SSSPR 2006

Structural inference of sensor-based measurements

Hong Kong, August 2006

Robert P.W. Duin

Artsiom Harol, Thomas Landgrebe, Piotr Juszczak, Pavel Paclik, Elzbieta Pekalska,

Dick de Ridder, Marina Skurichina, David M.J. Tax, Serguei Verzakov

Delft University of Technology

The Netherlands

1

Structural Inference

How can we learn from structure?

Learning : Input + knowledge → more knowledge

Input : Examples, observations

Learning from examples : Generalization

Learning from structure : Generalization from structure

Concepts versus Observations

Plato:

The concepts live somewhere inside us. We awake them by our observations.

It is our struggle to interpret new observations in terms of existing concepts.

Aristotle:

We learn the world by observing it. In this process we struggle with creating concepts, i.e. generalizations of observations.

Structural versus Statistical Pattern Recognition

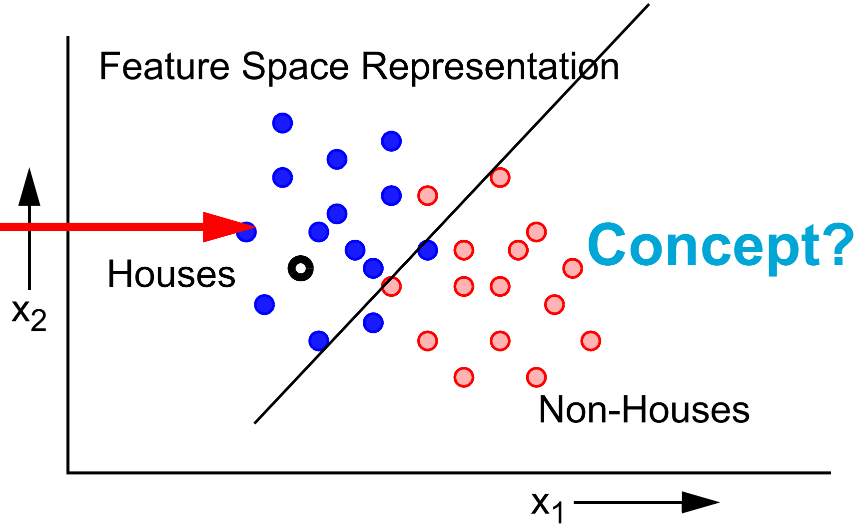
Concept of a house

Are these houses?

Platonic approach:

Aristotalian approach:

Examples of houses



Concept?

Structural Inference

Logic Inference:	Infer knowledge by means of logic
	All men are mortal Socrates is a man → Socrates is mortal
Statistical Inference:	Infer knowledge by means of statistics

Structural Inference

Logic Inference: Infer knowledge by means of logic

All men are mortal

Socrates is a man

→Socrates is mortal

Statistical Inference: Infer knowledge by means of statistics

---

Grammatical Inference: Infer a grammar by means of logic

Structural Inference: Infer a structure by means of statistics

Structural Inference

Logic Inference: Infer knowledge by means of logic

All men are mortal

Socrates is a man

→Socrates is mortal

Statistical Inference: Infer knowledge by means of statistics

---

Grammatical Inference: Infer a grammar by means of logic

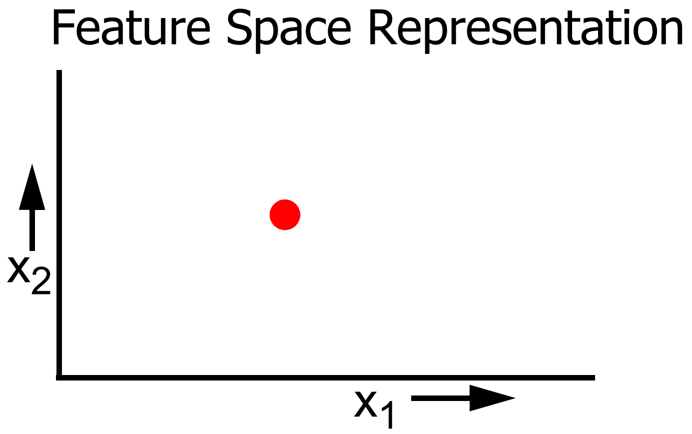
Structural Inference: Infer a structure by means of statistics

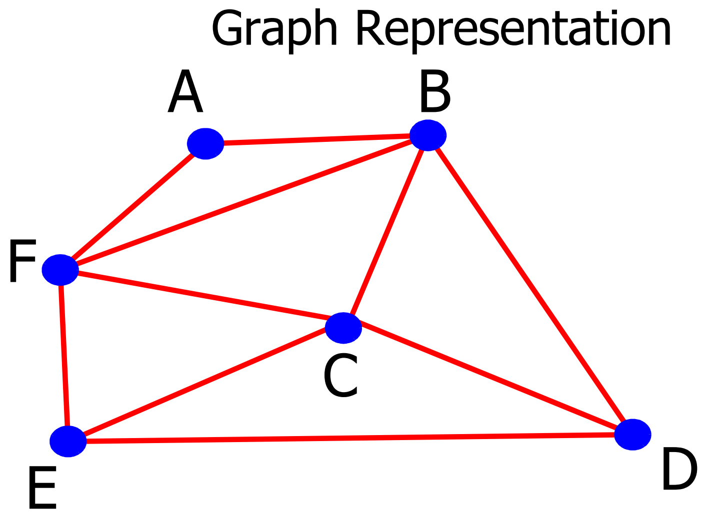
---

Is there something like:

Structural Inference: Infer knowledge by means of structure?

Representations





Features: sizes, colors, angles
Relations: lost
Generalization: easy, many examples needed
overlapping classes

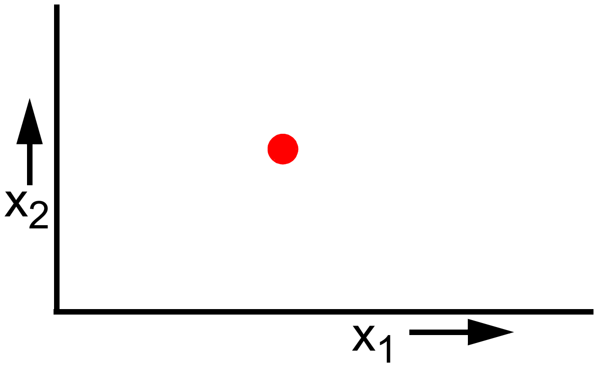
Relations: windows, door, roof
Features: node attributes
Generalization: difficult
often no class overlap

Is there something in between?

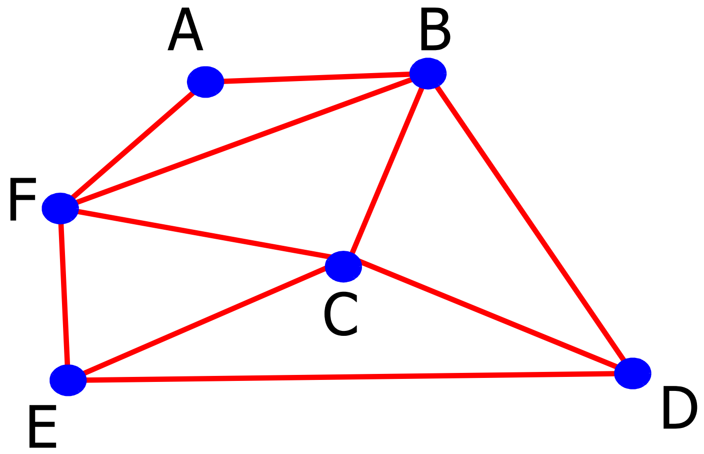
If we want to gain something, we should be prepared to lose something

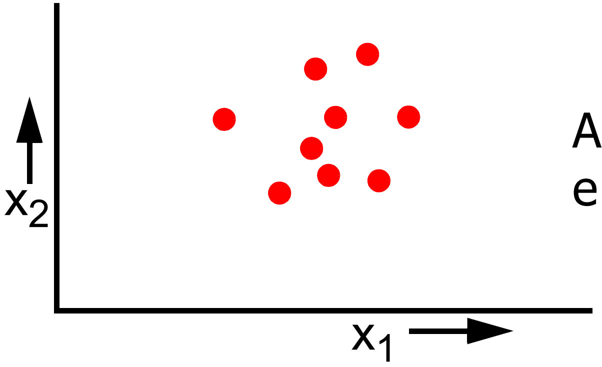
Intermediate Representations?

Feature Space Representation



Graph Representation

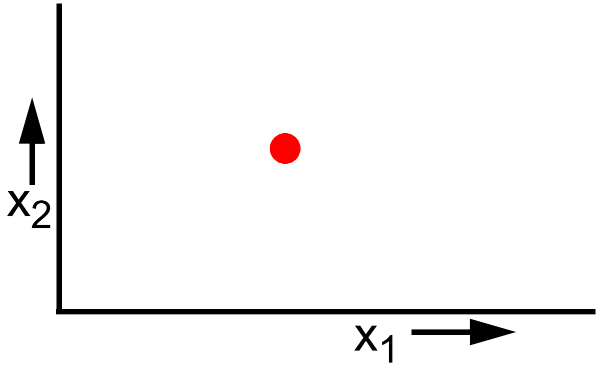




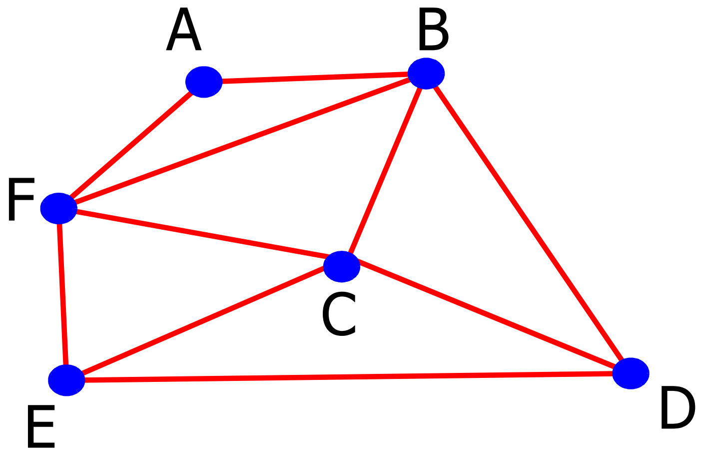
A set of points
e.g. feature representations of all regions

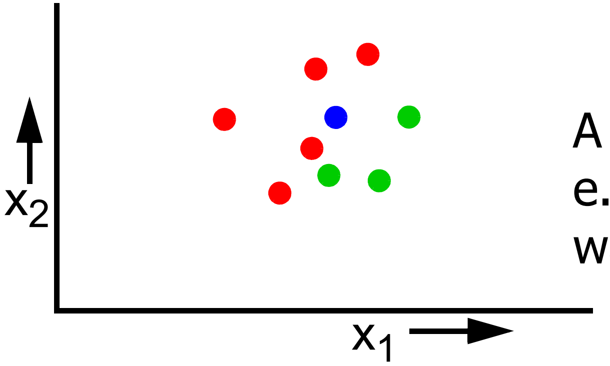
Intermediate Representations?

Feature Space Representation



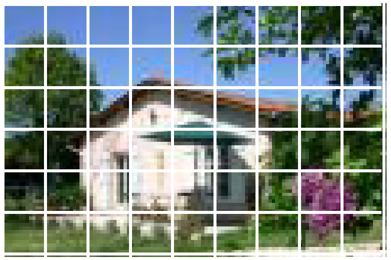
Graph Representation



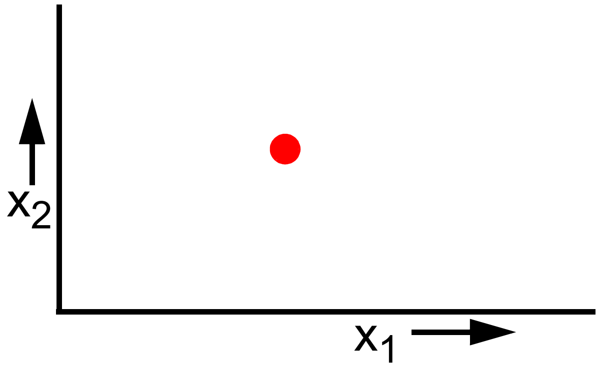


A labeled set of points
e.g. feature representations of
walls, windows, doors, roofs

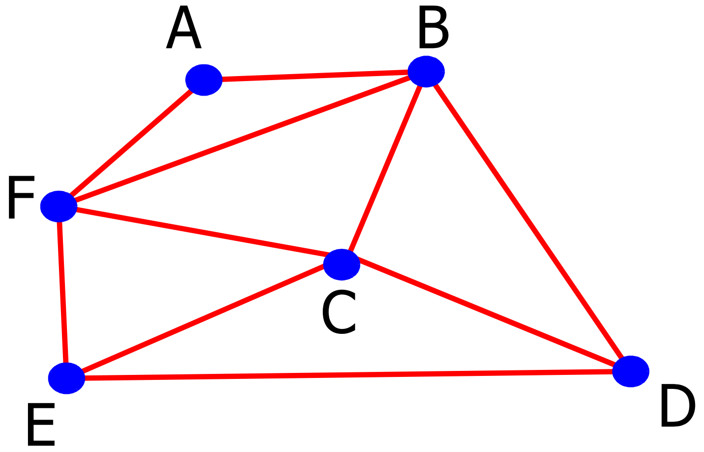
Intermediate Representations?

