Generalized dissimilarity representations for pattern recognition

Mauricio Orozco Alzate
UNIVERSIDAD NACIONAL DE COLOMBIA SEDE MANIZALES
`http://www.docentes.unal.edu.co/morozcoa/docs/`

In cooperation with Robert P.W. Duin
DELFT UNIVERSITY OF TECHNOLOGY

Talk at Pontificia Universidad Javeriana, Cali
OCTOBER 9, 2009

Outline

2009/10/09

- Introduction, problem statement and objectives
- Dissimilarity representations and generalizations
- Generalization of dissimilarity representations by feature lines and feature planes
  - Nearest feature classifiers
  - Generalizing dissimilarity representations by feature lines and feature planes
- Volcanic seismology and pattern recognition
  - Dissimilarity-based classification of seismic signals
  - Clustering of seismic signals for detecting mislabeling
  - Band selection of seismic spectra
- Conclusions, open questions and future work
- Publications

This presentation

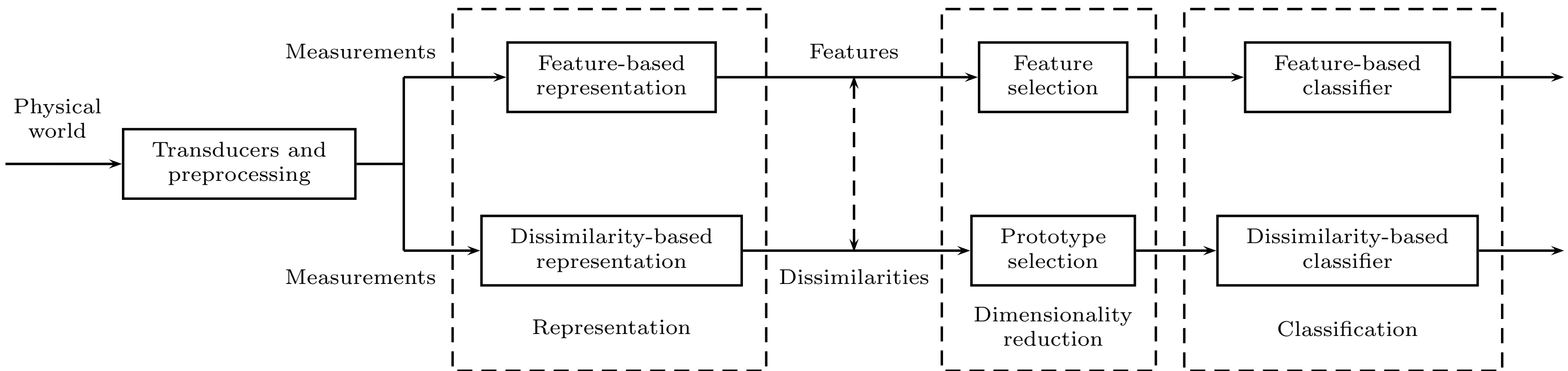
Framework: Pattern recognition. In particular, statistical pattern recognition.

Main part: Generalization procedure for the so-called dissimilarity-based approach. It covers topics related to our proposed procedure for generalizing dissimilarity representations by using the principles of nearest feature lines and planes. The strengths and weaknesses of our generalization approach are discussed and demonstrated with experiments.

Second part: Collection of our research endeavors in classification of seismic-volcanic signals, not just from the dissimilarity-based perspective but also using other pattern recognition techniques.

The issue of representation

In Statistical PR, the issue of how to properly represent measurements, to be provided to the subsequent stages of the system, has been traditionally dealt by defining a set of **features** arranged as a vector: $\Rightarrow \boldsymbol{x} = (x_1, x_2, \cdots, x_m)$

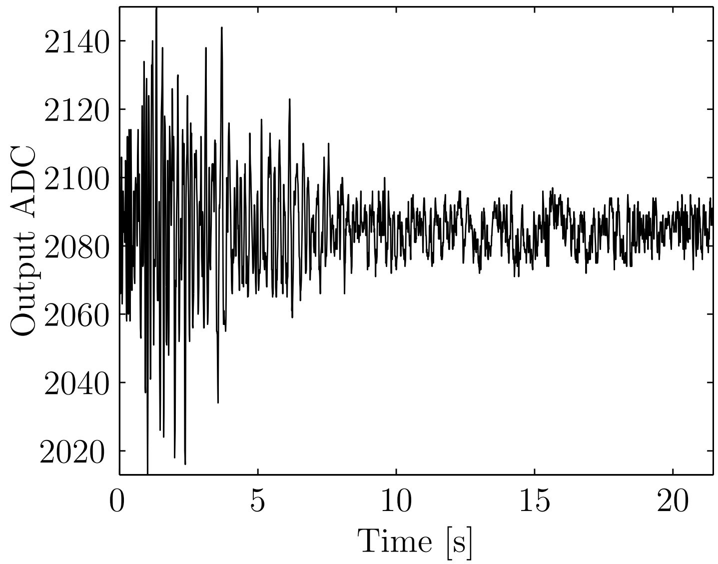


An alternative approach for representing objects in PR is the so-called dissimilarity-based approach. It consists in representing the objects in terms of their **(dis)similarities** to a set of prototypes; consequently, it is a relative representation, i.e. based on the relations between the objects.$\Rightarrow$
$D(x, R) = [d(x, p_1), d(x, p_2), \ldots d(x, p_n)]$, $R = \{p_1, p_2, \ldots, p_n\}$.

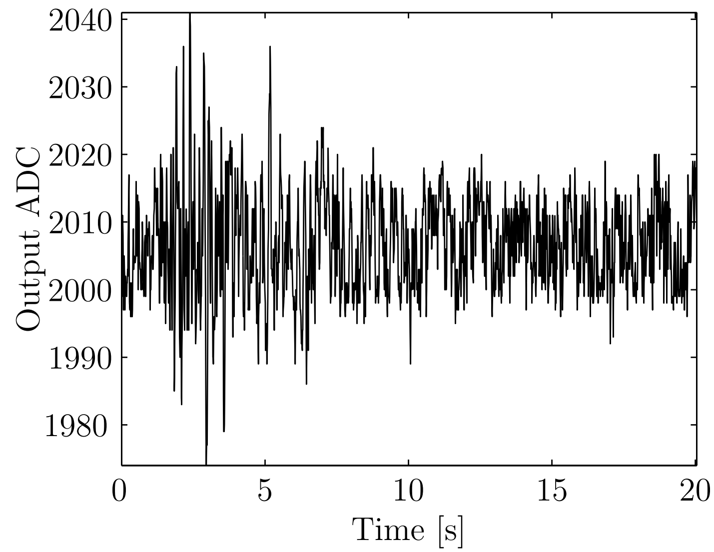
A practical example

2009/10/09

(a) Volcano tectonic earthquake (VT)



(b) Long period earthquake (LP)



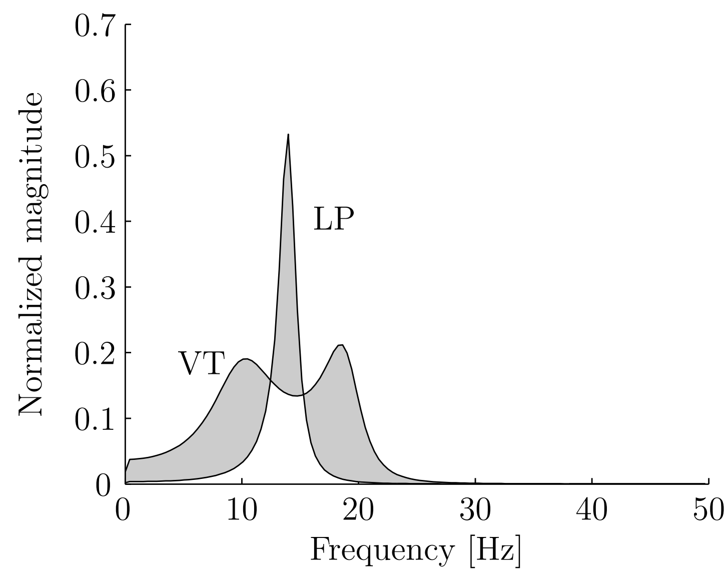
Time domain

⇓

Frequency domain (spectrum)

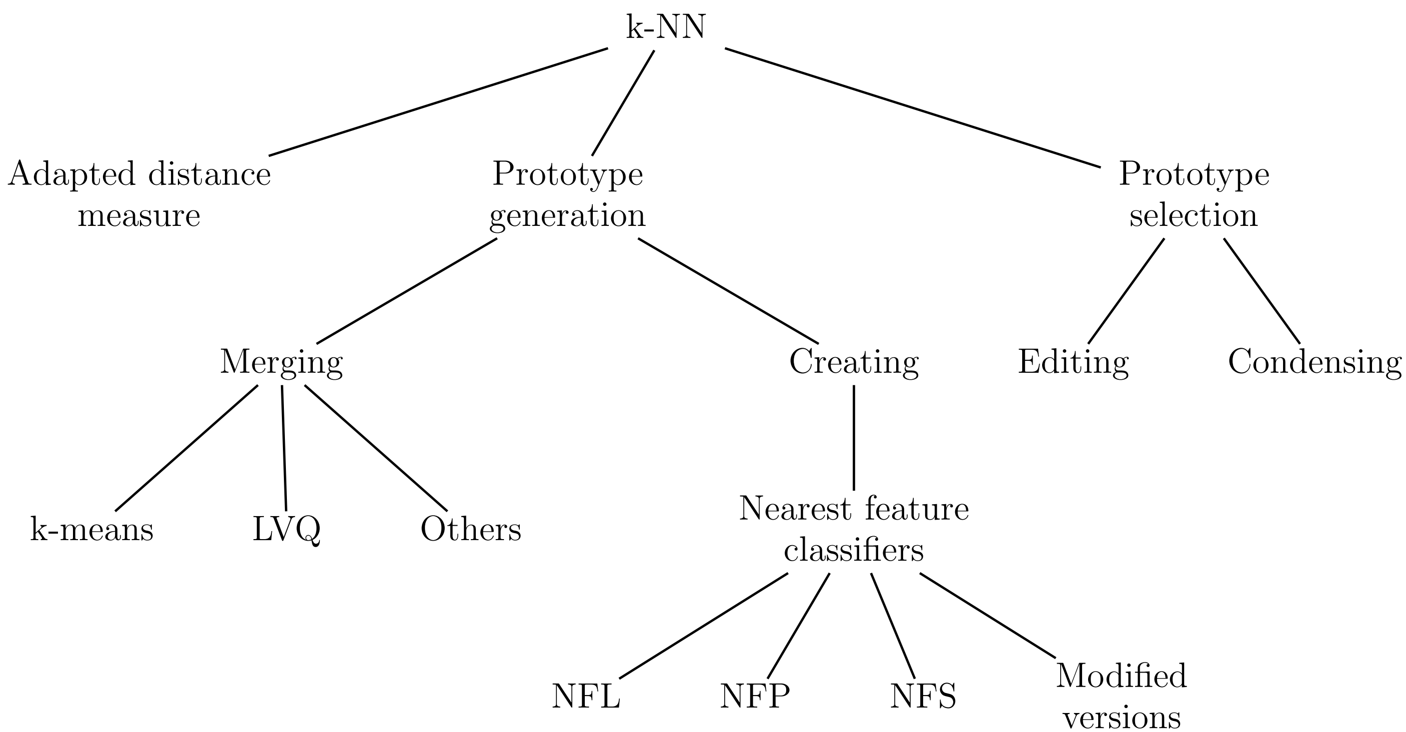
⇓

Dissimilarity representation



Further considerations in representation

- Curse of the dimensionality $\Rightarrow$ peaking phenomenon
- Feature extraction/selection
- Prototype generation/selection



Motivation for studying representation issues

Some reasons are [Duin and Pękalska, 2005]:

- Representation was almost neglected and reduced to the demand of having good features provided by some expert.
- The learning is often believed to start at the given feature vector space
- Representation is application or domain-dependent (not easy!)
- If the starting point of a pattern recognition problem is not well defined, this cannot be improved later in the process of learning.

A radical novelty in the field of PR is just possible by radically new forms of representation [Goldfarb, 2004b, Goldfarb, 2004a]

Importance of studying dissimilarity representations

Using dissimilarity representations is motivated by a number of reasons [Pękalska and Duin, 2005]:

- The notion of similarity is inherent to the concept of a class
- Dissimilarities can be interpreted as distances in vector spaces
- Dissimilarities can be directly computed from row or pre-processed measurements as well as from a weighted combination of features or from non-numerical representations such as graphs and grammars.
- The dissimilarity approach is much more flexible than the one based on features
- The dissimilarity approach is aimed at unifying the statistical and structural approaches

Problem statement

In spite of the recent research advances on dissimilarity representations, the work is not completed yet; particularly, meaningful transformations and manipulations of dissimilarity representations are still an open and promising field to be further researched.

Manipulations to enrich the original dissimilarity representations might be useful; e.g. by using a geometrical, function-based or model-based generalizations [Bicego et al., 2004, Kim, 2006, Orozco-Alzate et al., 2007, Orozco-Alzate et al., 2008a].

Generalization:
Object x is defined as a set of dissimilarities respect to a collection of objects $R := \{p_1, p_2, \ldots, p_n\}$ and expressed as a vector $D(x, R) = [d(x, p_1), d(x, p_2), \ldots, d(x, p_n)]$, is generalized by considering a new set R_g composed by objects lying in another space, e.g. lines, planes, functions or models. Subindex g stands for *generalized* representation.

Problem statement

Generalization of dissimilarity representations:

Previous attempts:

- Approach by using hidden Markov models [Bicego et al., 2004]
- Approach by pre-clustering [Kim, 2006]

They are aimed to produce a condensed dissimilarity representation, that is, to attack the small sample size problem.

Our proposal: Under representational limitations, our approach is intended to enrich a given dissimilarity representation by using the principle of the nearest feature rules. $\Rightarrow$ We end up in higher dimensional spaces.

Our main purpose is enriching the representation and exploiting it by geometric generalizations.

Objectives

- To develop an approach for generalizing dissimilarity representations by applying the principle of the nearest feature rules; in order to enrich a given dissimilarity representation under representational limitations.

- To identify cases for which generalized dissimilarity representations might be advantageous or, conversely, cases for which generalized dissimilarity representations might not be useful.

- To compare the structure of dissimilarity representations against the generalized ones

- To explore procedures for learning from such generalized dissimilarity representations, compared to learning from the non-generalized ones as well as compared to the nearest neighbor method.

Objectives

- Since generalized dissimilarity representations are, in general, higher dimensional spaces, we also consider a number of alternatives for dimensionality reduction and/or prototype selection.

In addition:

- New applications should be considered in order to describe other pattern recognition problems where both dissimilarity representations and generalized dissimilarity representations might be advantageous.

- Particularly, due to our cooperation with the Volcanological and Seismological Observatory at Manizales, we investigate the problem of (dissimilarity based) classification of seismic-volcanic signals.

Methodology

- Study of properties of the nearest feature classifiers: complexity, asymptotic behavior, experimental comparison
- Development of the generalization procedure by feature lines and feature planes: mathematical formulation
- Construction of classifiers in generalized dissimilarity spaces
- Experiments on artificial and real-world data (biosignals and seismic data)
- Comparison, discussion of results and conclusions

Outline

- Introduction, problem statement and objectives
- Dissimilarity representations and generalizations
- Generalization of dissimilarity representations by feature lines and feature planes
  - Nearest feature classifiers
  - Generalizing dissimilarity representations by feature lines and feature planes
- Volcanic seismology and pattern recognition
  - Dissimilarity-based classification of seismic signals
  - Clustering of seismic signals for detecting mislabeling
  - Band selection of seismic spectra
- Conclusions, open questions and future work
- Publications

Dissimilarity representations

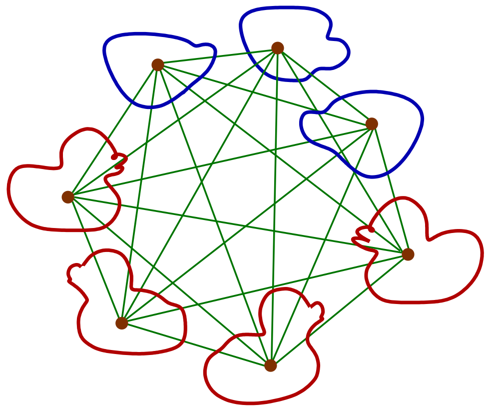


Illustration taken from [Pękalska and Duin, 2005]

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

Relative representation based on a dissimilarity measure d_{ij} between the raw data of objects i and j.

