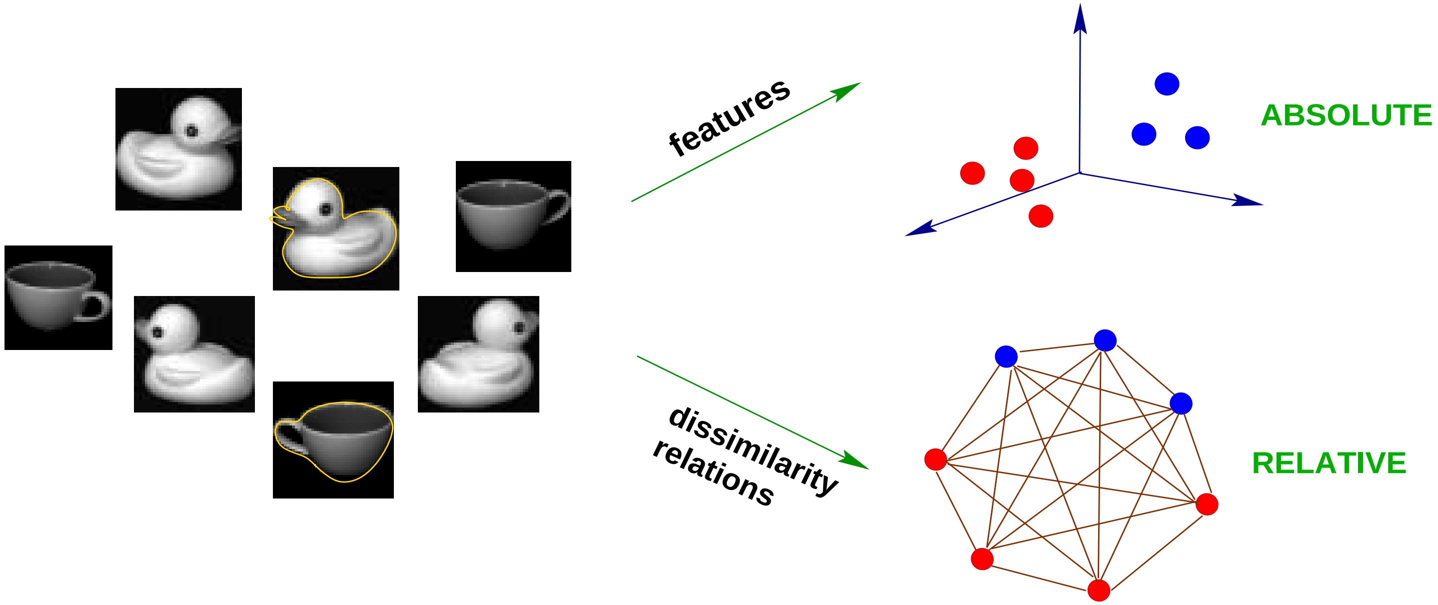
Spanish Pattern Recognition Network

Dissimilarity **representations** for pattern recognition



Robert P.W. Duin, Elzbieta Pękalska and Pavel Paclik

Delft University of Technology

Madrid, 5 September 2003

Contents

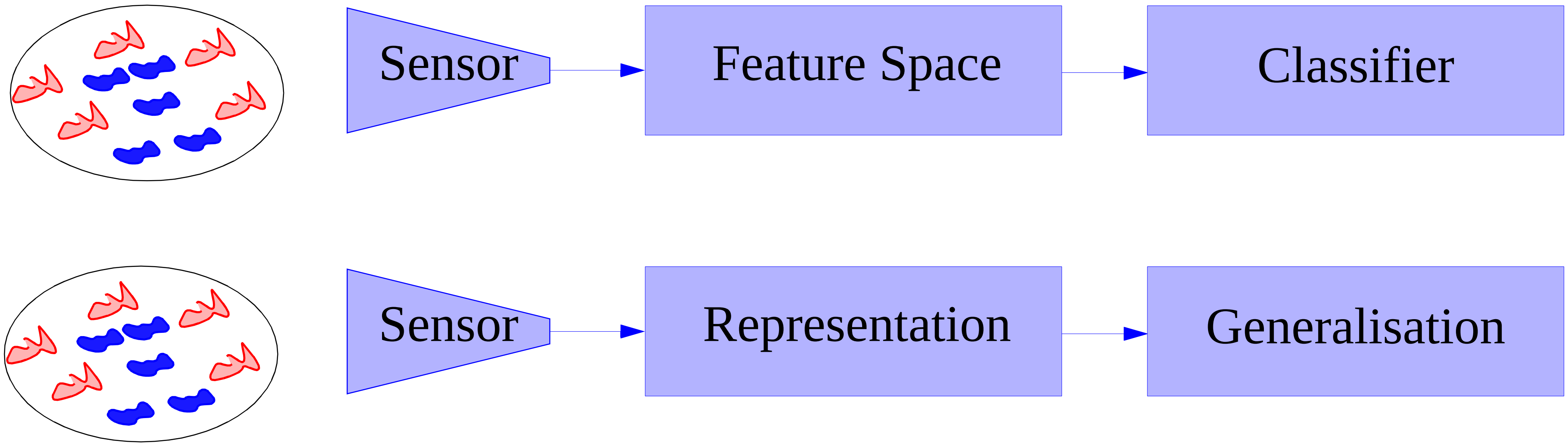
$\Rightarrow$ Representations

Dissimilarity Representations

Approaches

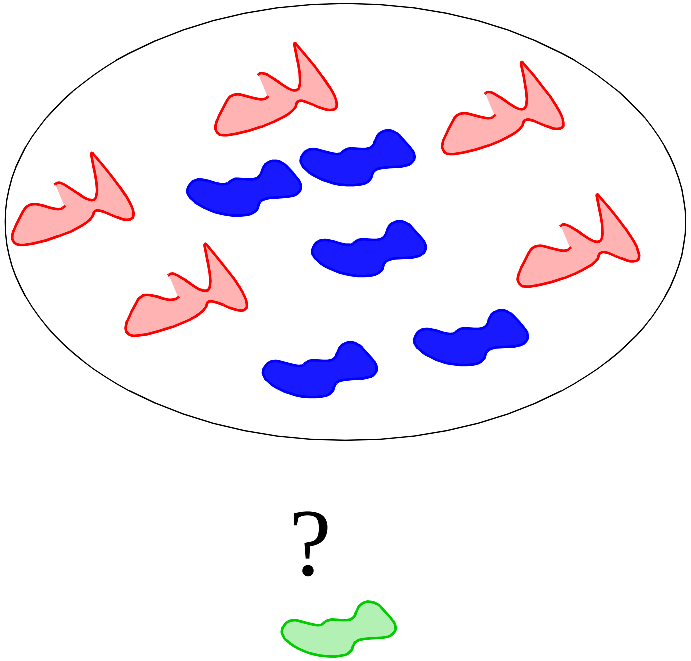
Examples

What is Pattern Recognition?



Other representations and generalisations?

What is Generalisation?



Generalisation is the ability to make statements on unknown properties of new objects (e.g. their class) on the basis of a set of examples.

This can be done on the basis of:

- (dis)similarities

- probabilities

- domains, decision functions

Demands for a Representation

The representation should enable generalisation:

Computation of:

dissimilarities, probabilities, domains, decision functions

Measure objects $\Rightarrow$ numbers

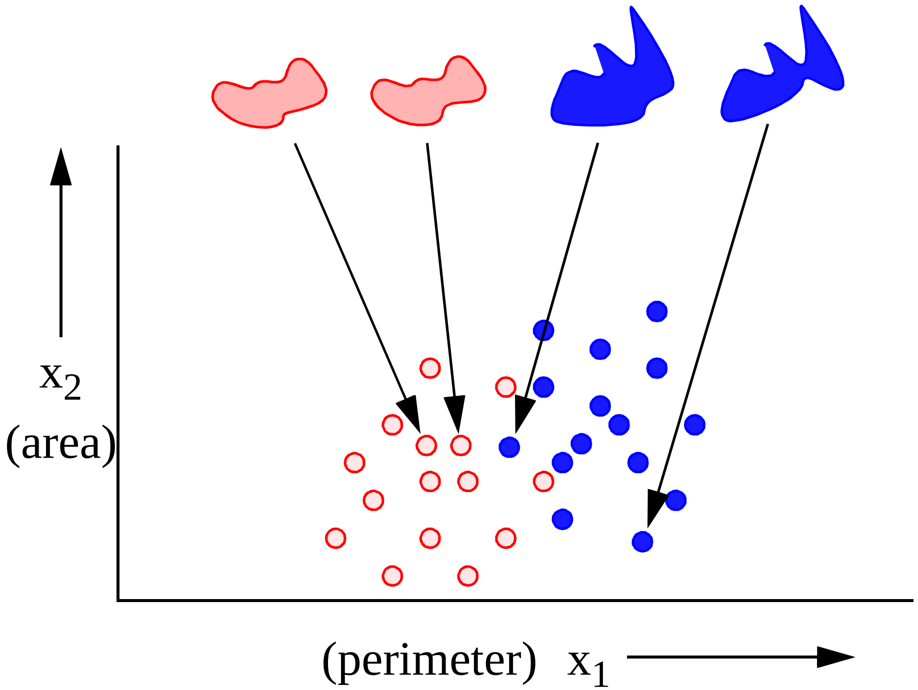
Compactness needed

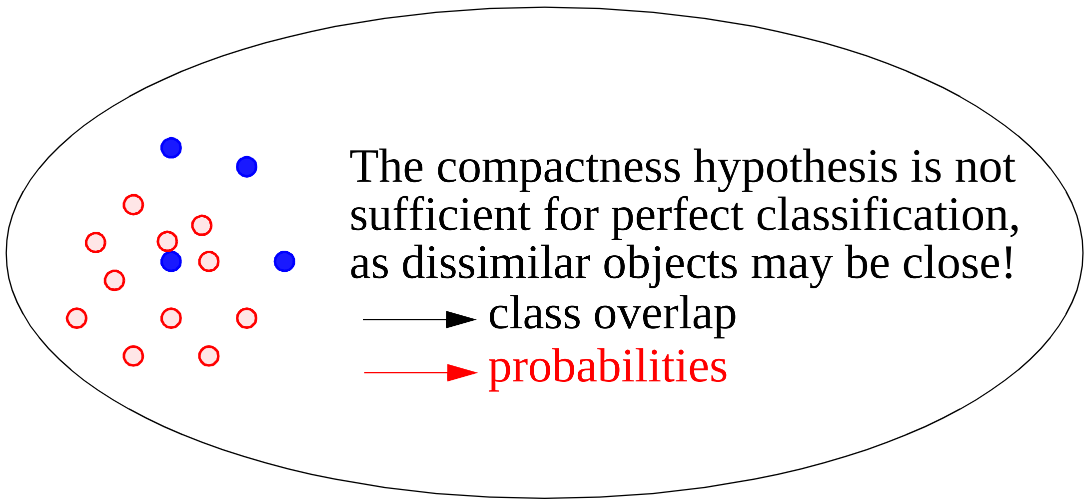
The Compactness Hypothesis

Representations of real world similar objects are close.

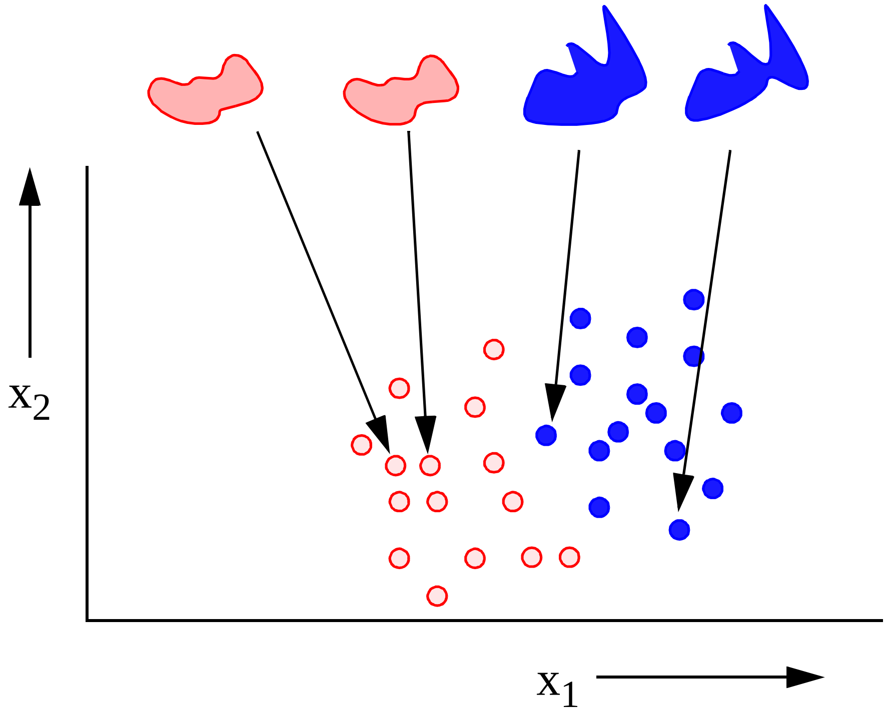
There is no ground for any generalisation (induction) on representations that do not obey this demand.

