Pattern Recognition for Spectral Imaging

Robert P.W. Duin

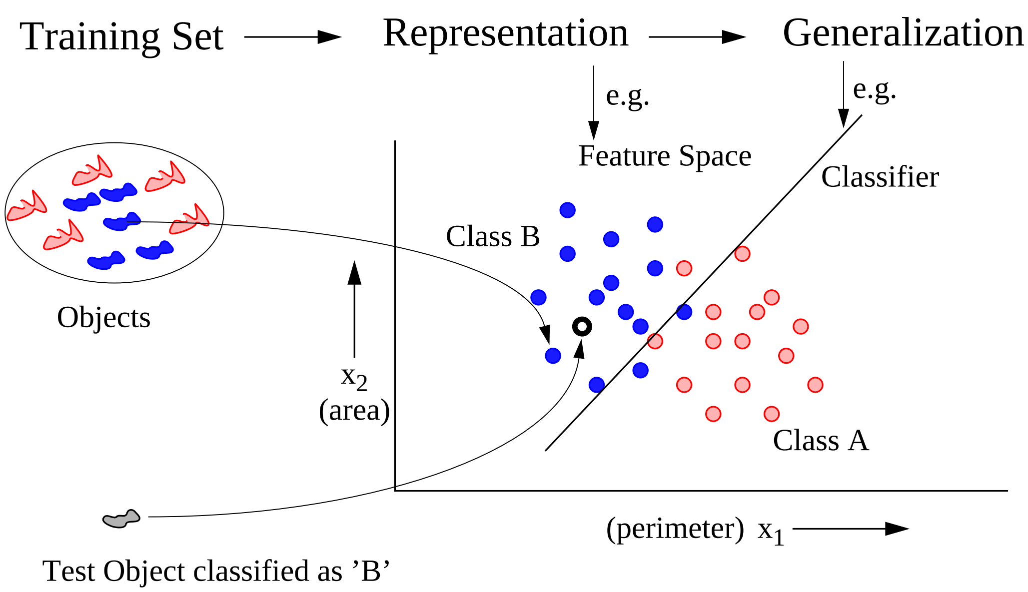
Pattern Recognition Group
Delft University of Technology
The Netherlands

duin@ph.tn.tudelft.nl

Graz, 3 April 2003

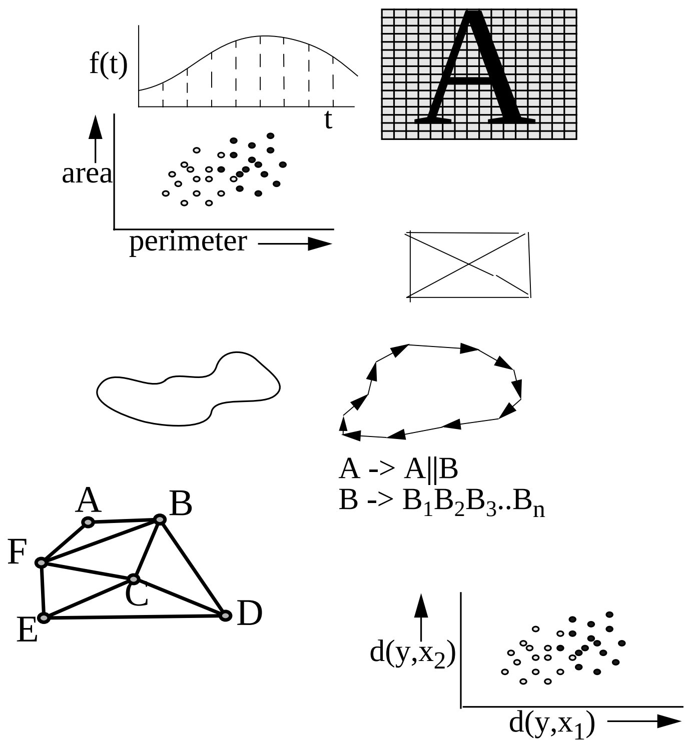
Contents

Pattern Recognition System



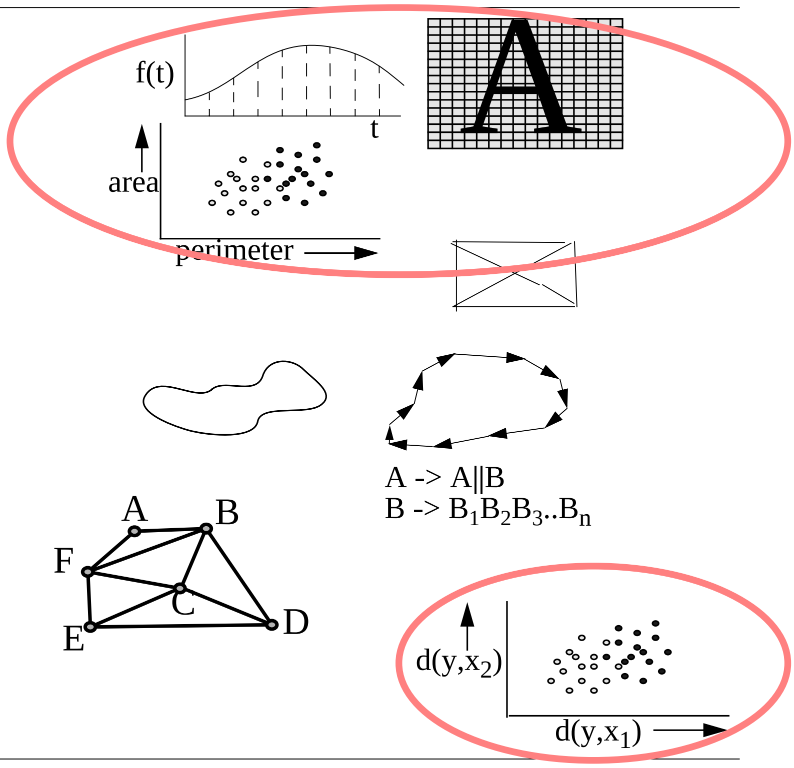
Possible Object Representations

- Measurement samples
- Feature vector
- Sets of segments or primitives
- Outline samples (shape)
- Symbolic structures
- (Attributed) graphs
- Dissimilarities

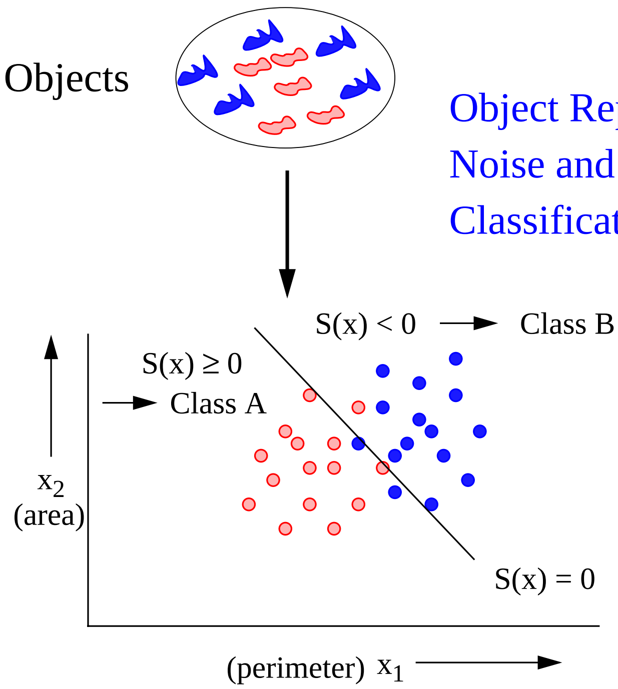


Object Representations

- Measurement samples
- Feature vector
- Sets of segments or primitives
- Outline samples (shape)
- Symbolic structures
- (Attributed) graphs
- Dissimilarities



