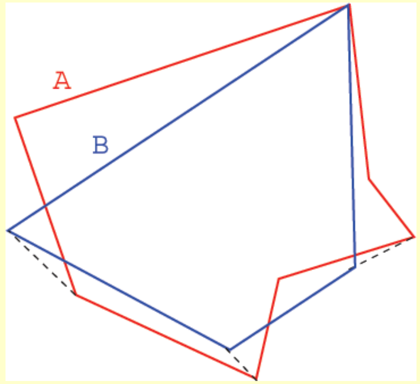
Quadrilaterals



Heptogons







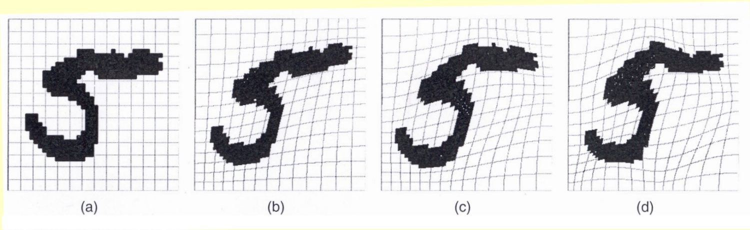
$$d(A,B) = \underset{a \in A}{\mathrm{mean}}(d(a,B))$$

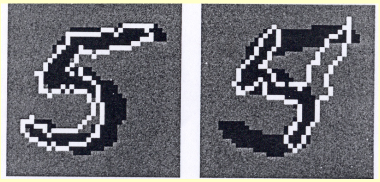
$$d(B,A) = \underset{b \in B}{\mathrm{mean}}(d(b,A))$$

$$d(A,B) = \max\{d(A,B), d(B,A)\}$$

See Dubuisson and Jain, ICPR12, 1994

Example: Digit Classification by Deformable Templates (Zongker Data)

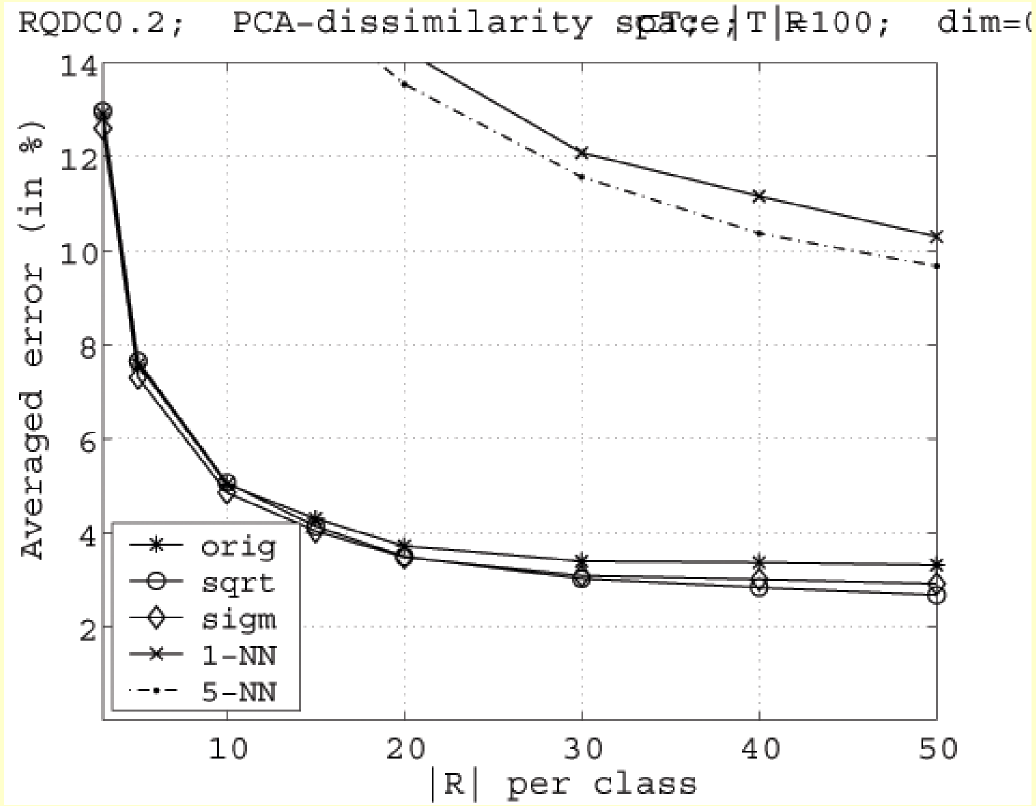




Matching new objects x to various templates y

$$class(x) = class(\arg min_y(D(x, y)))$$

A.K. Jain, D. Zongker, PAMI, 1997.



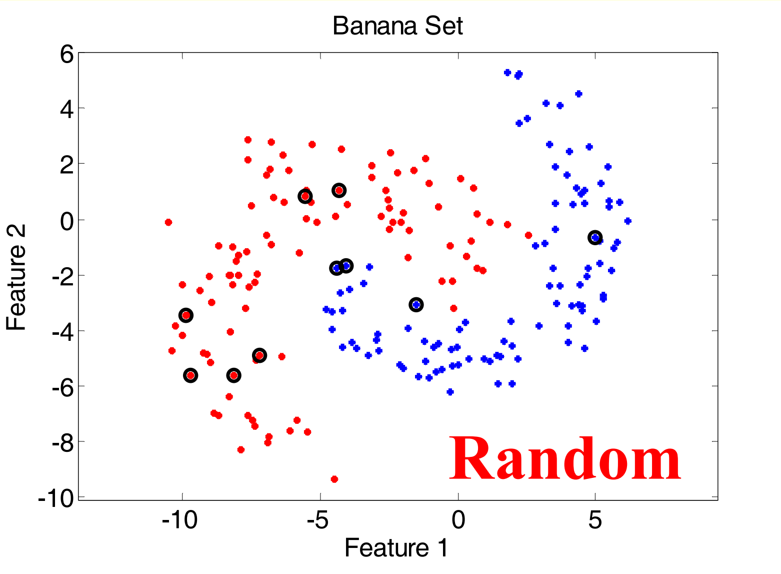
Dissimilarity based classifiers using a randomly selected representation set (prototypes) compared to the nearest neighbor rule for a 10-class digit classification problem.