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





True Representations



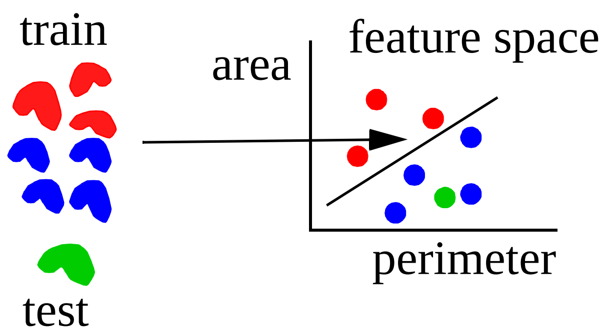
Similar object should be close and dissimilar objects should be distant

→ domains

Representation Principles

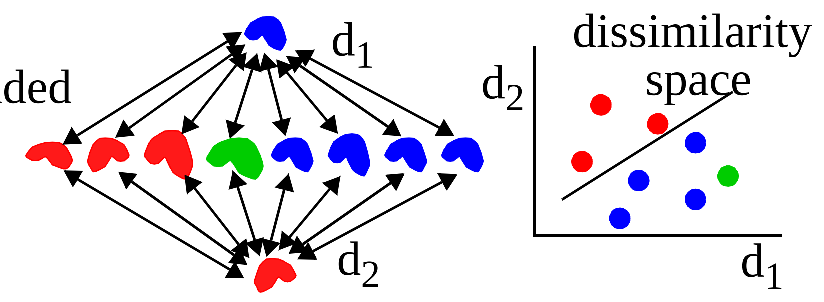
<u>Absolute</u>, by features of objects

distributions, connectivity neglected



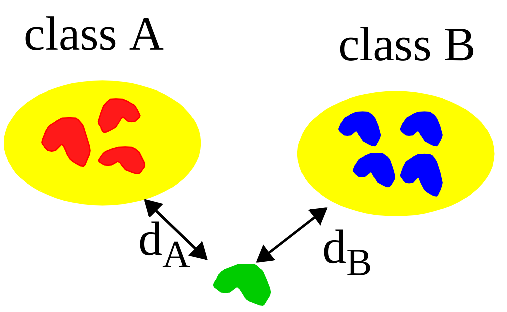
<u>Relative</u>, by dissimilarities between objects

distributions or domains, connectivity possibly included



<u>Conceptual</u>, by dissimilarities between objects and classes

domains, structural, connectivity included



Examples of Relative Representation

Distance between objects

in terms of measurements, in terms of models

Distance between an object and a set of objects

Distance between an object and a class

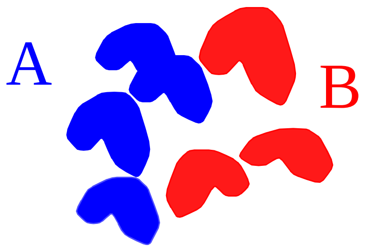
Distance between an object and a classifier

Related to combining classifiers

Contents

Dissimilarity Representation

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 Neighbour rule (template matching) just finds:

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

without using D_T. Can we do any better?

Dissimilarity - 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

Dissimilarity Spaces - Examples

Pre-Topological		~~continuity~~
Topological	Mahalanobis	~~definiteness~~
Pre-Metric	Lp, $p < 1$ weighted edit-distance	~~triangle inequality~~
Metric	L1	
Euclidean	L2, RMSE	

Compactness always needed

Why Dissimilarity Spaces?

Many (exotic) dissimilarity measures are used in pattern recognition

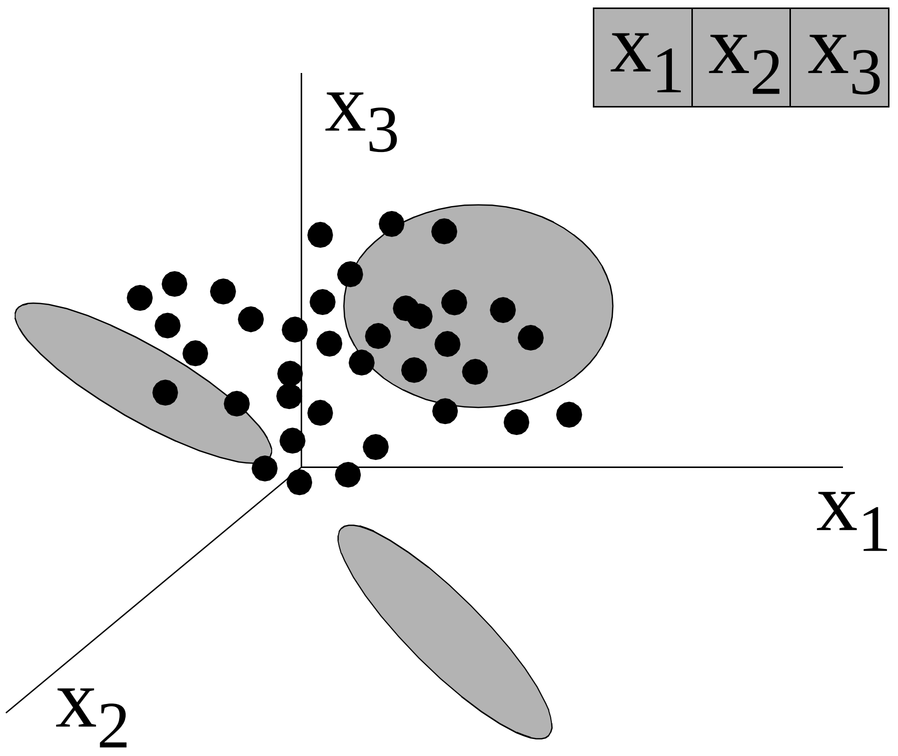
- they may solve the connectivity problem (e.g. pixel based features)
- they may offer a way to integrate structural and statistical approaches e.g. by graph distances.

Prospect of zero-error classifiers by avoiding class overlap

Better rules than the nearest neighbour classifier appear possible (more accurate, faster)