Training set of N objects: $T = \{x_1, x_2, \ldots, x_N\}$.

Representation set of n prototypes: $R = \{p_1, p_2, \ldots, p_n\}$. Often $R \subseteq T$.

A set S of new test objects is provided in terms of their dissimilarities to R: $D(S, R)$.

Dissimilarity representations

- A number of studies showed advantages of learning from dissimilarity representations instead of learning from feature representations: [Pękalska and Duin, 2002, Pękalska and Duin, 2005].

- More general than feature-based representations; in fact, the notion of dissimilarity is more fundamental than that of a feature.

- For dissimilarities the geometry is contained in the definition, giving the possibility to include physical background knowledge; in contrast, feature-based representations usually suppose a Euclidean geometry.

Classification in dissimilarity spaces

Traditional (feature-based) 1-NN rule is equivalent to find label($\text{argmin}_{\text{trainset}}(d_i)$) in $D(\cdot, T)$.

The dissimilarity space approach

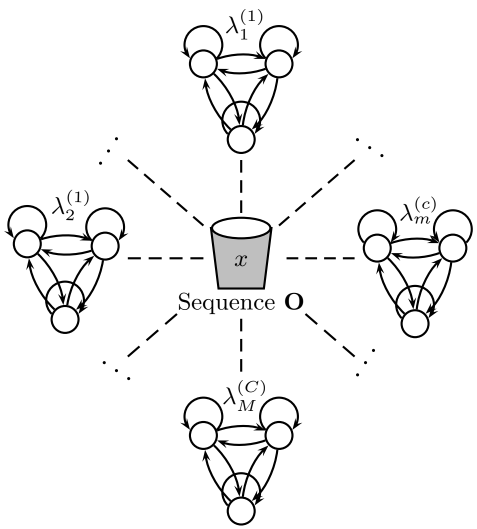
$$D(T,R) = \begin{array}{c c} & \begin{array}{ccccc} p_1 & p_2 & p_3 & \cdots & p_N \end{array} \\ \begin{array}{c} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_N \end{array} & \begin{pmatrix} d_{11} & d_{12} & d_{13} & \cdots & d_{1n} \\ d_{21} & d_{22} & d_{23} & \cdots & d_{2n} \\ d_{31} & d_{32} & d_{33} & \cdots & d_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ d_{N1} & d_{N2} & d_{N3} & \cdots & d_{Nn} \end{pmatrix} \end{array}, \quad d_{jk} = D(x_j, p_k).$$

- Consider dissimilarities as "features".
- Select "features", i.e. representation objects (prototypes)
- Build a trained classifier, e.g. a normal density-based one, on top of the representation $D(T, R)$.
  - Summation-based distances are often approximately (clipped) normally distributed.

Wider definition: representations based on dissimilarities with functions of (or models built by) objects

- In general, representation objects used for building those functions or models do not need labels
- Representation might even be artificially created, selected by an expert or belong to different classes than the ones studied
- In spite of the potential to omit labels, to the best of the author's knowledge, all the current generalization procedures —including ours— make use of them.

Generalization using hidden Markov models



$$D(T, R_\lambda) = \begin{array}{c} \\ \boldsymbol{O}_1 \\ \boldsymbol{O}_2 \\ \boldsymbol{O}_3 \\ \vdots \\ \boldsymbol{O}_N \end{array} \begin{array}{c} \begin{array}{ccccccc} \lambda_1^{(1)} & \lambda_2^{(1)} & \cdots & \lambda_m^{(c)} & \cdots & \lambda_M^{(C)} \end{array} \\ \begin{pmatrix} d_{11} & d_{12} & \cdots & d_{2m} & \cdots & d_{1M} \\ d_{21} & d_{22} & \cdots & d_{3m} & \cdots & d_{2M} \\ d_{31} & d_{32} & \cdots & d_{4m} & \cdots & d_{3M} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ d_{N1} & d_{N2} & \cdots & d_{Nm} & \cdots & d_{NM} \end{pmatrix} \end{array}$$

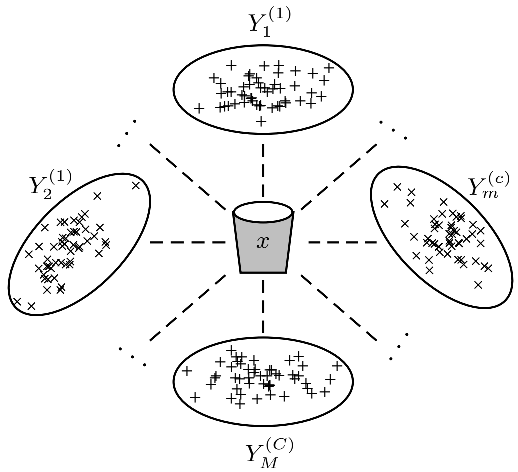
$$D(\boldsymbol{O}, R_\lambda) = [d_1 \quad d_2 \quad \cdots \quad d_m \quad \cdots \quad d_M]$$

Aimed to extend the dissimilarity-based paradigm to the HMM-based classification [Bicego et al., 2004].

Likelihoods $P\left(\boldsymbol{O}|\lambda^{(c)}\right)$ and/or $P(\boldsymbol{O}|\lambda_i^{(c)})$ interpreted as similarities.

$$d_{ij} = d(\boldsymbol{O}_i, \boldsymbol{O}_j) = \frac{\log P(\boldsymbol{O}_i|\lambda_j)}{T_i}; \quad T_i\text{: length of seq. } \boldsymbol{O}_i$$

Generalization by clustering prototypes



$$D(T, R_Y) = \begin{array}{c} \\ x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_N \end{array} \begin{array}{c} \begin{array}{cccccc} Y_1^{(1)} & Y_2^{(1)} & \cdots & Y_m^{(c)} & \cdots & Y_M^{(C)} \end{array} \\ \begin{pmatrix} d_{11} & d_{12} & \cdots & d_{2m} & \cdots & d_{1M} \\ d_{21} & d_{22} & \cdots & d_{3m} & \cdots & d_{2M} \\ d_{31} & d_{32} & \cdots & d_{4m} & \cdots & d_{3M} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ d_{N1} & d_{N2} & \cdots & d_{Nm} & \cdots & d_{NM} \end{pmatrix} \end{array}$$

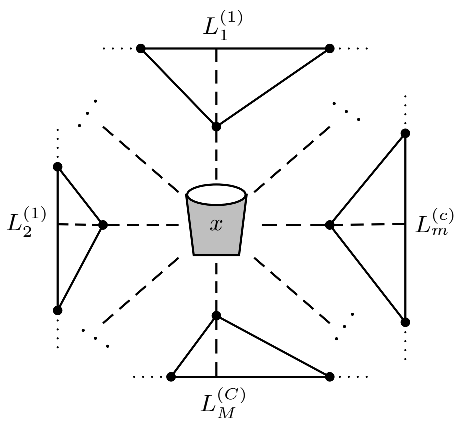
$$D(x, R_Y) = [d_1 \quad d_2 \quad \cdots \quad d_m \quad \cdots \quad d_M]$$

Attempt to reduce dimensionality by choosing means of clusters as representatives [Kim, 2006]

1. For each class, perform a clustering of R into a few subsets $Y_m^{(c)}$, $c = 1, \ldots, C$ and $i = m, \ldots, M$; then, compute the M mean vectors $\overline{Y}_i^c$ by averaging each cluster.

2. A dissimilarity based classifier is built in $D(T, R_Y)$.

Proposed generalization procedure



$$D(T, R_L) = \begin{array}{c} \\ x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_N \end{array} \begin{array}{c} \begin{array}{cccccc} L_1^{(1)} & L_2^{(1)} & \cdots & L_m^{(c)} & \cdots & L_M^{(C)} \end{array} \\ \begin{pmatrix} d_{11} & d_{12} & \cdots & d_{2m} & \cdots & d_{1M} \\ d_{21} & d_{22} & \cdots & d_{3m} & \cdots & d_{2M} \\ d_{31} & d_{32} & \cdots & d_{4m} & \cdots & d_{3M} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ d_{N1} & d_{N2} & \cdots & d_{Nm} & \cdots & d_{NM} \end{pmatrix} \end{array}$$

$$D(x, R_L) = [d_1 \quad d_2 \quad \cdots \quad d_m \quad \cdots \quad d_M]$$

- Our generalization method by feature lines and feature planes can be included in the family of model- or function-based generalization procedures.

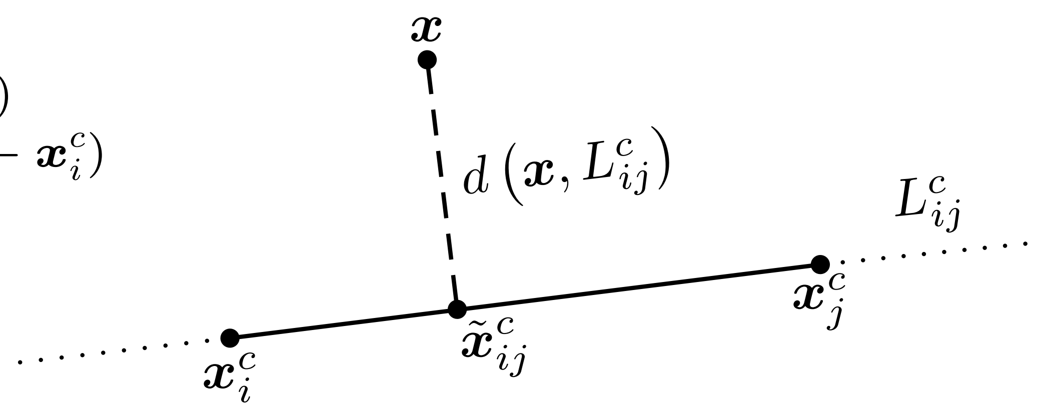
- Notice that the representation role is played in our case by a function generated by two representative objects, e.g. the so-called feature lines: $\{L_m^c\}$, which are the principle of one of the **nearest feature classifiers**.

Outline

- Introduction, problem statement and objectives
- Dissimilarity representations and generalizations
- Generalization of dissimilarity representations by feature lines and feature planes
  - Nearest feature classifiers
  - Generalizing dissimilarity representations by feature lines and feature planes
- Volcanic seismology and pattern recognition
  - Dissimilarity-based classification of seismic signals
  - Clustering of seismic signals for detecting mislabeling
  - Band selection of seismic spectra
- Conclusions, open questions and future work
- Publications

The nearest feature line rule

$$L_{ij}^c = \mathrm{sp}(\boldsymbol{x}_i^c - \boldsymbol{x}_j^c)$$
$$\tilde{\boldsymbol{x}}_{ij}^c = \boldsymbol{x}_i^c + \tau(\boldsymbol{x}_j^c - \boldsymbol{x}_i^c)$$
$$\tau = \frac{(\boldsymbol{x}-\boldsymbol{x}_i^c)\cdot(\boldsymbol{x}_j^c-\boldsymbol{x}_i^c)}{\|\boldsymbol{x}_j^c-\boldsymbol{x}_i^c\|^2}$$



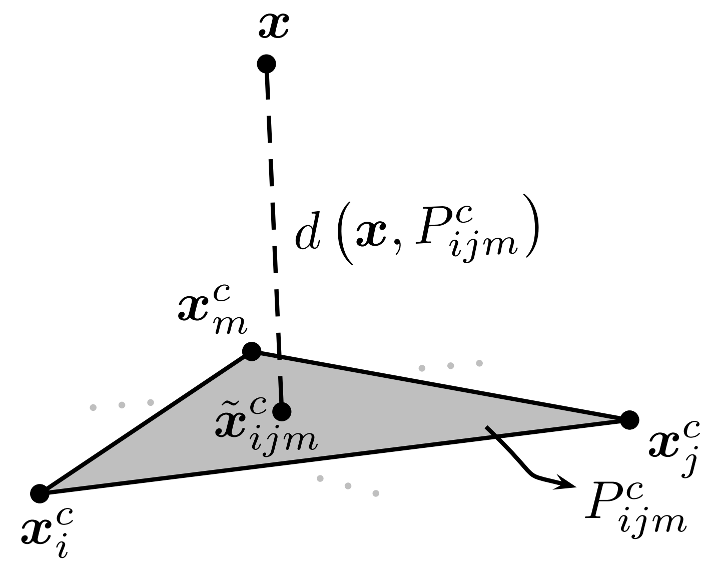
$$d(\boldsymbol{x}, L_{\hat{i}\hat{j}}^{\hat{c}}) = \min_{\substack{1\leqslant c\leqslant C,\ 1\leqslant i,j\leqslant n_C \\ i\neq j}} d(\boldsymbol{x}, L_{ij}^c)$$

Extension of the 1-NN rule! [Li and Lu, 1999]

The nearest feature line (NFL) rule generalizes each pair of prototype feature points belonging to the same class through a line L_{ij}^c: interpolation and extrapolation.

The nearest feature plane rule

$$P^c_{ijm} = \mathrm{sp}(\boldsymbol{x}^c_i, \boldsymbol{x}^c_j, \boldsymbol{x}^c_m)$$
$$\tilde{\boldsymbol{x}}^c_{ijm} = \boldsymbol{X}^c_{ijm}({\boldsymbol{X}^c_{ijm}}^T \boldsymbol{X}^c_{ijm})^{-1} {\boldsymbol{X}^c_{ijm}}^T \boldsymbol{x}$$
$$\boldsymbol{X}^c_{ijm} = [\boldsymbol{x}^c_i \; \boldsymbol{x}^c_j \; \boldsymbol{x}^c_m]$$



$$d(\boldsymbol{x}, P^{\hat{c}}_{\hat{i}\hat{j}\hat{m}}) = \min_{\substack{1 \leqslant c \leqslant C,\ 1 \leqslant i,j,m \leqslant n_C \\ i \neq j \neq m}} d(\boldsymbol{x}, P^c_{ijm})$$

Extension of NFL [Chien and Wu, 2002]

The nearest feature plane (NFP) rule generalizes each triplet of prototype feature points belonging to the same class through a plane P^c_{ijm}: interpolation and extrapolation.

Theoretical and experimental comparisons of the nearest feature classifiers

$$\begin{aligned}\mathrm{d}(\boldsymbol{x}, P^c_{ijm}) &\leqslant \min(\mathrm{d}(\boldsymbol{x}, L^c_{ij}), \mathrm{d}(\boldsymbol{x}, L^c_{jm}), \mathrm{d}(\boldsymbol{x}, L^c_{mi})) \\ &\leqslant \min(\mathrm{d}(\boldsymbol{x}, \boldsymbol{x}_{ci}), \mathrm{d}(\boldsymbol{x}, \boldsymbol{x}_{cj}), \mathrm{d}(\boldsymbol{x}, \boldsymbol{x}_{cm})).\end{aligned}$$

- Experimental comparison of the nearest feature classifiers [Orozco-Alzate and Castellanos-Domínguez, 2006]

- Quantification of the computational complexity of the nearest feature classifiers [Orozco-Alzate and Castellanos-Domínguez, 2007b]

⇒ The considerable increment in the number of distance calculations associated to NFP is not compensated by a valuable increase in classification performance.

⇒ NFP is often too expensive for practical applications.

⇒ NFL should be preferred in most cases.

Outline

- Introduction, problem statement and objectives
- Dissimilarity representations and generalizations
- Generalization of dissimilarity representations by feature lines and feature planes
  - Nearest feature classifiers
  - Generalizing dissimilarity representations by feature lines and feature planes
- Volcanic seismology and pattern recognition
  - Dissimilarity-based classification of seismic signals
  - Clustering of seismic signals for detecting mislabeling
  - Band selection of seismic spectra
- Conclusions, open questions and future work
- Publications

Dissimilarity representations and feature lines

Small sample case:

- Representational limitations: Lost of accuracy of the 1-NN rule!
- Alternative: we propose to used dissimilarity-based classifiers for good generalization and the concept of feature lines for enriching the original representation.