Feature Space - Discriminant Analysis



Object Representation in Feature Space

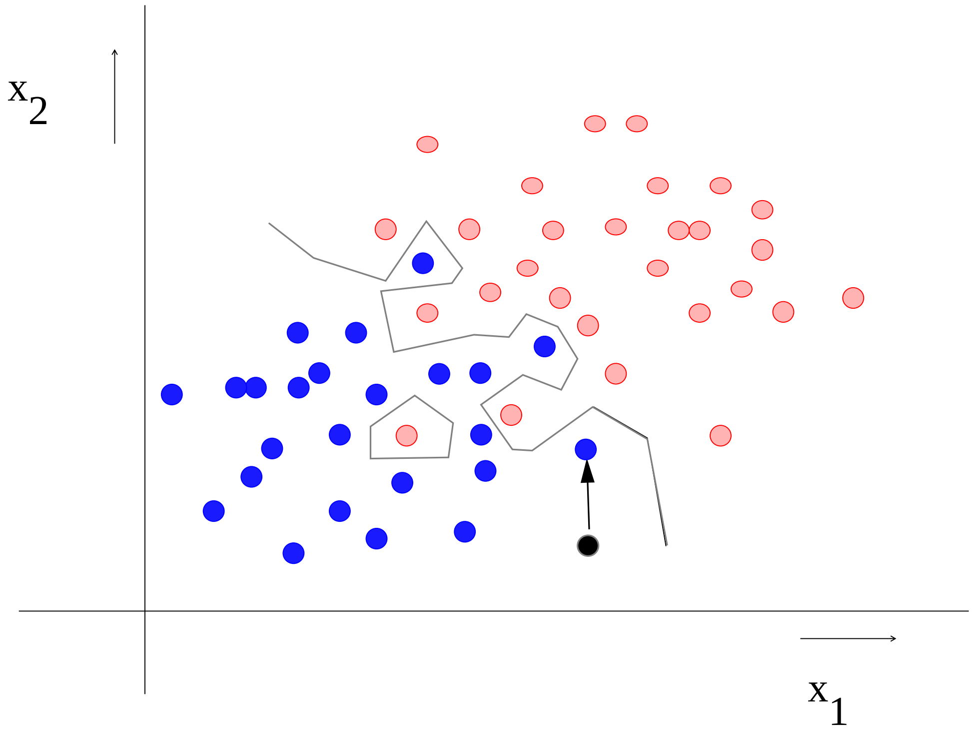
Noise and Biological Variations Cause Class Spread

Classification error due to class overlap

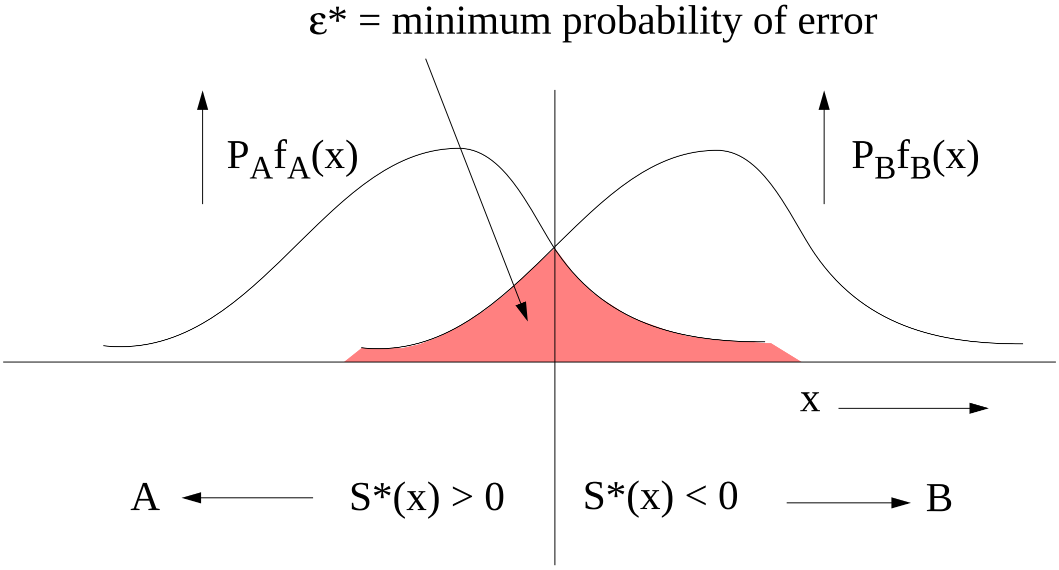
Find classifiers by

1. Distances
2. Densities
3. Error minimization

Nearest Neighbor Classifier

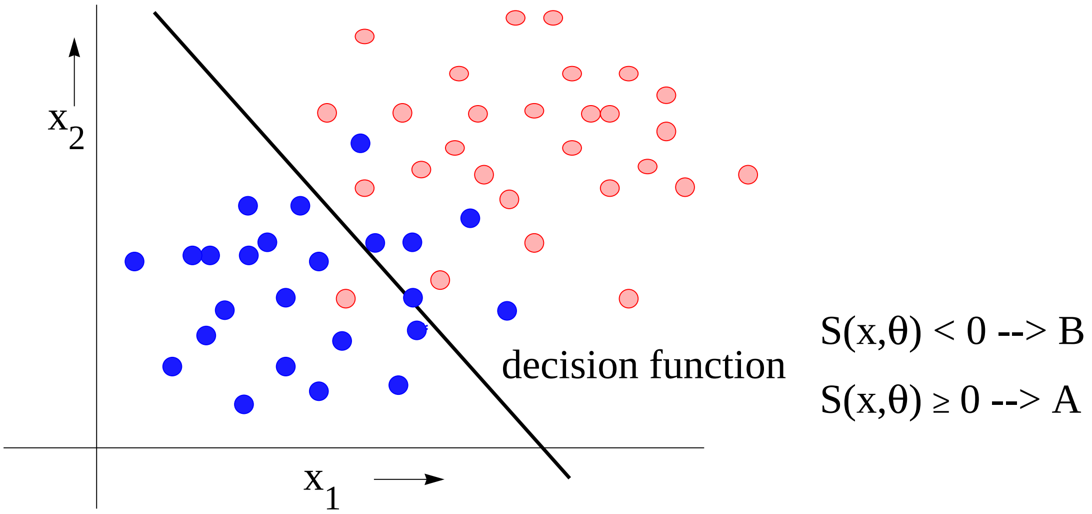


Probabilistic Approach: Bayes Rule



$$S^*(x) = P_A f_A(x) - P_B f_B(x) = 0$$

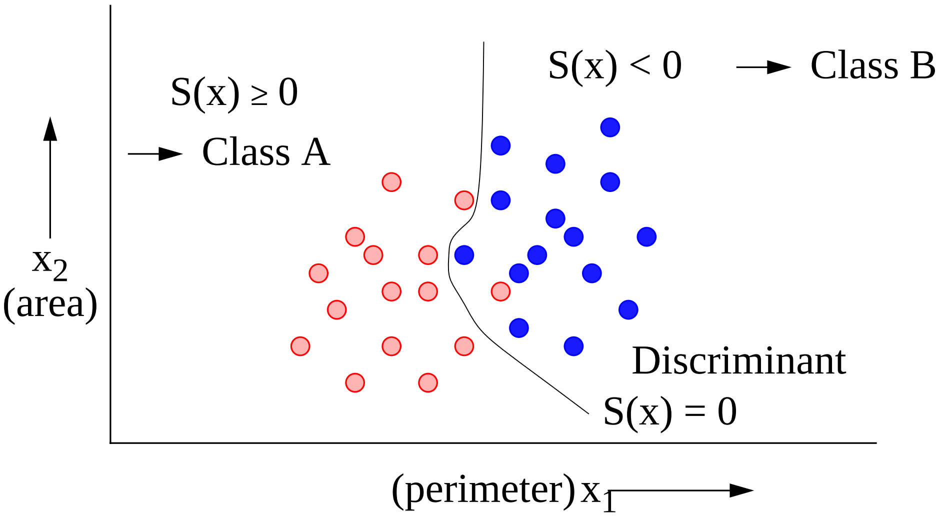
Error Minimization



Change the parameters θ of the decision function such that the error is minimized