Problems with the Pixel_Feature Representation - 1

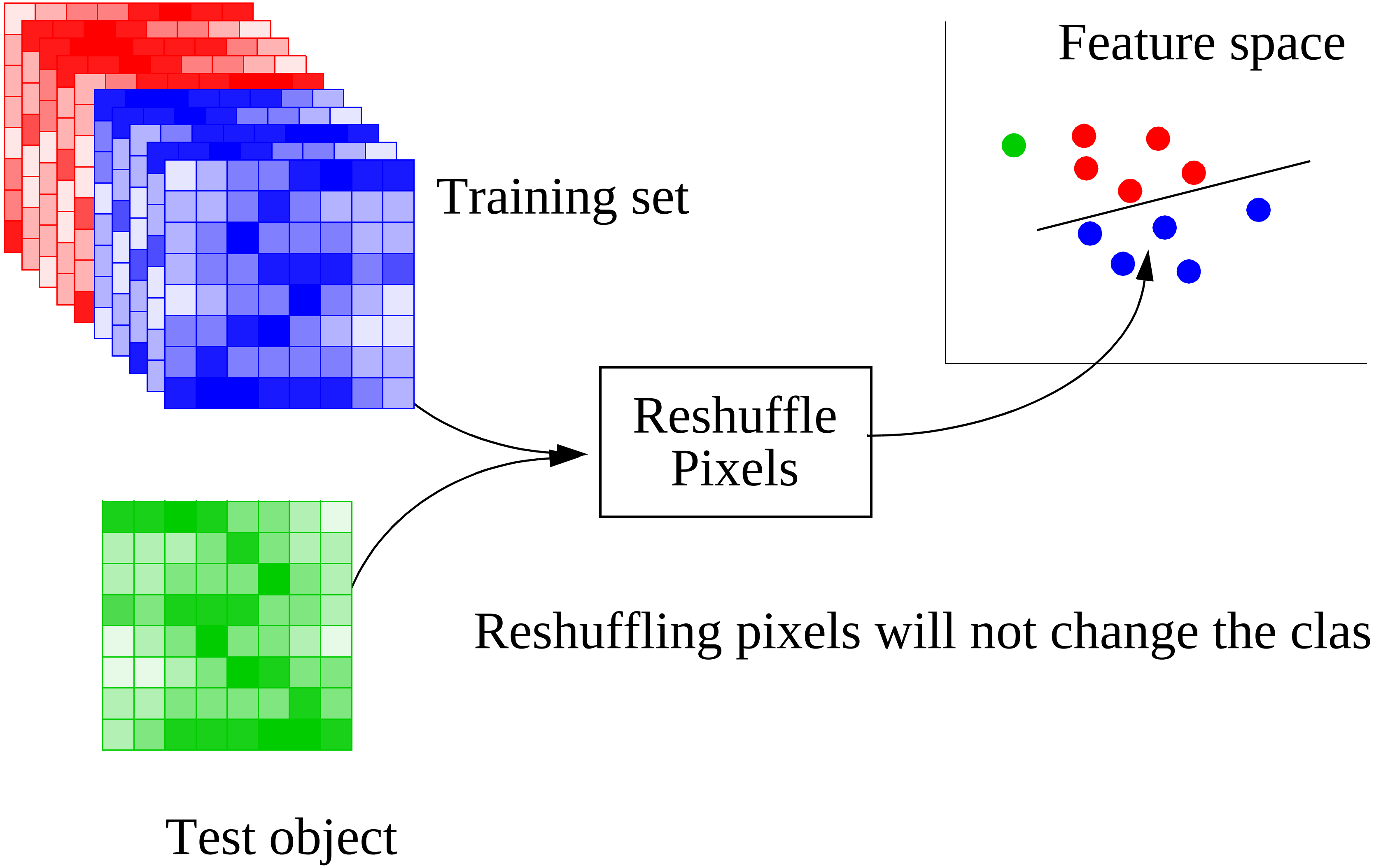
Spatial connectivity is lost



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

Problems with the Pixel_Feature Representation -2

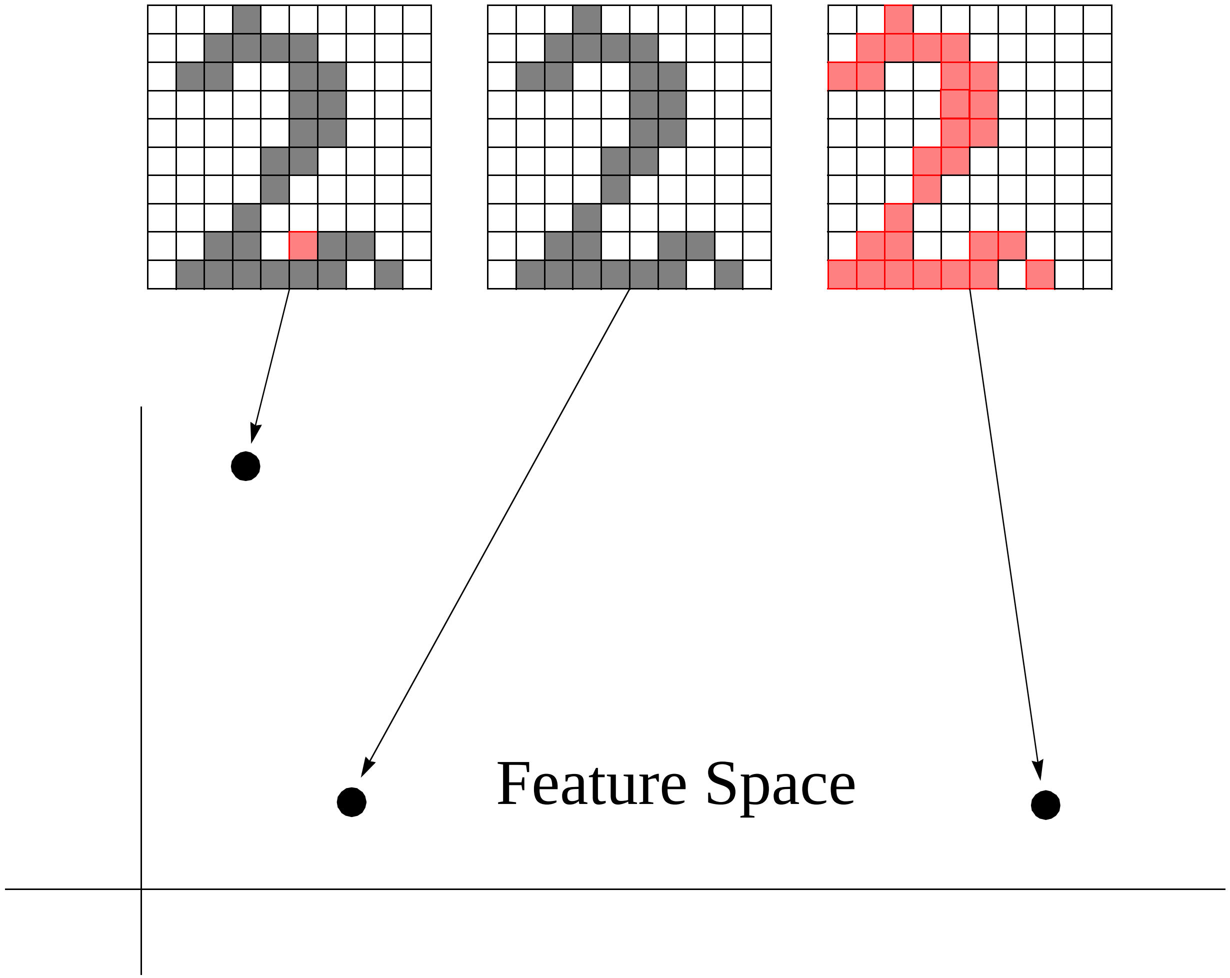
Spatial connectivity is lost



Reshuffling pixels will not change the classification

Problems with the Pixel_Feature Representation - 3

Representation jumps after small disturbances



Problems with the Pixel_Feature Representation - 4

Interpolation does not yield valid objects

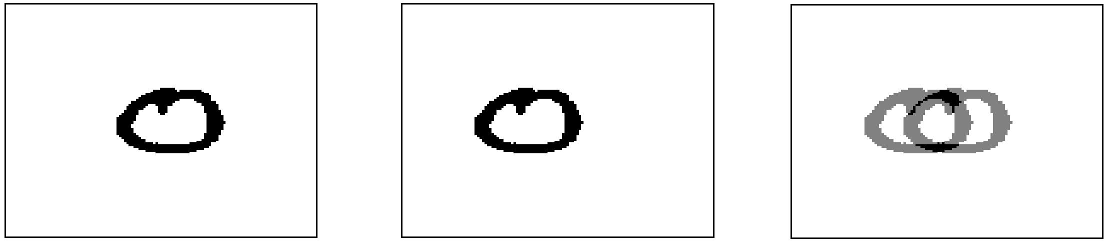
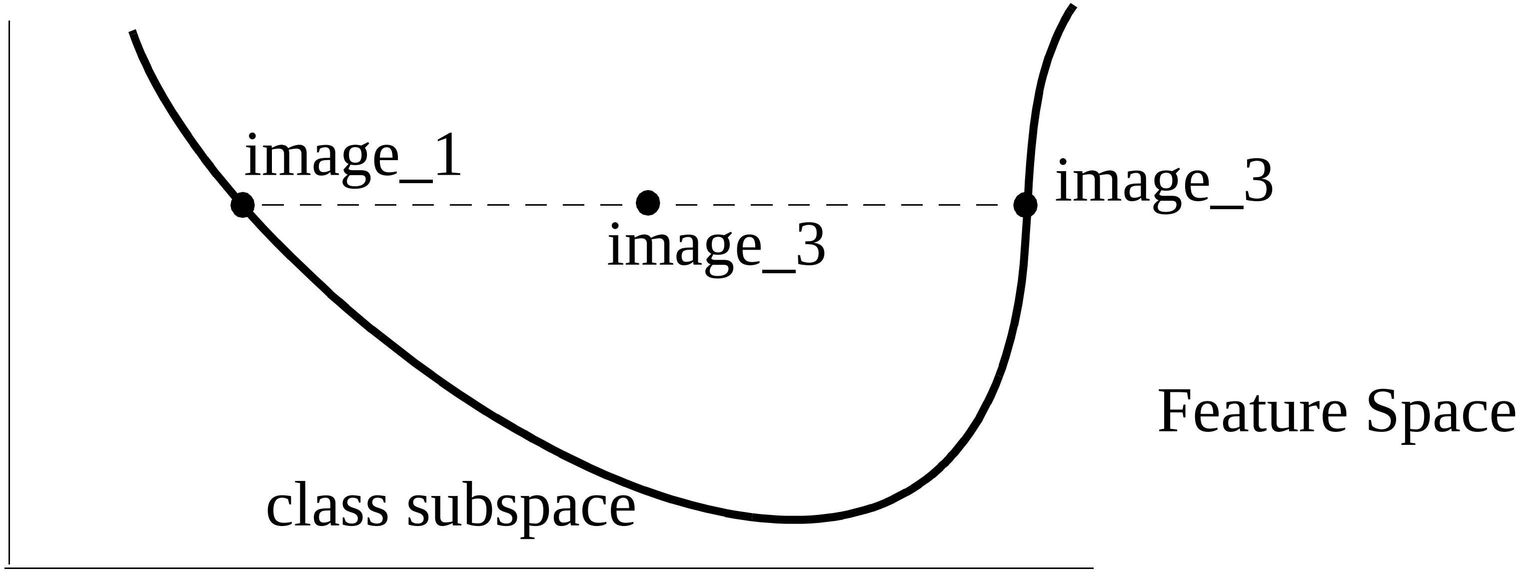
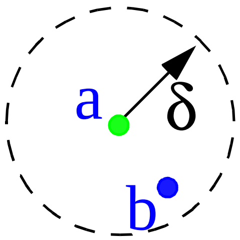


image_1 image_2 image_3



The Prospect of Dissimilarity based Representations: Zero Error



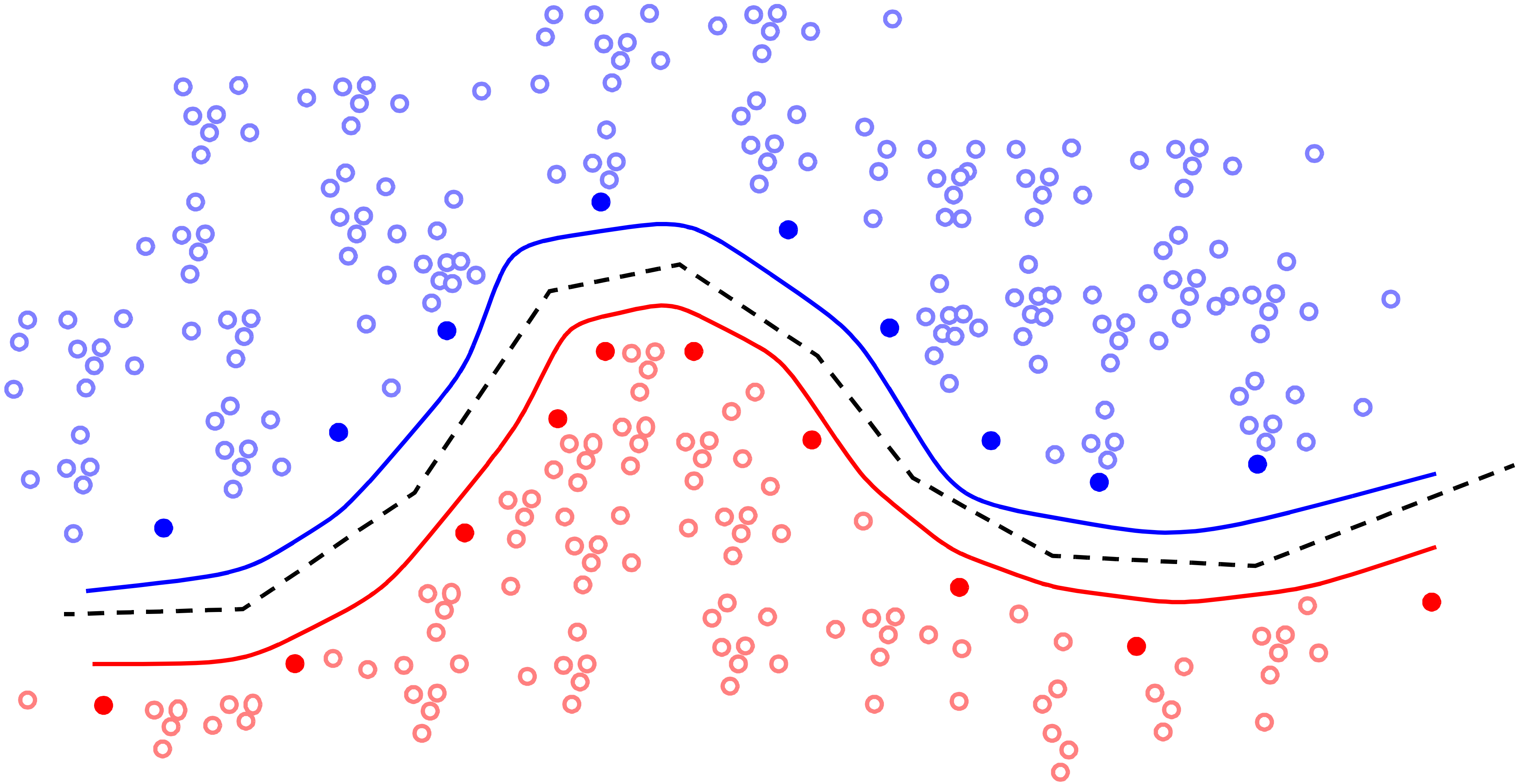
Let us assume that we deal with true representations:

$d_{ab} < \delta$ if and only if the objects a and b are very similar.

If δ is sufficiently small than a and b belong to the same class, as b is just a minor distortion of a (Assuming true representations).

However, as Prob(b) > 0, there will be such an object for sufficiently large training sets $\rightarrow$ zero classification error possible!

Zero-Error Classification



Contents

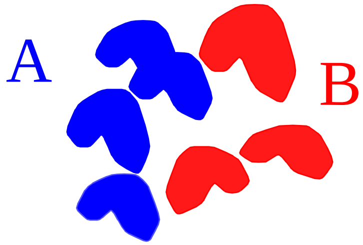
Representations

Dissimilarity Representations

$\Rightarrow$ Approaches

Examples

Approaches



Given labeled training set T

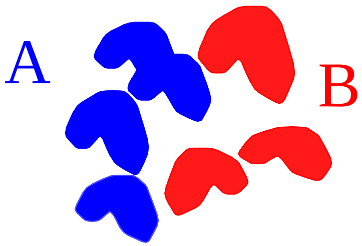
$$D_T = \begin{pmatrix} d_{11} & d_{12} & d_{13} & d_{14} & d_{15} & d_{16} & d_{17} \\ d_{21} & d_{22} & d_{23} & d_{24} & d_{25} & d_{26} & d_{27} \\ d_{31} & d_{32} & d_{33} & d_{34} & d_{35} & d_{36} & d_{37} \\ d_{41} & d_{42} & d_{43} & d_{44} & d_{45} & d_{46} & d_{47} \\ d_{51} & d_{52} & d_{53}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

Unlabeled object x to be classified

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

1. Nearest neighbour : find min(d_x)
2. Dissimilarity representation: use d_x as a feature vector.
3. Embedding: find a feature space for which D_T is correct

Approach 1: Nearest Neighbour Rule



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

Approach 2: Dissimilarity Representation

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

Consider dissimilarities as ’features’

$\Rightarrow$ n objects given by n features $\Rightarrow$ overtrained

$\Rightarrow$ select ’features’, i.e. representation objects, by

- regularisation
- systematic selection
- random selection

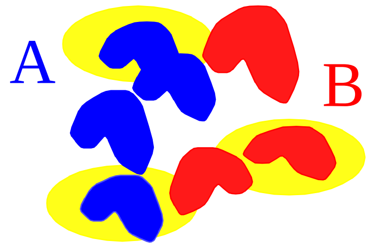
Approach 2: Dissimilarity Representation

Dissimilarities

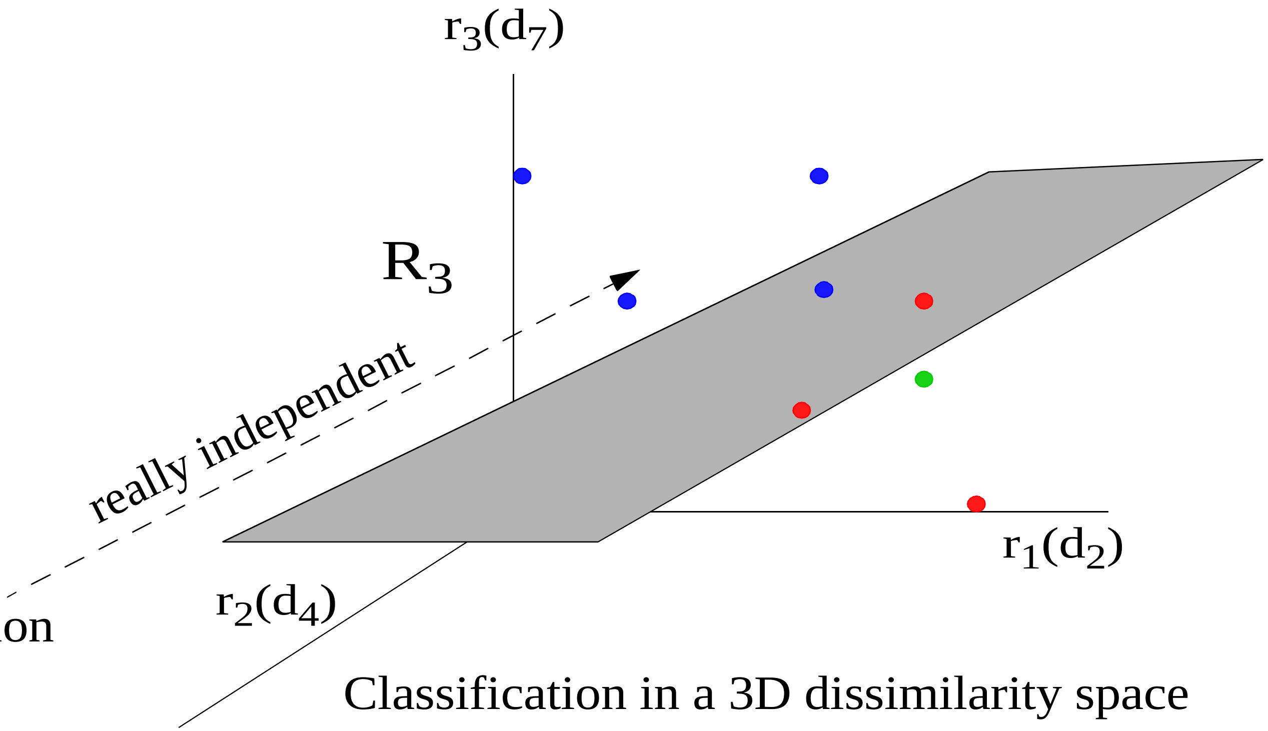
$$
\begin{array}{cc}
 & \begin{array}{ccccccc} & r_1 & & r_2 & & & r_3 \end{array} \\
D_T = & \begin{pmatrix}
d_{11} & d_{12} & d_{13} & d_{14} & d_{15} & d_{16} & d_{17} \\
d_{21} & d_{22} & d_{23} & d_{24} & d_{25} & d_{26} & d_{27} \\
d_{31} & d_{32} & d_{33} & d_{34} & d_{35} & d_{36} & d_{37} \\
d_{41} & d_{42} & d_{43} & d_{44} & d_{45} & d_{46} & d_{47} \\
d_{51} & d_{52} & d_{53} & d_{54} & d_{55} & d_{56} & d_{57} \\
d_{61} & d_{62} & d_{63} & d_{64} & d_{65} & d_{66} & d_{67} \\
d_{71} & d_{72} & d_{73} & d_{74} & d_{75} & d_{76} & d_{77}
\end{pmatrix} \\
d_x = & (d_1 \; d_2 \; d_3 \; d_4 \; d_5 \; d_6 \; d_7)
\end{array}
$$