Generalization of dissimilarity representations:

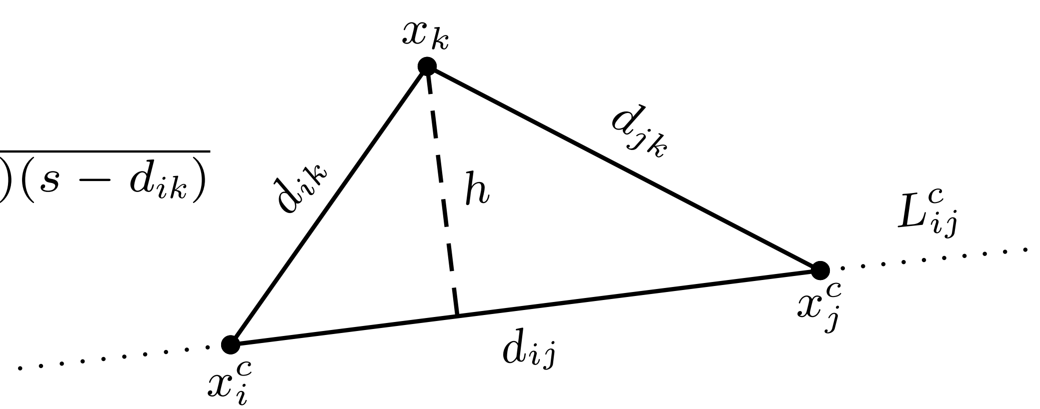
- To enhance the representation using feature lines.
- To achieve a better generalization, building a Bayesian classifier in the enhanced (generalized) representation, may improve the performance of both techniques when they are used separately.

An associated feature representation is not always available. How to compute distances to feature lines using only the information available at $D(T, R)$?

Distances to feature lines in terms of dissimilarities

Associated feature representation might not be available!

$$s = (d_{jk} + d_{ij} + d_{ik})/2$$
$$A = \sqrt{s(s - d_{jk})(s - d_{ij})(s - d_{ik})}$$
$$A = \frac{d_{ij}h}{2}$$



$$h = \frac{\sqrt{2(s(s - d_{jk})(s - d_{ij})(s - d_{ik}))}}{d_{ij}}$$

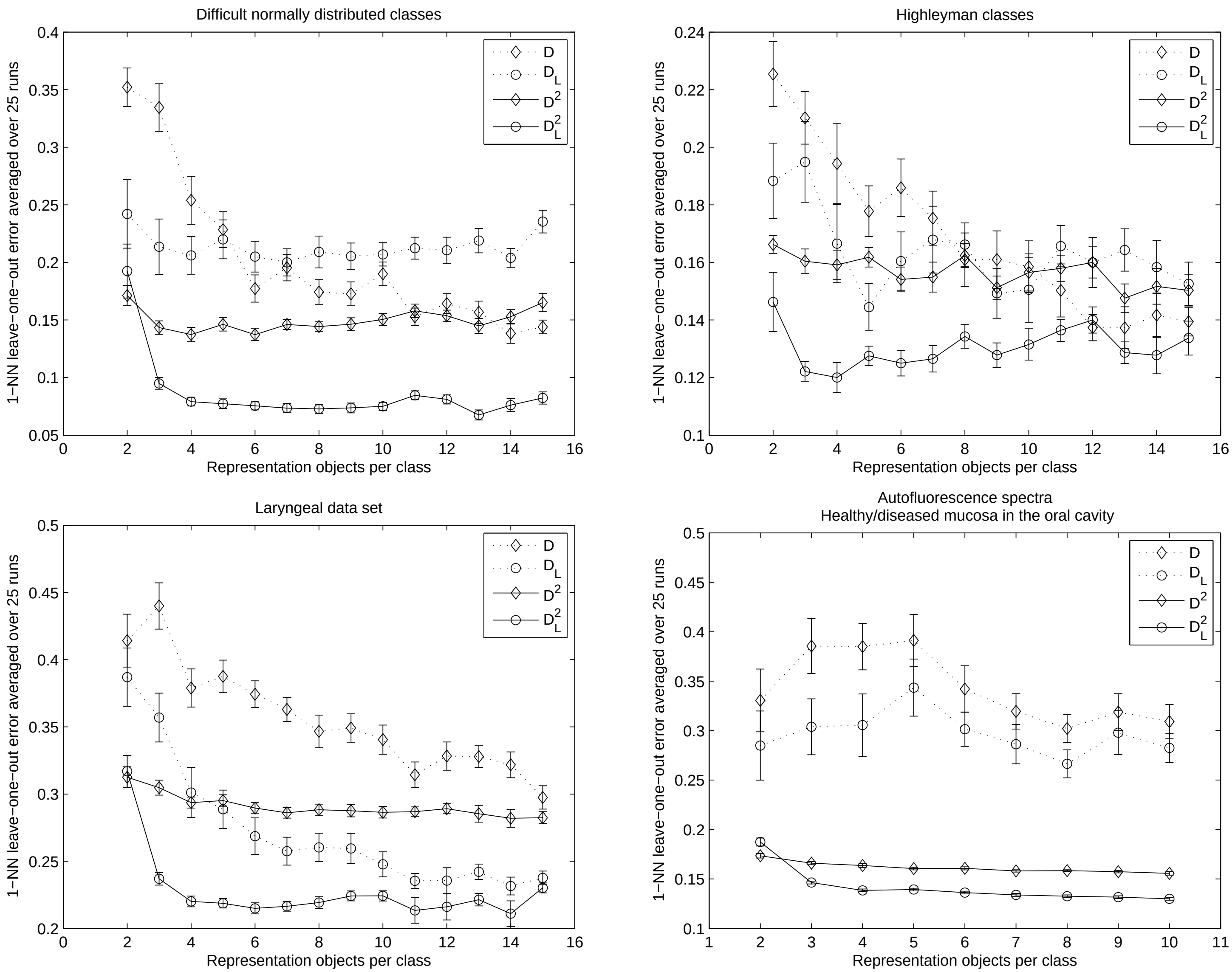
$$D(T,R) \Rightarrow D_L(T,R_L) = \begin{array}{c} \\ x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_N \end{array} \begin{array}{c} \begin{array}{ccccc} L_1 & L_2 & L_3 & \cdots & L_{n_L} \end{array} \\ \begin{pmatrix} d_{11} & d_{12} & d_{13} & \cdots & d_{1n_L} \\ d_{21} & d_{22} & d_{23} & \cdots & d_{2n_L} \\ d_{31} & d_{32} & d_{33} & \cdots & d_{3n_L} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ d_{N1} & d_{N2} & d_{N3} & \cdots & d_{Nn_L} \end{pmatrix} \end{array},$$

Exploratory experiment: potential goodness of generalizing dissimilarity representations

Leave-one-out 1-NN error on 4 different representations:

- Artificial and real-world data (biological)
- Dissimilarity representation derived by computing pairwise Euclidean distances, denoted by D
- Generalization of D by using feature lines, denoted by D_L
- A dissimilarity representation resulting of the computation of pairwise Euclidean distances between rows of D, denoted by D^2
- Generalized dissimilarity representation computed by pairwise Euclidean distances between rows of D_L, denoted by D_L^2

1-NN LOO errors for D, D_L, D^2 and D_L^2



Generalized dissimilarity representations

The information on a set S of new incoming objects is provided in terms of their distances to R_L, i.e., as a *generalized* dissimilarity matrix $D(S, R_L)$.

The number of lines increases combinatorially!: selection of feature lines should be based on a simple criterion $\Rightarrow$
i) random, *ii)* according their properties, e.g. their length.

Length based selection criterion

- Rank all the feature lines according to their length (d_{ij}).
- The initial representation set R_L is the shortest (longest) feature line. Then, the second shortest one is added to R_L and so on.

Experimental setup

Data sets

Difficult normally distributed classes. It corresponds to a two-dimensional and two-class dataset having very different class variances for the dimensions. Separation is thereby, for small sample sizes, difficult.

Highleyman classes. A two-dimensional and two-class dataset generated by the Highleyman distribution.

Wine data. The *Wine* data come from the UCI Repository and describe three types of wine by 13 features.

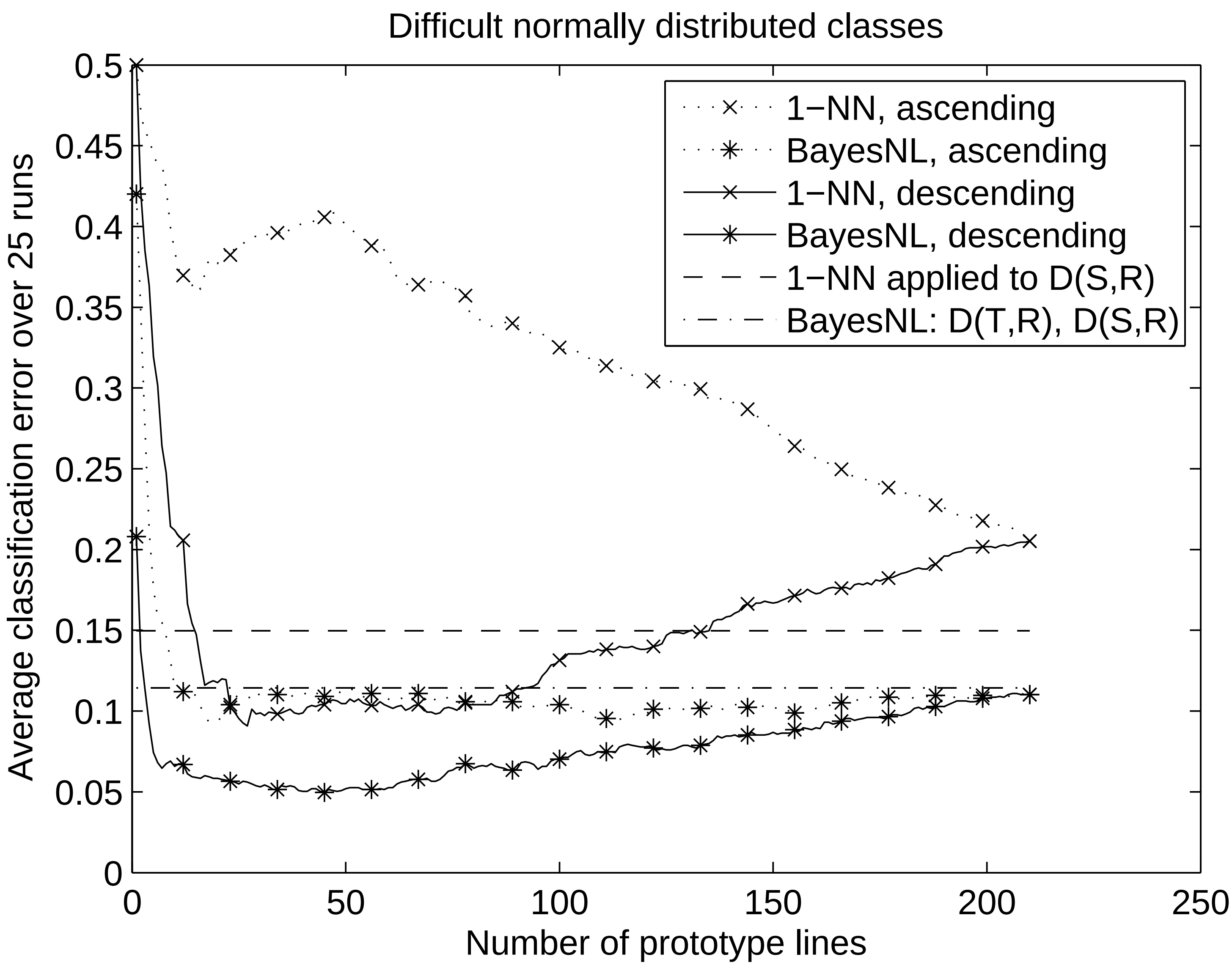
Laryngeal data. The *Laryngeal* dataset comes from the Bulgarian Academy of Sciences. Normal and pathological voices are described by 16 parameters in the time, spectral and cepstral domains.

Experimental setup

- Maximum number of prototypes considered: $r_c = 15$.
- Total number of feature lines: 315 (*Wine*, three-class problem) and 210 (other two-class data sets).
- For each repetition, a new representation set R is randomly chosen.
- The best results, that we used as a reference (horizontal line), do not necessarily correspond to the case $R = T$.
- Regularized linear Bayesian classifier (BayesNL): regularization parameter $\lambda = 0.01$

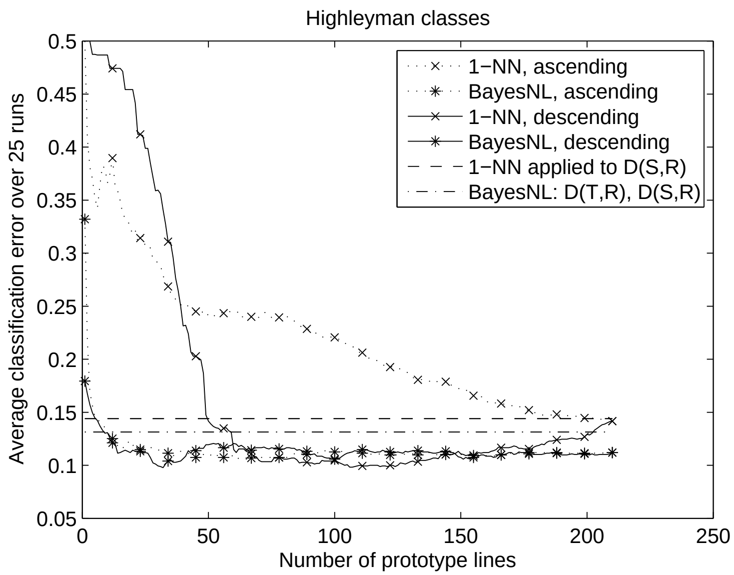
Results for artificial and real-world data

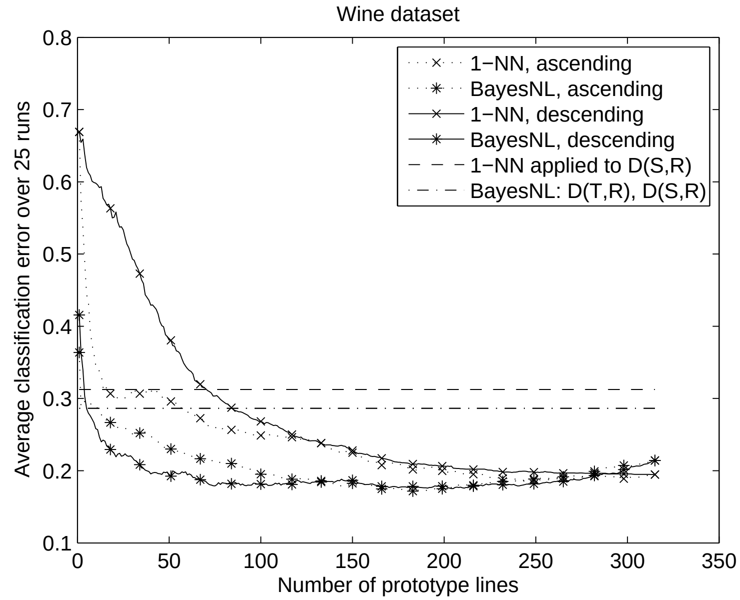
Feature lines incrementally included according to their length.

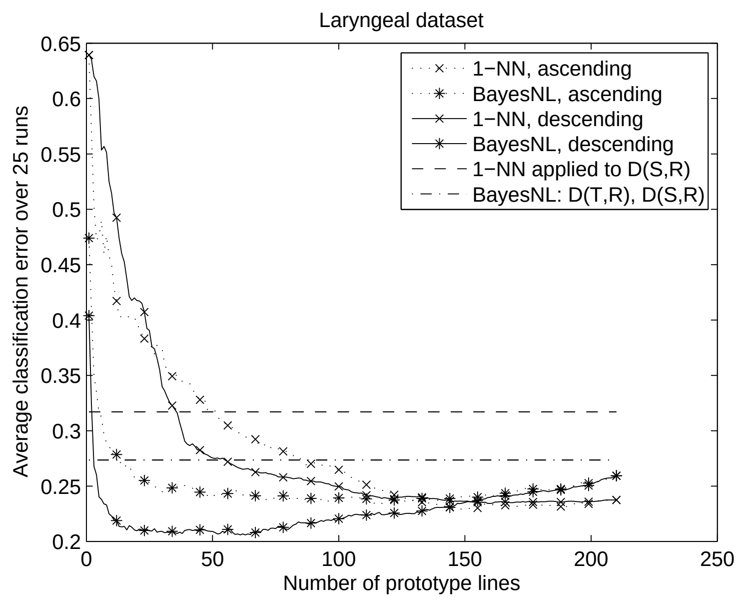


Results for artificial and real-world data

Feature lines incrementally included according to their length.