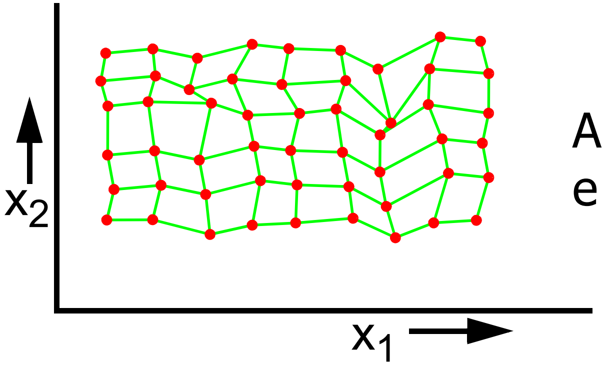


Feature Space Representation



Graph Representation

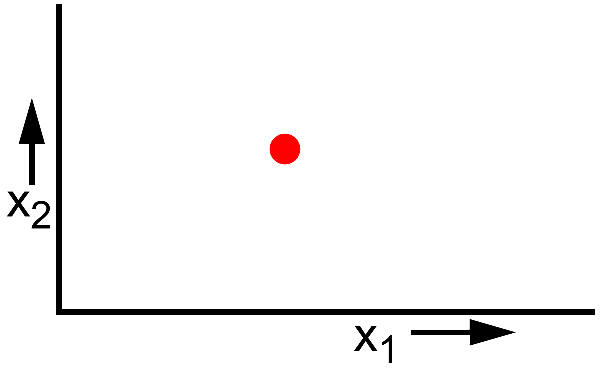




A connected set of points
e.g. feature representations of all image patches

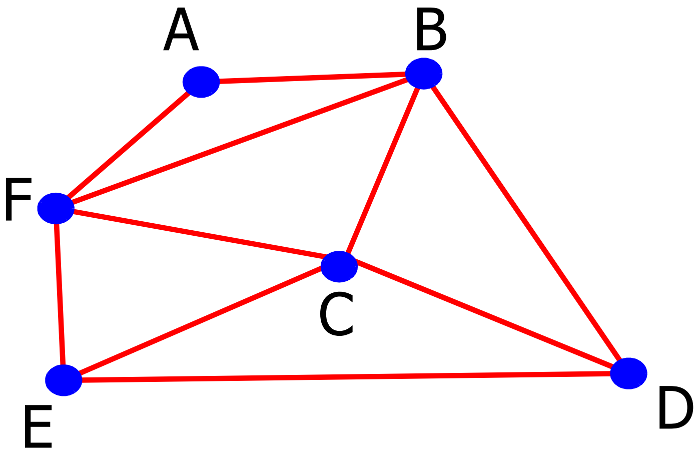
Intermediate Representations?

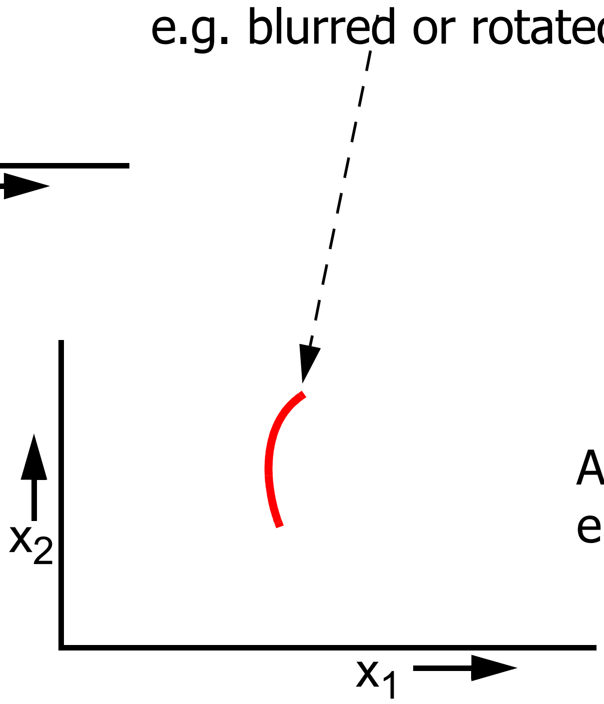
Feature Space Representation



e.g. blurred or rotated images

Graph Representation

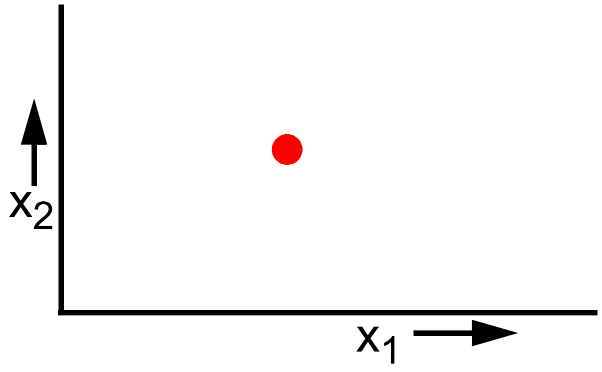




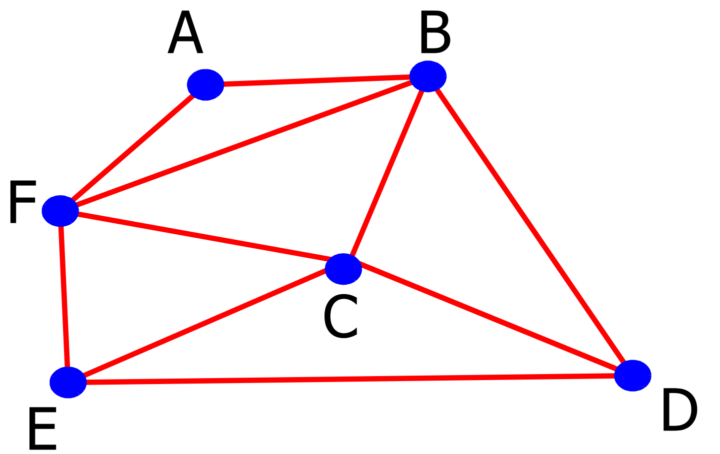
A stroke covering a series of feature representations
e.g. representing an invariant

Intermediate Representations?

Feature Space Representation



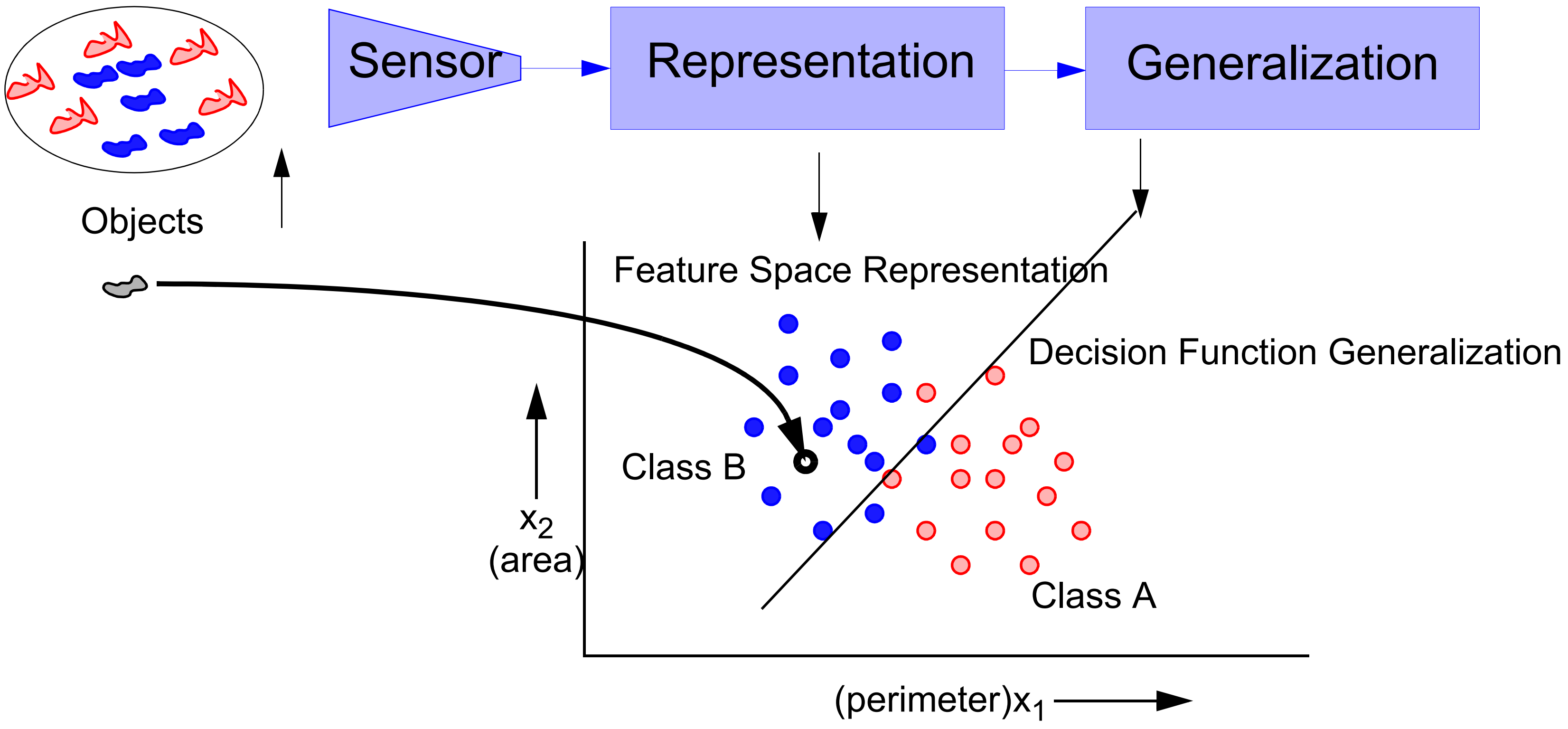
Graph Representation



Single point feature space representations are well supported. All others directly complicate the representation of classes and classification functions.

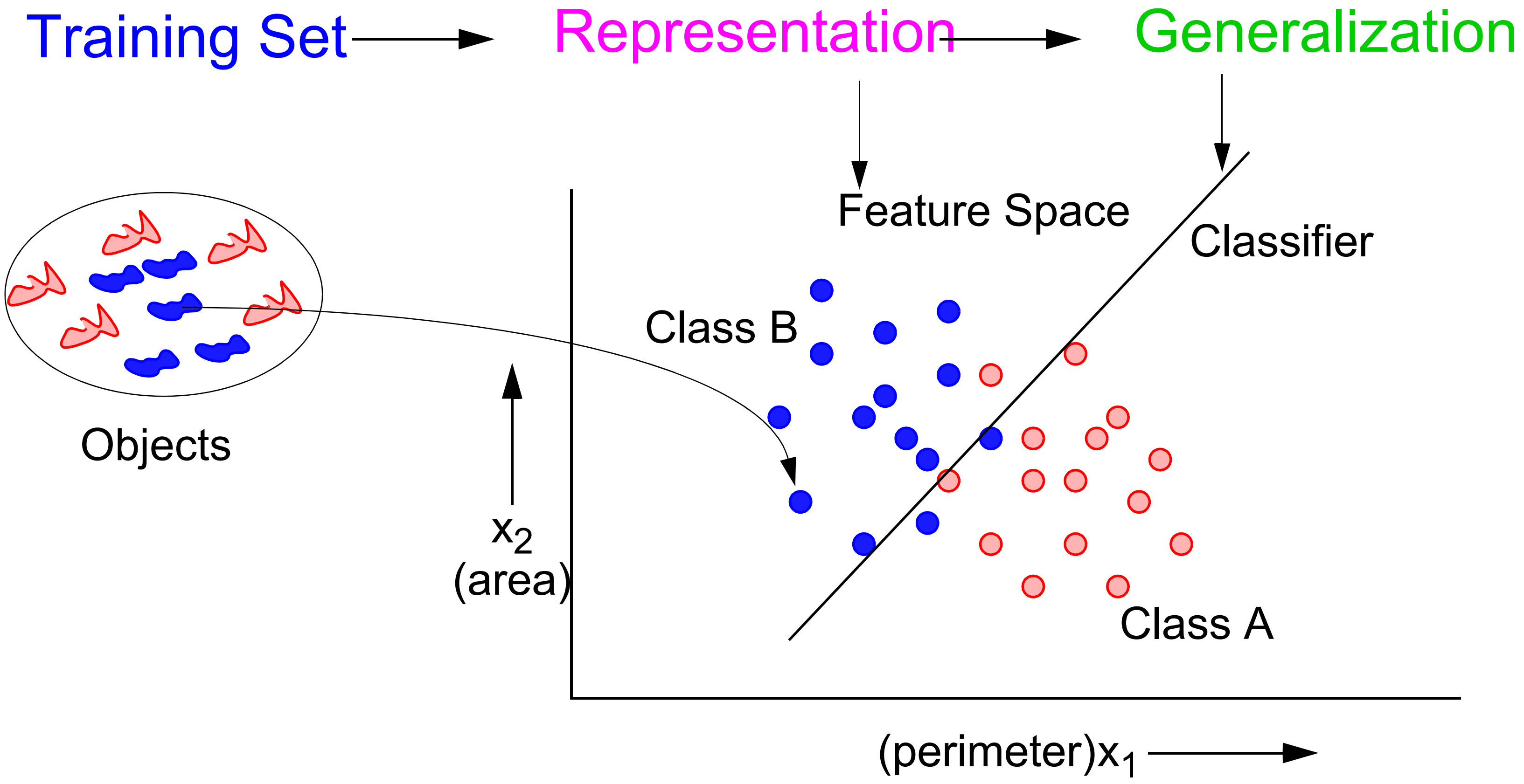
Is a smaller step possible?