Selection of 3 objects for representation

Given labeled training set T

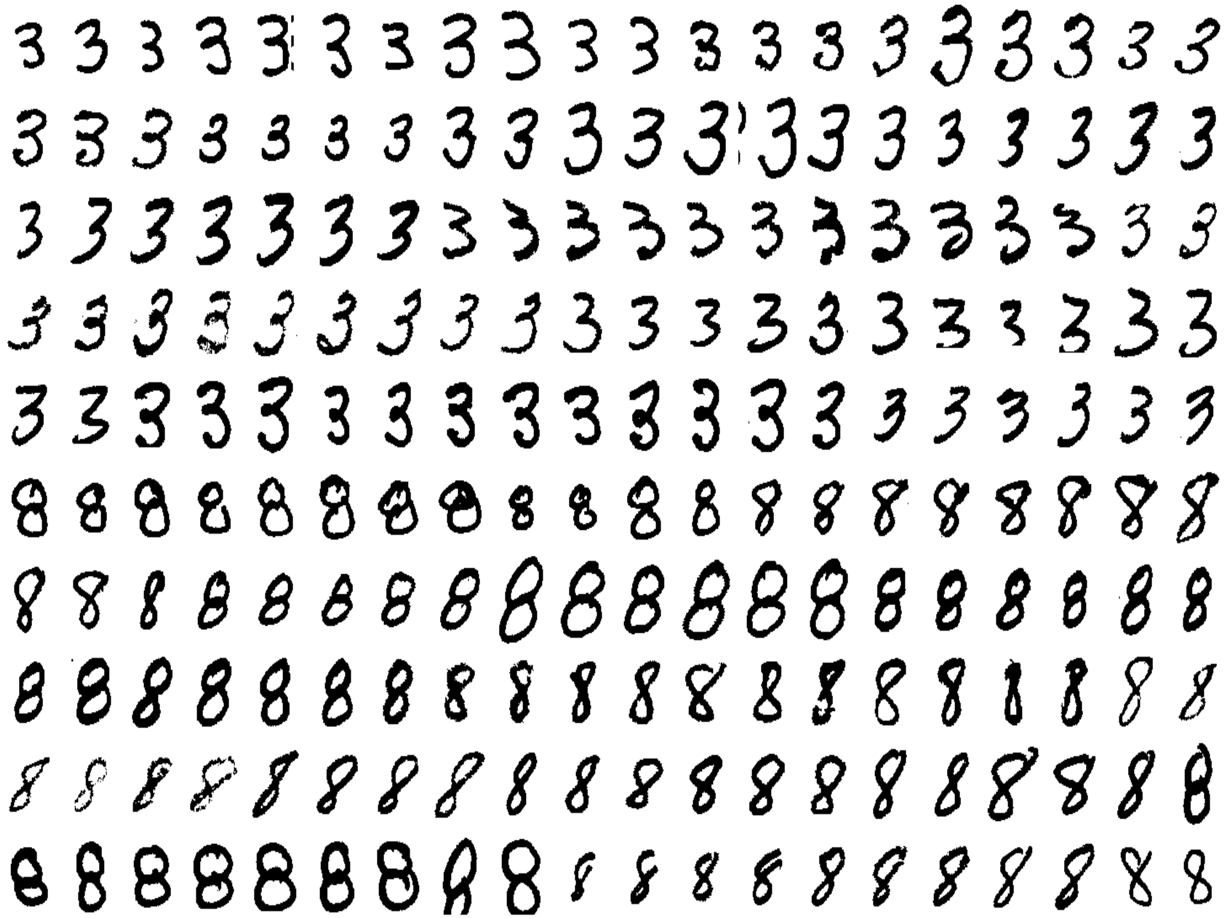


Unlabeled object x to be classified



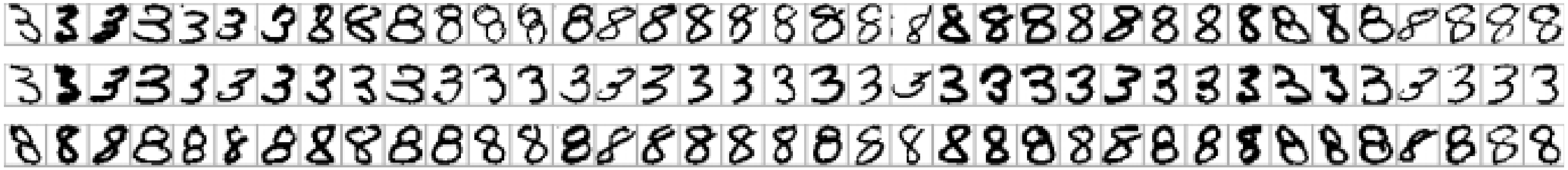
Classification in a 3D dissimilarity space

Example Dissimilarity Representation: NIST Digits 3 and 8



Examples of the raw data

Nearest Neighbour Errors



All digits ’3’ and ’8’ with an incorrect nearest neighbour.

Distance measure: Hamming distance in 32 x 32 resized images

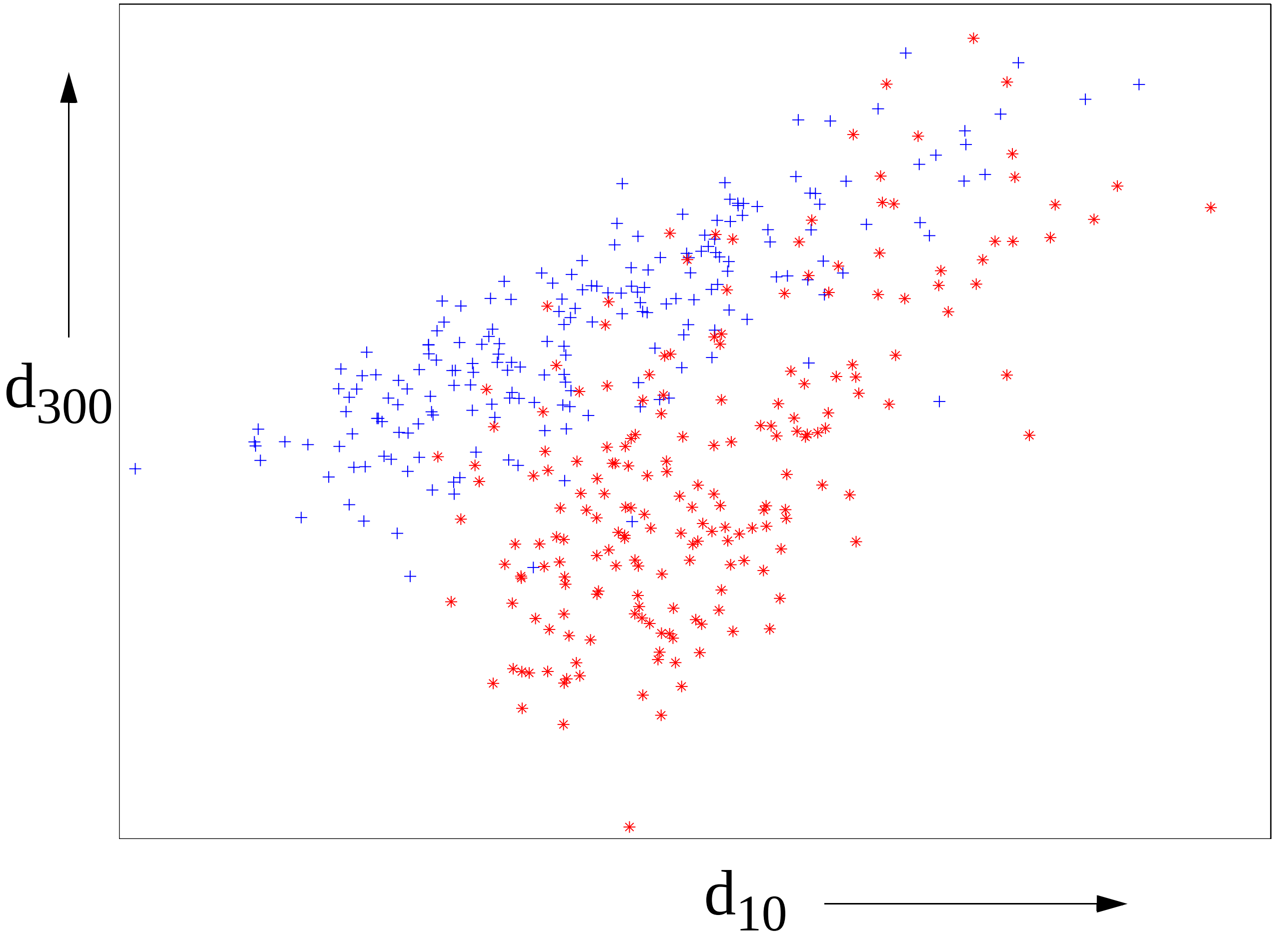
Row 1: the incorrectly classified objects

Row 2: the nearest neighbour in class ’3’

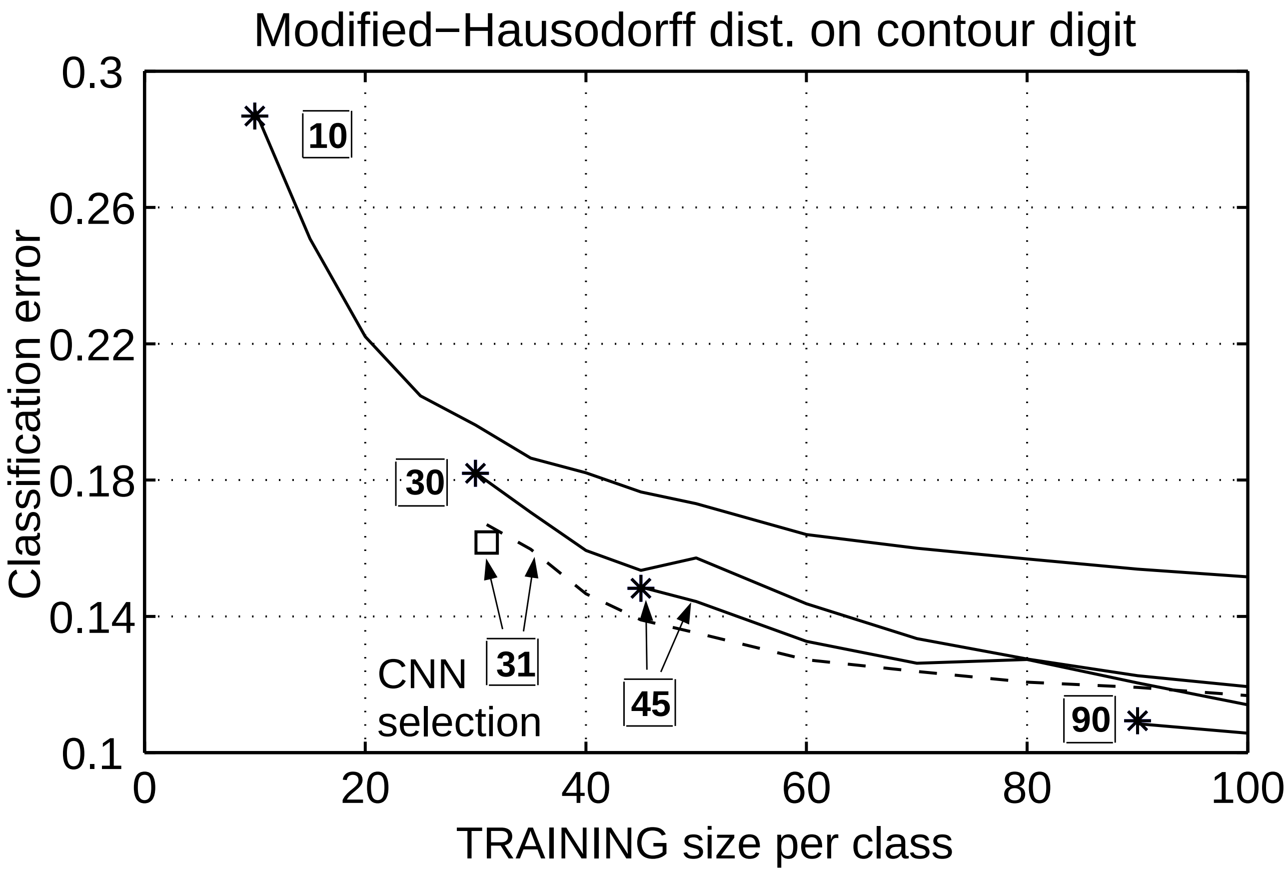
Row 3: the nearest neighbour in class ’8’