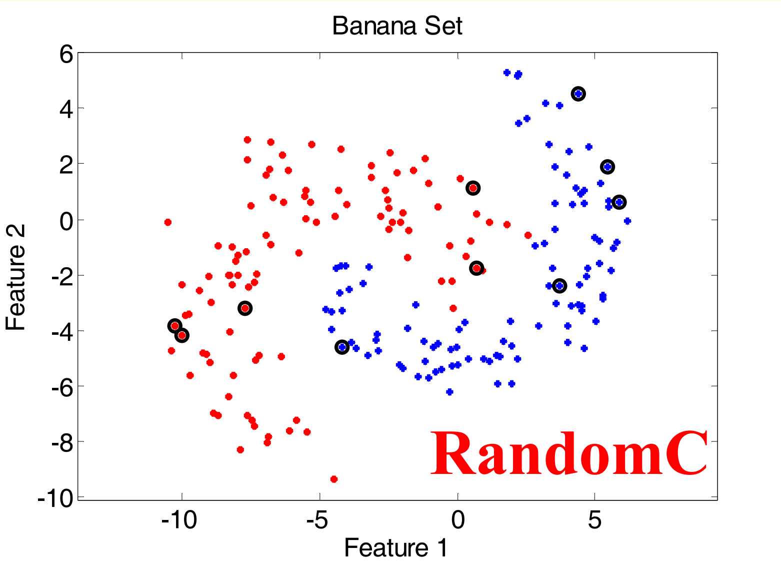
Prototype Selection

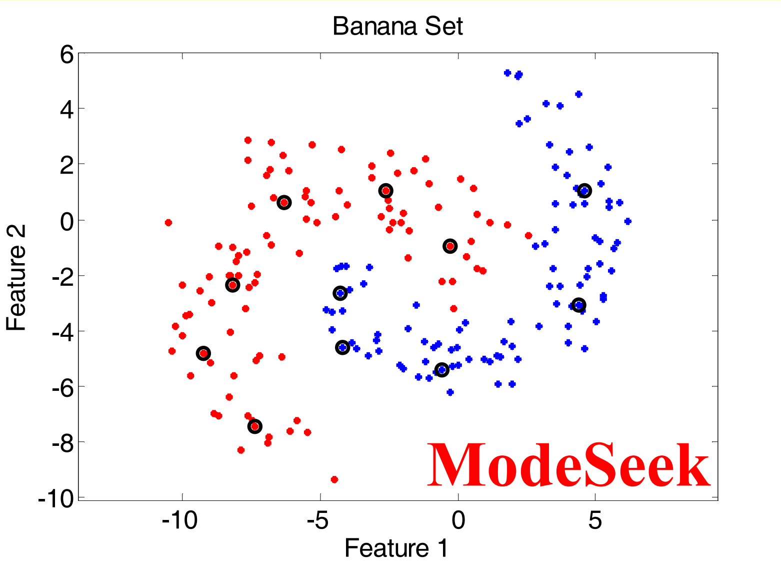
How should a representation set be constituted?

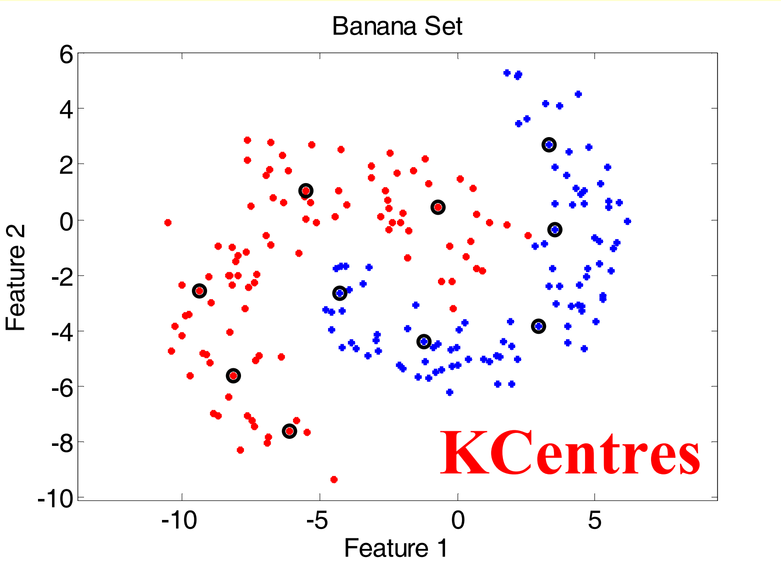
Random	Random selection
RandomC	Random selection per class
ModeSeek	Cluster analysis by mode seeking
KCentres	Cluster analysis by K-Centres
FeatSel	Forward Feature Selection
LinProg	Linear Programming Sparse Solution
KCentres-LP	K-Centres followed by Linear Programming Sparse
Edicon	Editing and Condensing