The Pattern Recognition System



Learning from examples
Finding concepts (classes) from observations

Feature Space

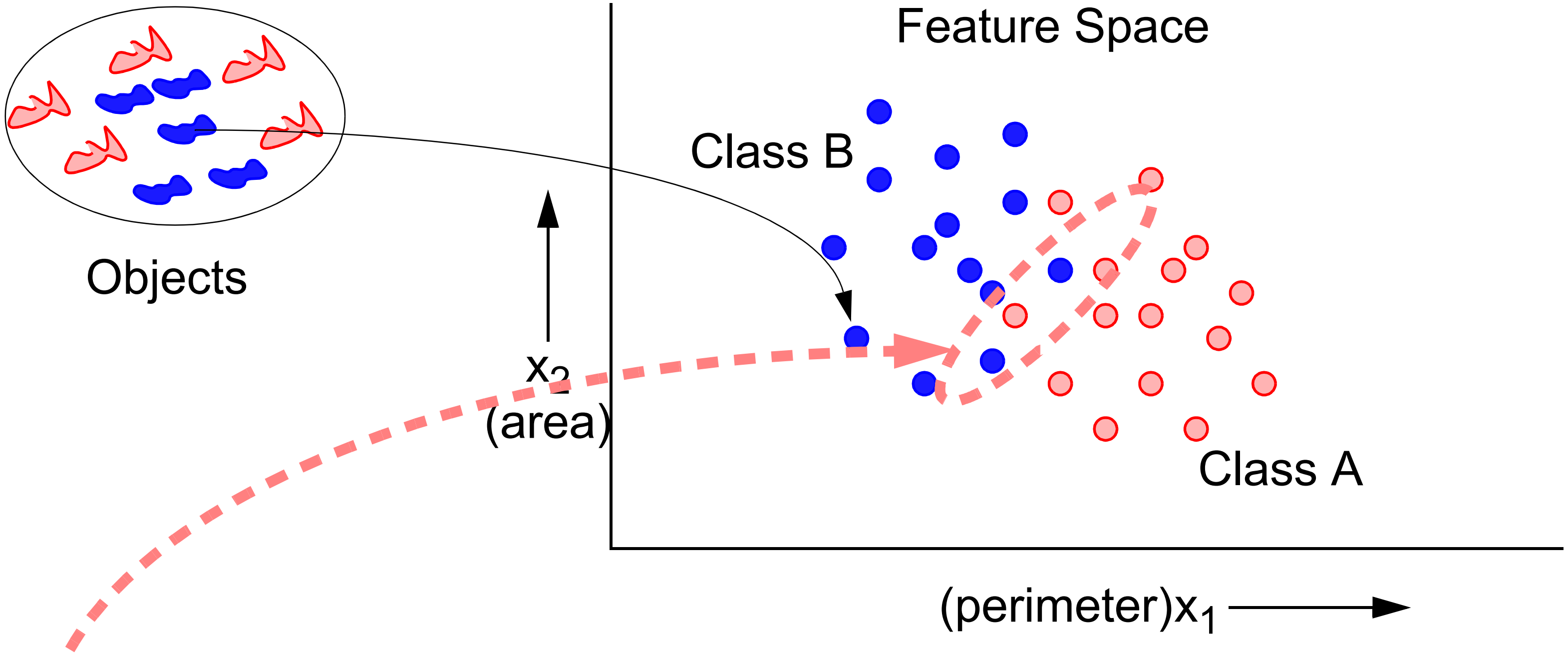


Knowledge → good features → (almost) separable classes

Lack of knowledge → (too many) bad features → hardly separable classes

Many features ~ Lack of knowledge, but objects better represented

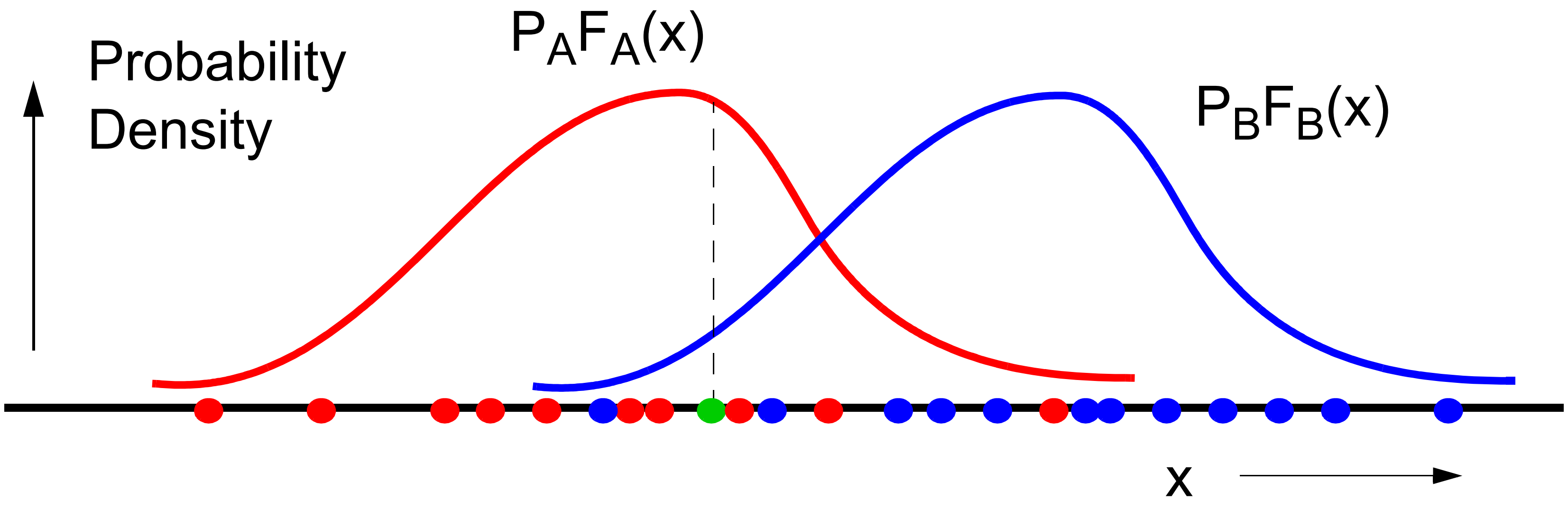
Features Reduce → Overlapping Classes



Due to reduction essentially different objects are represented identically

→ The feature representation needs a statistical (probabilistic) generalization

Features Reduce → Overlapping Classes → Probabilities



Best guess is to choose the most 'probable' class (→ small error).

→ Good for overlapping classes.

→ Assumes

- the existence of a probabilistic class distribution
- a representative set of examples.

No Feature Reduction

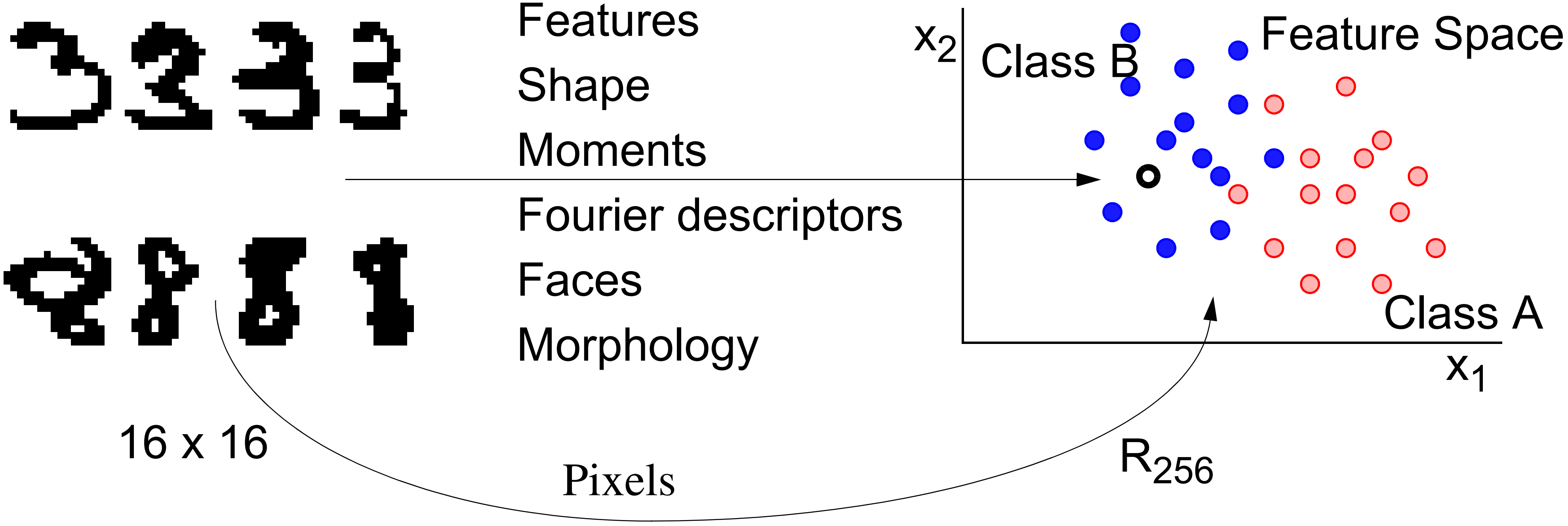
The feature representation enforces class overlap.
To be solved by a probabilistic approach.

However:

Are densities needed in high dimensional spaces?
Are classes densities?

→ Objects described by structures
→ Classes interpreted by structures

Pixel (Sample) Representation

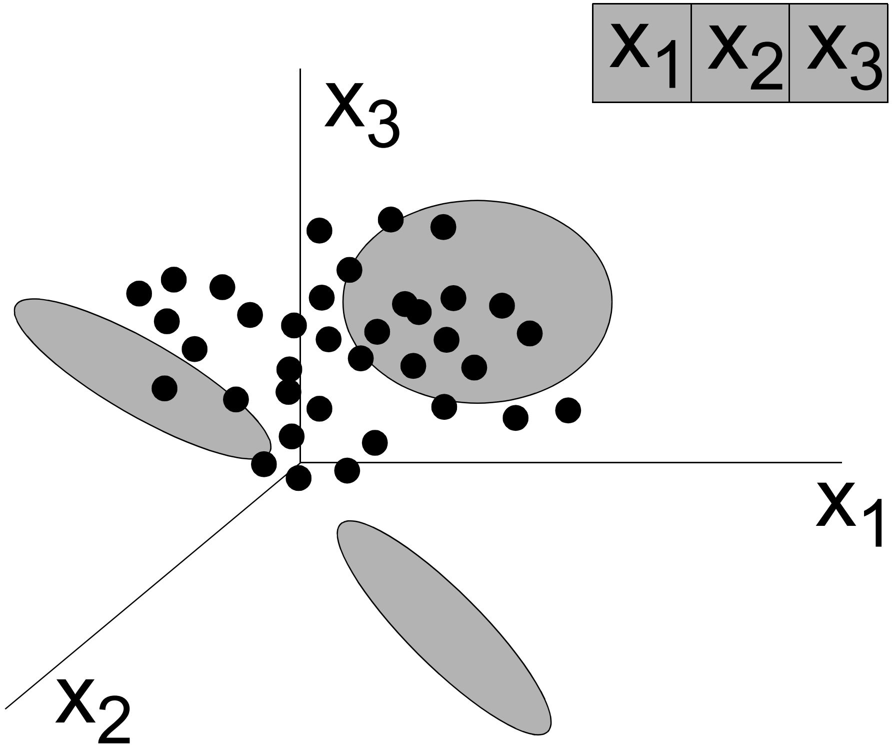


Features
Shape
Moments
Fourier descriptors
Faces
Morphology

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

The Connectivity Problem in the Pixel Representation

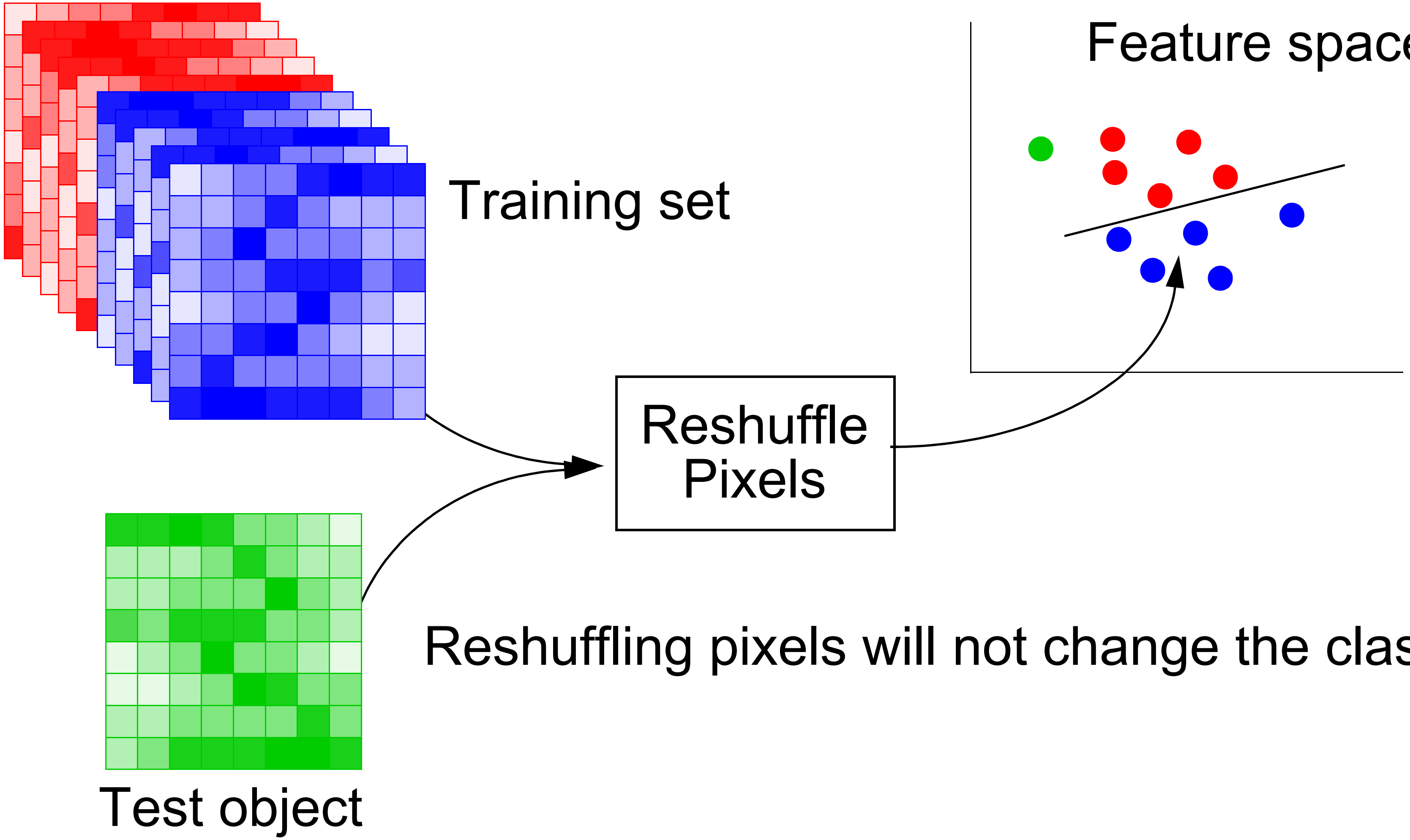
Spatial connectivity is lost



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

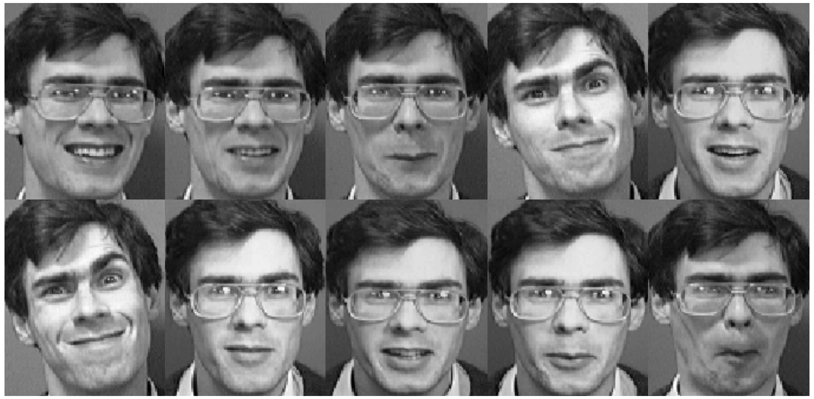
The Connectivity Problem in the Pixel Representation