Example of Dissimilarity Dissimilarity Space

NIST digits: Hamming distances of 2 x 200 digits



Dissimilarity Space based Classification ⇔ Nearest Neighbour Rule



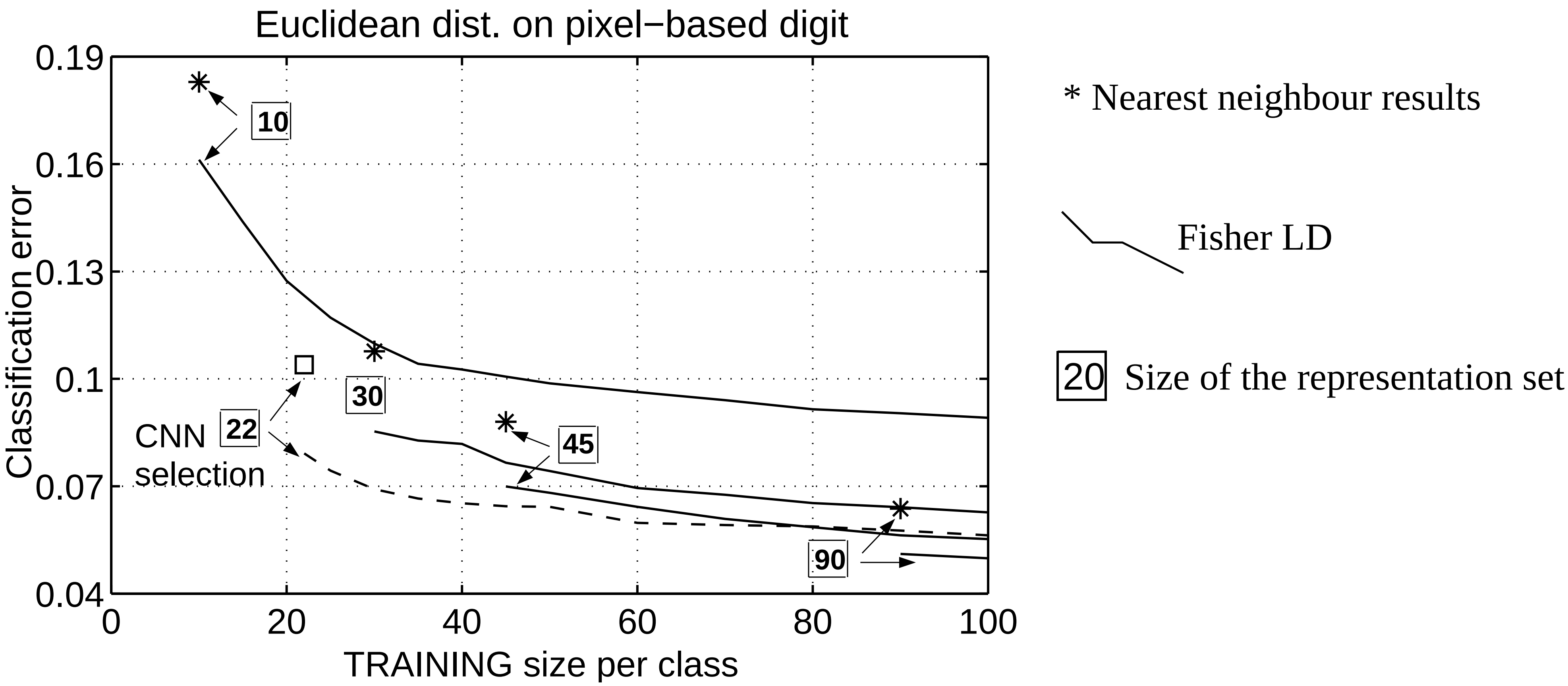
* Nearest neighbour results

Fisher LD

20 Size of the representation set

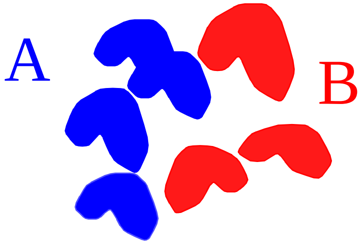
Dissimilarity based Classification outperforms Nearest neighbour Rule

Dissimilarity Space based Classification ⇔ Nearest Neighbour Rule



Dissimilarity based Classification outperforms Nearest Neighbour Rule

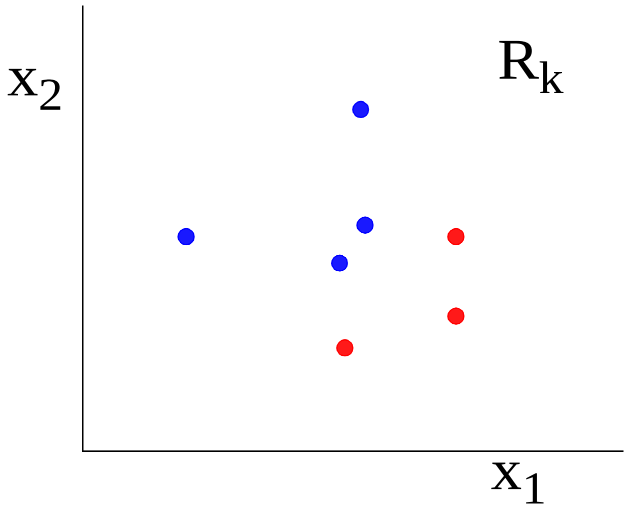
Approach 3: Embedding the Dissimilarity Representation



Training set

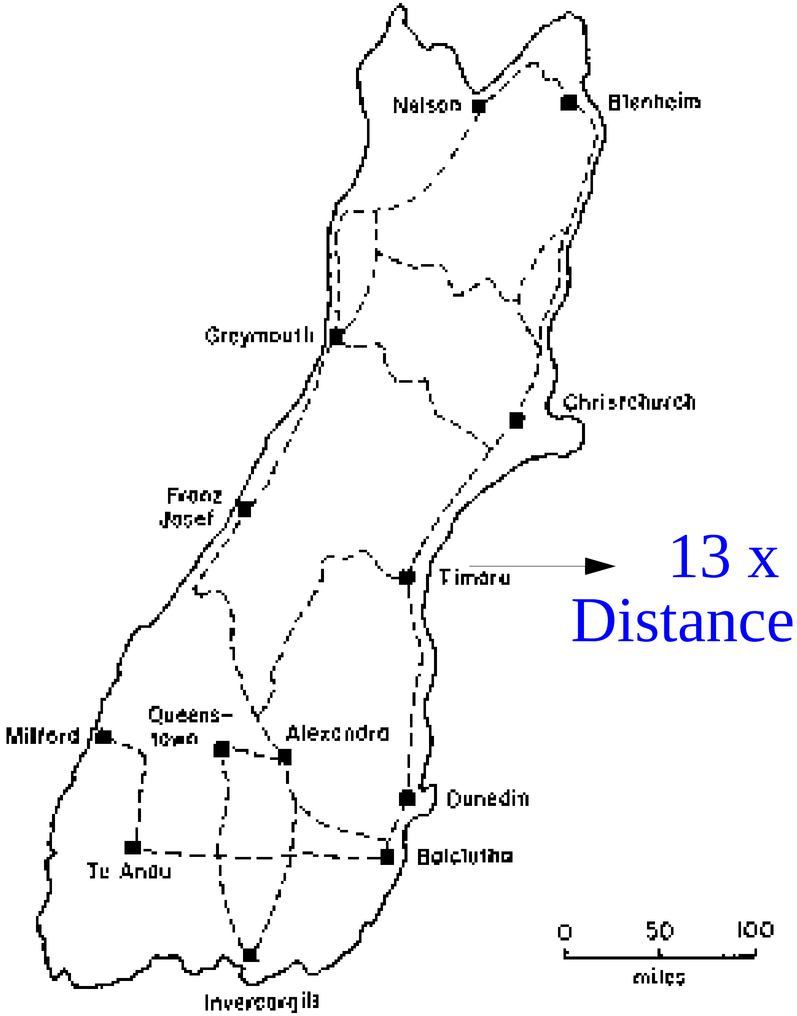
→ Dissimilarity matrix D → X?

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

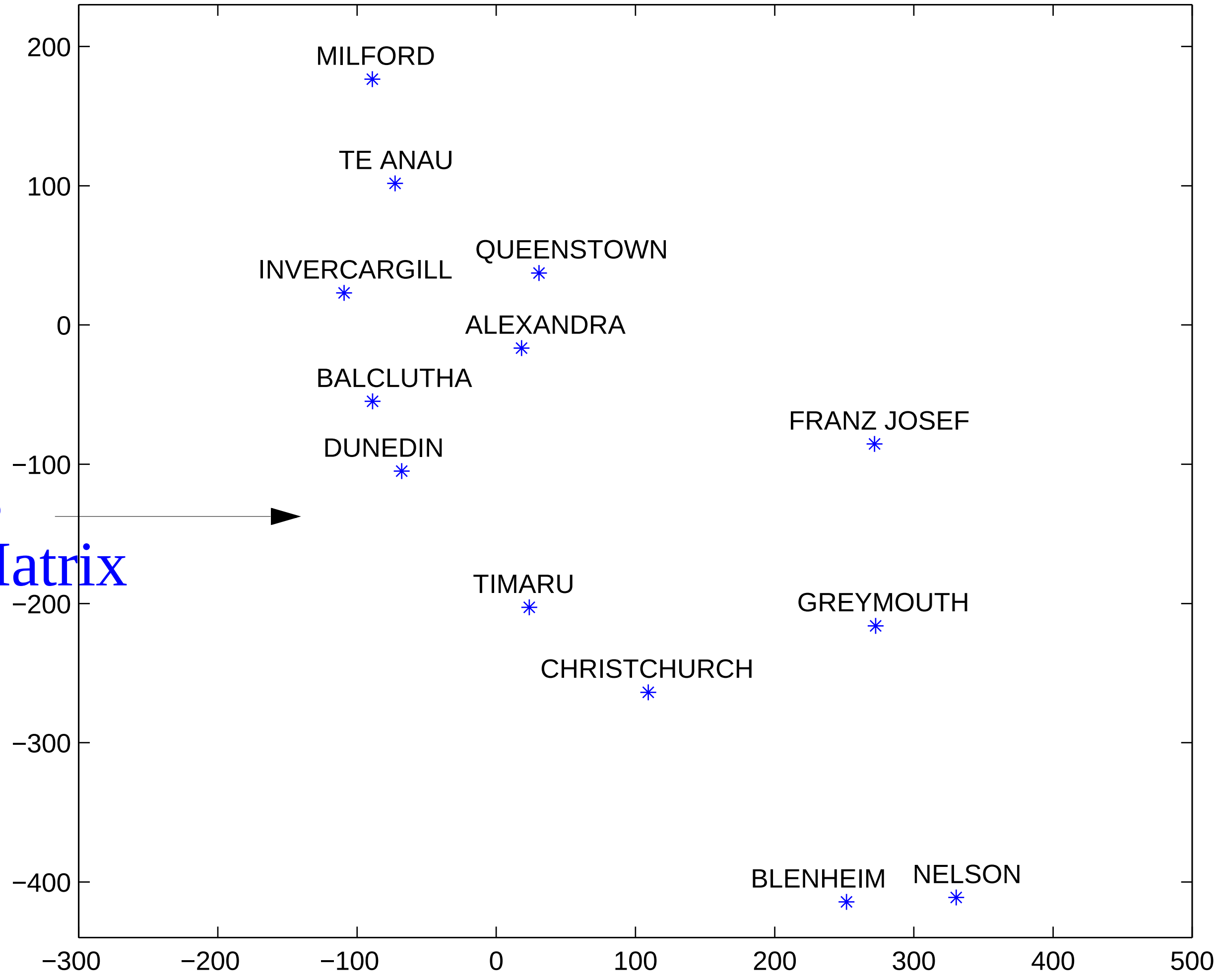


Euclidean distances → D

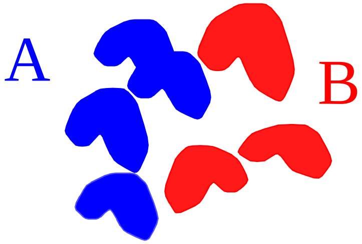
Euclidean Embedding: Multidimensional Scaling (MDS)



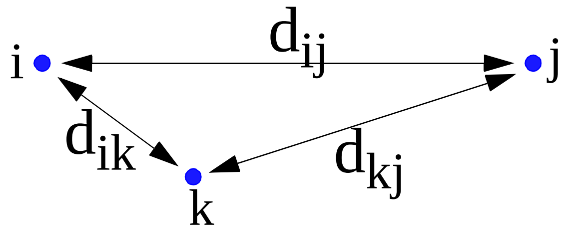
13 x 13
Distance Matrix



(Linear) Euclidean Embedding



Training set → Dissimilarity matrix D



If the dissimilarity matrix cannot be explained from a vector space,

e.g. Hausdorff and Hamming distance.

or if sometimes $d_{ij} > d_{ik} + d_{kj}$: triangulation inequality not satisfied.

embedding in Euclidean space not possible $\rightarrow$ Pseudo-Euclidean embedding

Pseudo-Euclidean Embedding

D is a given, imperfect dissimilarity matrix of training objects.