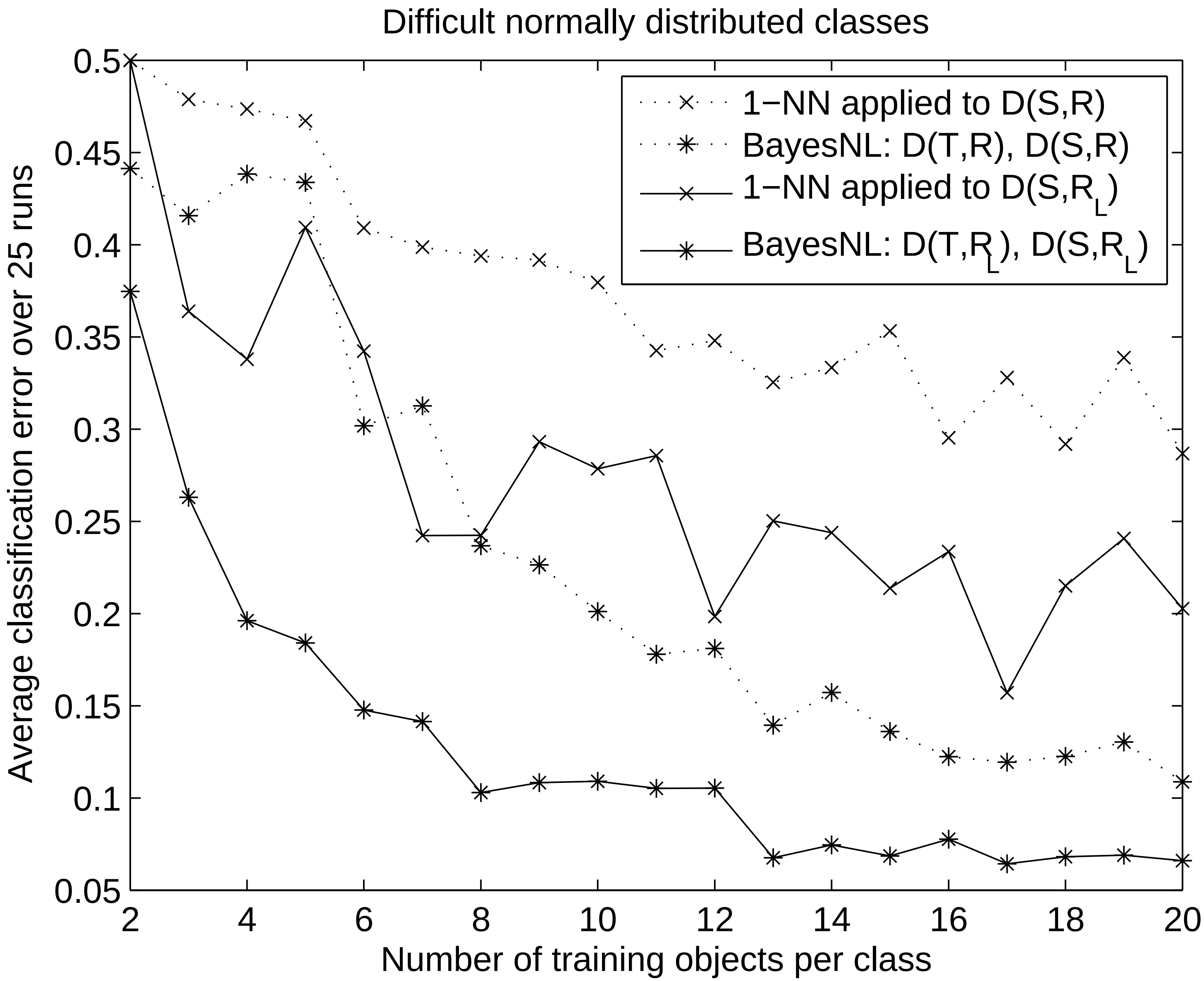






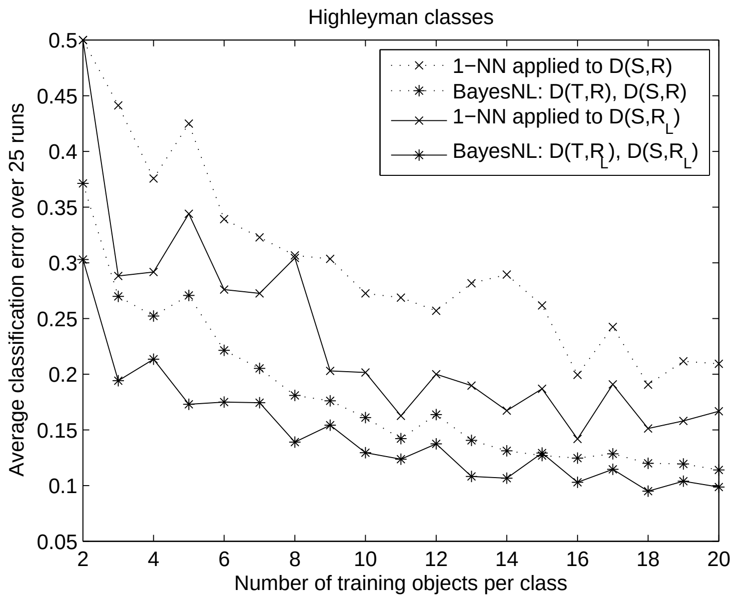
Results for artificial and real-world data

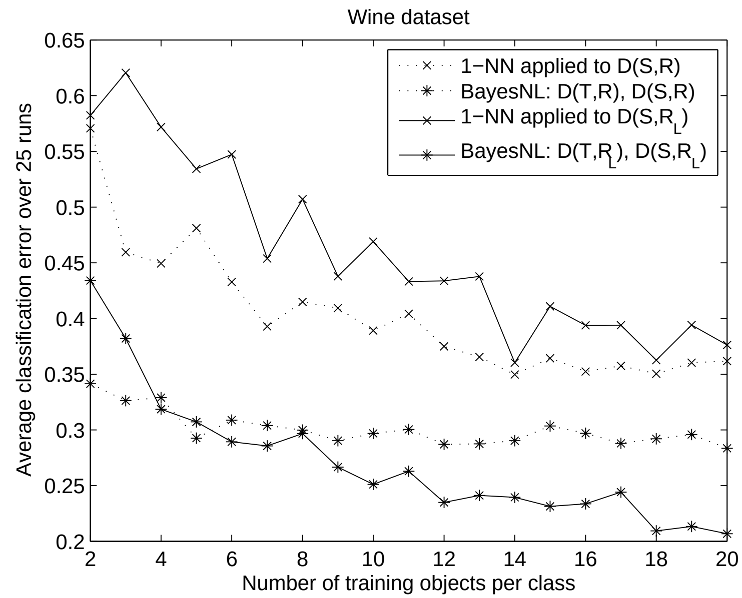
A rule-of-thumb of selecting $n_c C/5$ prototypes is used here.

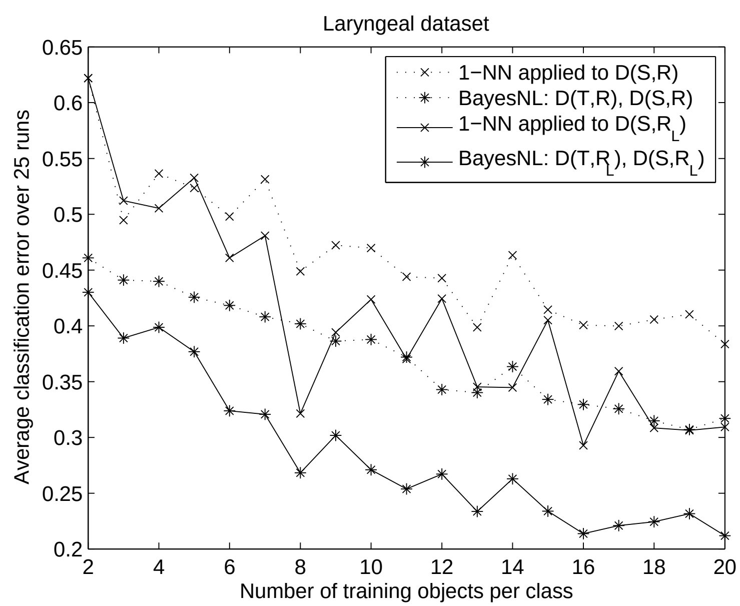


Results for artificial and real-world data

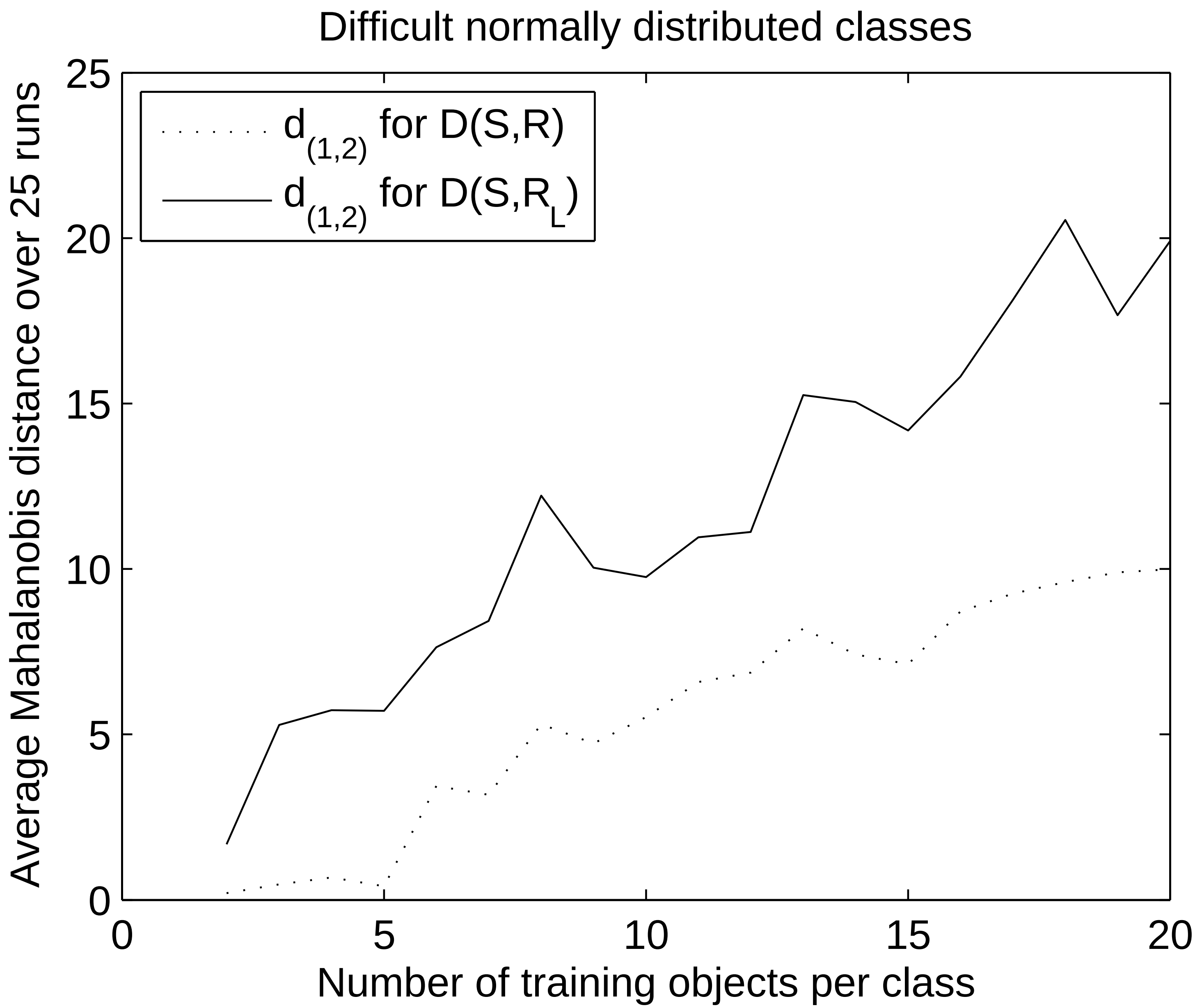
A rule-of-thumb of selecting $n_c C/5$ prototypes is used here.





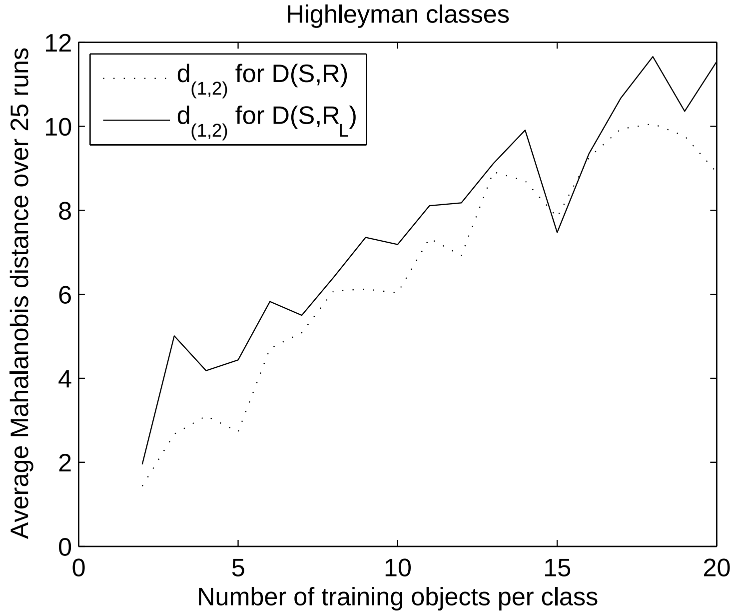


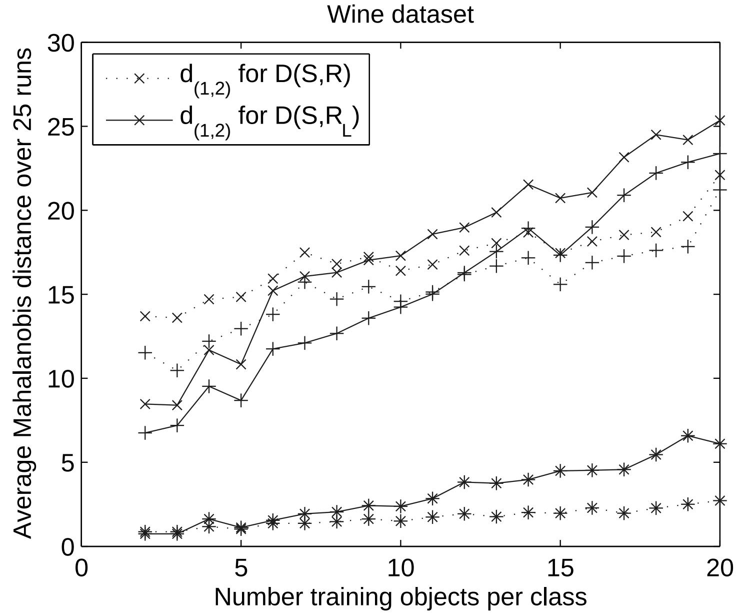
Mahalanobis distances

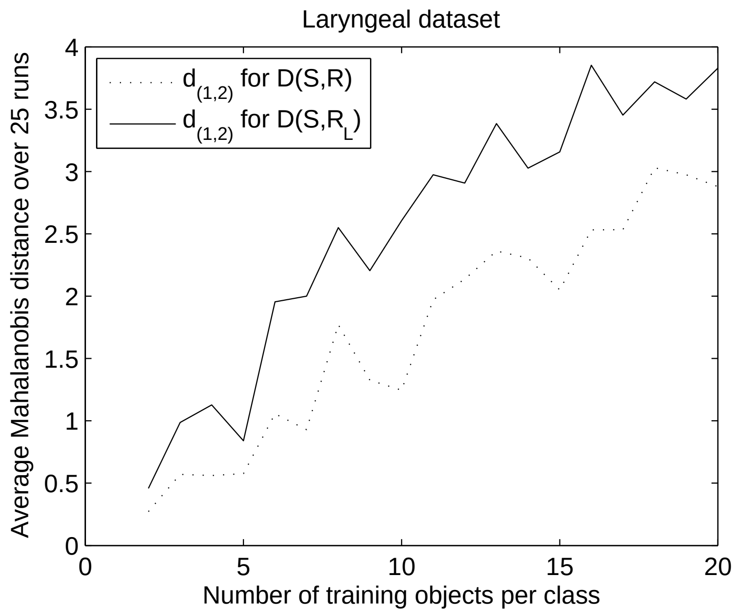


Mahalanobis distances

2009/10/09







Discussion: generalization by feature lines

- The generalization procedure, when using a random and a length-based selection of prototype feature lines, seems to be especially profitable for elongated (cigar-like) datasets.
- Just a few long feature lines are needed to describe correlated (elongated) data sets.
- Compared to the non-generalized dissimilarity representations, the generalized ones exploit more the intrinsic geometric information available at the pairwise dissimilarities, effectively finding an enriched representation.
- Our method is particularly advantageous for small sample size problems because in such sparse spaces, the feature lines are somewhat filling them.