Spatial connectivity is lost



Reshuffling pixels will not change the classification

High dimensional data often does not overlap

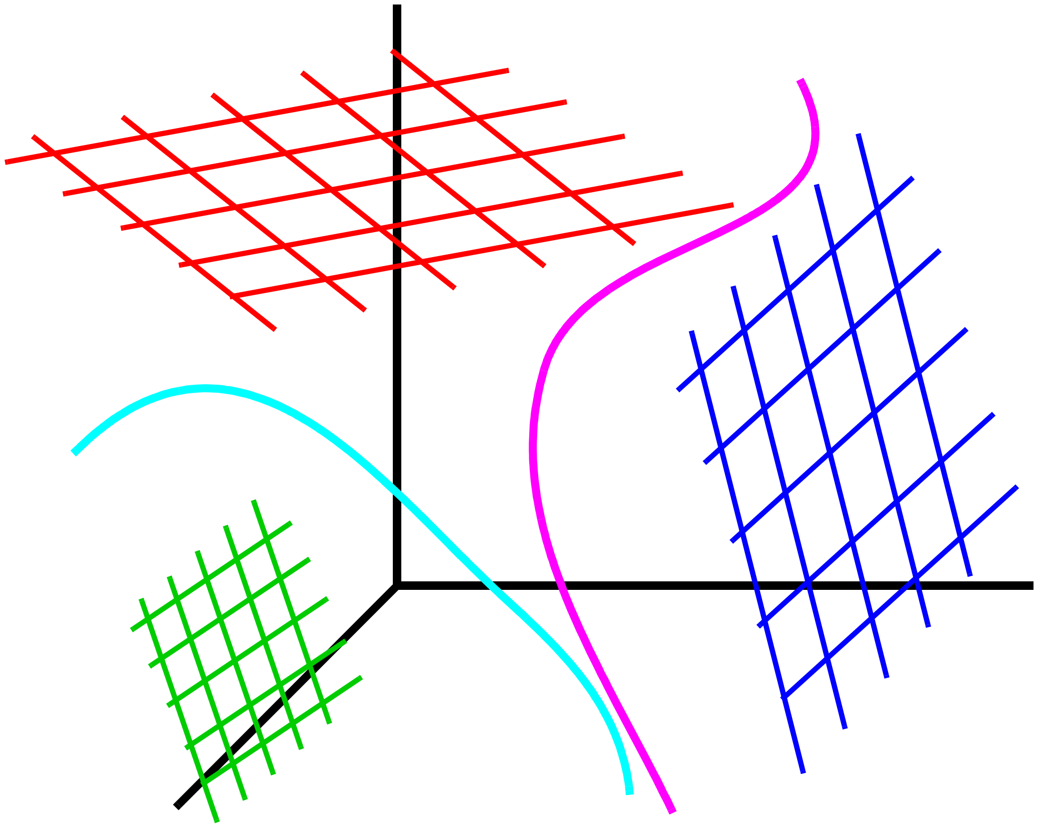


Complete feature representations, which enable the reconstruction of human recognizable, may yield separable classes.

There is no picture that could be member of different classes.

Structural information is not yet there.

Domain based classification

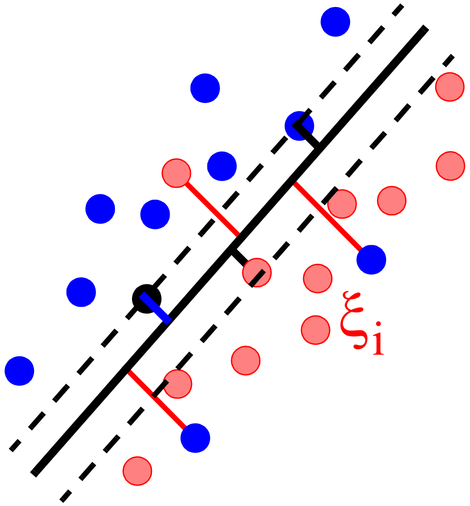


No well sampled training sets are needed.

Classifiers still to be developed.

Class structure ⟷ Object invariants

SVM



The Support Vector Classifier:

min$_\alpha$: $\|W\| + C\sum\xi_i$

min$_\alpha$: $\|W\|$: maximize the margin, minimize # support vectors: most simple

(**distance argument, Occam**)

min$_\alpha$: $\sum\xi_i$: minimize the total error

(**probabilistic argument, Bayes**)

SVM --> Domain based classifier

Objects

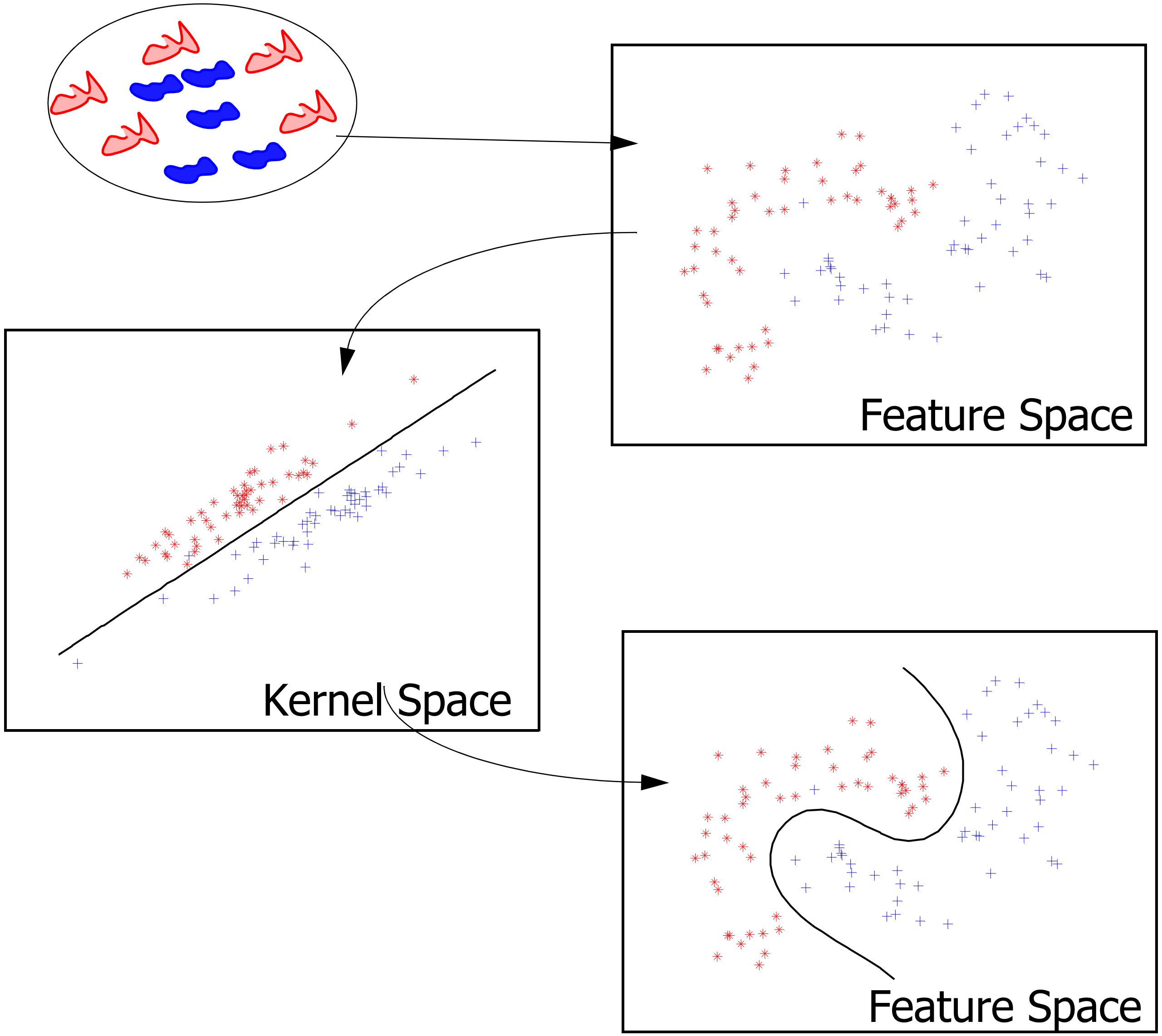
Feature Representation

Kernel: $K(x,y) = N_x(y,\sigma)$

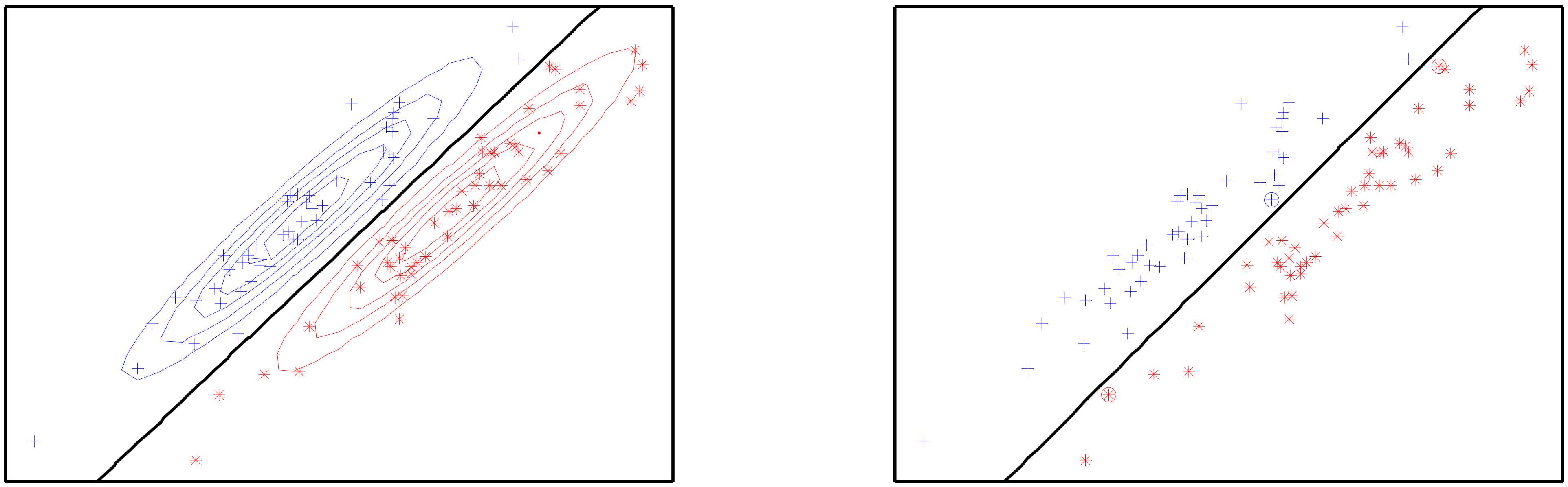
Kernel (Hilbert) Space

Linear SVM

Non-Linear Classifier



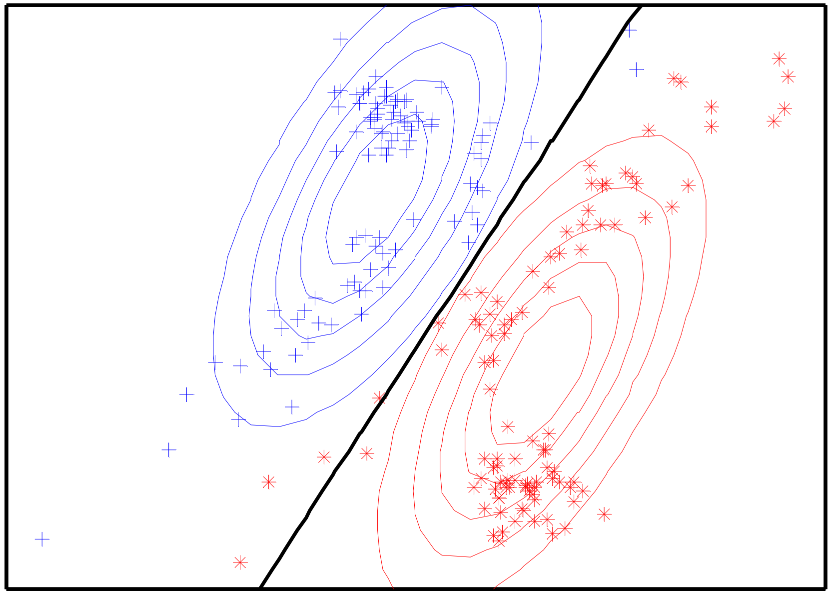
Fisher’s LD ⟷ SVM on Gaussian Data



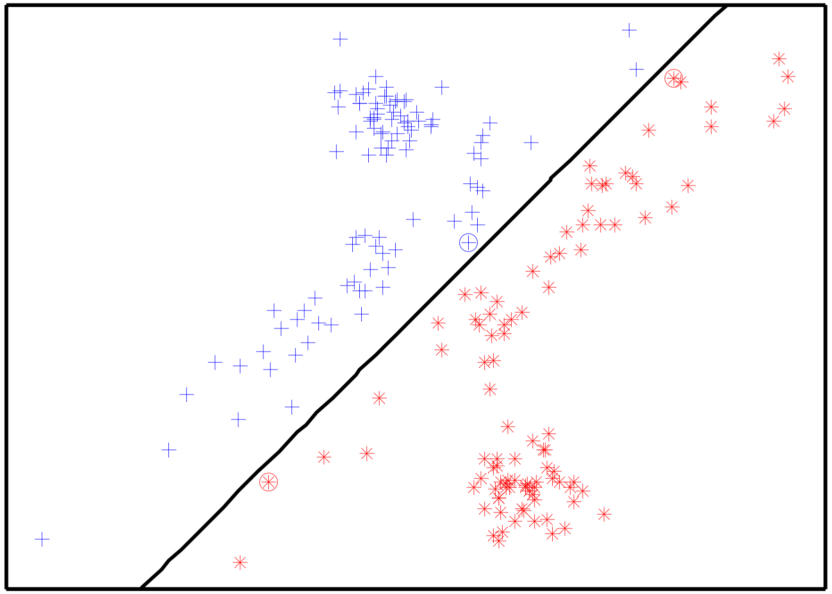
Fisher's LD
Sensitive for class distributions