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

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

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

Let M_k be a k x k diag. matrix, $M_k(i,i) = \mathrm{sign}(\Lambda_k(i,i))$ $X = Q_k |\Lambda_k|^{\frac{1}{2}} M_k$

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

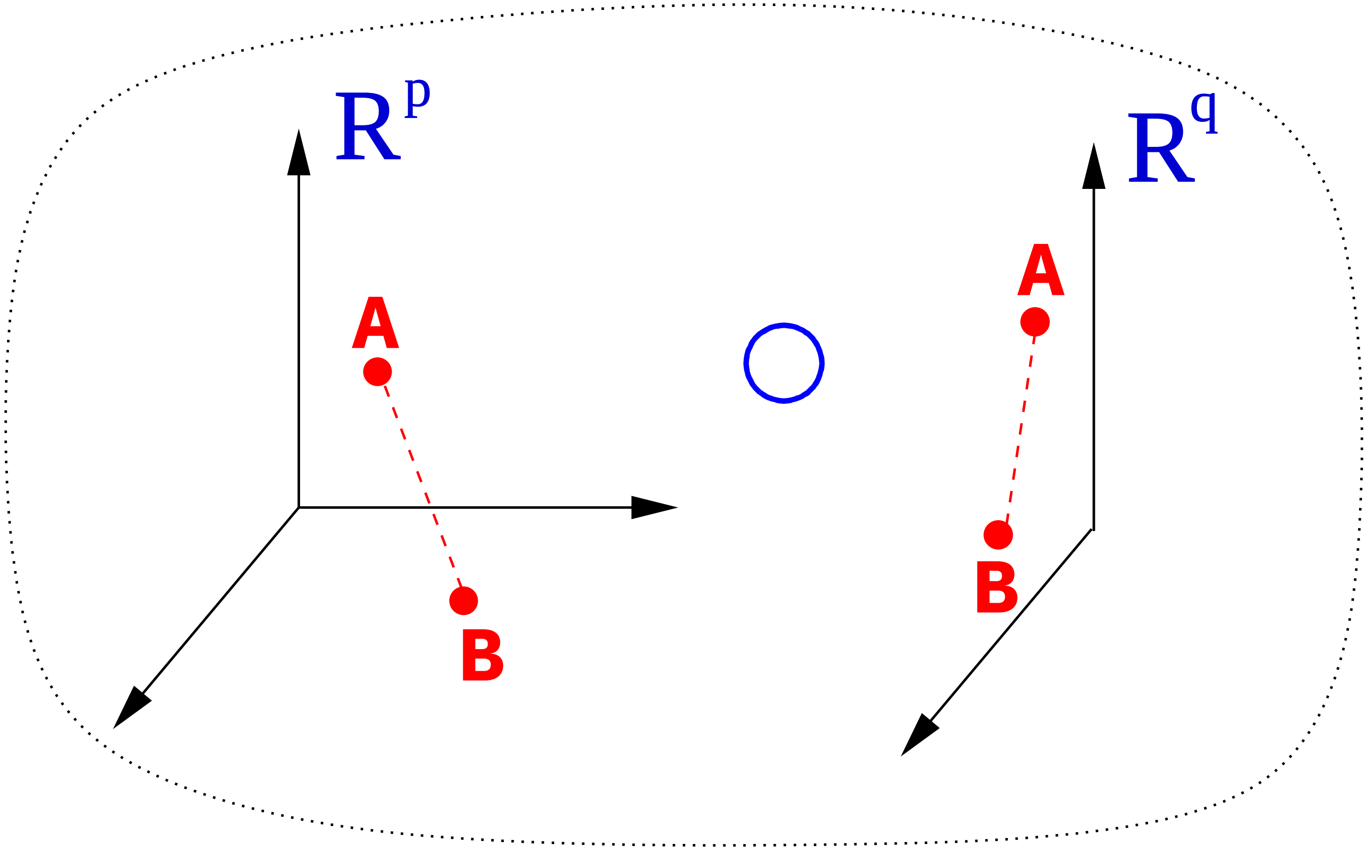
Let D_z be the dissimilarity matrix between new objects and the training set.

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

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

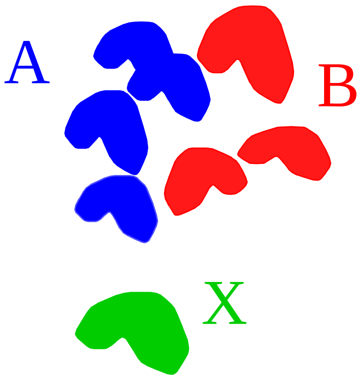
Distances in a Pseudo-Euclidean Space

$R^{p+q} =$



$$d^2_{R^{p+q}}(A,B) = d^2_{R^p}(A,B) - d^2_{R^q}(A,B)$$

Dissimilarity Based Classification

Training set 

Training set → Dissimilarity matrix D

Test object → Dissimilarities d_x with training set

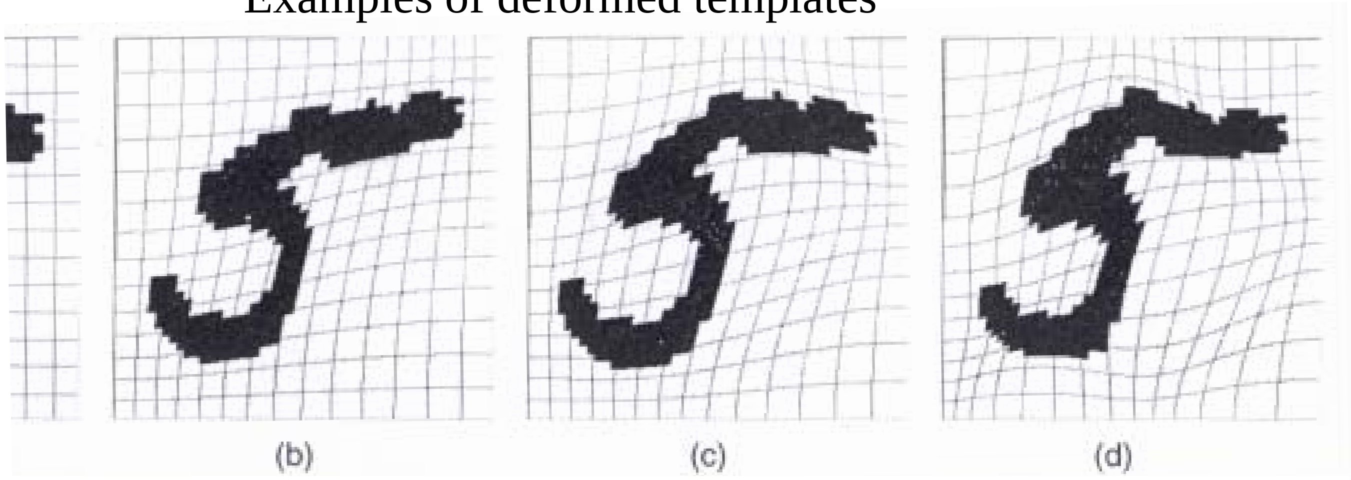
1. Nearest neighbour rule
2. Reduce training set to representation set
3. Embedding:

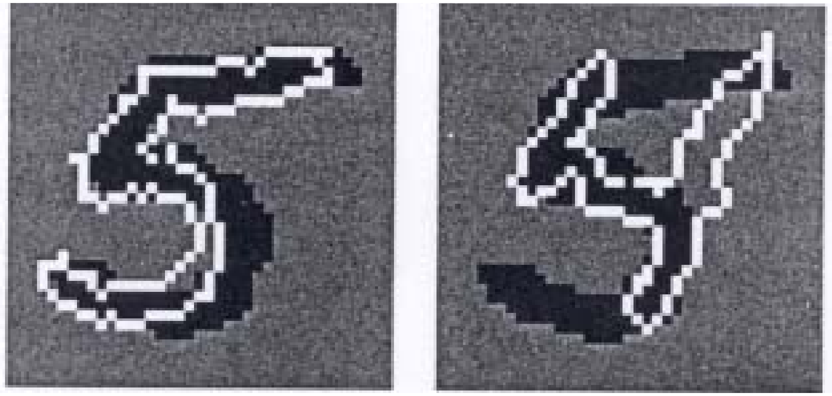
$$\left.\begin{array}{lll} & & \rightarrow \text{dissimilarity space} \\ \text{3. Embedding:} & \text{Select large } \Lambda_{ii} > 0 & \rightarrow \text{Euclidean space} \\ & \text{Select large } |\Lambda_{ii}| > 0 & \rightarrow \text{pseudo-Euclidean space} \end{array}\right\} \text{discriminant function}$$

Contents

Example 1: Zongker Data

Examples of deformed templates



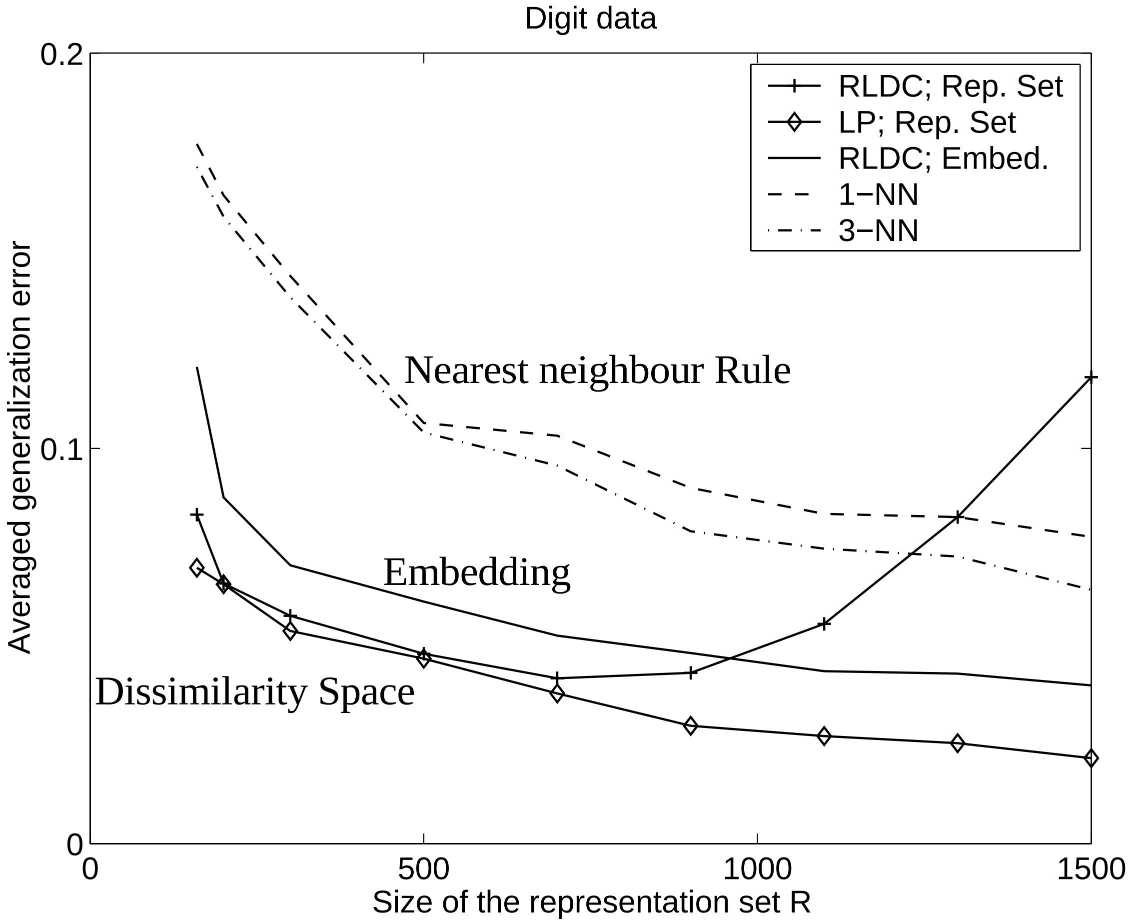


Matching new objects x to various templates y

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

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

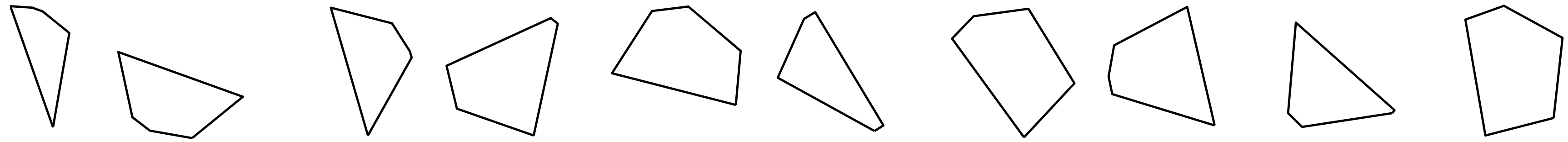
3 Approaches Compared for the Zongker Data



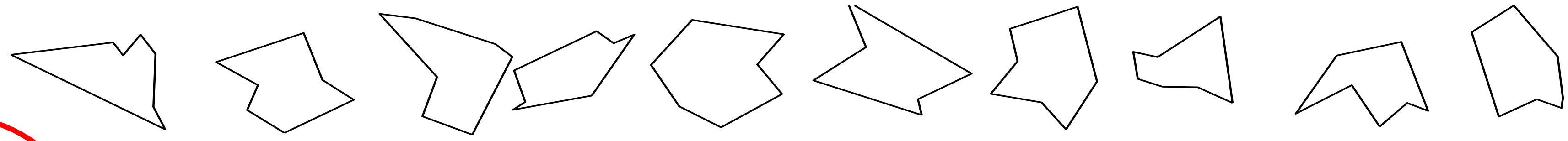
Dissimilarity Space better than Embedding better than Nearest Neighbour Rule

Example 2: Polygons

Convex
Pentagons:



Heptagons:



no class overlap
zero error

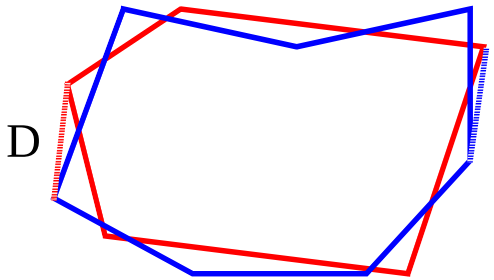
Minimum edge length: 0.1 of maximum edge length

Distance measures: Hausdorff $D = \max\{ \max_i(\min_j(d_{ij}))\ ,\ \max_j(\min_i(d_{ij}))\ \}$.

Modified Hausdorff $D = \max\{\text{mean}_i(\min_j(d_{ij}))\ ,\ \text{mean}_j(\min_i(d_{ij}))\ \}$. (no metric!)