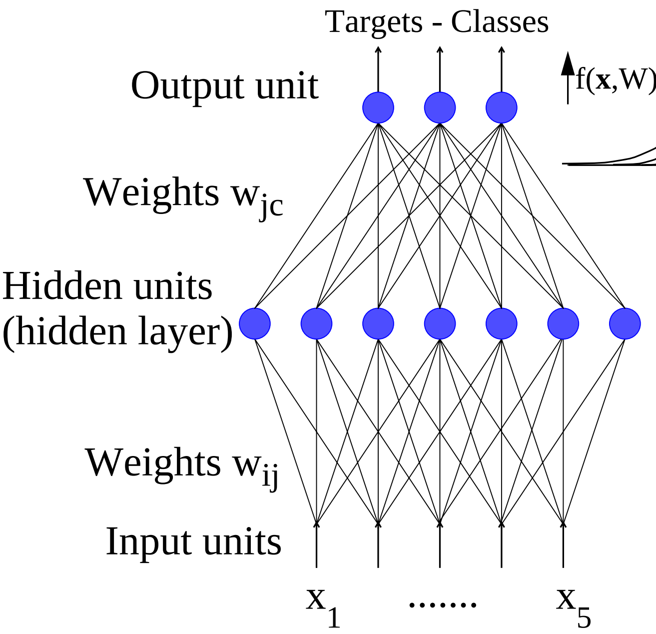
$$\text{Error criterion: } J(\theta) = \sum_{x \in \text{Training set}} C(S(x,\theta)), \text{ e.g. error counting}$$

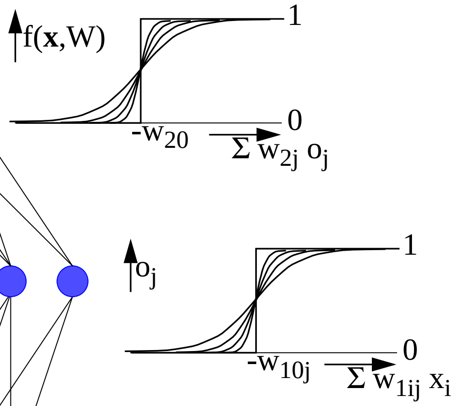
Non-Linear Classification



--> Neural Networks

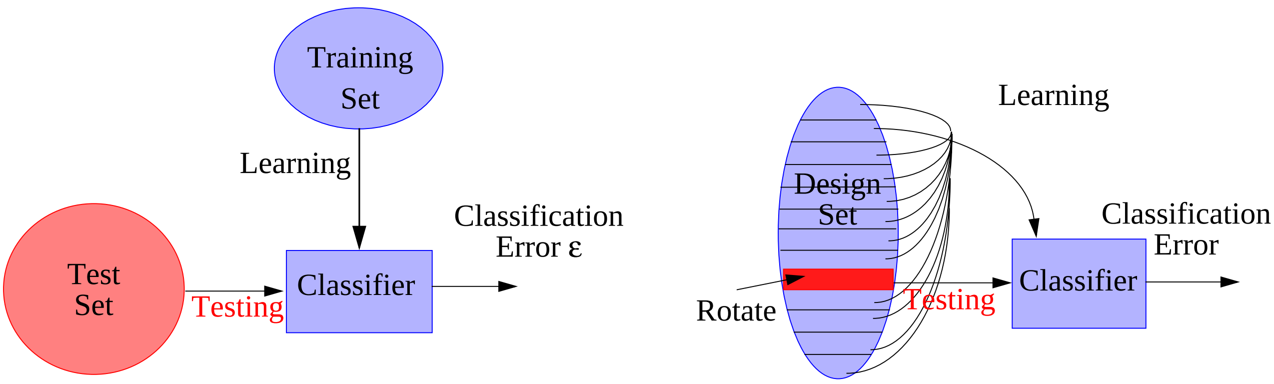
Neural Network Classifiers



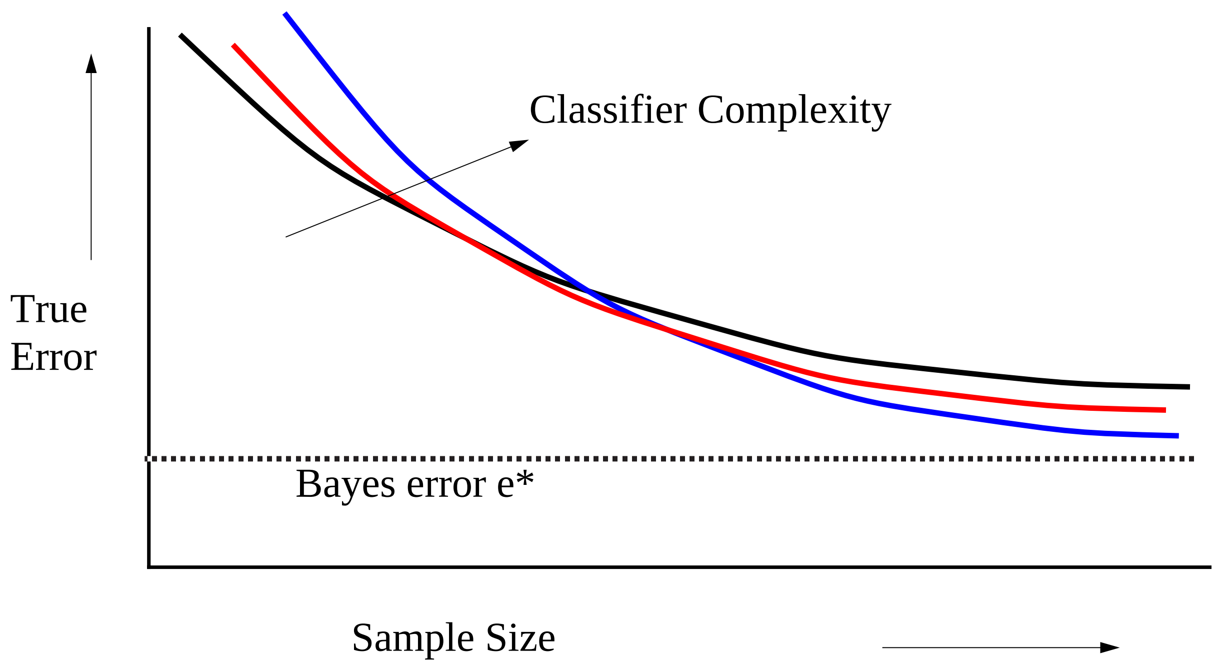


Weights in all layers
Normalized outputs (0,1) in all layers
More output units are possible
More hidden layers are possible.