Feature Space Examples Prototype Selection Procedures



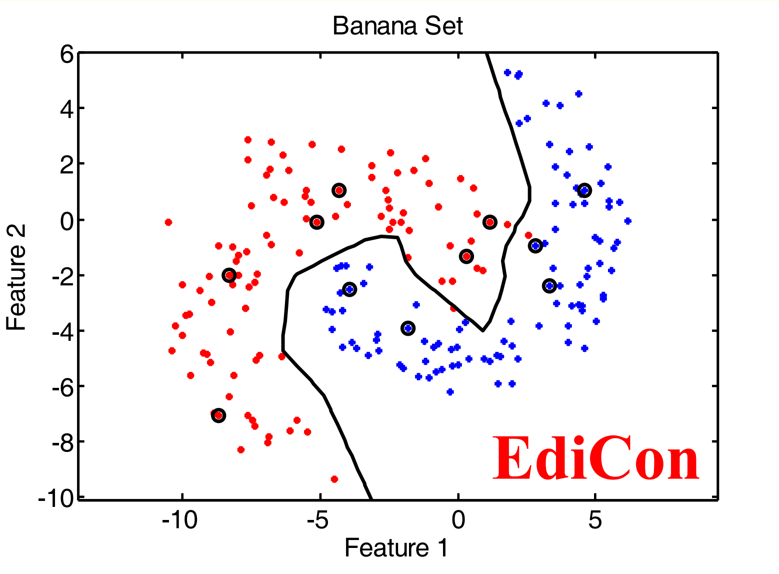


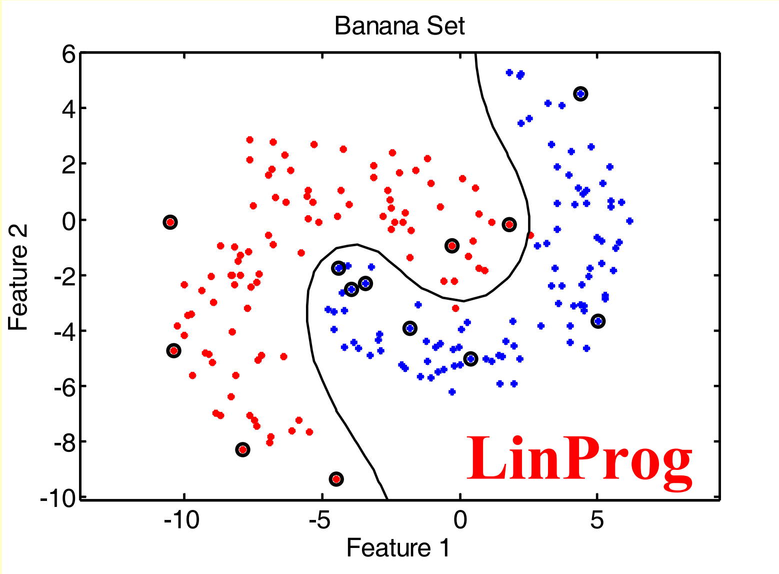


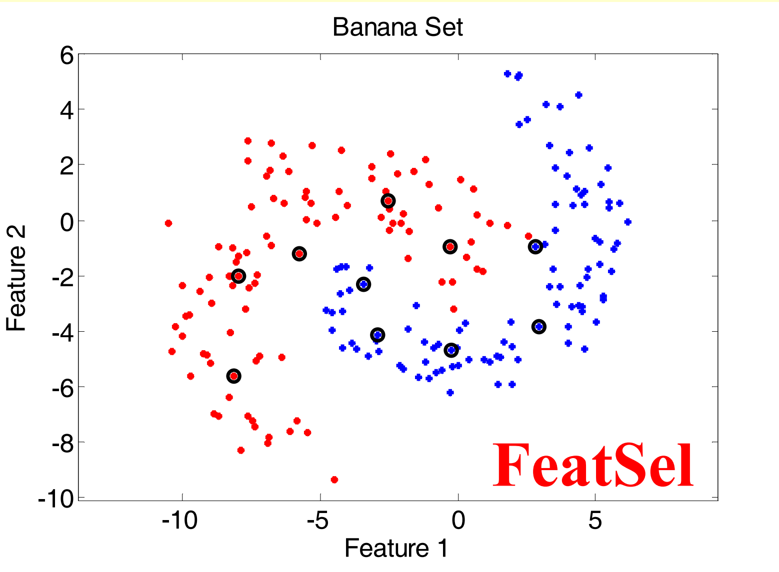


Note: Procedures are applied in dissimilarity space, **not** in feature space!