Further experiments and discussions: [Orozco-Alzate et al., 2007, Orozco-Alzate et al., 2008a].

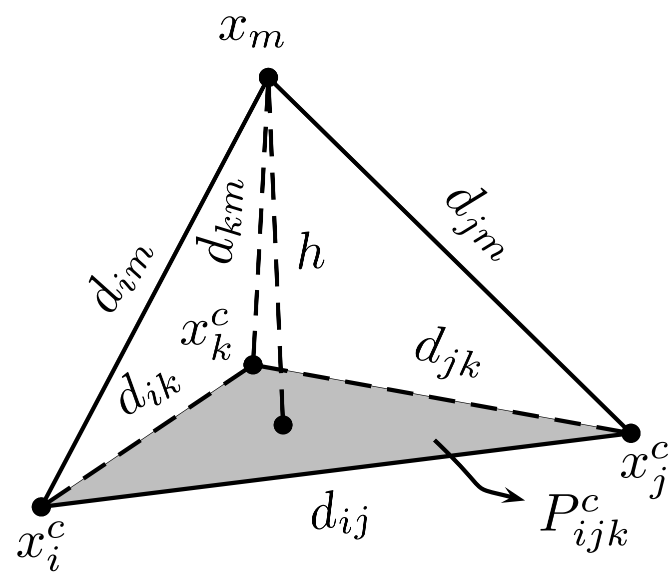
Distances to feature planes in terms of dissimilarities

Associated feature representation might not be available!

$$s = (d_{jk} + d_{ij} + d_{ik})/2$$

$$V = \frac{h\sqrt{s(s-d_{jk})(s-d_{ij})(s-d_{ik})}}{3}$$

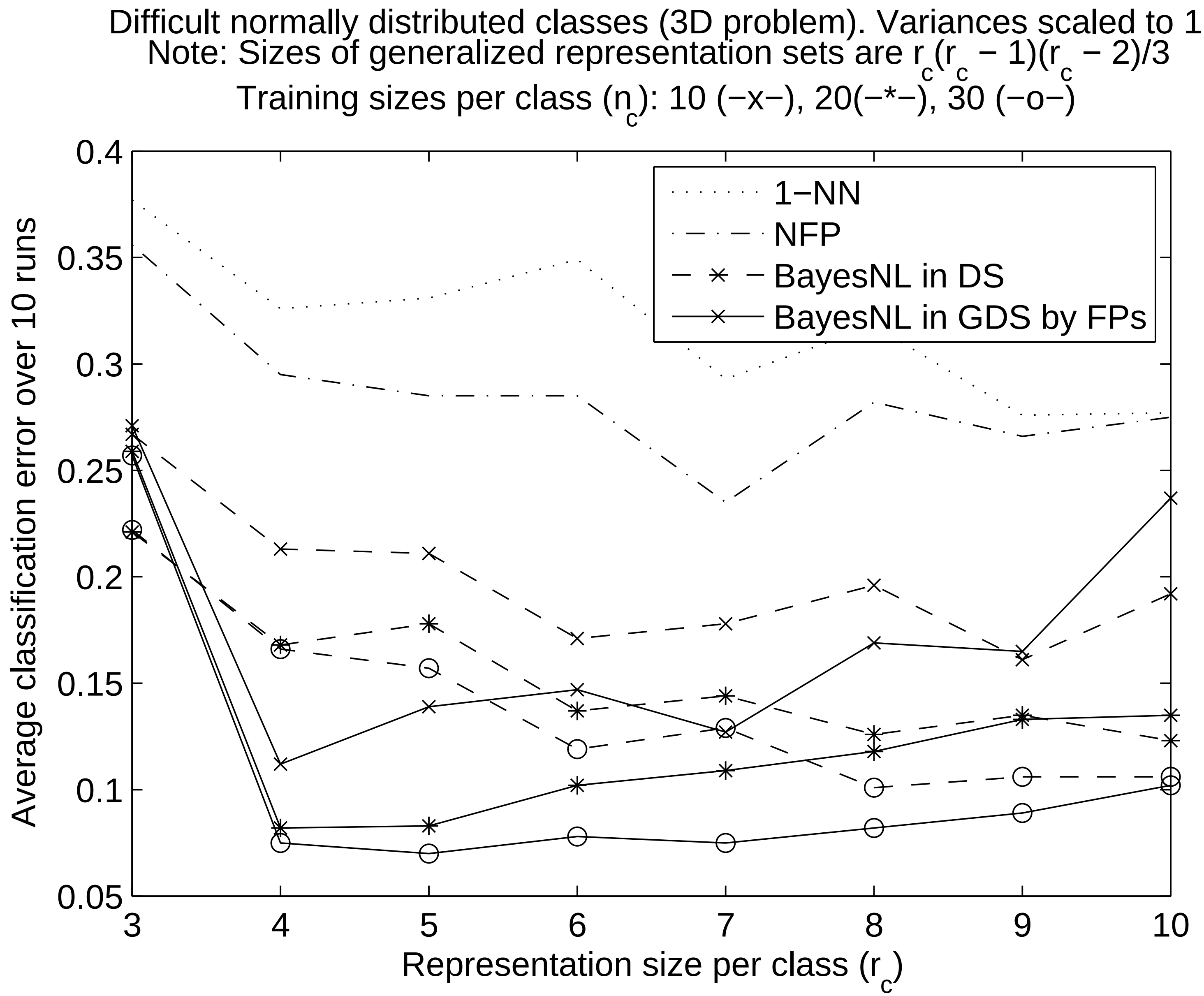
$$288V^2 = \begin{vmatrix} 0 & d_{ij}^2 & d_{ik}^2 & d_{im}^2 & 1 \\ d_{ij}^2 & 0 & d_{jk}^2 & d^2 & 1 \\ d_{ik}^2 & d_{jk}^2 & 0 & d_{km}^2 & 1 \\ d_{im}^2 & d_{jm}^2 & d_{km}^2 & 0 & 1 \\ 1 & 1 & 1 & 1 & 0 \end{vmatrix}$$



$$D(T,R) \Rightarrow D_P(T,R_P) = \begin{array}{c c} & \begin{array}{ccccc} P_1 & P_2 & P_3 & \cdots & P_{n_P} \end{array} \\ \begin{array}{c} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_N \end{array} & \begin{pmatrix} d_{11} & d_{12} & d_{13} & \cdots & d_{1n_P} \\ d_{21} & d_{22} & d_{23} & \cdots & d_{2n_P} \\ d_{31} & d_{32} & d_{33} & \cdots & d_{3n_P} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ d_{N1} & d_{N2} & d_{N3} & \cdots & d_{Nn_P} \end{pmatrix} \end{array},$$

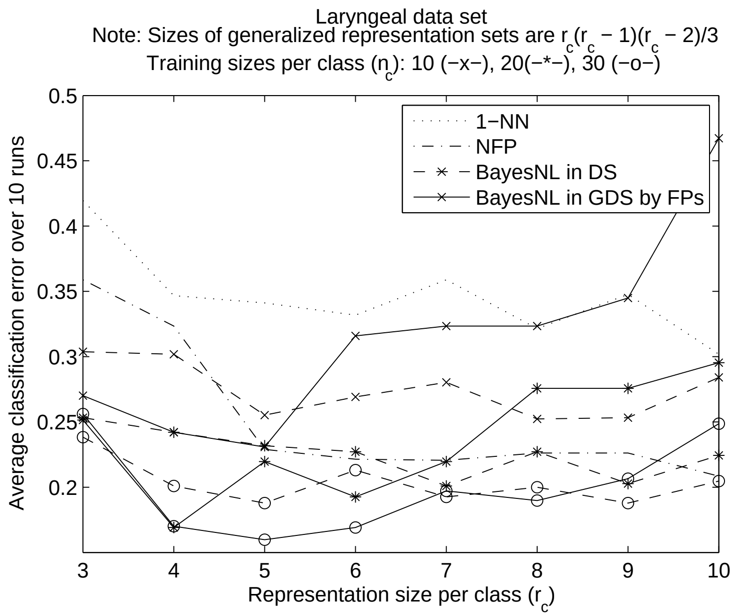
Results for artificial and real-world data

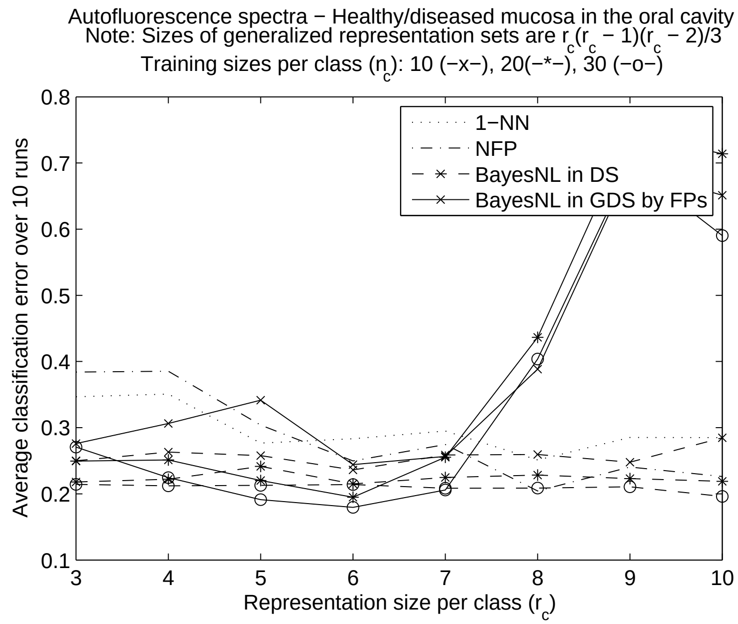
2009/10/09

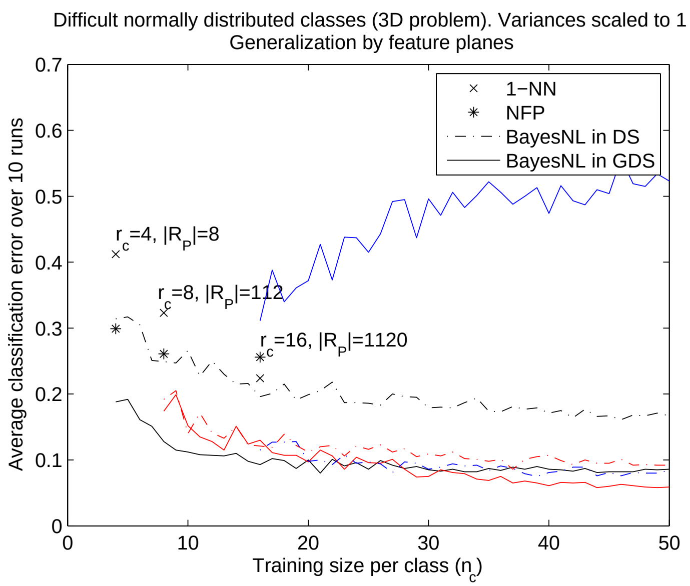


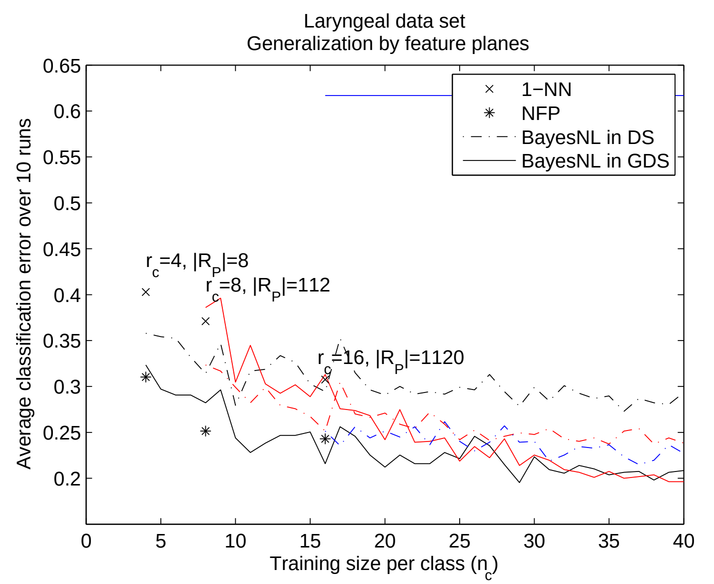
Results for artificial and real-world data

2009/10/09









Discussion: generalization by feature planes

- Results for generalized dissimilarity representations by feature planes are not significantly better than the results obtained for the original dissimilarity representation. Moreover, they get worse quickly due to the peaking phenomenon. It suggests that the generalization by feature lines should be preferred.

- Generalization by using feature planes rapidly yields innacuracies in the estimation of the class covariance matrices. However, the procedure appears beneficial for the smallest representation sizes.

Further experiments and discussions: [Orozco-Alzate et al., 2008a].

Length based selection criterion

- Ranking all the feature lines according to their length (d_{ij})
- Decide if a feature line is included in R_L or not according to that criterion
- Ascending method: initial representation set is the the shortest feature line. Then, the second shortest one is added and so on
- Descending method: Reverse case of te ascending method

Just a few long feature lines are needed to describe correlated (elongated) data sets [Orozco-Alzate et al., 2007]

Length based selection criterion

Hypothesis: Middle-length feature lines might be better to describe slightly non-linear subspaces, i.e. curved manifolds.

Method :

- Start the selection in the middle of the ranking.
- Include in R_L the feature line exactly placed in in the middle of the sorted list.
- Take feature lines placed at its left and right sides alternately.

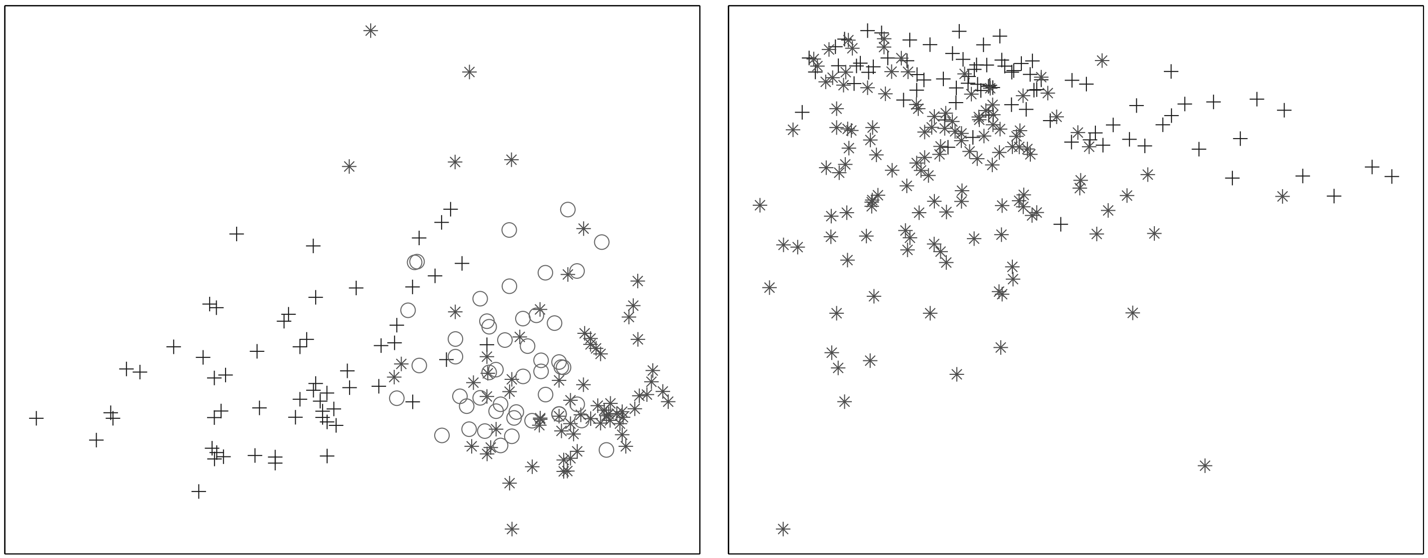
$$\Downarrow$$

Piece-wise description!