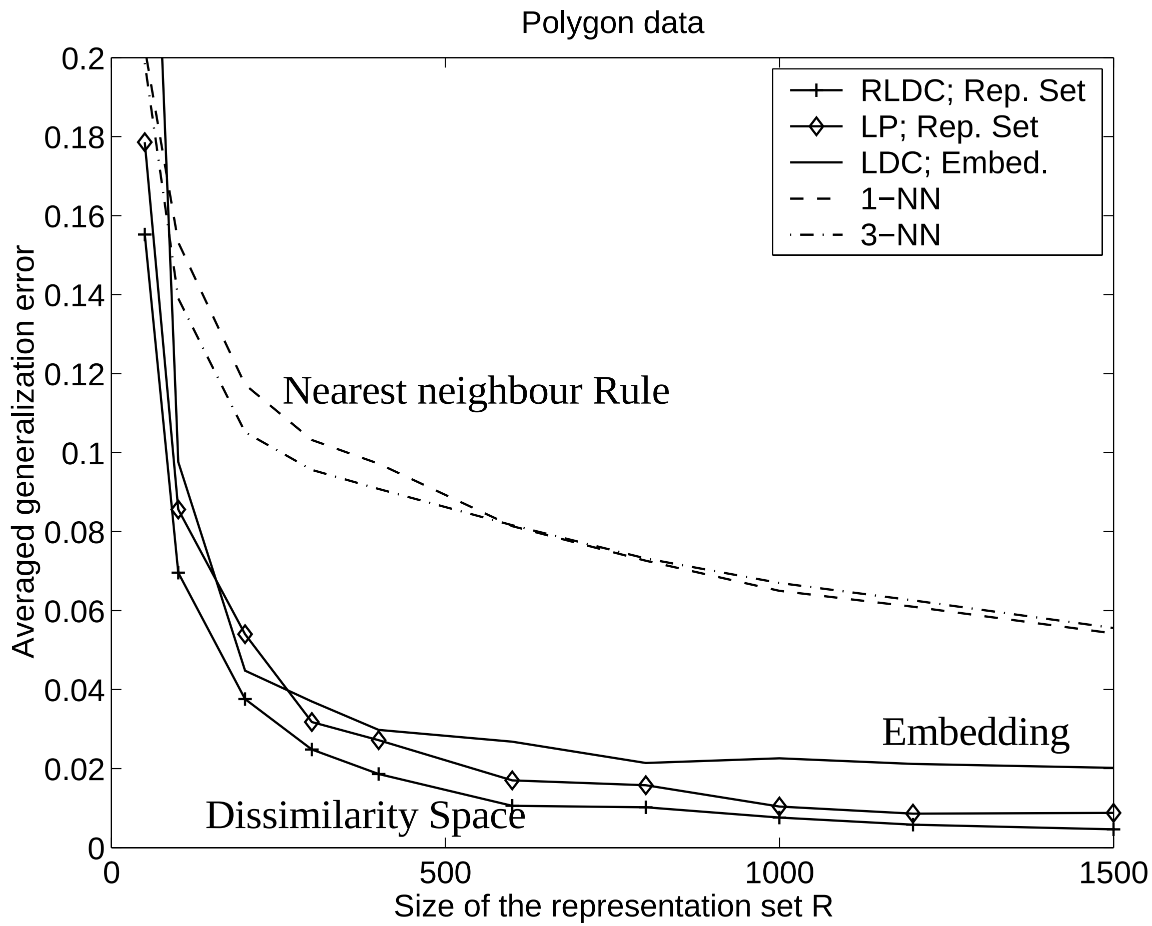
d_{ij} = distance between vertex i of polygon_1 and vertex j of polygon_2.

Polygons are scaled and centered.



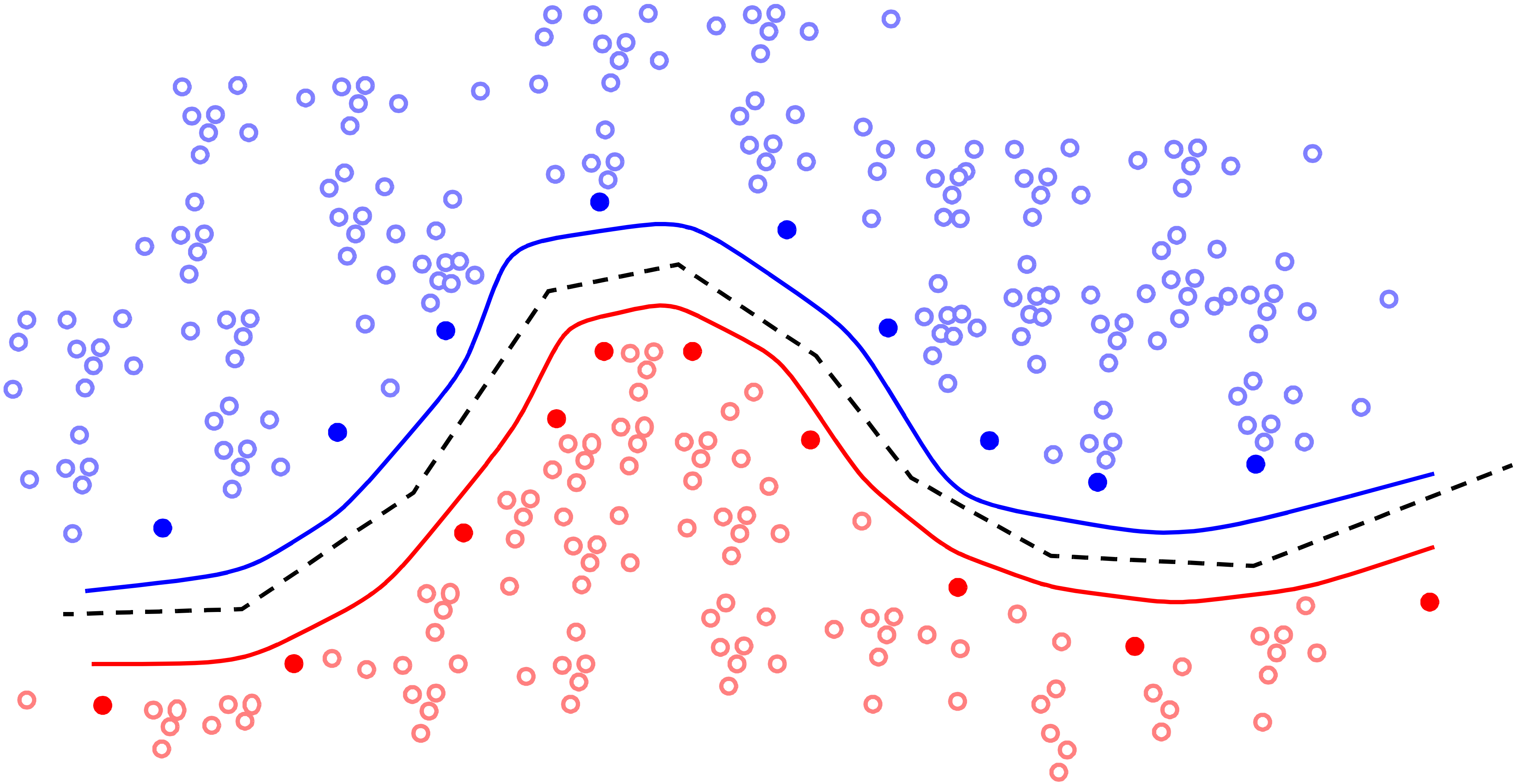
find the largest from the smallest vertex distance

Dissimilarity Based Classification of Polygons

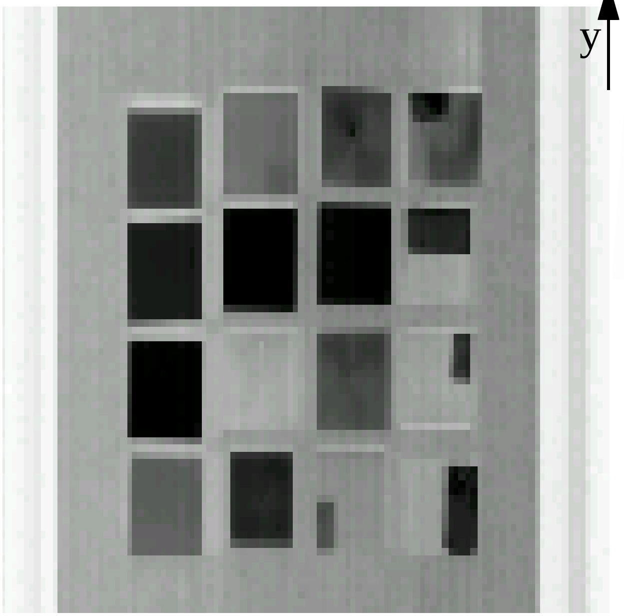


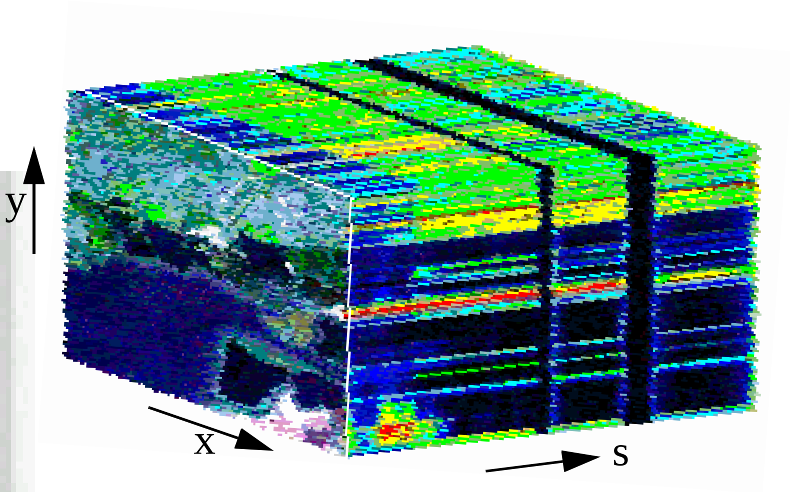
Zero error difficult to reach

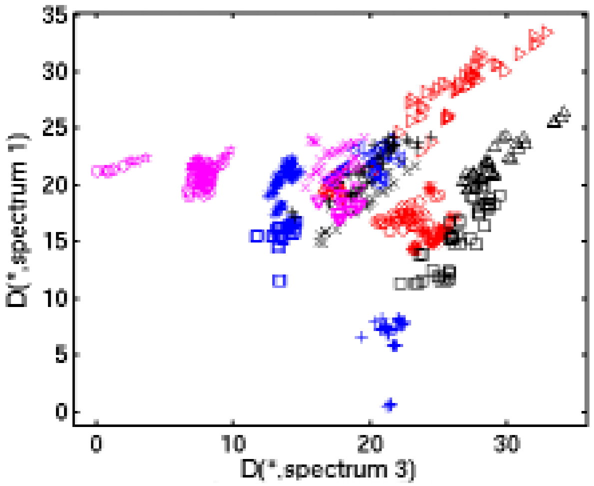
Zero-Error Classification by a Small Representation Set