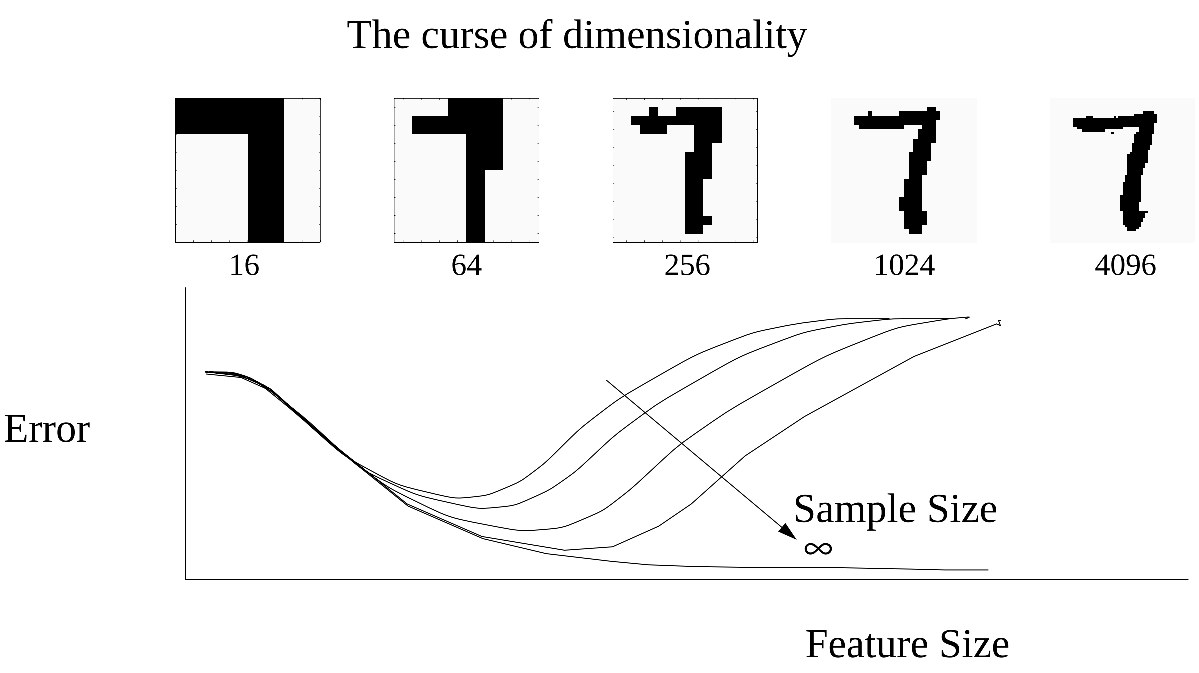
Classifier Evaluation



Learning Curves



Feature Curves: Peaking Phenomenon



Feature Selection

Reduction K Feature to N < K Features:

1. Define an evaluation criterion for some feature set:
   e.g. Divergence, Mahalanobis Distance, Classification Error.
2. Define a Strategy:
   e.g. Individual Selection, Forward, Backward, Branch & Bound, Floating.
3. Run.

For large features sets:

very time consuming to learn.

suboptimal.

doubtful whether a small set of features will do.

Feature Extraction

Find a small set of (non)linear combinations of given features:

1. Define an criterion, e.g. variance, Fisher distance.
2. Select the optimal combinations in parallel,
   directly, e.g. by eigenvalue decomposition
   iteratively, e.g. by some optimization procedure
   or sequentially, one by one.

For large feature sets:

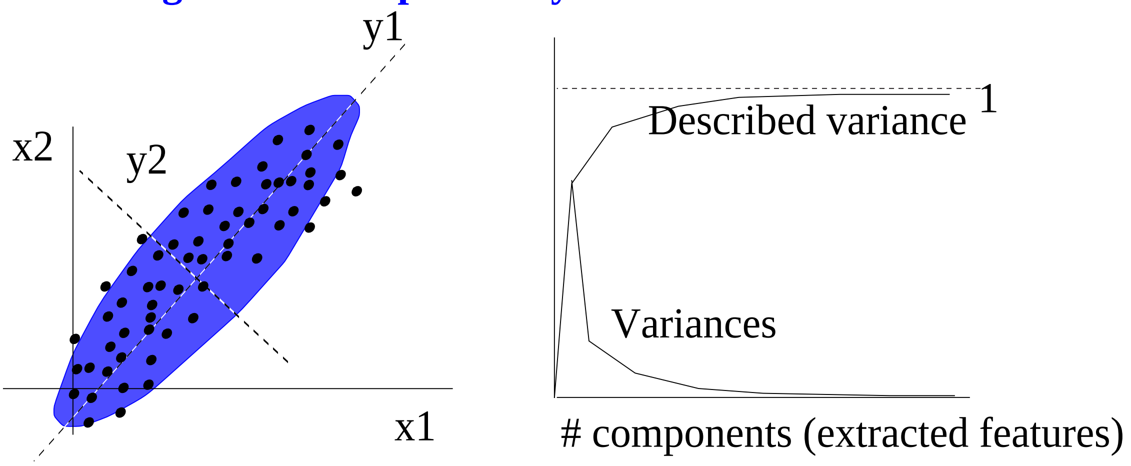
very time consuming to execute

Feature Extraction by Principal Component Analysis

Maximizes the preserved variance

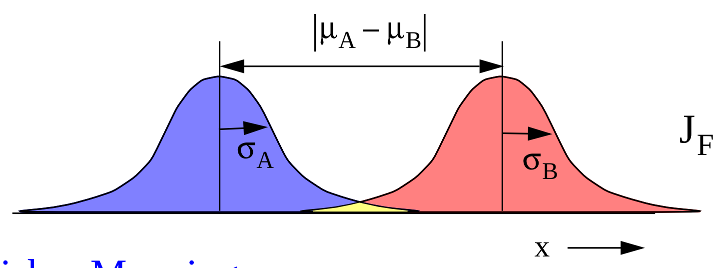
Not feasible for thousands of features

Neglects class separability



Fisher Mapping (1)

e.g. Fisher Criterion:



$$J_F(x) = \frac{|\mu_A - \mu_B|^2}{\sigma_A^2 + \sigma_B^2}$$

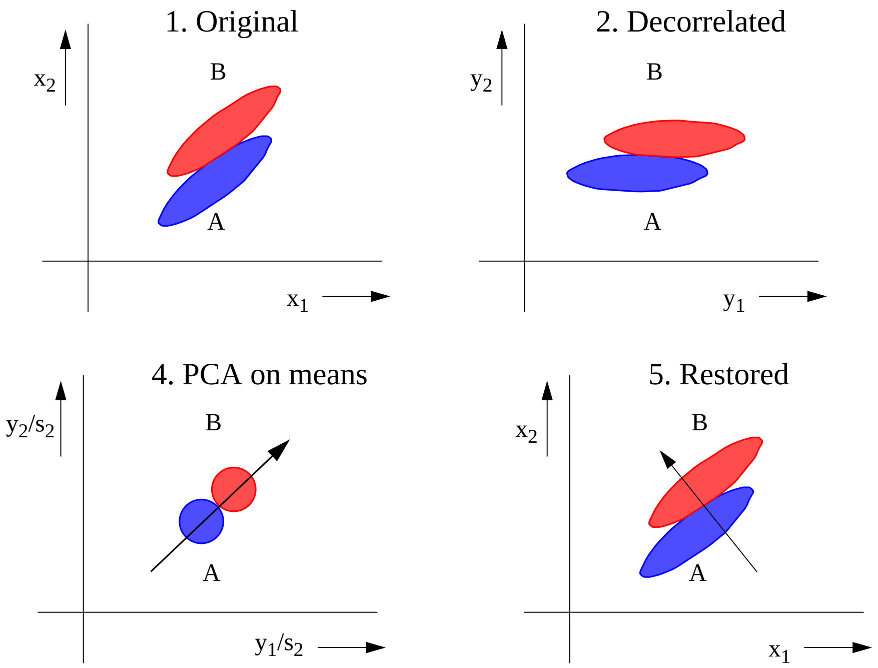
Fisher Mapping:

Find the direction that maximizes the sum of all pairwise Fisher criteria over the set classes.

Repeat perpendicular to established directions.