Experimental setup

2009/10/09

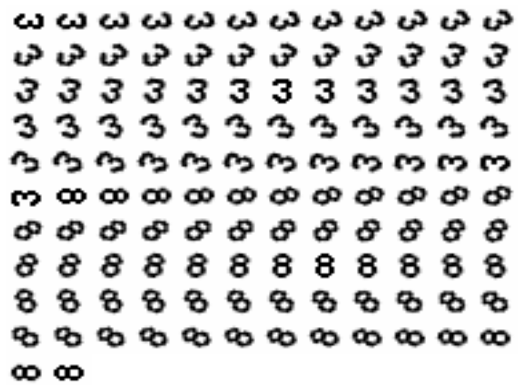
Data sets



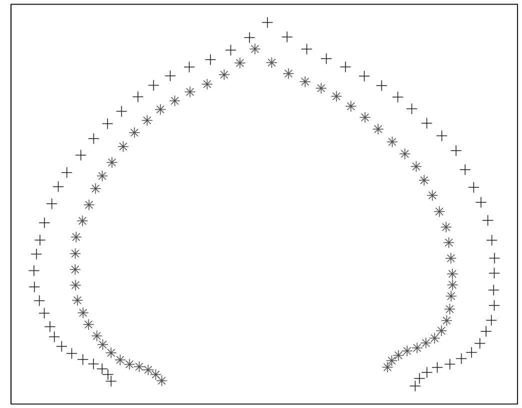
(a) *Wine* data set (b) *Laryngeal* data set

Figure 1: Scatter plots using a classical multidimensional scaling

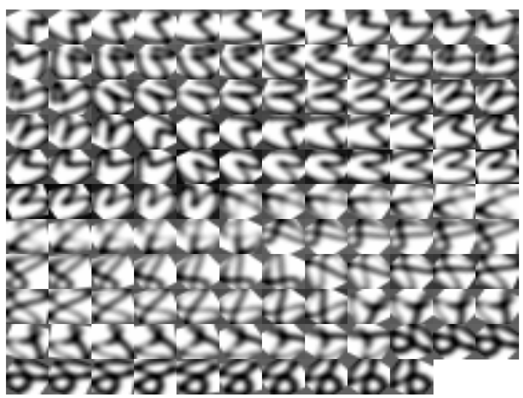
Experimental setup



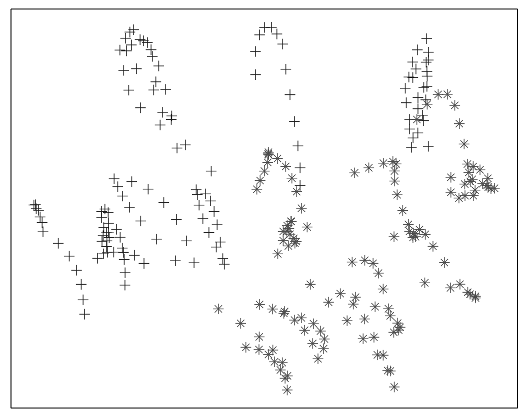
(a) Rotated machine-printed digits



(b) Scatter plot using classical MDS



(c) Rotated handwritten digits



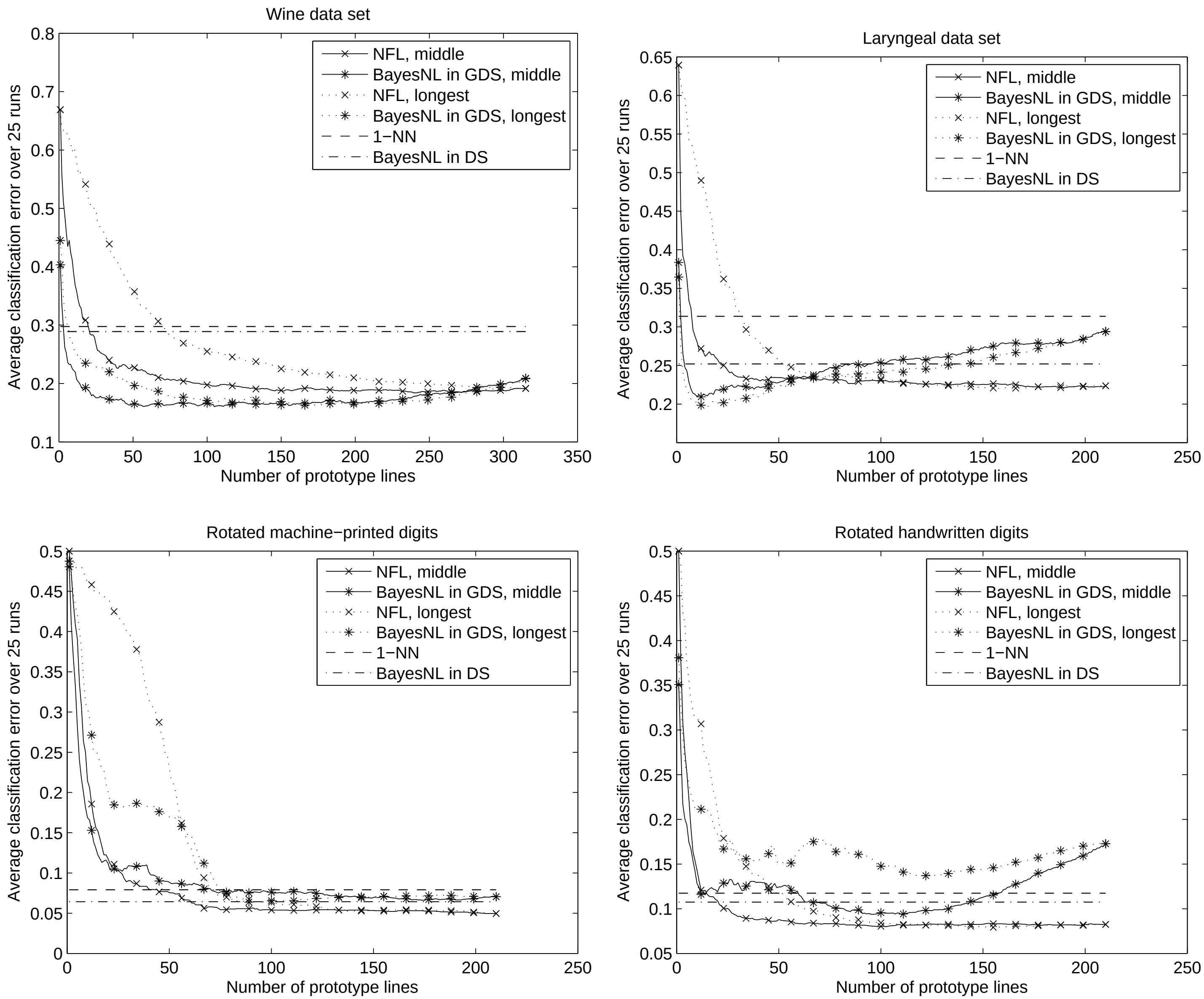
(d) Scatter plot using classical MDS

Figure 2: Digits '3' and '8' rotated (a)-(b) between -90° and 90° with steps of 3° and (c)-(b) between -60° and 60° with steps of 10°.

Experimental setup

- Maximum number of prototypes considered: $r_c = 15$.
- Total number of feature lines: 315 (*Wine*, three-class problem) and 210 (other two-class data sets).
- For each repetition, a new representation set R is randomly chosen.
- The best results, that we used as a reference (horizontal line), do not necessarily correspond to the case $R = T$.
- Regularized linear Bayesian classifier (BayesNL): regularization parameter $\lambda = 0.01$

Experimental results



Discussion: middle-length feature lines

- BayesNL classifier in GDS outperforms both the 1-NN rule and the BayesNL in DS for the *Wine* and *Laryngeal* data sets

- For the digit recognition problems, the NFL rule outperforms the dissimilarity-based classifiers as well as the 1-NN rule

- The benefit of using the middle-length feature lines, is consistently observed (solid curves are mostly below the dotted ones)

- Feature lines deal better with non-scaled data (*Wine* data set)

Discussion: middle-length feature lines

- Middle-length feature lines are beneficial for the NFL rule while the longest lines provide a better description for the BayesNL classifier (*Laryngeal*)

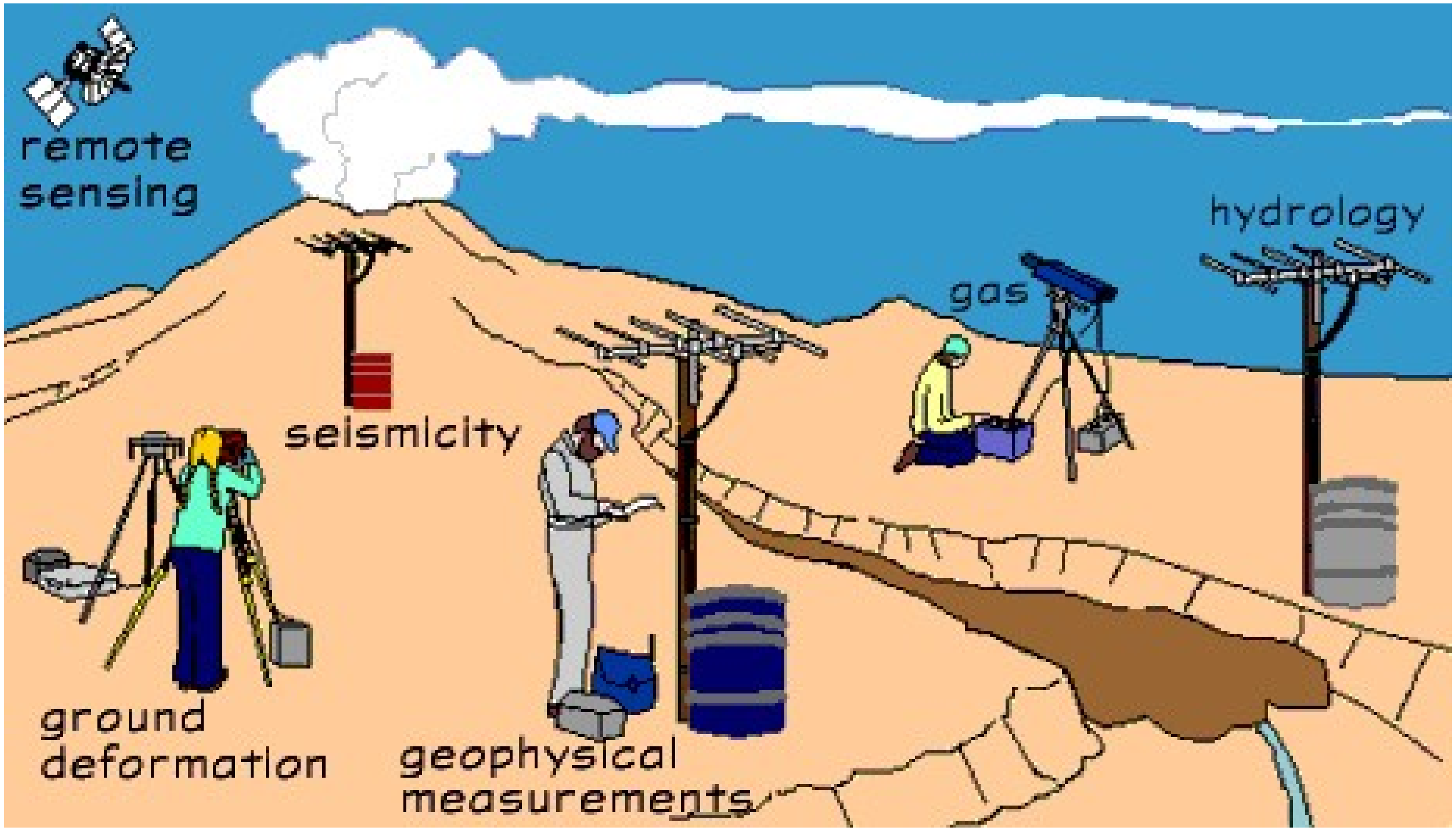
- Few middle-length feature lines may describe curved subspaces better than a small number of the longest feature lines

- The middle-length feature lines may provide a better piecewise description of the structure of the data because they are less likely to cross the territory of the other class than the longest feature lines.

Outline

- Introduction, problem statement and objectives
- Dissimilarity representations and generalizations
- Generalization of dissimilarity representations by feature lines and feature planes
  - Nearest feature classifiers
  - Generalizing dissimilarity representations by feature lines and feature planes
- Volcanic seismology and pattern recognition
  - Dissimilarity-based classification of seismic signals
  - Clustering of seismic signals for detecting mislabeling
  - Band selection of seismic spectra
- Conclusions, open questions and future work
- Publications

Volcano-monitoring techniques



Monitoring Nevado del Ruiz volcano

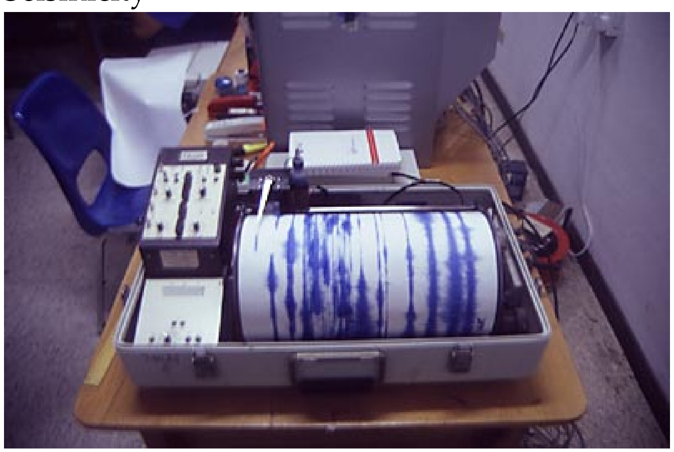
2009/10/09

Due to the tremendous disaster of 1985, the government created the volcanological observatories; before that, Colombian volcanoes were not daily and sufficiently monitored.

Ground deformation

Radon gas

Seismicity



Sulfur emissions

Nevado del Ruiz volcano

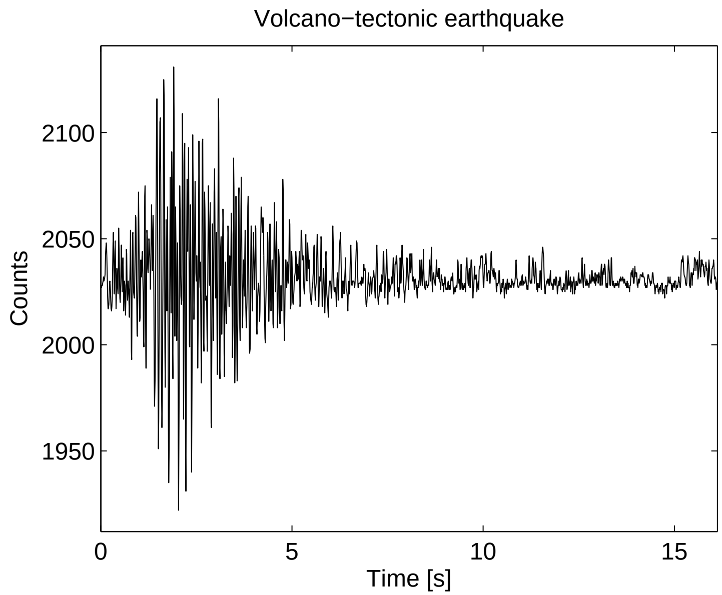
- Volcanological and Seismological Observatory at Manizales (VSOM) staff currently classifies volcanic earthquakes by visual inspection; such a method supposes a great deal of workload for the seismic analysts..

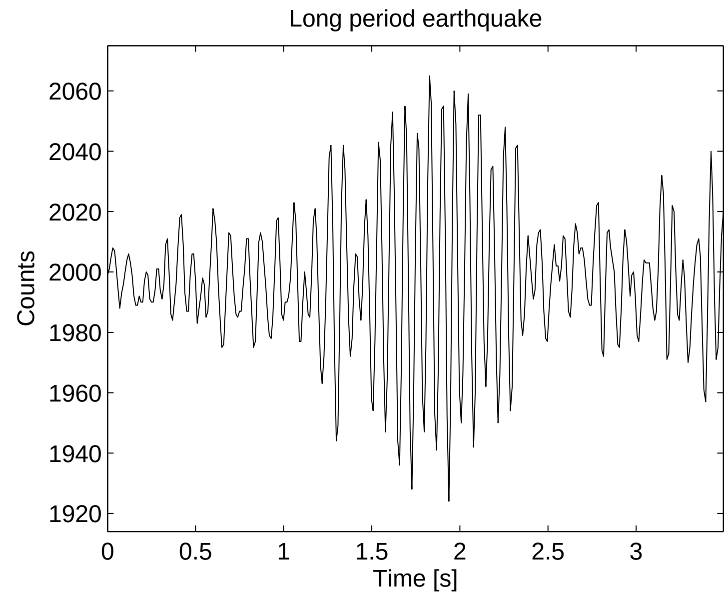
- An automatic classification tool dramatically reduces this arduous task and also turns classification reliable and objective, removing errors associated to tedious evaluations and changing of personnel.