Feature Space Examples Prototype Selection Procedures



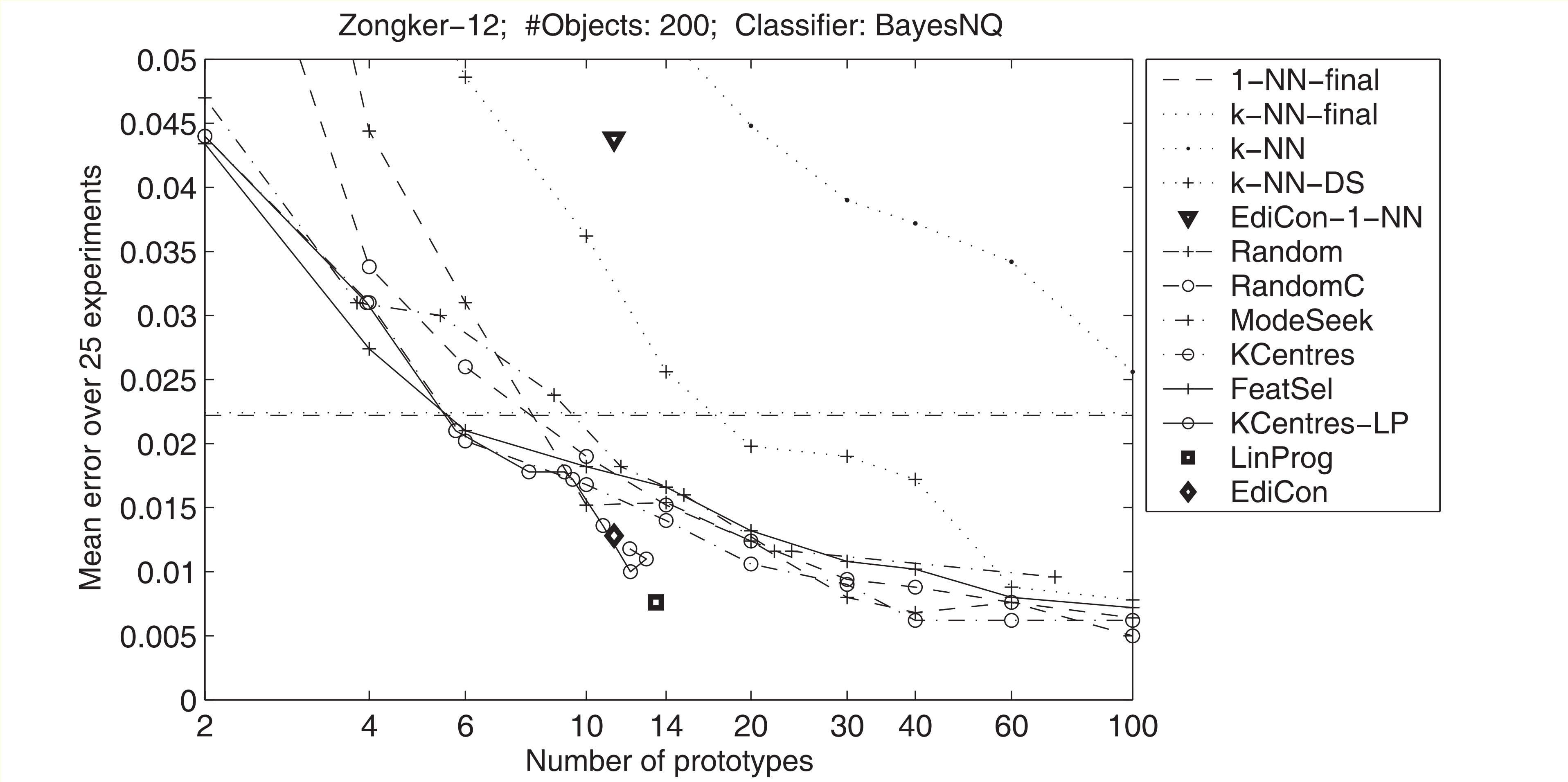




FeatSel treats the dissimilarity space representation D(T,T) as a feature matrix and selects the best 'features' (columns) using the LOO 1-NN error based on the original distances as a criterion.

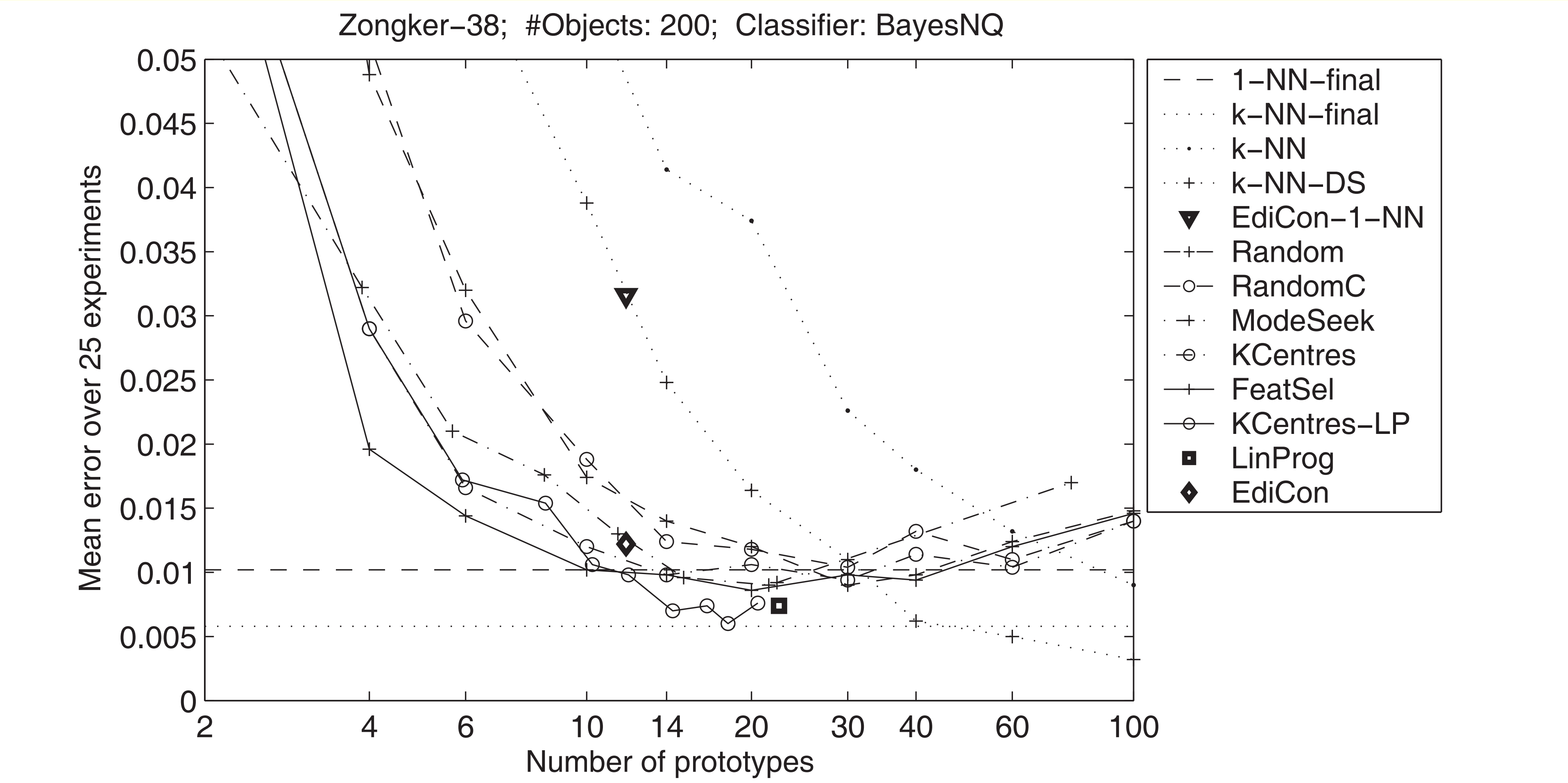
Note: Procedures are applied in dissimilarity space, **not** in feature space!