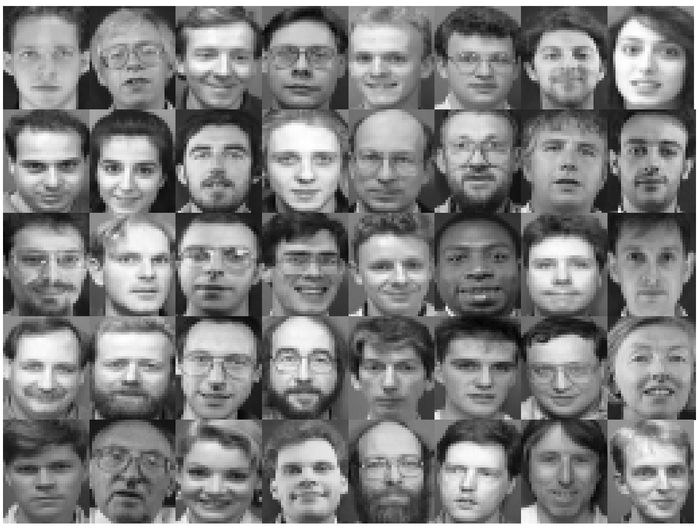
In total c-1 directions are found for c classes.

Fisher Mapping (2)



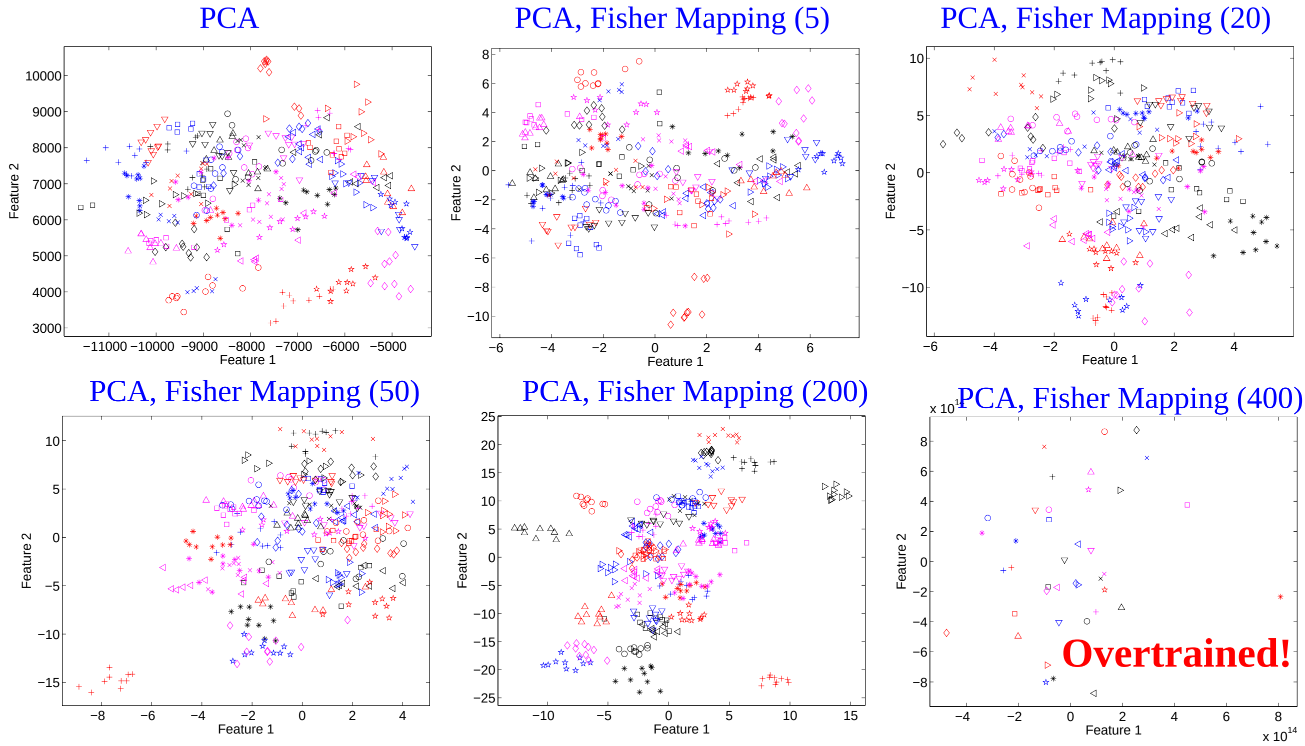
Eigenvectors of $S_w^{-1} S_B$

Face Recognition Example



40 classes, 10 images per class, 92 x 112 pixels

PCA <--> Fisher Mapping

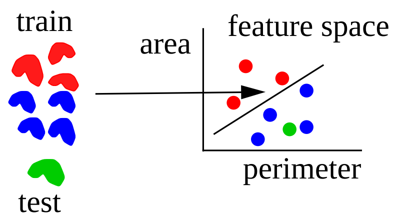


Representations

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

distributions, connectivity neglected

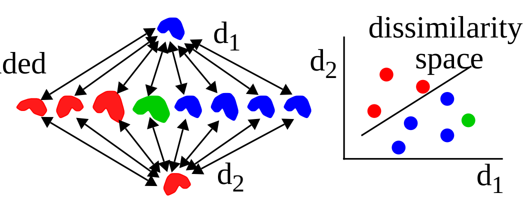
traditional



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

distributions or domains, connectivity possibly included

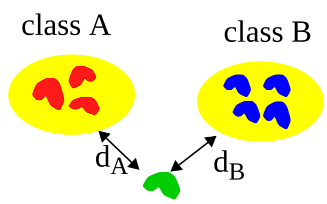
proposed by us in 1997



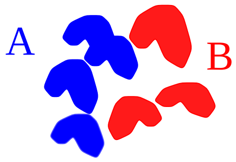
<u>Conceptual</u>, by dissimilarities between objects and classes

domains, structural, connectivity included

new, inspired by Goldfarb



The Dissimilarity Representation



Given labeled training set T

Unlabeled object x to be classified

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

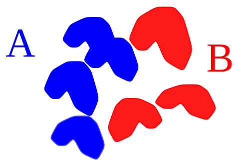
$$d_x = (d_1 \; d_2 \; d_3 \; d_4 \; d_5 \; d_6 \; d_7)$$

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

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

without using DT. Can we do any better?

Approaches: Nearest Neighbor Rule



Given labeled training set T

Unlabeled object x to be classified

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

$$\text{class}(x) = \text{label} (\text{argmin}(d_i))$$

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

Approaches: Dissimilarity Space

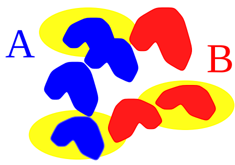
Dissimilarities

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

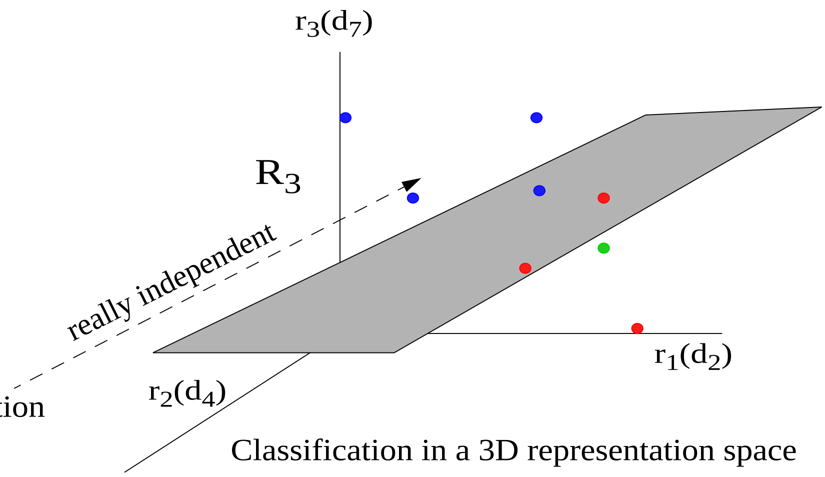
$$d_x = (\, d_1 \;\; d_2 \;\; d_3 \;\; d_4 \;\; d_5 \;\; d_6 \;\; d_7 \,)$$