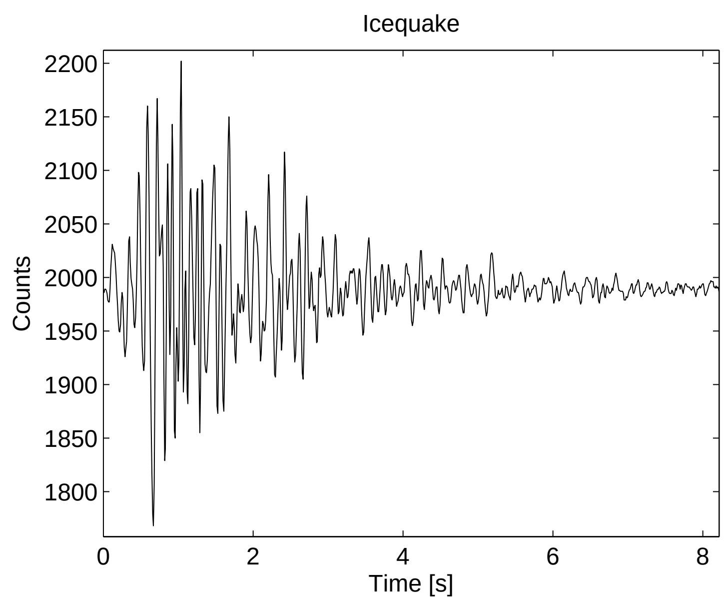
Dataset

- Three classes considered: Volcano-Tectonic (VT) earthquakes, Long-Period (LP) earthquakes and Icequakes (IC).
- Signals from two stations (Olleta crater station and Glacier station) have been selected for the experiments because, according to the VSOM staff experience, they are a reference for the volcanic and ice-related events.
- Stations are located 4.08 km and 1.8 km from the active crater respectively.
- Signals were digitized at 100.16 Hz sampling frequency by using a 12 bits analog to digital converter.
- Number of events per class: VT (483), LP (580), IC (782)

Dataset (Cont.)





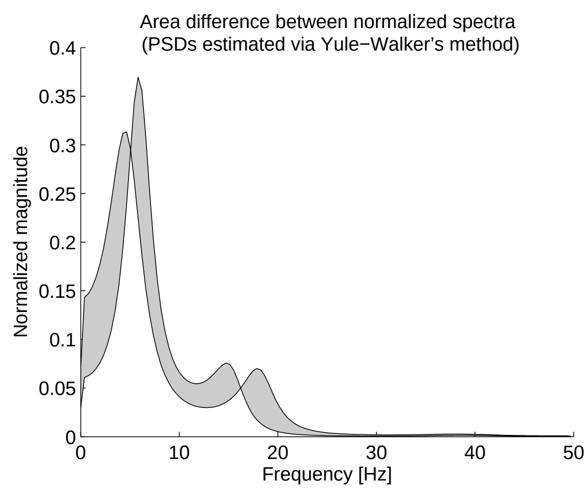


Deriving dissimilarity representations

- Calculating the spectrum:
  1. N-point Fast Fourier Transform (FFT).
  2. Parametric estimation of the power spectral density (PSD) by using the Yule-Walker AR method.

- Removing DC bias and normalization.

- Computing dissimilarities between spectra:
  1. Pointwise Euclidean distance (L_2-norm).
  2. Area difference between non-overlapping parts (L_1-norm).



Classifiers

- The k Nearest Neighbor classifier (k-NN): employed as a reference for performance comparison, $k = 1$.

- Linear normal density based classifier (LDC):

$$f(D(x,R)) = \left[D(x,R) - \frac{1}{2}\left(\boldsymbol{m}_{(1)} + \boldsymbol{m}_{(2)}\right)\right]^T \times C^{-1}\left(\boldsymbol{m}_{(1)} - \boldsymbol{m}_{(2)}\right) + \log\frac{P_{(1)}}{P_{(2)}}$$

- Quadratic normal density based classifier (QDC):

$$f(D(x,R)) = \sum_{i=1}^{2}(-1)^i\left(D(x,R) - \boldsymbol{m}_{(i)}\right)^T \times C_{(i)}^{-1}\left(D(x,R) - \boldsymbol{m}_{(i)}\right) + 2\log\frac{p_{(1)}}{p_{(2)}} + \log\frac{\left|C_{(1)}\right|}{\left|C_{(2)}\right|}$$

Regularized version (C singular):

$$C_{reg}^{\lambda} = (1-\lambda)\,C + \lambda\mathrm{diag}(C), \quad \lambda = 0.01$$

Experimental results

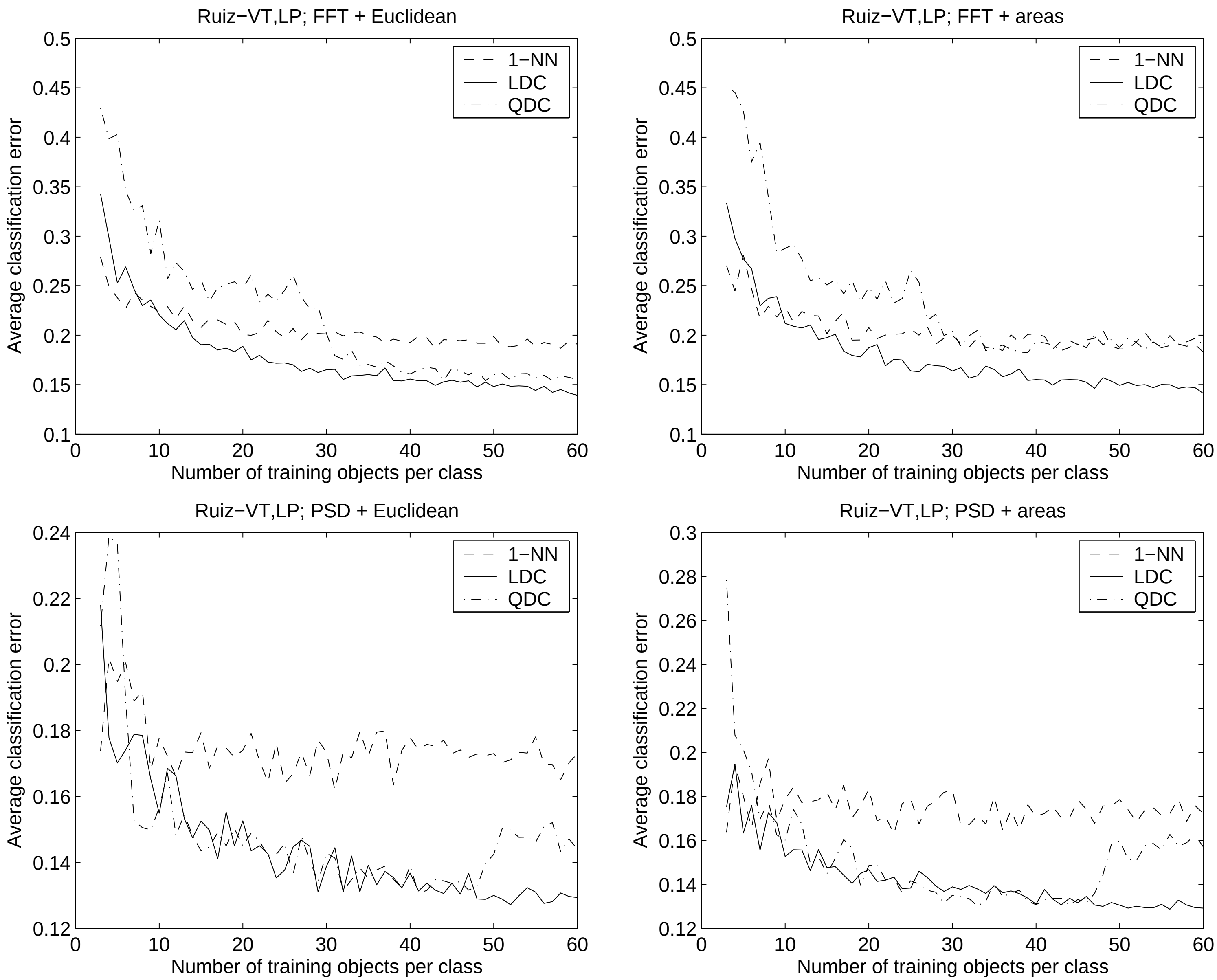
- Classification problems conducted:
  - Two-class problem: *Ruiz-VT,LP*
  - multi-class problem: *Ruiz-all*

- Experiments were performed 25 times for randomly chosen training and test sets.

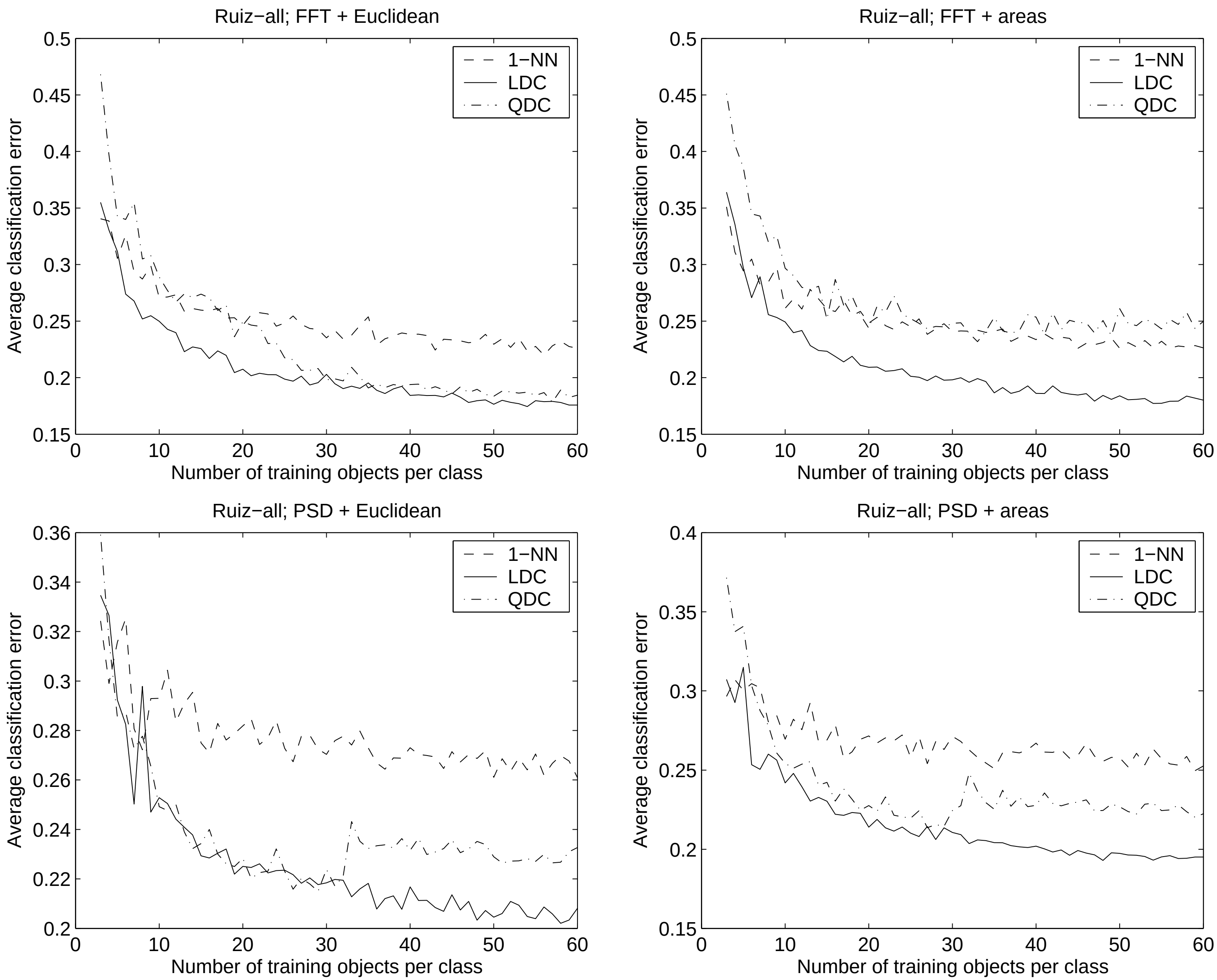
- The entire training set has been used as the representation set.

- Figures present the generalization errors as a function of the number of training objects randomly chosen.

Experimental results (*Ruiz-VT,LP*)



Experimental results (*Ruiz-all*)



Discussions

- The two-class *Ruiz-VT,LP* problem seems the easiest because it contains signals recorded and identified at the same station (Olleta crater station); in consequence, it is expected that sensor and noise conditions are the same, influencing the subsequent steps for representation and classification. It is well know that, in general, multi-class problems are more difficult.

- For the two-class problem, experiments based on parametric PSD estimation outperform those based on the FFT. It makes sense because event lengths are, in general, short and, consequently, a parametric spectral estimation yields a higher resolution; in addition, the autoregressive methods (AR) tend to adequately describe spectra of peaky data, which is precisely the nature of seismic volcanic signals

Discussions (Cont.)

- Experiments confirm that Bayesian classifiers outperform the 1-NN classifier, when a sufficient number of prototypes is provided. The LDC constructed on the different dissimilarity representations, for both *Ruiz-VT,LP* and *Ruiz-all* problems, always outperforms the 1-NN rule.

- LDC accuracies for the *Ruiz-VT,LP* problem vary between 85% and 87% when the average classification error curve reaches a steady state. Similarly, classification accuracies for the *Ruiz-all* problem vary between 81% and 84%.

- QDC shows a loss of accuracy when certain number of prototypes is provided. Therefore, a further study on a proper regularization for the QDC should be conducted in order to obtain an improvement of this classifier.

Discussions (Cont.)

- A re-analysis of the original a-priori classification is needed in order to to confirm the labels assigned by the experts.

- LDC accuracies for the *Ruiz-VT,LP* problem vary between 85% and 87% when the average classification error curve reaches a steady state. Similarly, classification accuracies for the *Ruiz-all* problem vary between 81% and 84%.

Outline

- Introduction, problem statement and objectives
- Dissimilarity representations and generalizations
- Generalization of dissimilarity representations by feature lines and feature planes
  - Nearest feature classifiers
  - Generalizing dissimilarity representations by feature lines and feature planes
- Volcanic seismology and pattern recognition
  - Dissimilarity-based classification of seismic signals
  - Clustering of seismic signals for detecting mislabeling
  - Band selection of seismic spectra
- Conclusions, open questions and future work
- Publications

Clustering of seismic signals for detecting mislabeling

- In many applications of pattern recognition, it is extremely difficult or expensive, or even impossible, to reliably label a training sample with its true category [Jain et al., 2000].

- Performance of learning methods is subject to the availability of a representative and a priori well classified training set.

- A considerable number of seismic events are erroneously attributed to other classes. As a result, a remarkable improvement in classification accuracy was obtained when the revised data set is used [Langer et al., 2006].