Zongker Data, Digits 1,2

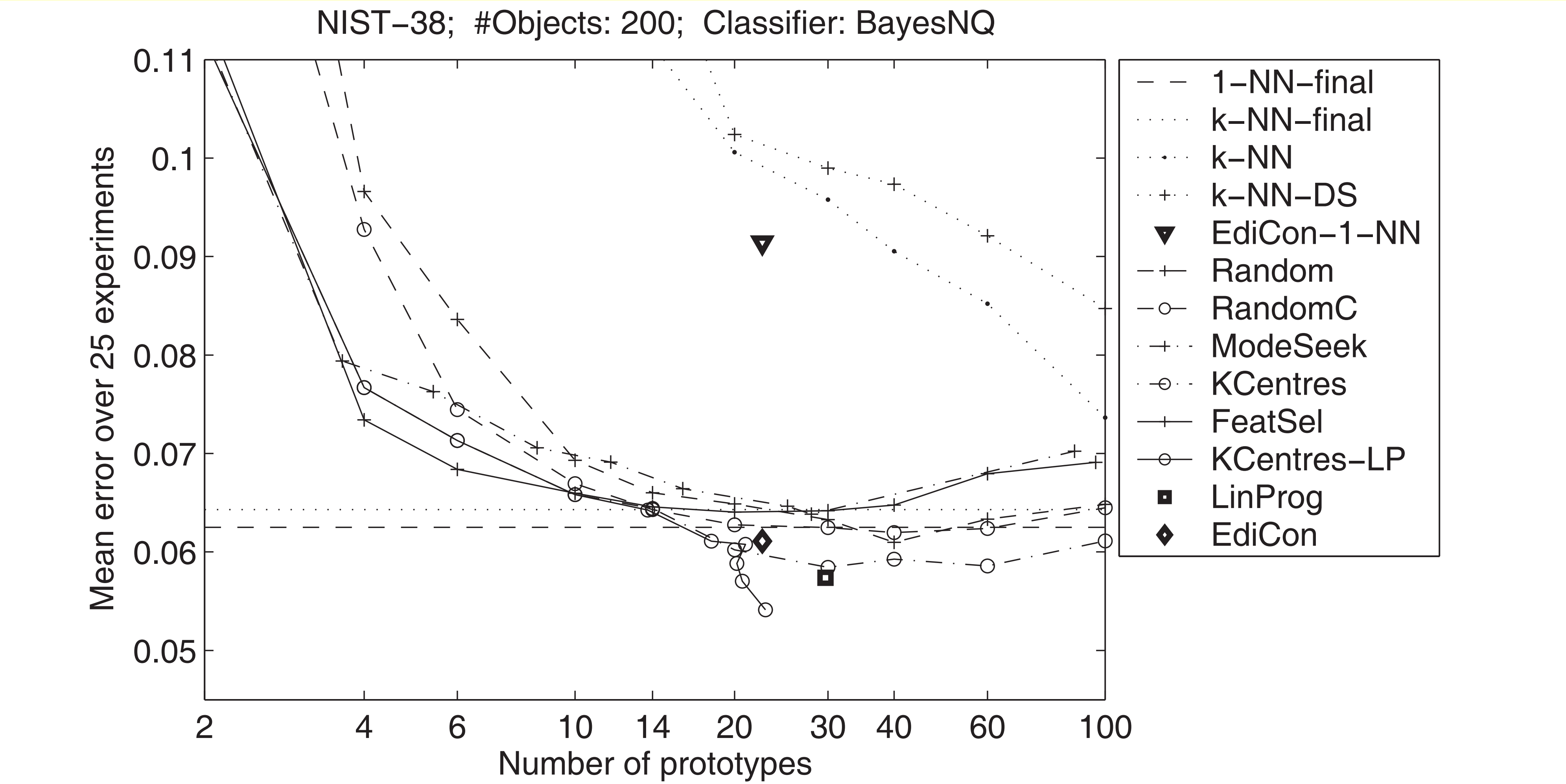


The averaged error (over 25 experiments) of the quadratic Bayes classifier (*) in dissimilarity spaces of various dimensionalities, based on a series of selection procedures. For comparison a number of nearest neighbor results (+) are presented.