Selection of 3 objects for representation

Given labeled training set T

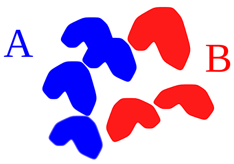


Unlabeled object x to be classified



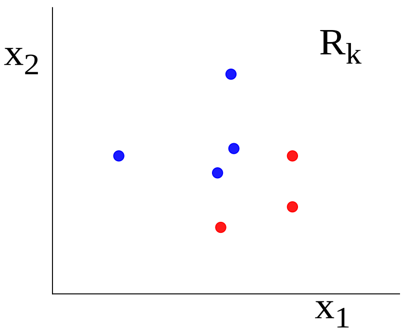
Classification in a 3D representation space

Embedding the Dissimilarity Representation



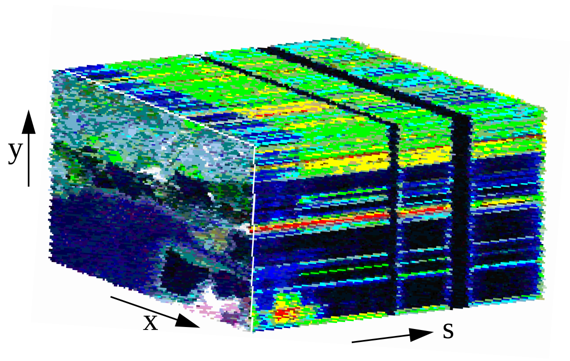
Training set → Dissimilarity matrix D → X?

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



Euclidean distances → D

Spectral Images



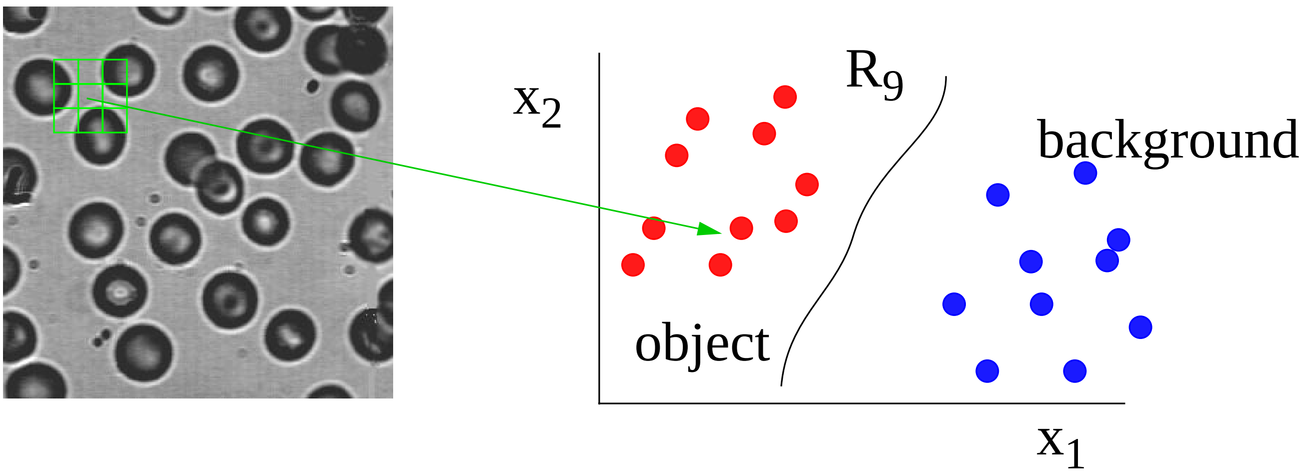
Pixel-band representation: very high-dimensional (~ 10 000 000)

Probably very redundant: need for other representations

Image pixels can be well represented by their spectra

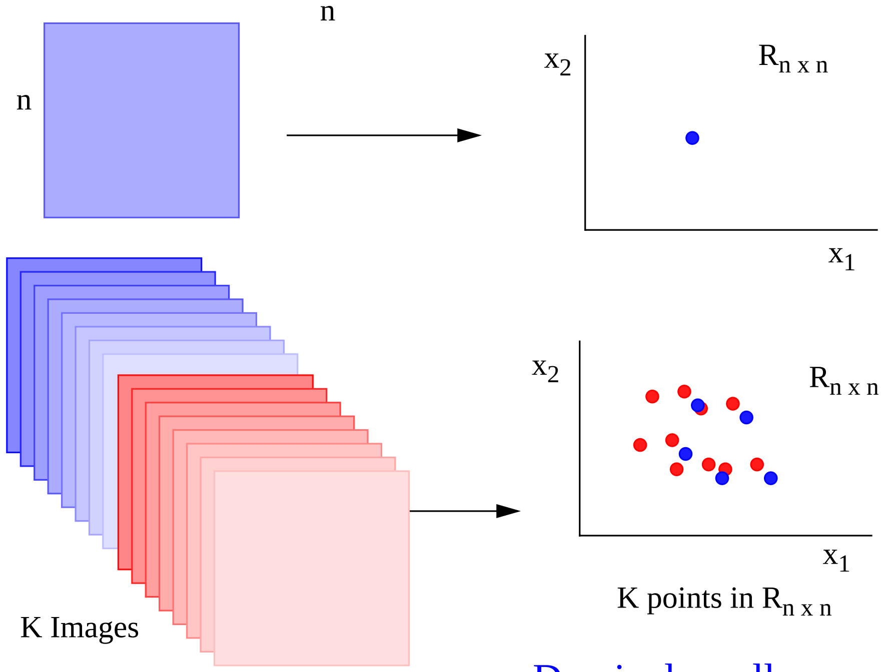
(unlike grey value images)

Pattern Recognition Task 1: Image Segmentation



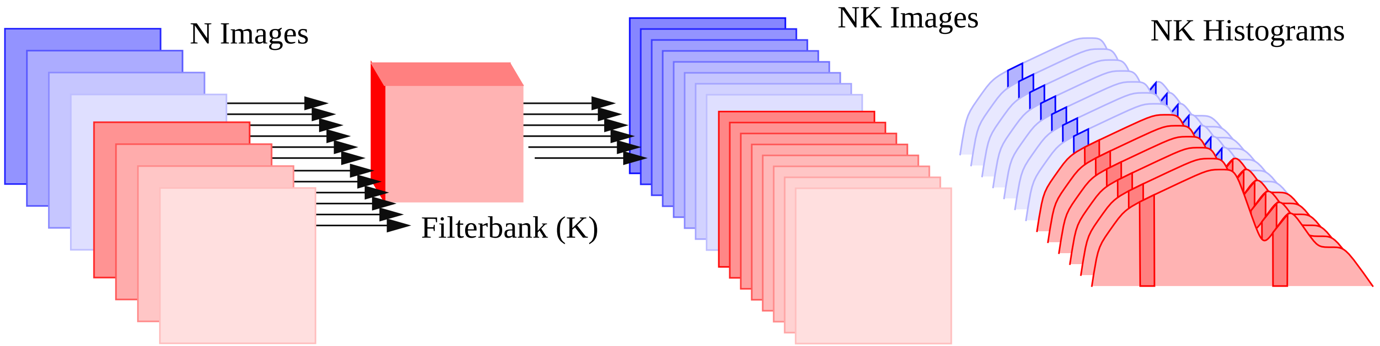
In grey value images the pixel neighborhood is <u>needed</u> for its characterization!
Does it help in spectral images??