SVM
Sensitive for between-class boundaries

Fisher’s LD <--> SVM on Non-Gaussian Data

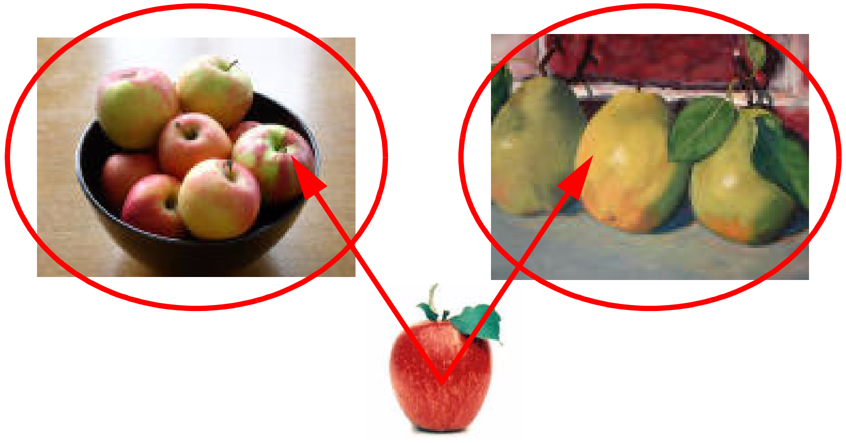


Fisher's LD
Sensitive for class distributions



SVM
Sensitive for between-class boundaries

Generalization Strategies: Densities versus Distances

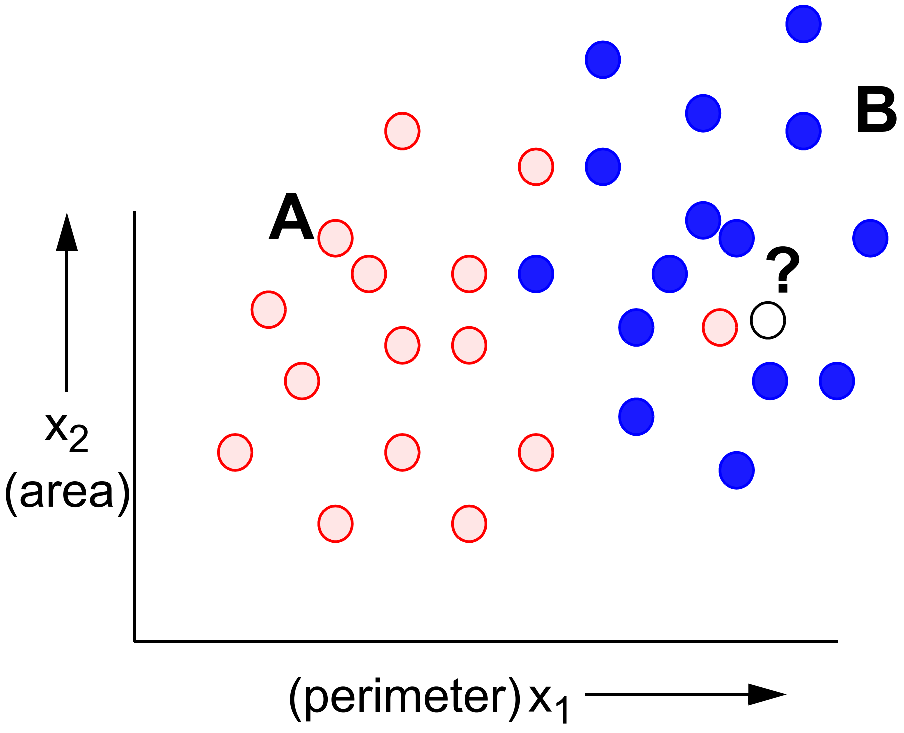


Why is this an apple and not a pear?

1. It is more similar to known apples than to known pears.

2. Most objects that look like this appear to be apples and not pears.

Densities versus Distances



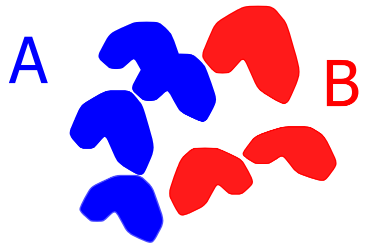
? to be classified as

A - because it is most closest to an object A

B - because the local density of B is larger

Dissimilarity Representation

Define dissimilarity measure dij between raw data of objects i and j



Given labeled training set T

Unlabeled object x to be classified

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

$$\mathbf{d}_x = (d_1 \;\; d_2 \;\; d_3 \;\; d_4 \;\; d_5 \;\; d_6 \;\; d_7)$$

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

Dissimilarity Space

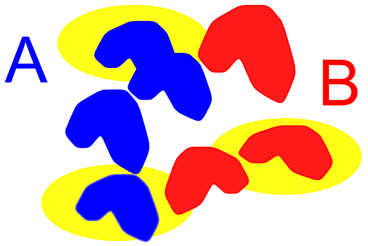
Dissimilarities

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

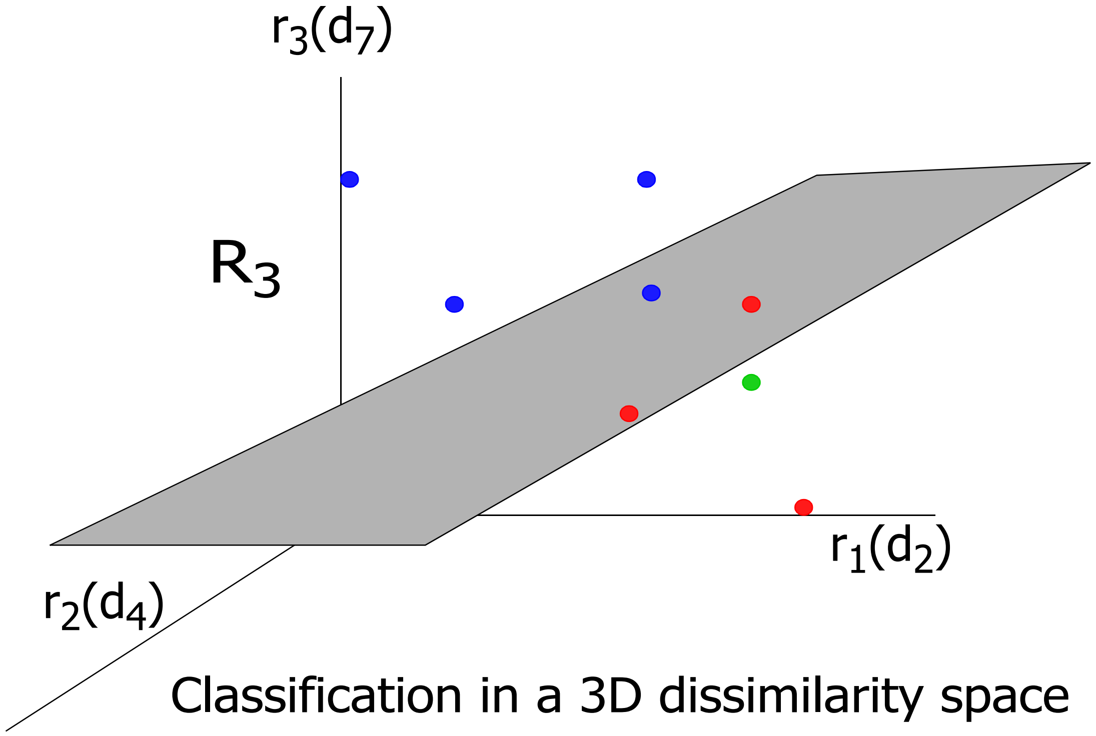
$$\mathbf{d}_x = (d_1 \; d_2 \; d_3 \; d_4 \; d_5 \; d_6 \; d_7)$$

Selection of 3 objects for representation

Given labeled training set T

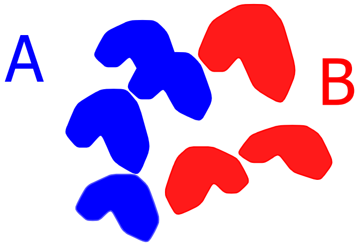


Unlabeled object x to be classified



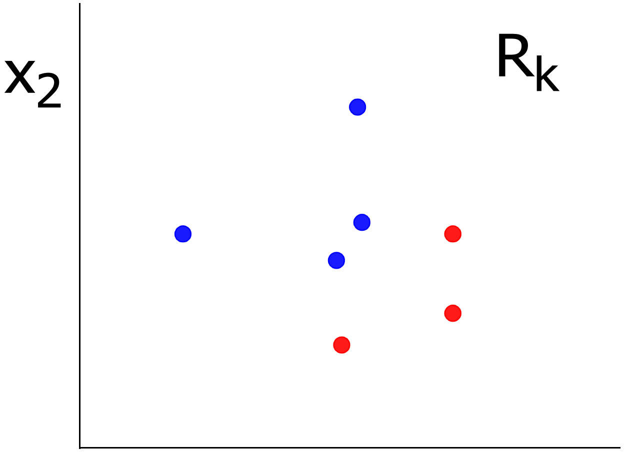
Classification in a 3D dissimilarity space

Embedding



Training set → Dissimilarity matrix D → X?

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



Euclidean distances → D

(Pseudo-)Euclidean Embedding

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

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

Eigenvalue Decomposition $(B = Q \Lambda Q^T)$p,

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, $\Im_k(i,i) = \mathrm{sign}(\Lambda_k(i,i))$ $X = Q_k |\Lambda_k|^{\frac{1}{2}} \Im_k$

$L_k(i,i) < 0$ → Pseudo-Euclidean

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

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

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

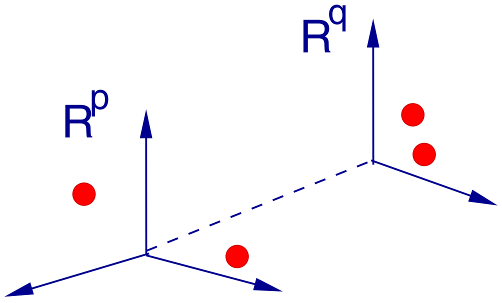
Pseudo-Euclidean Space (Krein Space)

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

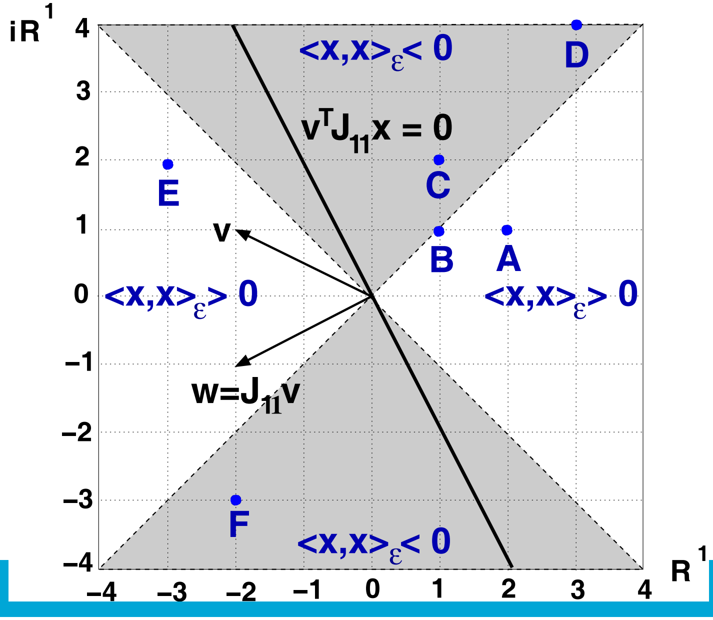
A pseudo-Euclidean space ε with signature (p,q), k =p+q, is a non-degenerate inner product space $\Re^k = \Re^p \oplus \Re^q$ such that:

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

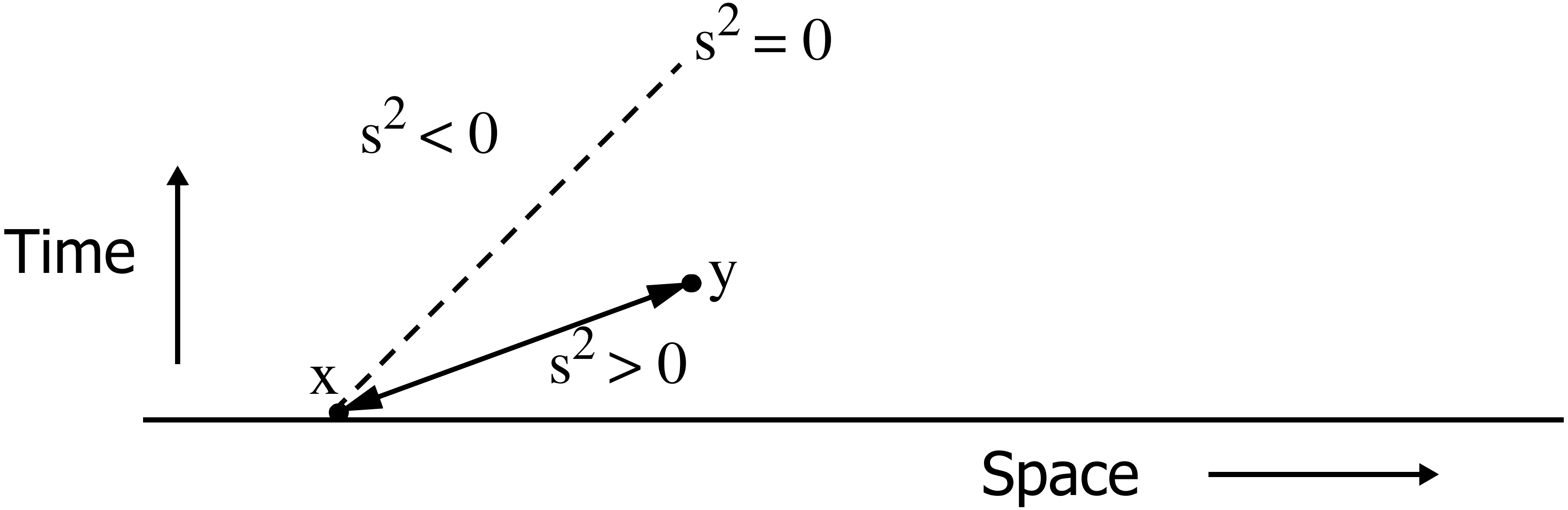
$$d_\varepsilon^2(x, y) = \langle x-y, x-y\rangle_\varepsilon = d_p^2(x, y) - d_q^2(x, y)$$