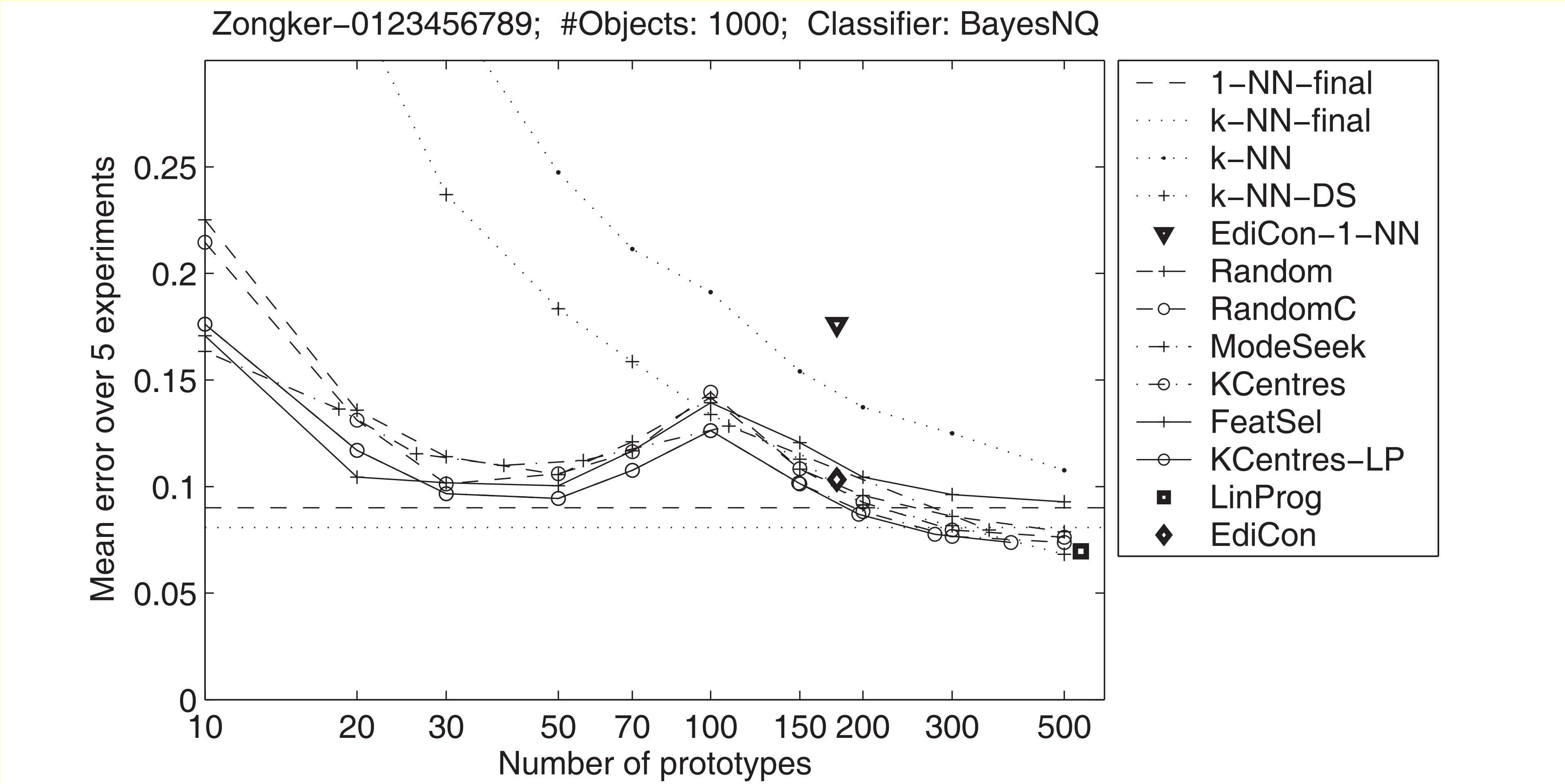
Zongker Data, Digits 3,8



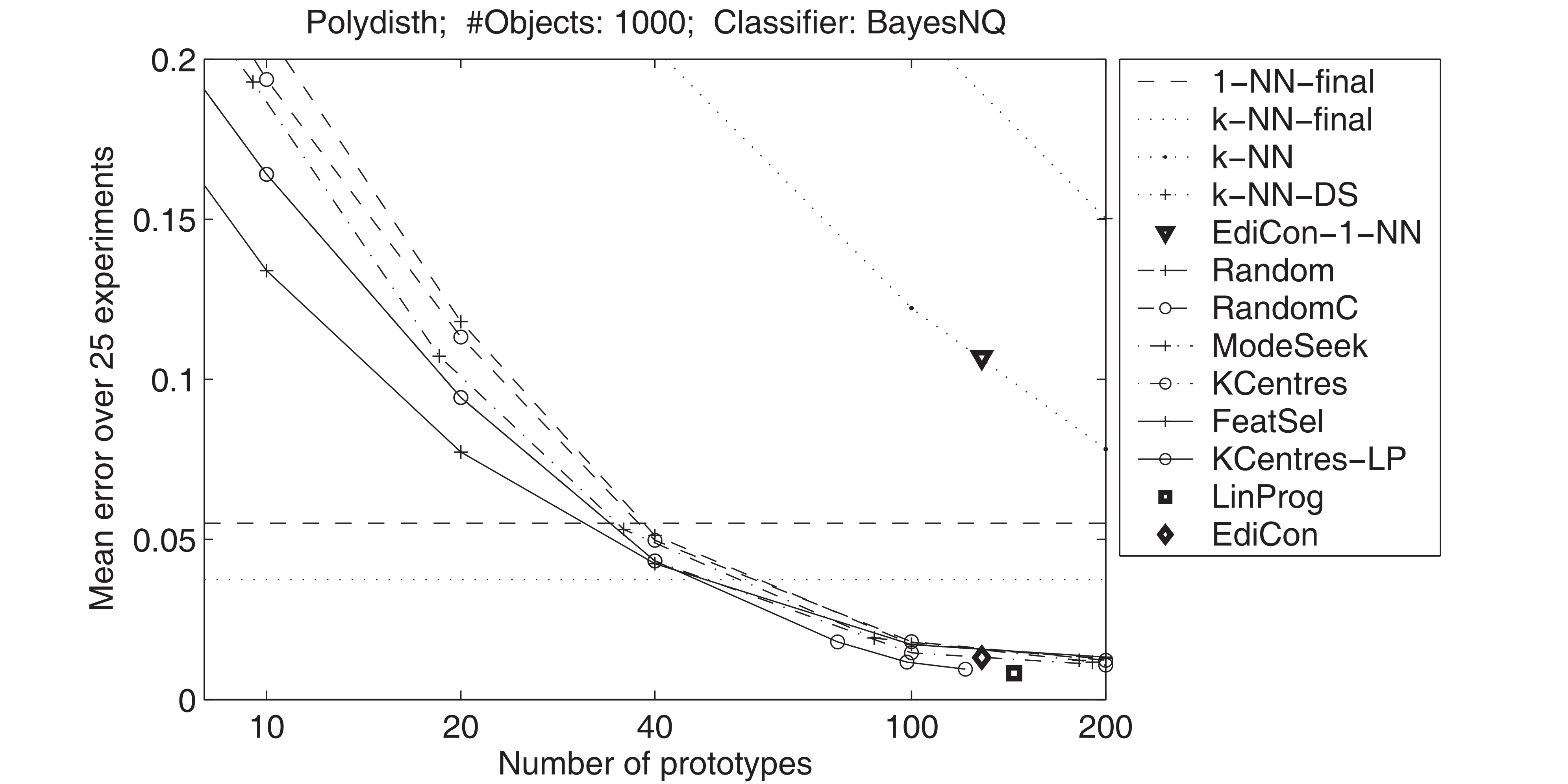
NIST Digits 3,8 - Euclidean Distances



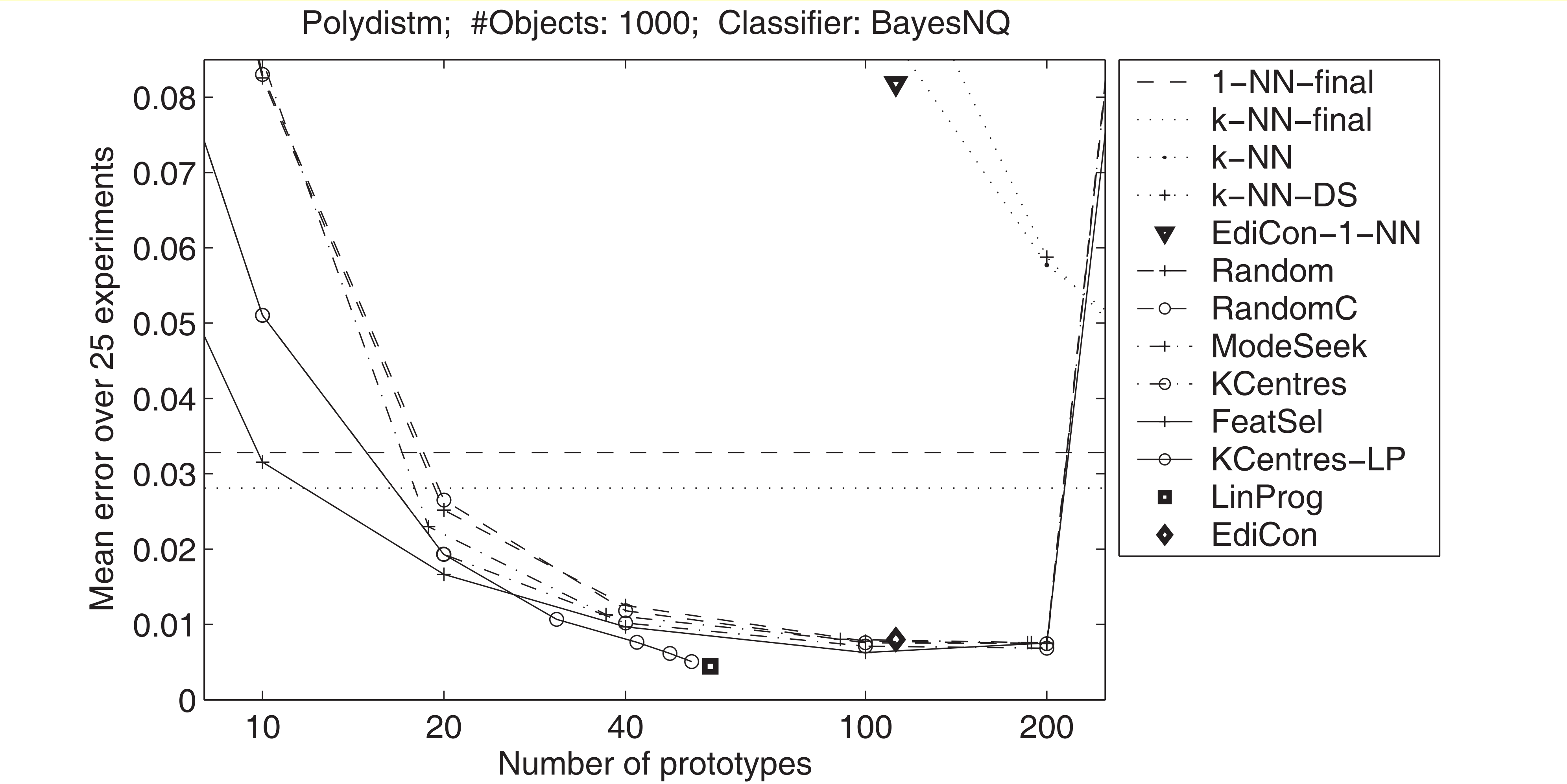
Zongker Data, All Digits



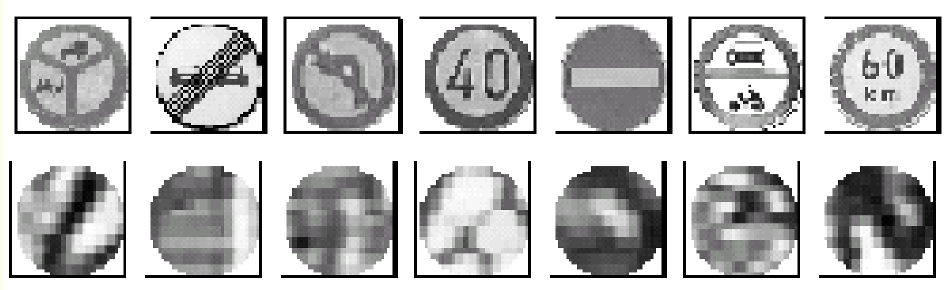
Polygon Classification by Hausdorff Distance