Pattern Recognition Task 2: Image Recognition

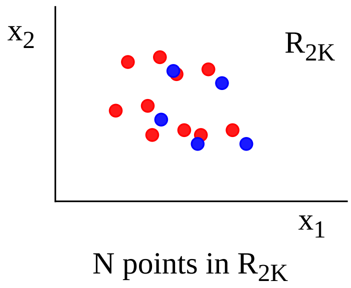


Do pixels really correspond?

Image Recognition (2)

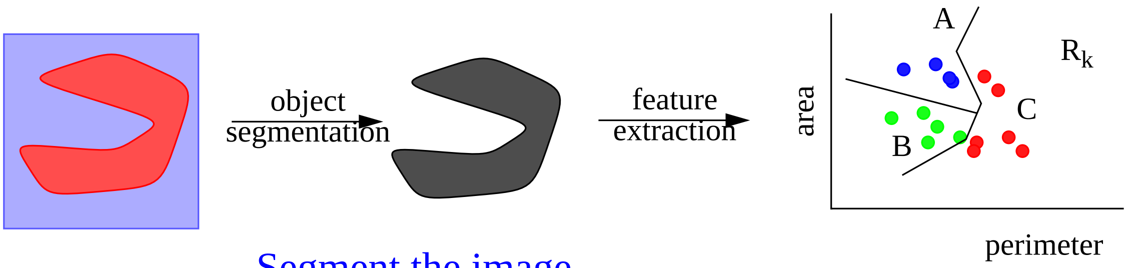


1. Characterize pixels by their neighborhood (feature filters)
2. Characterize feature filters by their histogram over the image
3. Characterize images by some bands in the histograms (here 2)



In spectral imaging the first step is not needed.
Each spectral band is already a pixel feature.

Object Recognition

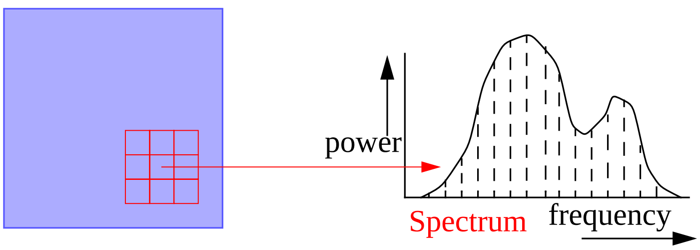


Segment the image

Find features for the segment

Classify

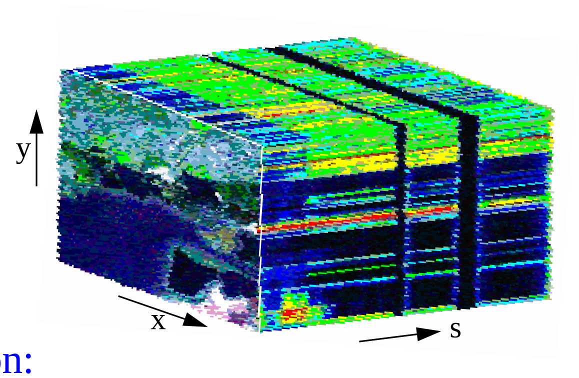
Spectral Pixel Representation



Pixel Representation:

By the amplitudes of all bands in its spectrum.
By characteristics (features) of its spectrum.
By the similarities of its spectrum with other spectra.
By the one of these extended with some neighborhood properties.

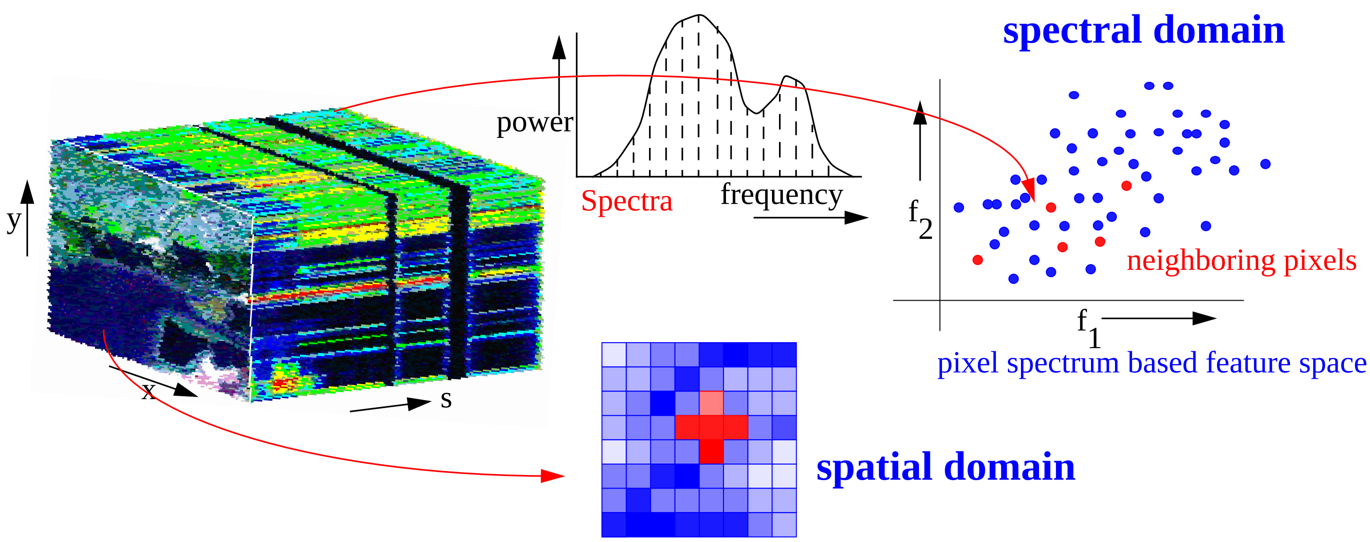
Spectral Image Representation



Spectral Image Representation:

Pixel based:	by all pixels
Histogram based:	by frequencies of particular pixel characteristics
Object based:	by characteristics of image segments
Image based:	by similarities with other images

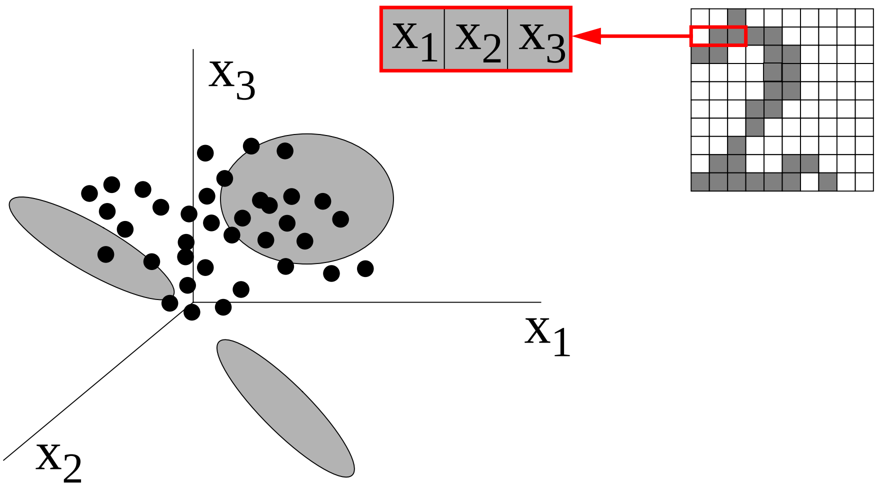
Spatial and Spectral Connectivity



In sample based representations the connectivity is lost between the representation domains, but also between neighboring samples.