$d^2(x,y) = d_p^2(x,y) - d_q^2(x,y)$



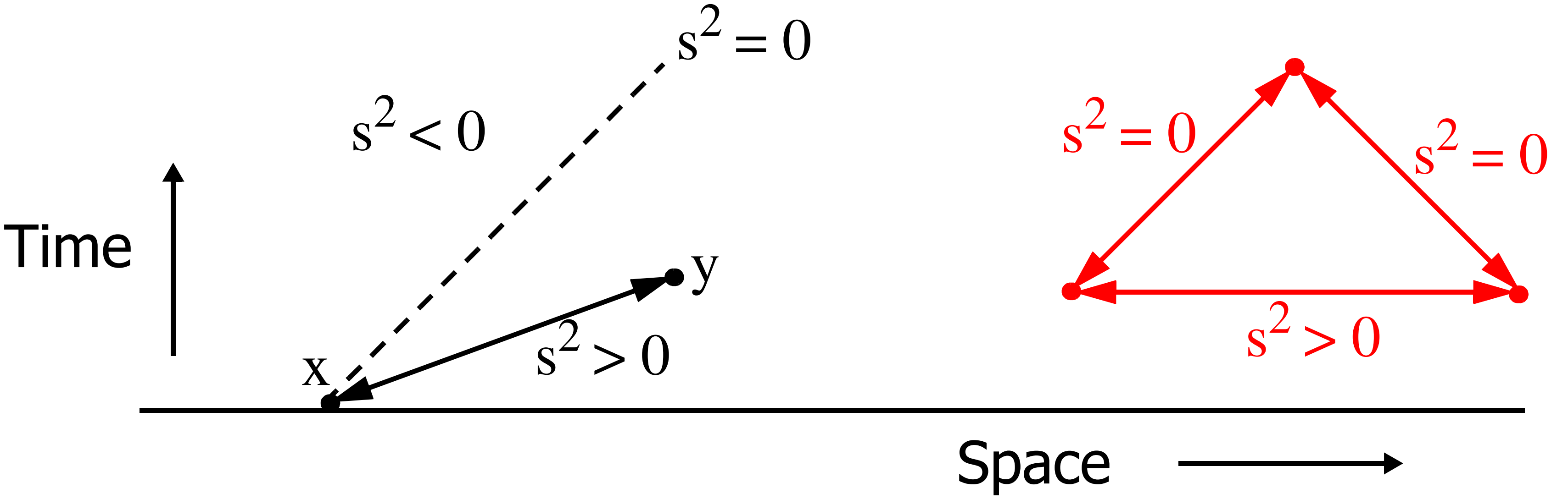
Minkowsky Space → Pseudo-Euclidean Space



SpaceTime distance:

$$s^2 = \sum_{i=1}^{3} (x_i - y_i)^2 - (t_x - t_y)^2$$

Minkowsky Space → Pseudo-Euclidean Space

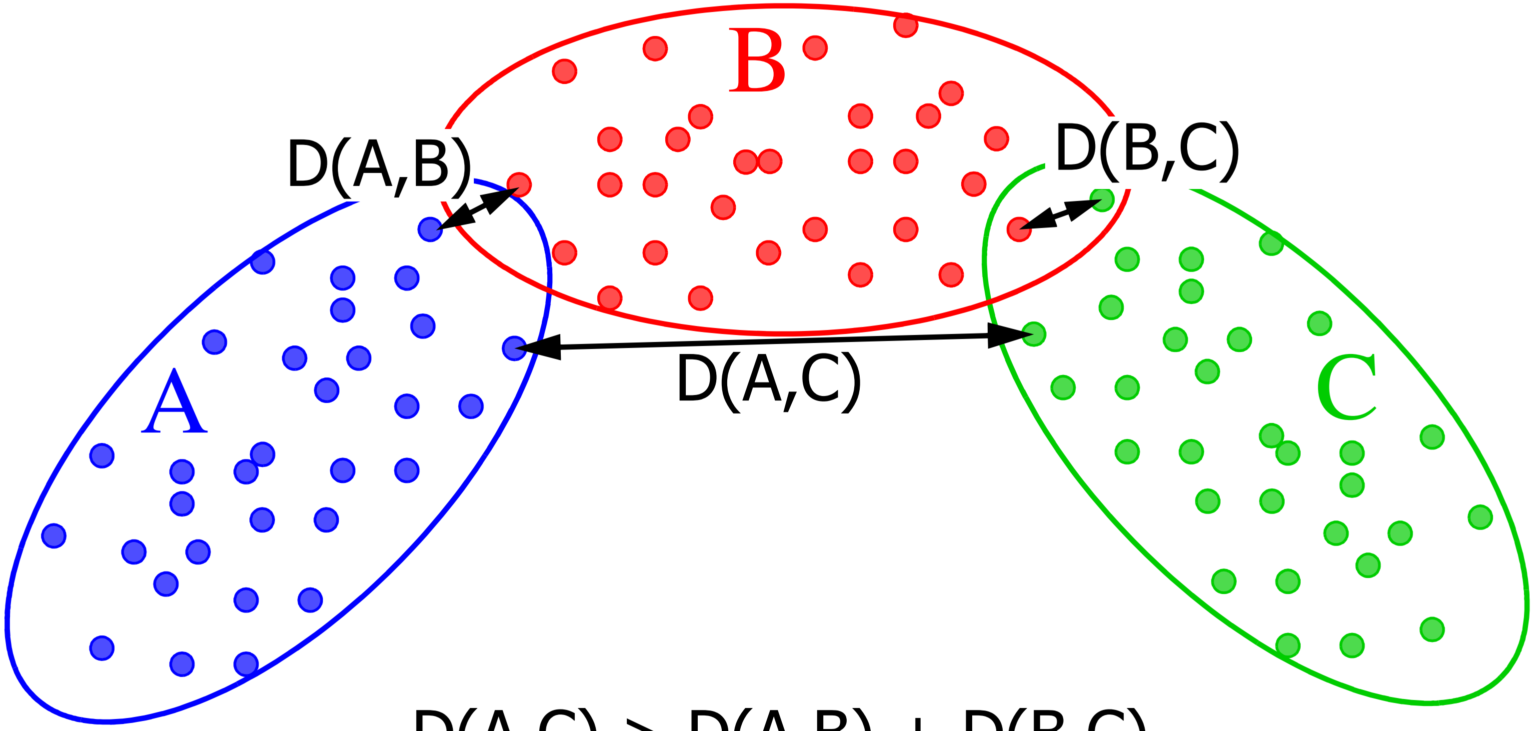


non-metric !!

SpaceTime distance:

$$s^2 = \sum_{i=1}^{3} (x_i - y_i)^2 - (t_x - t_y)^2$$

Non-Metric Dissimilarities in Pattern Recognition

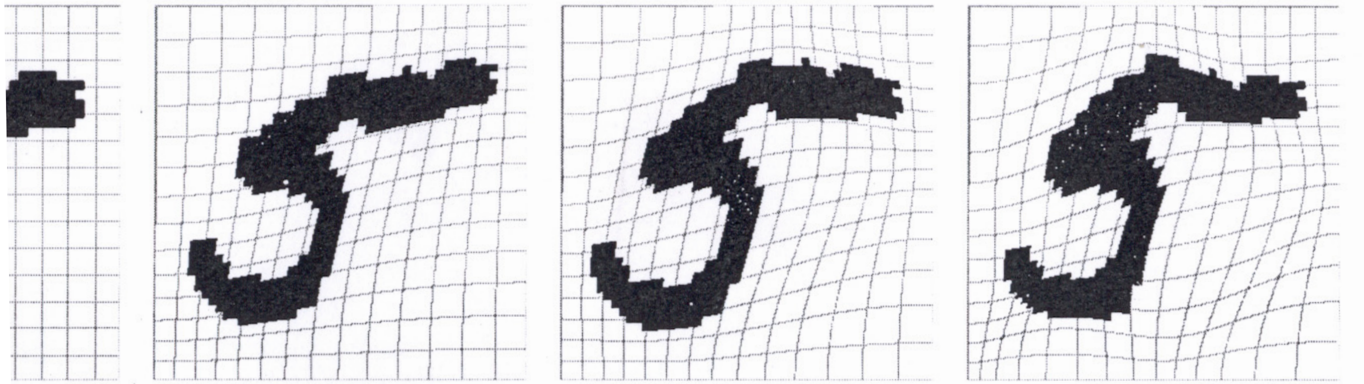


D(A,C) > D(A,B) + D(B,C)

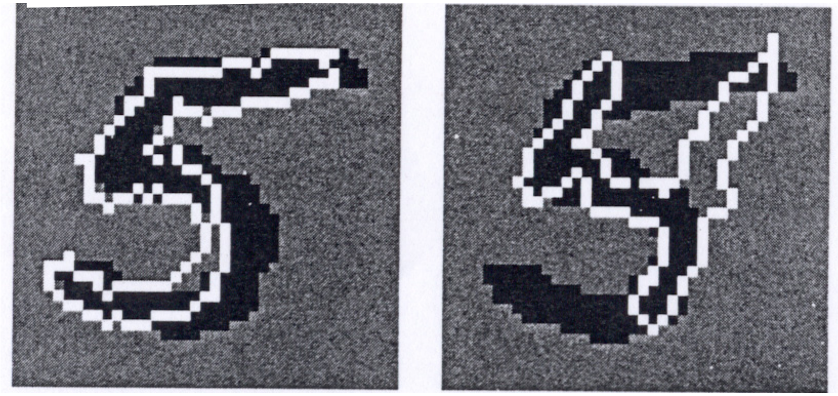
The single-linkage cluster distance is non-metric

Dissimilarity based Structural PR - Transformations

Examples of deformed templates



(b) (c) (d)

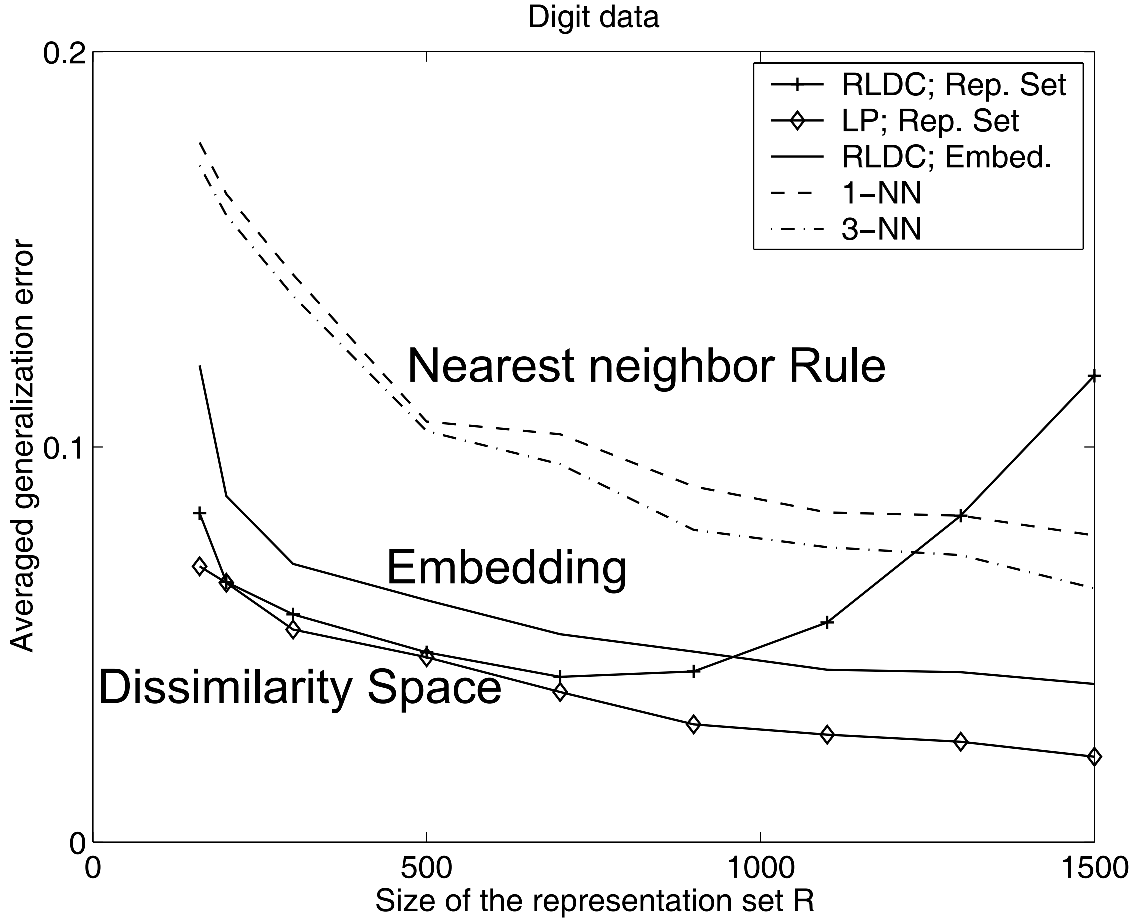


Matching new objects x to various templates y

$$\text{class}(x) = \text{class}(\text{argmin}_y(D(x, y)))$$

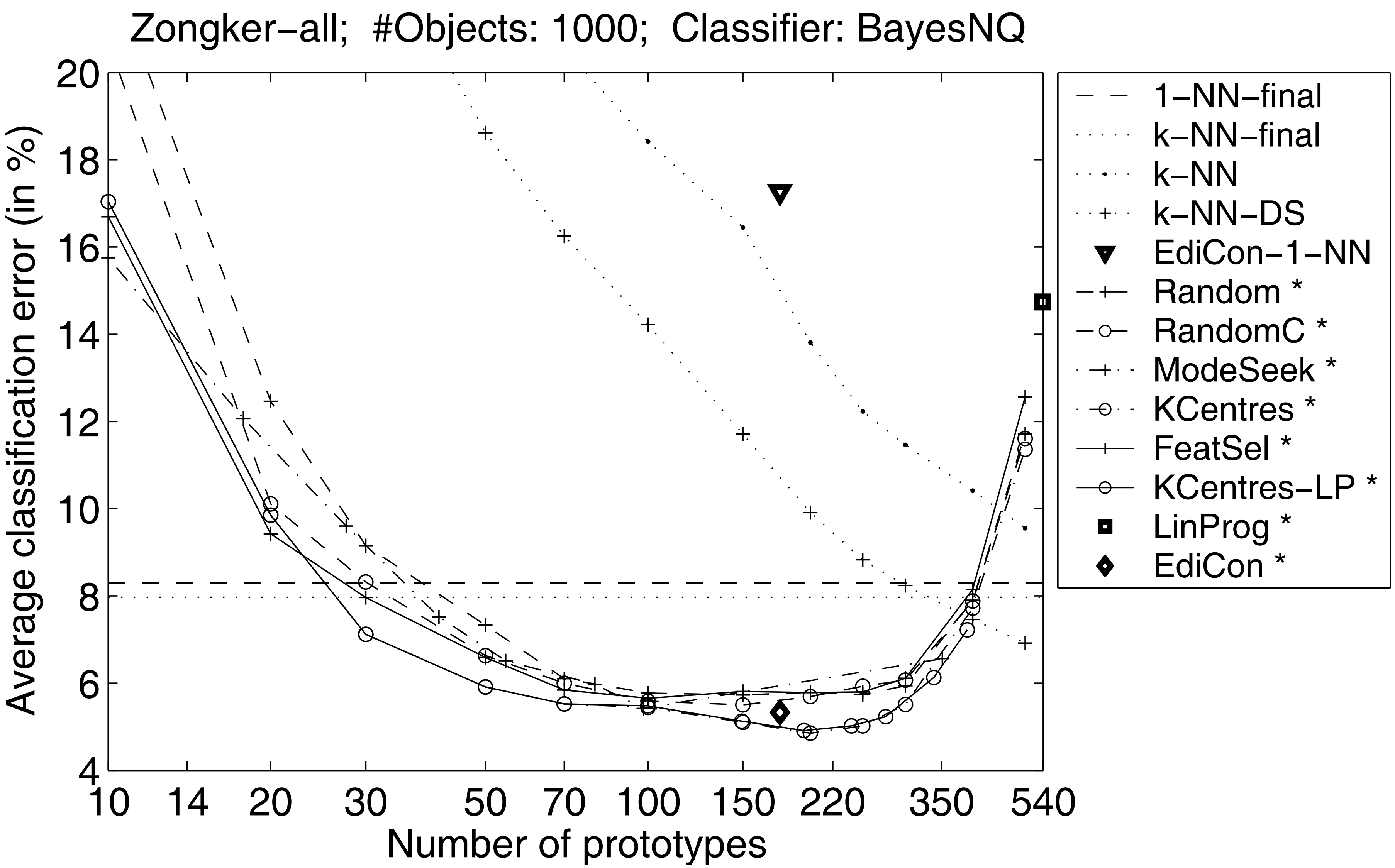
A.K. Jain, D. Zongker, Representation and recognition of handwritten digit using deformable templates, *IEEE-PAMI*, vol. 19, no. 12, 1997, 1386-1391.