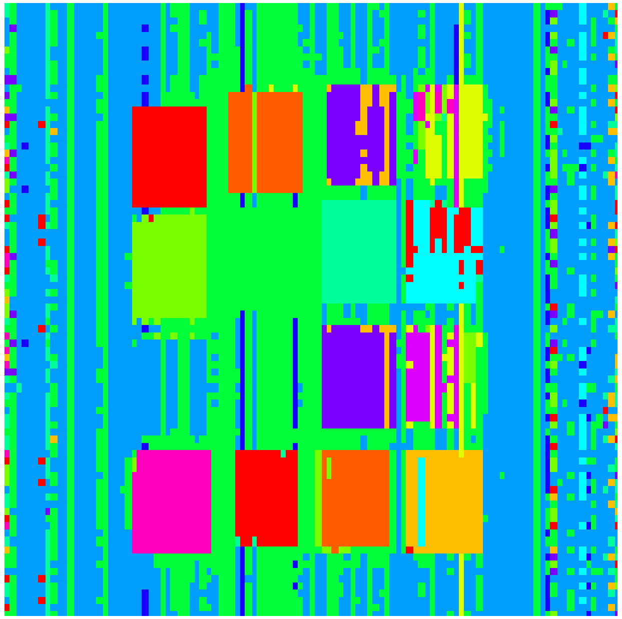


Example 3: Plastic Recognition by Hyperspectral Imaging

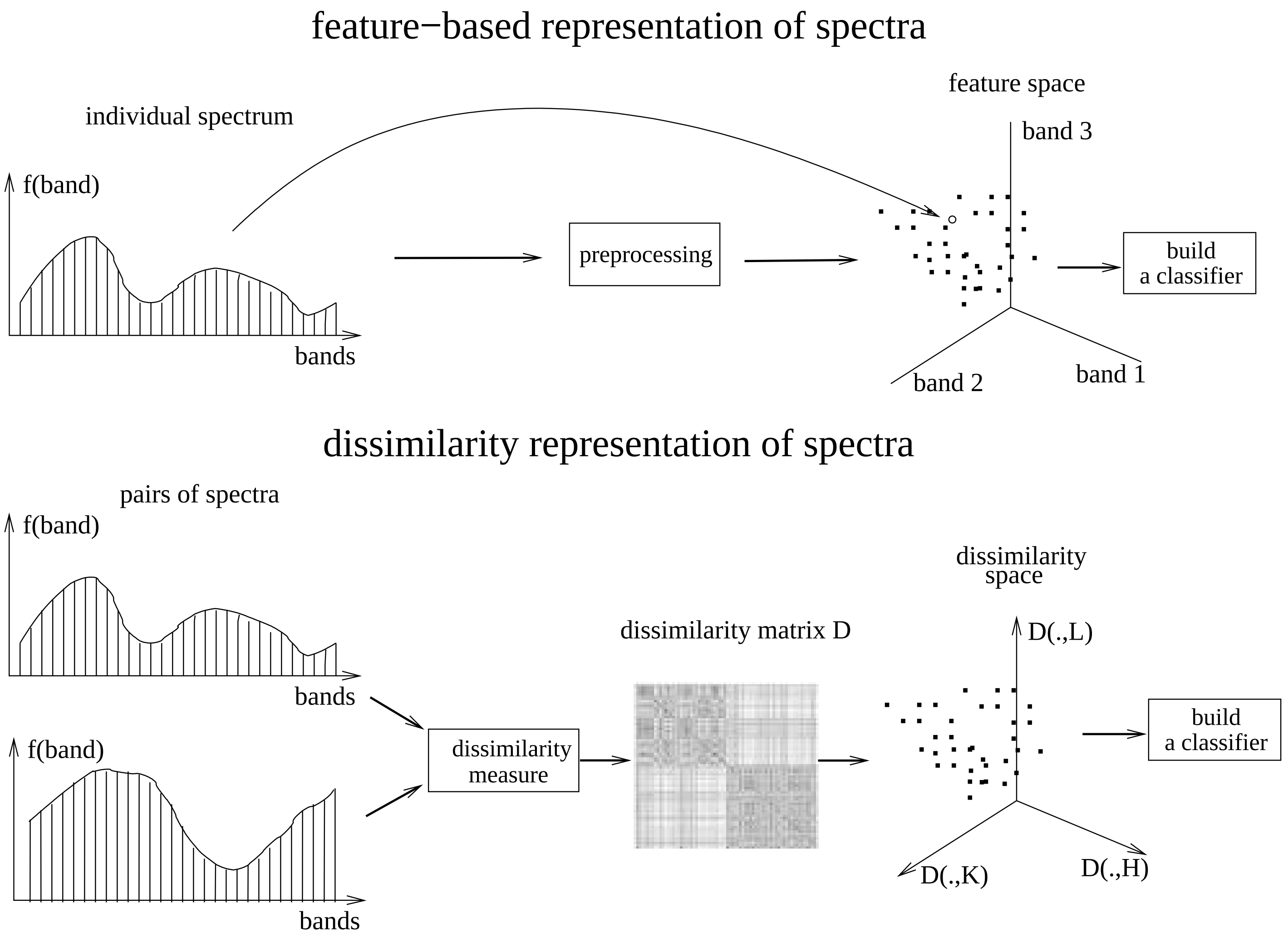




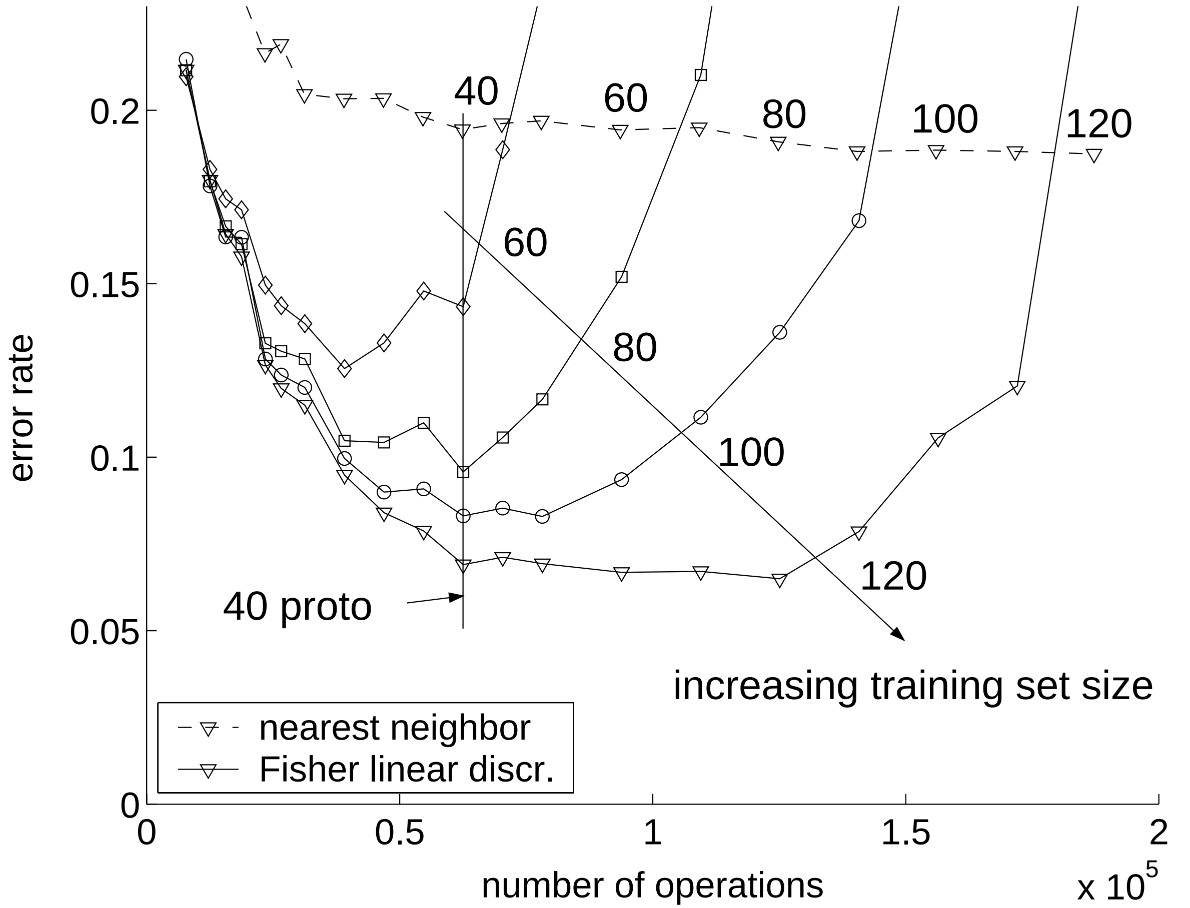




Classification of Spectra in Dissimilarity Space

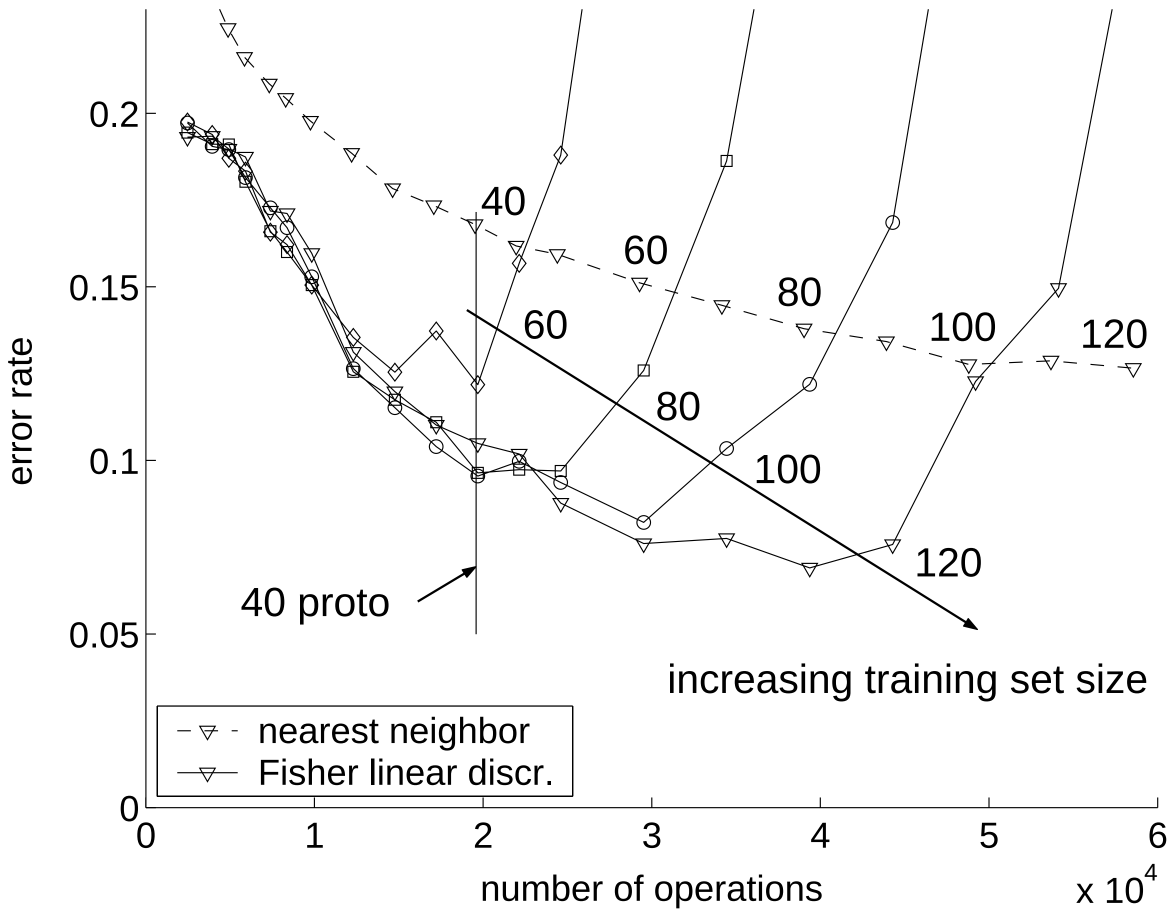


Results Dissimilarity Measure 1: Spectral Angle Mapper

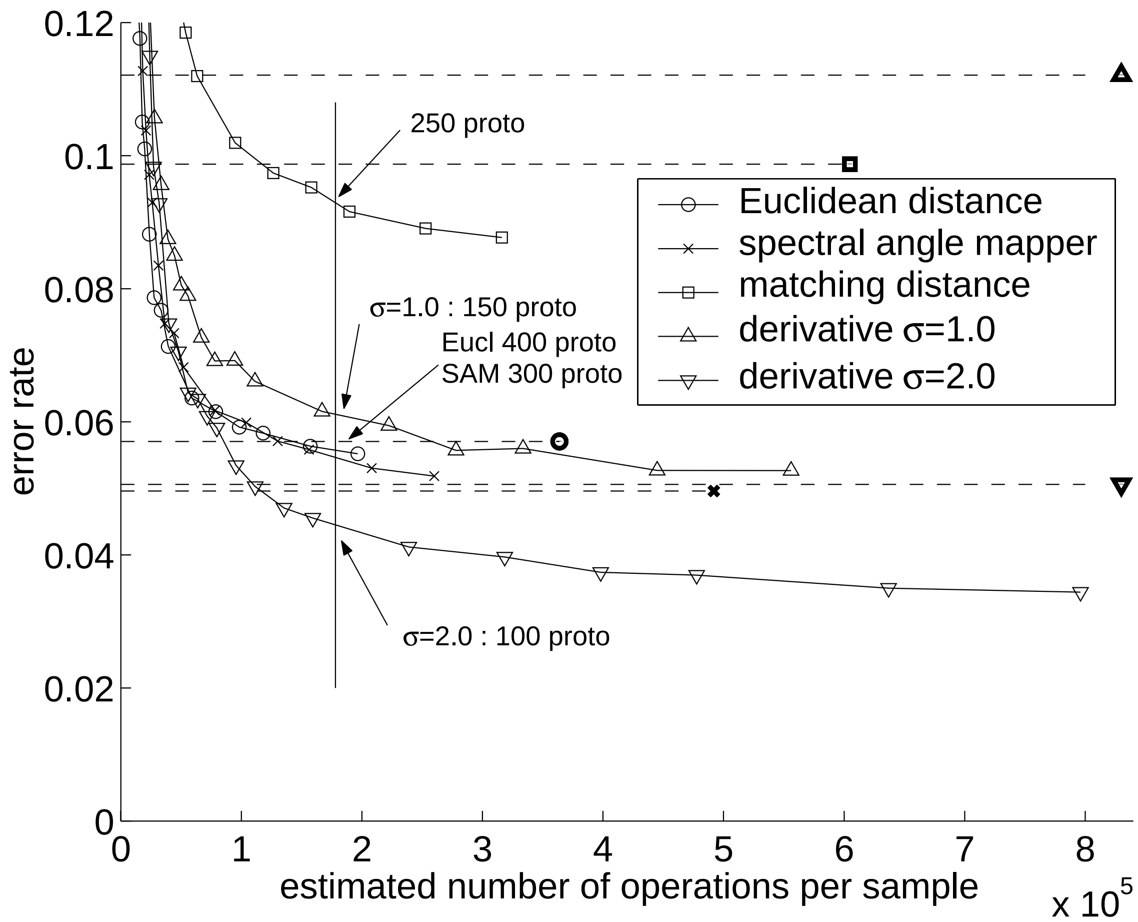


Better and faster than Nearest Neighbour

Results Dissimilarity Measure 2: Derivative Based Distances



Better and faster than Nearest Neighbour



Better and faster than Nearest Neighbour

Conclusions

Dissimilarity based representation is an alternative for features.

Classifier can be built in

- in dissimilarity spaces by selecting a representation set
- in (pseudo-)Euclidean spaces by embedding

Possible advantages:

- other type of expert knowledge (dissimilarities instead of features)
- larger training sets may compensate bad dissimilarity measures
- good performance, usually better than Nearest Neighbour !!
- zero-error is in principle possible, in practice very hard to achieve
- control of computational complexity

Finally

The study of representation

is a key area for pattern recognition