Problems with the Pixel Based Image Representation - 1

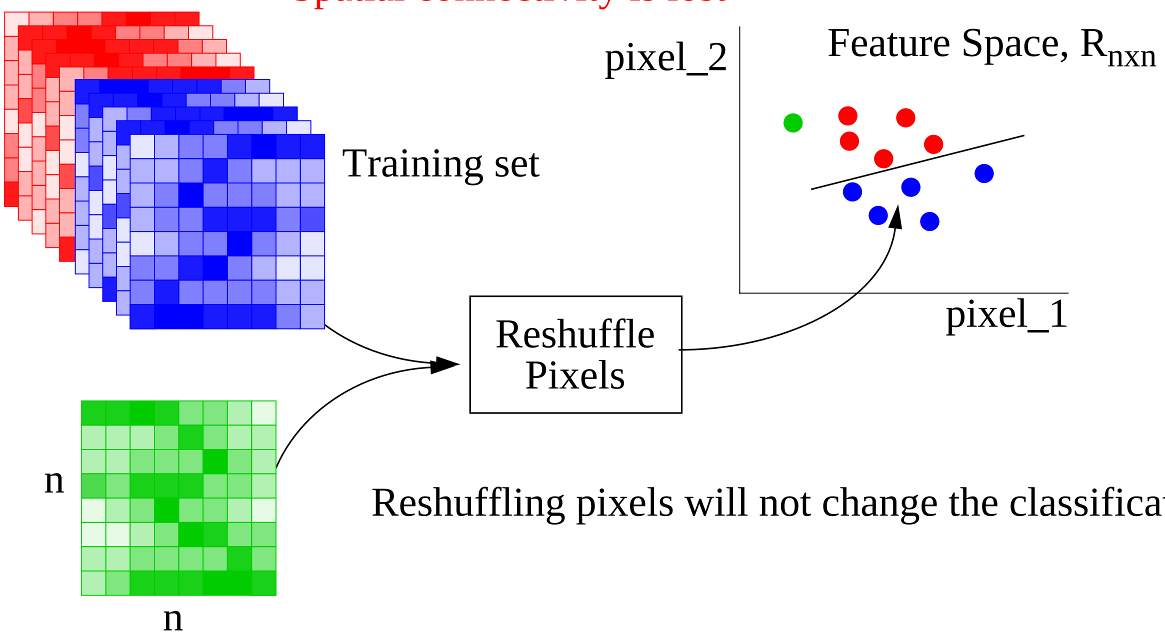
Spatial connectivity is lost



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

Problems with the Pixel Based Image Representation -2

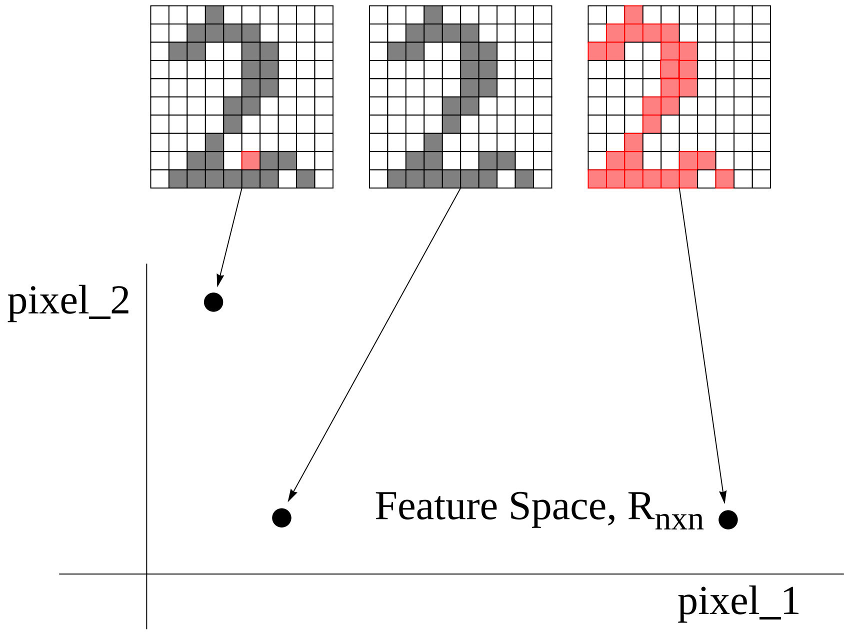
Spatial connectivity is lost



Reshuffling pixels will not change the classification

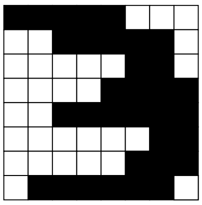
Problems with the Pixel Based Image Representation - 3

Representation jumps after small disturbances

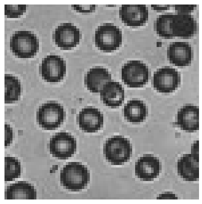


Non-pixel Based Representations Needed for Connected Measurements

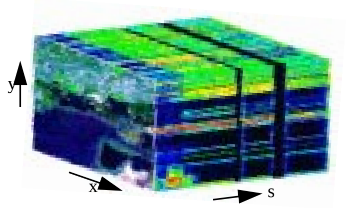
binary images

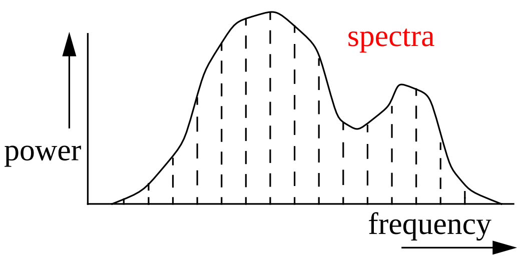


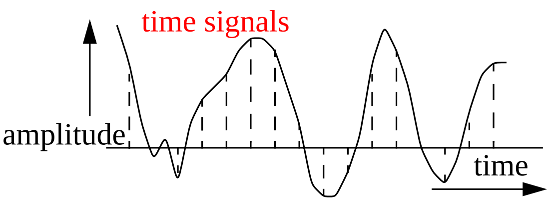
grey value images



hyperspectral images







What are representations that:

- use of the connectedness
- enable the integration of more knowledge

How can they be optimized for

- knowledge?
- for the data?
- for recognition (learning and testing)

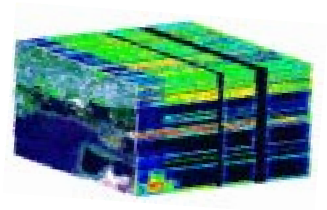
Performance in accuracy (and time)

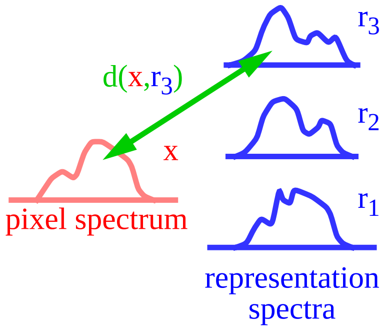
Dissimilarities for Spectral Images

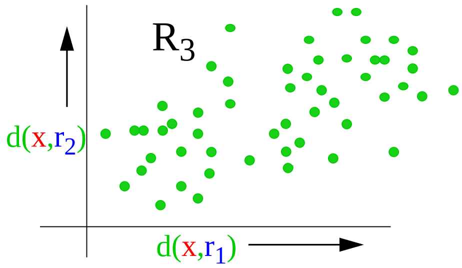
Relative, dissimilarity representations for solving the connectivity problem:

- select an appropriate dissimilarity d(x,r) measure between spectra
- select a few (e.g. 3) pixels or a few standard spectra (r) for representation
- compute all dissimilarities d(x,r) with all pixel-spectra x

--> pixels in R_r for segmentation (clustering) or image recognition







Discussion

Spectral images have well defined pixels,

so less need for characterization by their neighborhood.

Amount of data is high, but possibly redundant.

Preservation of the connectivity of spectra and images during the analysis

is recommended, but still not established.

Dissimilarity representations are promising.