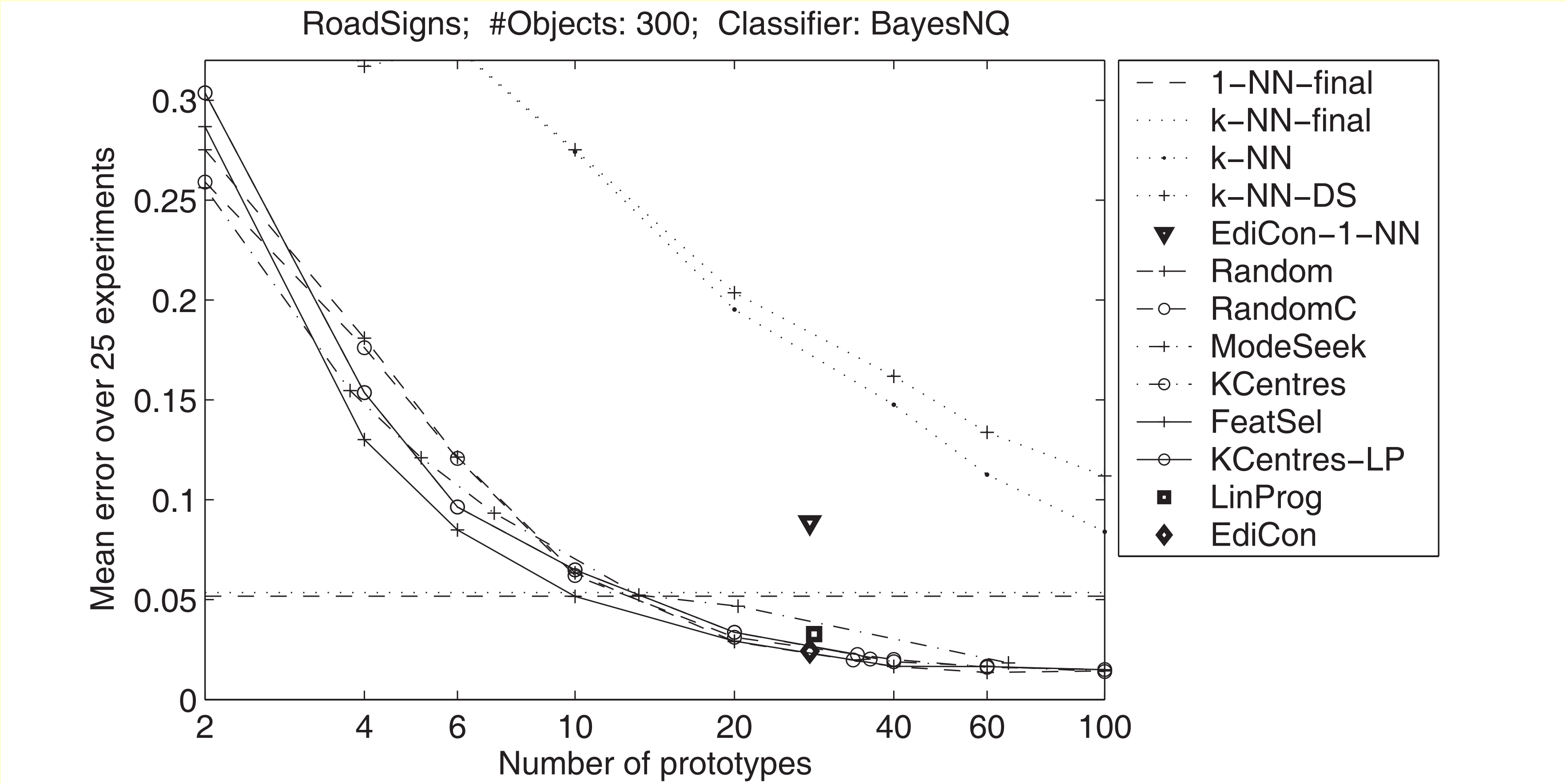


Polygon Classification by Modified Hausdorff Distance



Road Sign Recognition by Partial Templates





Conclusions

Dissimilarity based pattern recognition needs just a small number of prototypes for building a good dissimilarity space.

Representation sets of 5-10% of the training set may perform equally well as k-NN

Larger representation sets may perform significantly better than kNN

The method is especially good for non-metric and non-Euclidean distance measures.

For small sets of prototypes a systematic approach is best.

For larger sets cluster analysis has to be preferred.

After that random selection is sufficient.