Three Approaches Compared for the Zongker Data



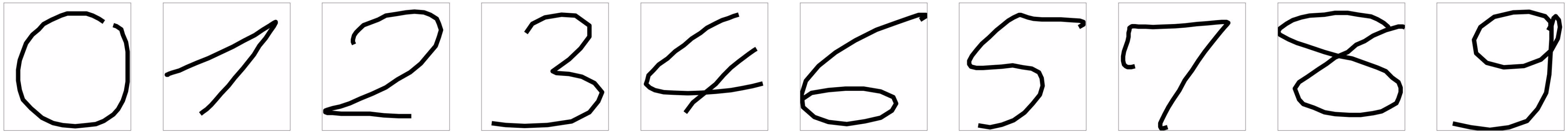
Dissimilarity Space better than Embedding better than Nearest Neighbor Rule

Prototype Selection on Zongker Data in Dissimilarity Space

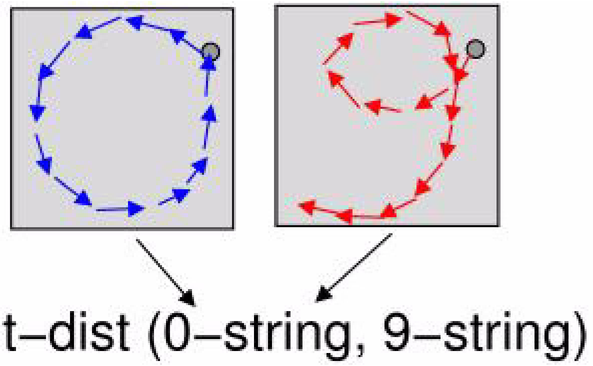


Edit Distances for Strings

Pen-written digits, (H. Bunke, Bern)



An edit distance between the string contours is applied with fixed insertion and deletion costs, and the Euclidean distance substitution cost.



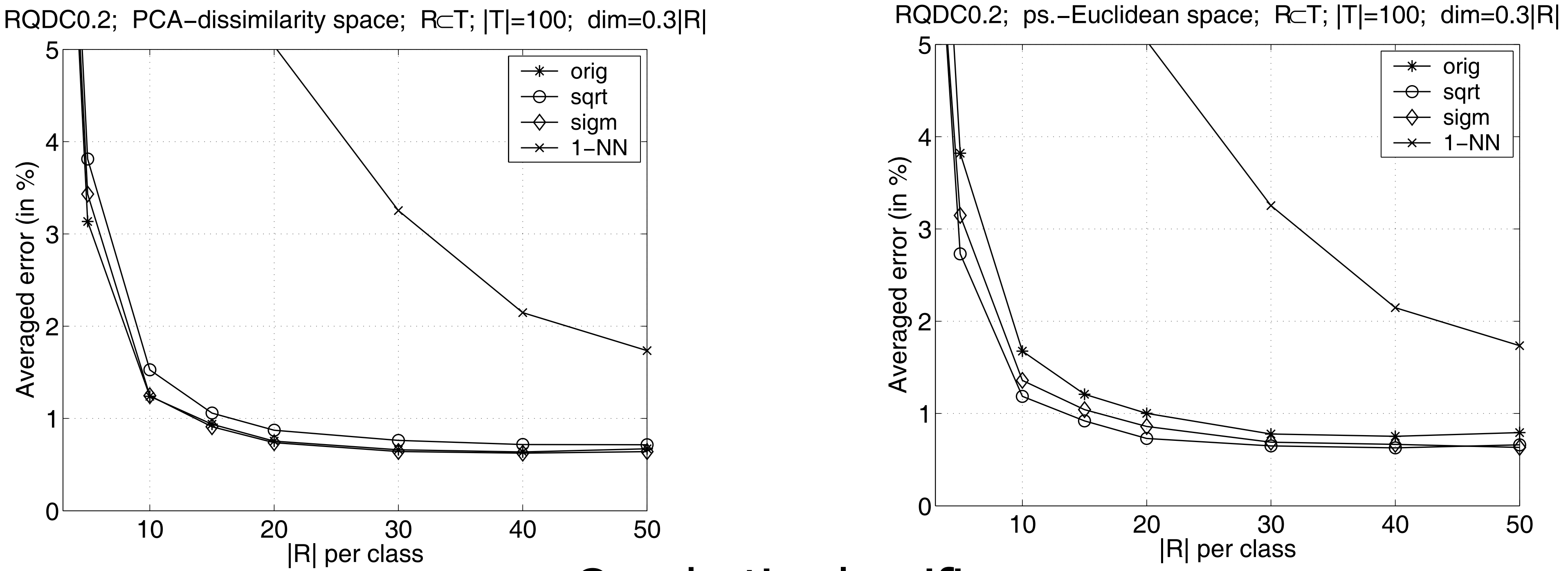
The measure is moderately non-metric and non-Euclidean.

Classification Results for Pen-written Digits

Representation D(T,R),

Fixed training set: |T| =10 x 200, R ⊂ changes.

Fixed test set: |S| = 10 x 200. Dimensionalities of the spaces: 0.3 |R|.



Quadratic classifiers

Dissimilarity Space

Pseudo-Euclidean Space

Dissimilarity Representation $\longleftrightarrow$ Kernels

Let R = $\{\mathbf{x_1}, \mathbf{x_2}, \ldots , \mathbf{x_n}\}$

Kernel classifier: $f(\mathbf{x}) = \mathbf{w}^T K(\mathbf{x}, R) + w_0 = \sum_{w_i \in SV} w_i K(\mathbf{x}, \mathbf{x_i}) + w_0$

Linear classifier in a dissimilarity space:

$$f(D(\mathbf{x}, R)) = \mathbf{w}^T D(\mathbf{x}, R) + w_0$$

If D has an Euclidean behavior, then

$$K = -\frac{1}{2} J D^{(2)} J \text{ or } K = \exp\{-D^{(2)} / \sigma^2\}$$

are Mercer kernels: SVM can be directly applied, Otherwise:

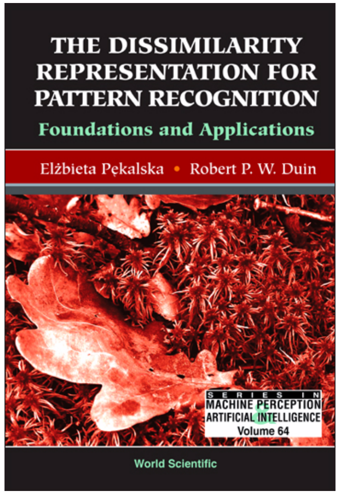
- Transform D appropriately or regularize K
- Treat K as kernels in pseudo-Euclidean space: indefinite SVM

Dissimilarity Based Classifiers ⟷ SVMs

Dissimilarity classification	SVM
arbitrary dissimilarity measure	Mercer kernels
full training set	reduced
complexity reduction	included
any classifier	structural risk minimization

Dissimilarity Based Classifiers ⟷ SVMs

Dissimilarity classification	SVM
arbitrary dissimilarity measure	Mercer kernels
full training set	reduced
complexity reduction	included
any classifier	structural risk minimization



E. Pekalska and R.P.W. Duin, *The Dissimilarity Representation for Pattern Recognition, Foundations and Applications,*
World Scientific, Singapore, 2005, 607 p. ISBN 981-256-530-2

Non-Euclidean is the normal

Many measures used in PR are non-Euclidean:

Mahalanobis, Single linkage, Modified Haussdorf, Weighted edit,

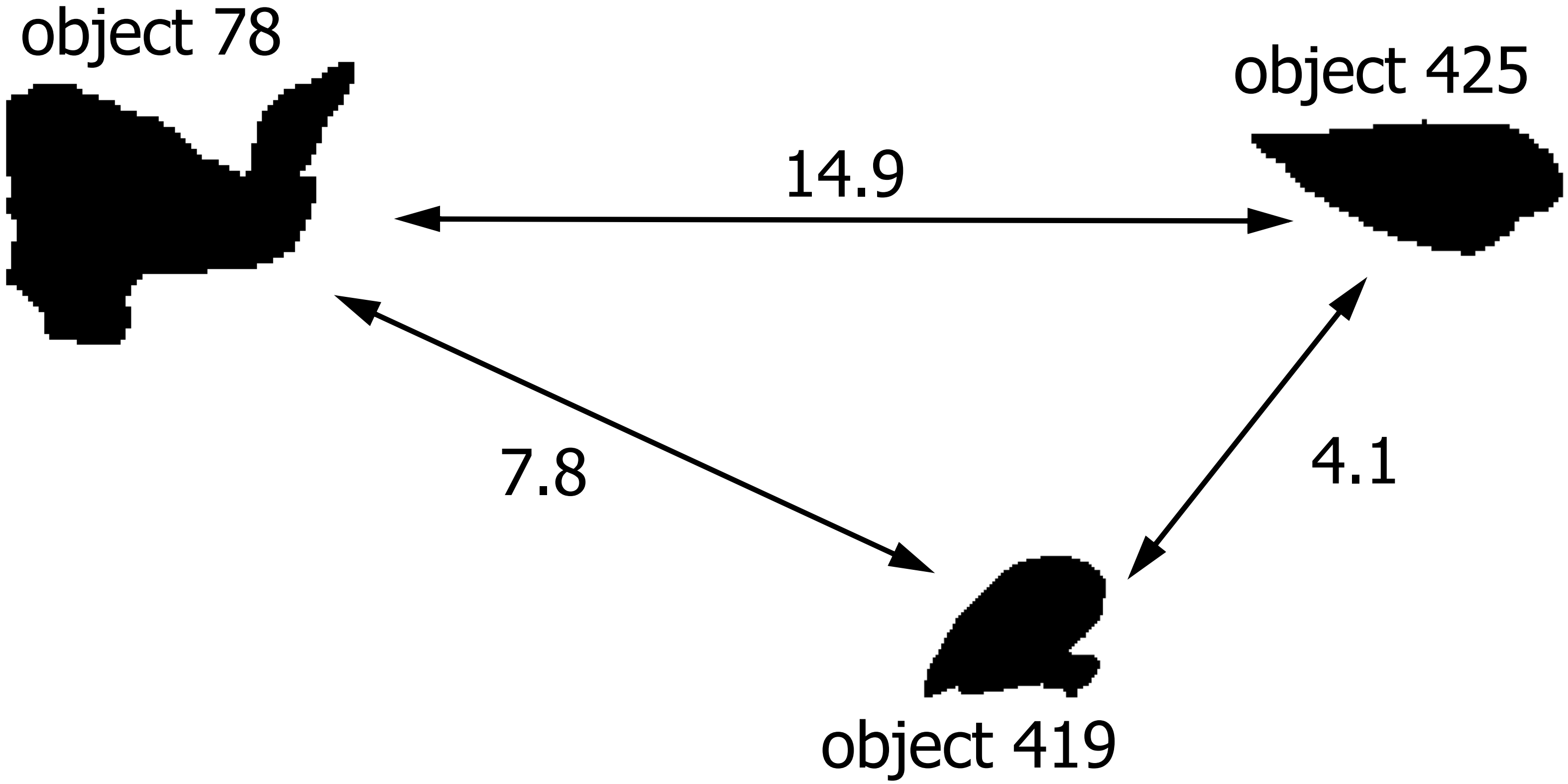
City block, Max norm, Cosine, Divergence, ...

No problem for dissimilarity spaces

Problem for embedding (Pseudo-Euclidean Space)

Problem for SVM (indefinite kernels)

Chicken pieces example



Edit_distance (length = 29, angle = 45)
non-metric result

What is the cause of non-metric data?

1. Distance measure is intrinsically non-metric
2. Some distances are estimated too large

   due to suboptimal path minimization (correction?)
3. Some distances are estimated too short

   e.g. due to projections or occlusions (correction?)
4. Accidental usage of a non-metric measure

   where a metric measure may do well (detectable?)

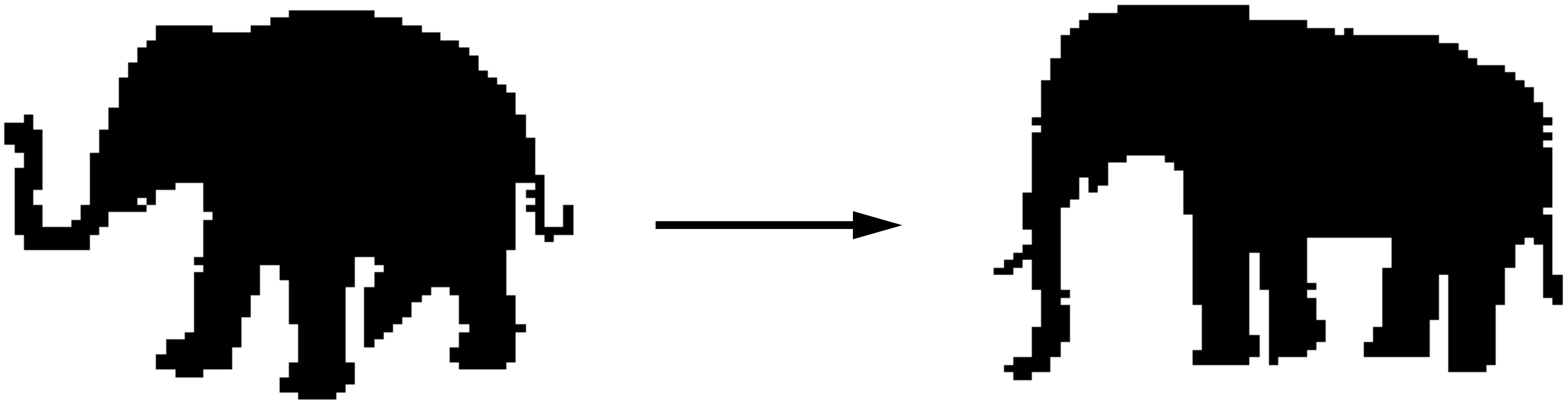
Structural inference by dissimilarity representations

By using dissimilarity representations we can train classifiers based on structural distance measures, but do we really learn from structure?

More should be understood about

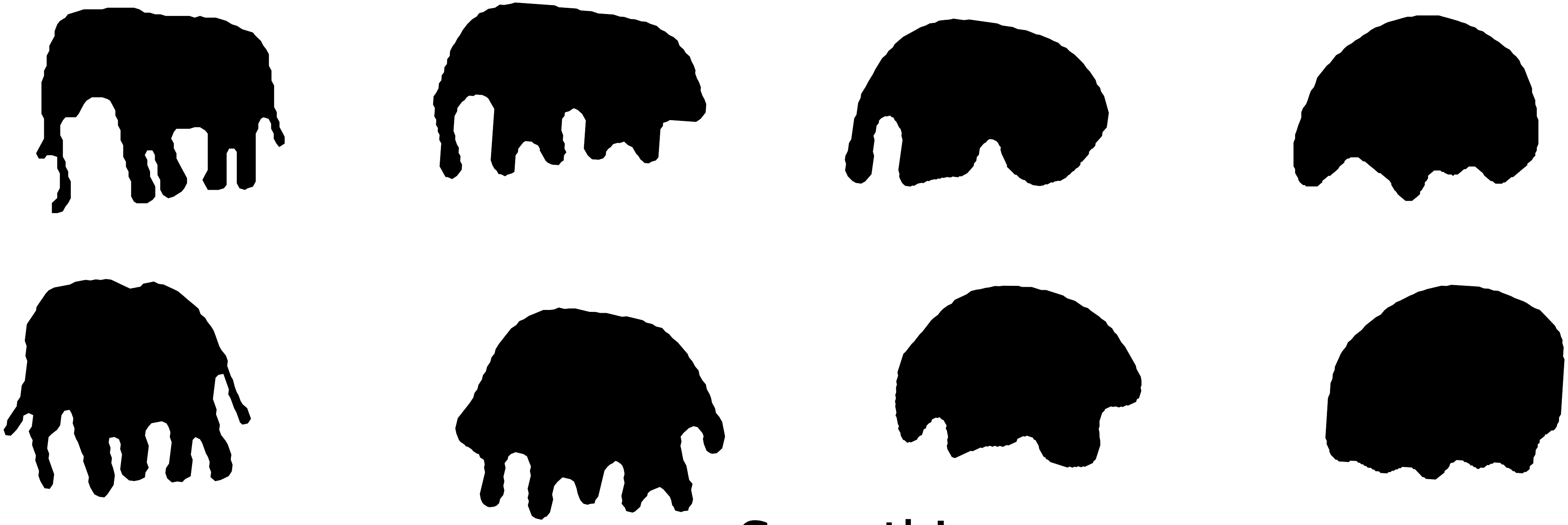
- non-Euclidean topology
- Pseudo-Euclidean embedding
- handling indefinite kernels

Structural inference: morphing



If we know the space of elephants, we can morph (edit) one elephant into another, with only elephants in between.

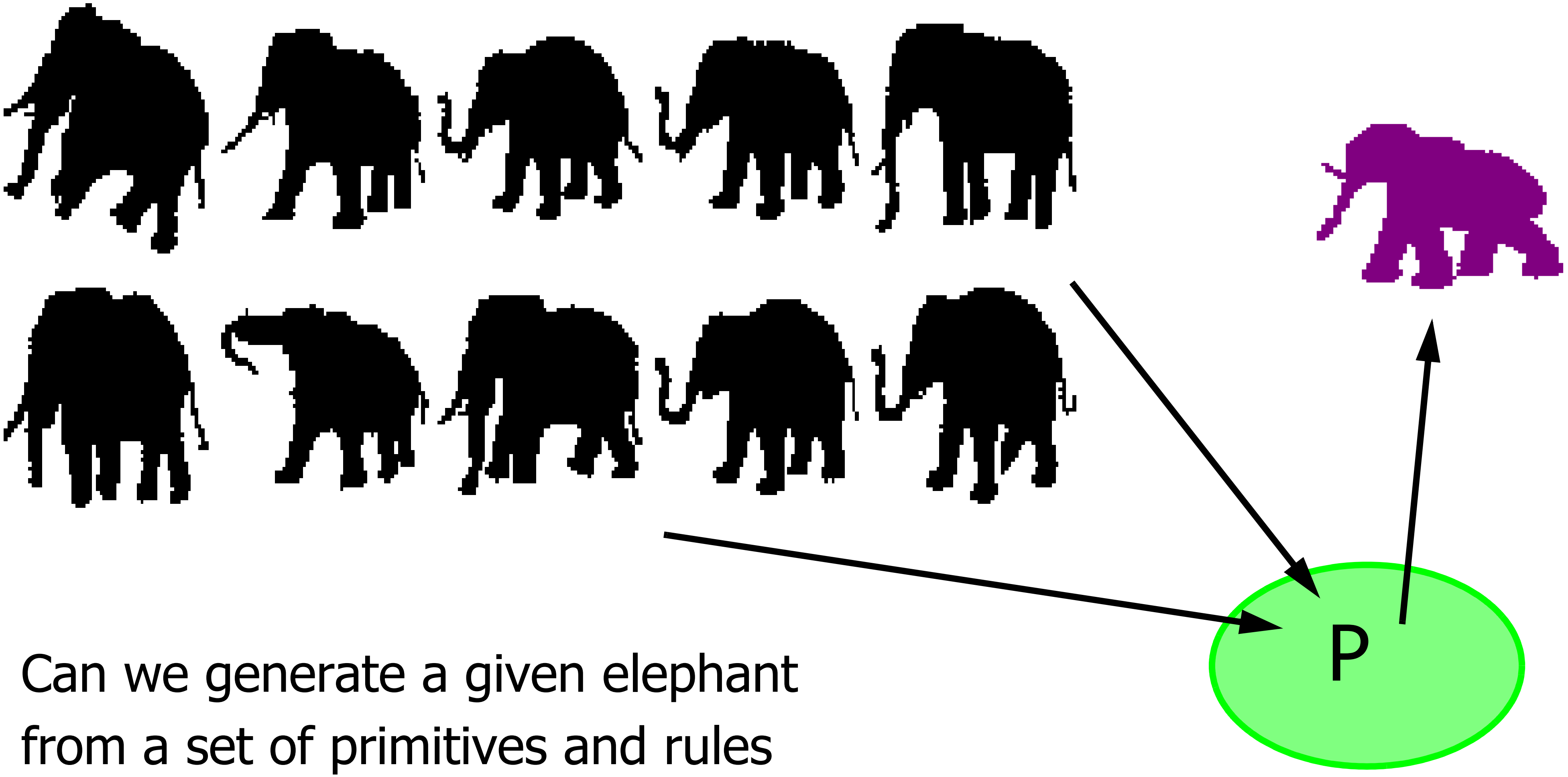
Structural inference: simplification



Smoothing

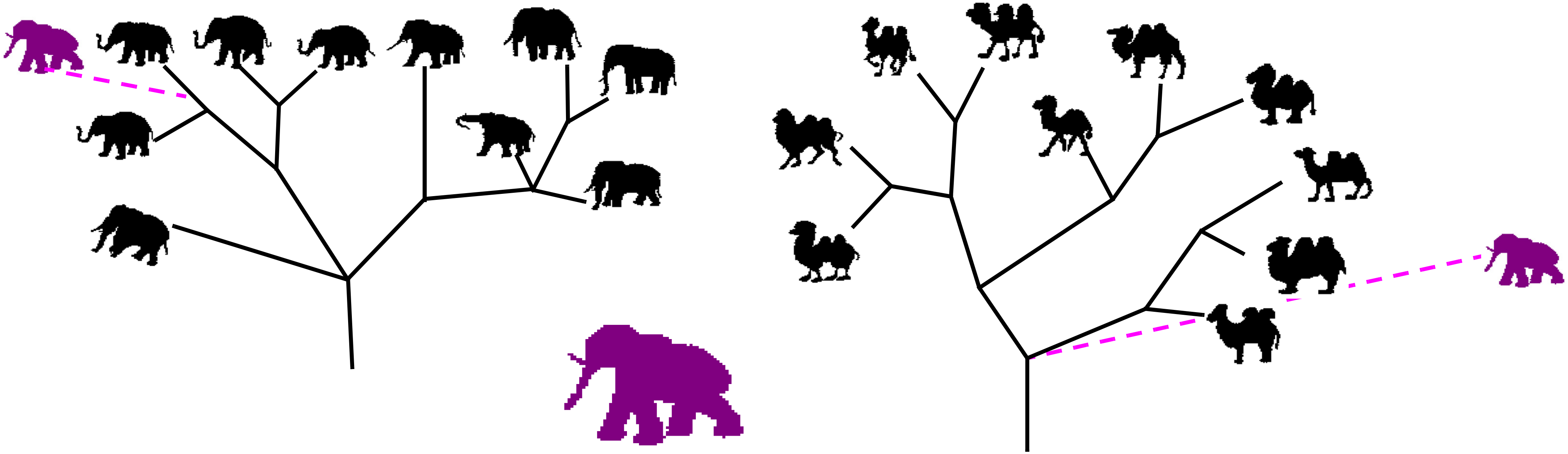
How much smoothing is needed to make shapes similar?

Structural inference: generation



Can we generate a given elephant from a set of primitives and rules induced from a training set?

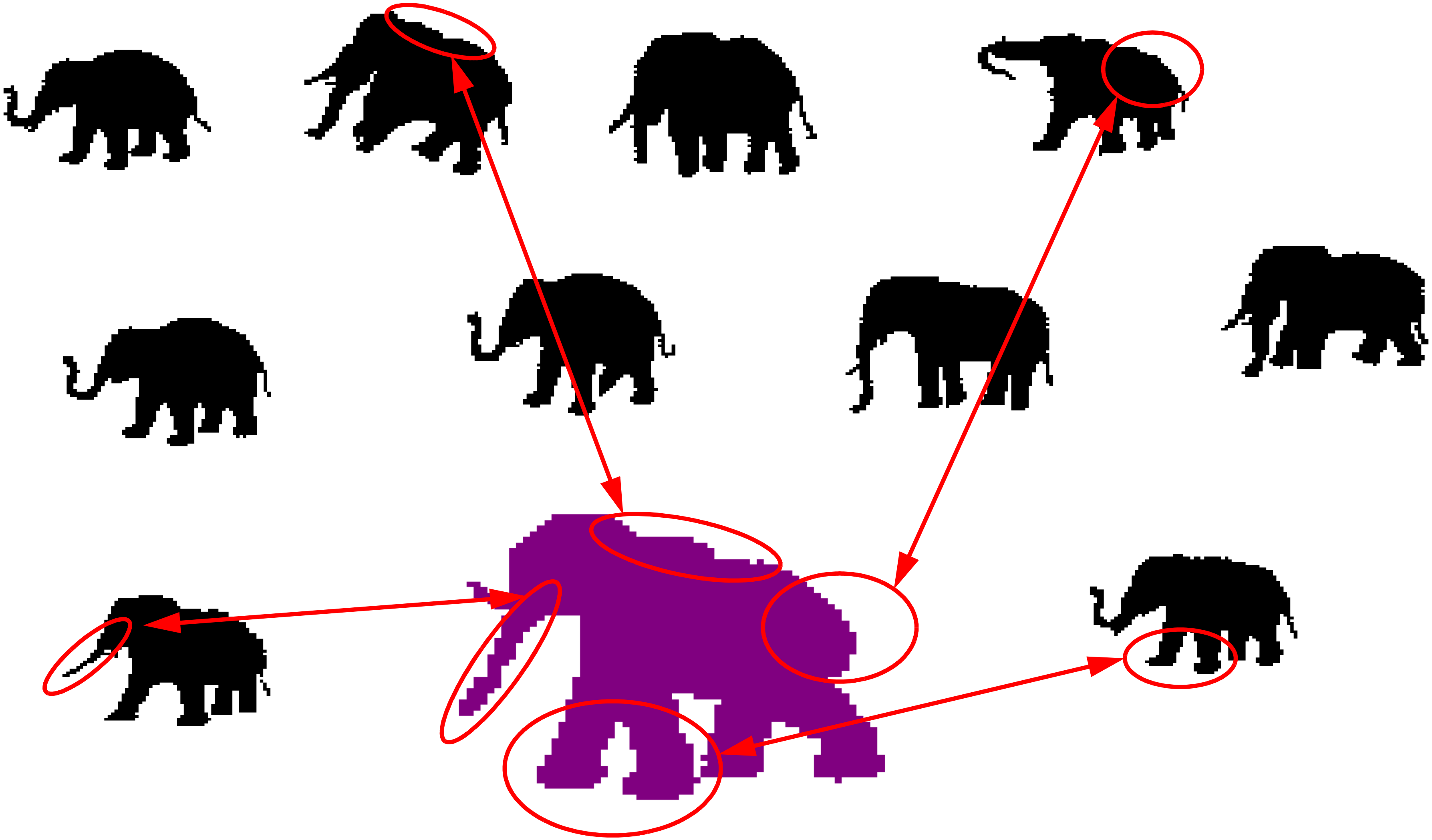
Goldfarb's Evolving Transformation System



Lev Goldfarb's Evolving Transformation System (ETS):

- Generate for each class how the objects could evolve from common primitive objects
- Test how new objects could most easily evolve from the generated trees

Structural inference: part by part



Conclusions

A small feature set reduces, causes overlapping classes, needs a probabilistic approach and bears no structural information.

A complete feature set may yield separable classes, so domain classification may be possible.

A dissimilarity based approach performs this and may use structural information in the dissimilarity measure.
Non-metric problems may have to be solved.

For real structural inference a similarity based approach seems more appropriate.

done under construction to do