In order to detect mislabelled events, we propose the use of clustering techniques on the dissimilarity representations. Further details in [Orozco-Alzate and Castellanos-Domínguez, 2007a]

Outline

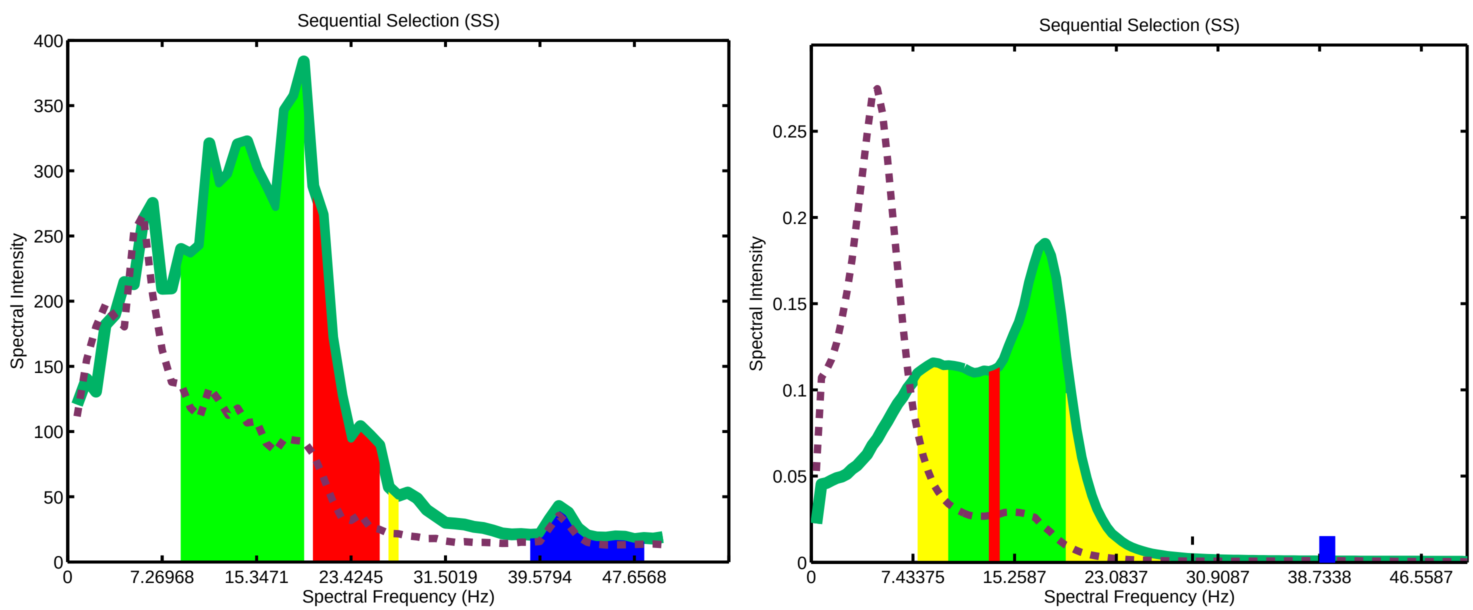
- Introduction, problem statement and objectives
- Dissimilarity representations and generalizations
- Generalization of dissimilarity representations by feature lines and feature planes
  - Nearest feature classifiers
  - Generalizing dissimilarity representations by feature lines and feature planes
- Volcanic seismology and pattern recognition
  - Dissimilarity-based classification of seismic signals
  - Clustering of seismic signals for detecting mislabeling
  - Band selection of seismic spectra
- Conclusions, open questions and future work
- Publications

Band selection of seismic spectra

2009/10/09

Spectral information is largely redundant by nature, i.e. its intrinsic dimensionality is low.

Solution: Reducing data redundancy and finding an understandable and reduced number of discriminative regions by using spectral band selection/extraction techniques.



Further details available at: [Orozco-Alzate et al., 2008b].

Outline

- Introduction, problem statement and objectives
- Dissimilarity representations and generalizations
- Generalization of dissimilarity representations by feature lines and feature planes
  - Nearest feature classifiers
  - Generalizing dissimilarity representations by feature lines and feature planes
- Volcanic seismology and pattern recognition
  - Dissimilarity-based classification of seismic signals
  - Clustering of seismic signals for detecting mislabeling
  - Band selection of seismic spectra
- Conclusions, open questions and future work
- Publications

Conclusions

In this presentation, we have tried to answer the following questions; most of them as stated in our doctoral research proposal:

1. Is it possible to say something about the asymptotic behavior of the nearest feature classifiers?

2. What is the complexity and the error-complexity trade-off of the nearest feature classifiers?

3. What are the cases for which generalized dissimilarity representations might be advantageous or, conversely, for which ones generalized dissimilarity representations might not be useful?

4. How to generalize very small datasets?

Conclusions

5. Does enriching the representation spoil the generalization?

6. How are generalized dissimilarity representations more informative? How informative is each column?

7. Are these generalized representations better in general (e.g. more compact)?

8. Is the application of classifying seismic volcanic signals addressable by dissimilarity based methods?

Conclusions

- When an arbitrarily large number of prototypes is available, they are representative enough and, consequently, a significant difference between the asymptotic behavior of k-NN and the nearest feature classifiers is not expected.

- The increment in the number of distance calculations associated to NFP is, in general, not compensated by a valuable increase in classification performance.

- We proposed a procedure for enriching a given dissimilarity representation by using feature lines and planes as bases to build the dissimilarity space. The output of our procedure is the so-called generalized dissimilarity space and is aimed to overcome the representational limitations of small sets of representation objects.

Conclusions

- In general, generalizing a given dissimilarity representation is particularly beneficial for correlated (elongated) data sets because the generalization procedure exploits the linear geometric relationships between the objects.

- Structures having possibly moderate non-linearities may also be benefited from the generalization.

- Limitations in the applicability of our approach can be attributed to two inaccuracies which affect the NFL and NFP techniques: extrapolation and interpolation.

- Even though the number of feature lines and planes increases rapidly, distances computations to them are arithmetically inexpensive. In contrast, object comparisons to derive larger dissimilarity representations may be prohibitive.

Conclusions

- Just a few long feature lines are needed to describe correlated data sets, in comparison with the number of short feature lines required to reach a similar performance.

- We proposed to use the length of the line as a alternative selection criterion. According to our experiments, middle-length feature lines may provide an accurate representation for slightly curved subspaces than the description provided by the longest lines.

- We conducted a number of experiments for classifying seismic volcanic signals. In particular, we followed a dissimilarity-based approach, deriving a dissimilarity representation from an initial spectral-based one.

Conclusions

- We also explored the use of clustering techniques for detecting mislabeled events in the same application. As suggested by the researchers from the Volcanological and Seismological Observatory, misclassifications are committed quite often due to both inexperience and tedious work.

- We addressed the problem of selecting spectral bands of seismic volcanic signals. Such techniques effectively allow us to find a small number of discriminative spectral regions. The physical interpretation of the selected peaks are beyond of this study but, certainly, a study in such direction would increase the impact of spectral selection techniques in the real application.

Open questions and future work

- Further research studies in dimensionality reduction for lines and planes as an alternative to regularization.
  - Selection procedures based on geometric properties such as the length of the base of the triangle or the area of the base of the tetrahedrons.
  - Rectification procedure.
  - Sparse classifiers to implicity select feature lines or feature planes (eigenspectrum).

- The problem of classifying seismic volcanic signals is far of being solved: multiple classifier systems, adaptive pattern recognition systems and further studies in spectral based representations.

Outline

- Introduction, problem statement and objectives
- Dissimilarity representations and generalizations
- Generalization of dissimilarity representations by feature lines and feature planes
  - Nearest feature classifiers
  - Generalizing dissimilarity representations by feature lines and feature planes
- Volcanic seismology and pattern recognition
  - Dissimilarity-based classification of seismic signals
  - Clustering of seismic signals for detecting mislabeling
  - Band selection of seismic spectra
- Conclusions, open questions and future work
- Publications

Publications

Books:

1. *Entrenamiento de Sistemas en la Identificación Automática de Patologías*, by Álvaro Ángel Orozco-Gutiérrez, César Germán Castellanos-Domínguez, Julio Fernando Suárez-Cifuentes, and Mauricio Orozco-Alzate. **Publicaciones Universidad Tecnológica de Pereira** (UTP), To appear.

Chapter books:

2. "Trends in Nearest Feature Classification for Face Recognition – Achievements and Perspectives," by M. Orozco-Alzate and C.G. Castellanos-Domínguez, in Face Recognition: State of the Art in Cognitive and Computational Processes, Vienna, Austria: **I-TECH Education and Publishing**, To appear.
3. "Nearest Feature Rules and Dissimilarity Representations for Face Recognition Problems," by M. Orozco-Alzate and C. G. Castellanos-Domínguez, in Face Recognition, pp. 337–356, ser. Artificial Intelligence, K. Delac and M. Grgic, Eds. Vienna, Austria: **I-TECH Education and Publishing**, June 2007.

Publications

Journal publications:

4. "A generalization of dissimilarity representations using feature lines and feature planes", by M. Orozco-Alzate, R.P.W. Duin ,and C. G. Castellanos-Domínguez, in ***Pattern Recognition Letters (Elsevier)***, to appear.
5. "Clustering on dissimilarity representations for detecting mislabelled seismic signals at Nevado del Ruiz volcano", by Mauricio Orozco-Alzate, and César Germán Castellanos-Domínguez, in ***Earth Sciences Research Journal***, vol. 11, no. 2, pp. 131–138, December 2007.
6. "Dissimilarity based classification of seismic signal and Nevado del Ruiz volcano," by M. Orozco-Alzate, M. E. García-Ocampo, R. P. W. Duin, and C. G. Castellanos-Domínguez, in ***Earth Sciences Research Journal***, vol. 10, no. 2, pp. 57–65, December 2006.
7. "Comparison of the nearest feature classifiers for face recognition", by M. Orozco-Alzate and C. G. Castellanos-Domínguez, in ***Machine Vision and Applications (Springer)***, vol. 17, no. 5, pp. 279–285, October 2006.

Publications

Conferences and symposia:

8. "Spectral characterization of volcanic earthquakes at Nevado del Ruiz Volcano using spectral band selection/extraction techniques," by M. Orozco-Alzate, M. Skurichina, and R. P. W. Duin, in Progress in Pattern Recognition, Image Analysis and Applications. Proceedings of the 13th Iberoamerican Congress on Pattern Recognition **CIARP 2008**, ser. Lecture Notes in Computer Science, J. RuizShulcloper and W. G. Kropatsch, Eds., vol. 5197, IAPR. Havana, Cuba: Springer, September 2008, pp. 708–715.
9. 'Generalizing Dissimilarity Representations Using Feature Lines," by M. Orozco-Alzate, C. G. Castellanos-Domínguez and R. P. W. Duin, in Progress in Pattern Recognition, Image Analysis and Applications. Proceedings of the 12th Iberoamerican Congress on Pattern Recognition **CIARP 2007**, ser. Lecture Notes in Computer Science, L. Rueda, D. Mery and J. Kittler, Eds., vol. 4756, IAPR. Viña del Mar-Valparaiso, Chile: Springer, November 2007, pp. 370–379.

Publications

Conferences and symposia:

10. "On selecting middle-length feature lines for dissimilarity-based classification," M. Orozco-Alzate, R. P. W. Duin and C. G. Castellanos-Domínguez, in XII Simposio de Tratamiento de Señales, Imágenes y Visión Artificial, **STSIVA 2007**, Universidad del Norte, September 2007.
11. "Quantifying the computational complexity of the nearest feature classifiers," in Segundo Simposio Regional IEEE de Electrónica y Aplicaciones Industriales **SREAI 2006**, pp. 1–5, Manizales, Colombia, October 2006.
12. "Dissimilarity-based classification of seismic signals at Nevado del Ruiz volcano," In **2nd Latin-American Congress of Seismology**, Bogotá, Colombia, August 2006.

Acknowledgments

Scholarship Program for Outstanding Postgraduate Students from Universidad Nacional de Colombia

TU Delft Research Grant

Research Groups:

- Control & Digital Signal Processing Group, Universidad Nacional de Colombia Sede Manizales.
- Information and Communication Theory Group, Delft University of Technology. The Netherlands.

Spectral representation and classification

General problem: In order to use the forecasting potential of LP events fully, one must first be able to distinguish their signatures from those of VT earthquakes, a task made difficult by the extreme heterogeneity of volcanic media

Differences in spectral content allow a discrimination of different types of volcanic earthquakes

$$\Downarrow$$

- Spectral-based classification is a natural approach to face the problem.
- Some physical insight into magma properties may be derived from a spectral study.

Spectral representation and classification

The always increasing capacity of sensory systems and data storage provides high quality and densely sampled spectral measurements.

Particular problem: Improvement in data acquisition challenges the performance of analysis algorithms, e.g. classifiers of volcanic events, that have to deal with a huge amount of data to be processed.

Fact: Spectral information is largely redundant by nature, i.e. its intrinsic dimensionality is low.

Solution: Reducing data redundancy and finding an understandable and reduced number of discriminative regions by using spectral band selection/extraction techniques.

Data

Signals were selected from data collected by the OVSM monitoring network deployed on the Nevado del Ruiz volcano

Composition: 483 VT events and 580 LP events.

Monitoring network: composed by $15 \sim 19$ stations (seismometers) deployed in the volcanic complex.

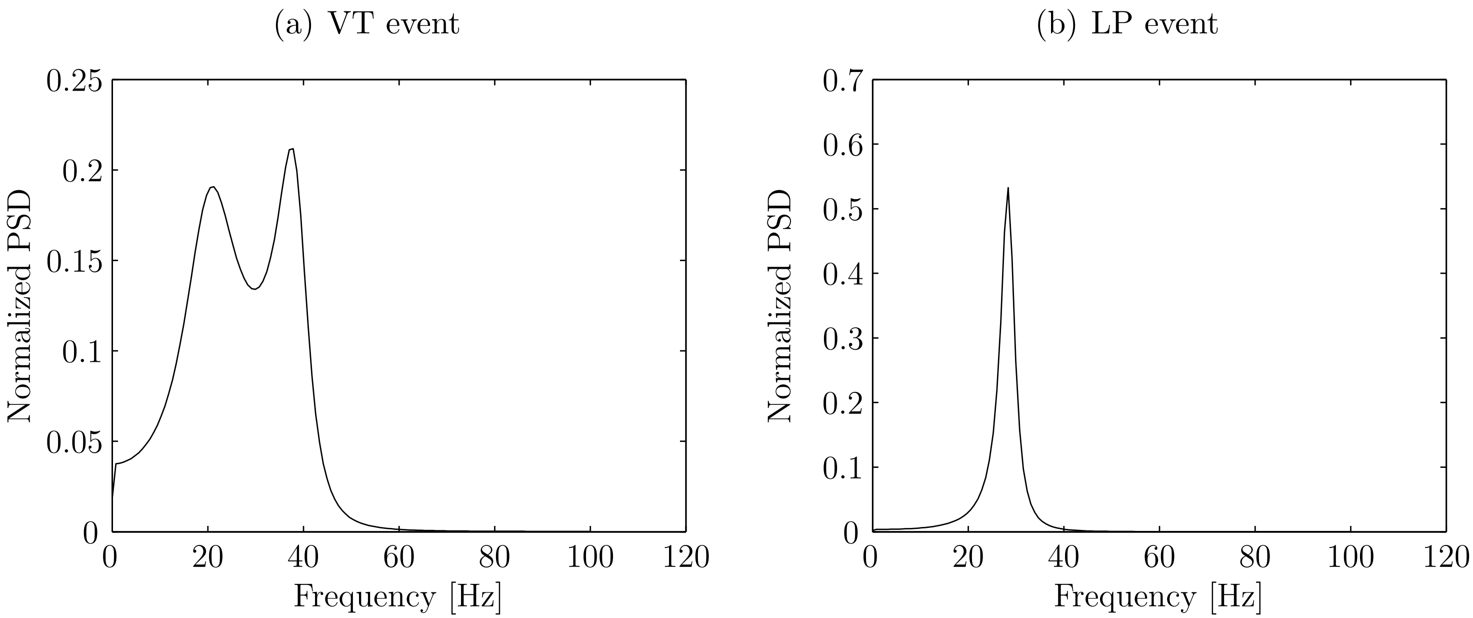
Selected station: *Olleta crater* station. It is considered a reference for the volcanic-related events. Besides, by considering signals recorded at the same location, we reduce the influence of the so-called path effect.

Location: 4.08 km from the active crater.
Digitization: 100.16 Hz sampling rate, 12 bits ADC.

Spectral representations

Fast Fourier Transform (FFT): N-point data-based spectral estimation.

Power Spectral Density (PSD): model-based spectral estimation. Particularly, a Yule-Walker AR method was used.



Spectral band selection methods

To evaluate a discriminative capacity of the extracted spectral regions, we use the Mahalanobis Distance (MD) between data classes as a discriminant measure (criterion):

$$MD = (\mu_A - \mu_B)' \left(p\Sigma_A + (1-p)\,\Sigma_B\right)^{-1} (\mu_A - \mu_B)\,,$$

where μ_A, μ_B and Σ_A, Σ_B are the means and the covariance matrices of data classes A and B, respectively; p is the prior probability of the data class A.

The larger MD, the larger discriminative capacity between data classes.

We use the mean function, i.e. taking the average of spectral intensities in the region, to reduce the dimensionality of each considered spectral region to a single value representation.

Spectral band selection methods

The following spectral band extraction/selection techniques are considered in this study:

1. *GLDB-TD*: Top-down multiresolution feature extraction algorithm
2. *GLDB-BU*: Bottom-up generalized local discriminant bases algorithm
3. *SP*: Sequential partitioning
4. *SPE*: Sequential partitioning and elimination of uninformative spectral bands
5. *SS*: Sequential selection
6. *SSN*: Sequential selection of non-overlapping discriminative spectral regions
7. *FP*: Floating partition

A detailed description of each method is available in [Orozco-Alzate et al., 2008b]

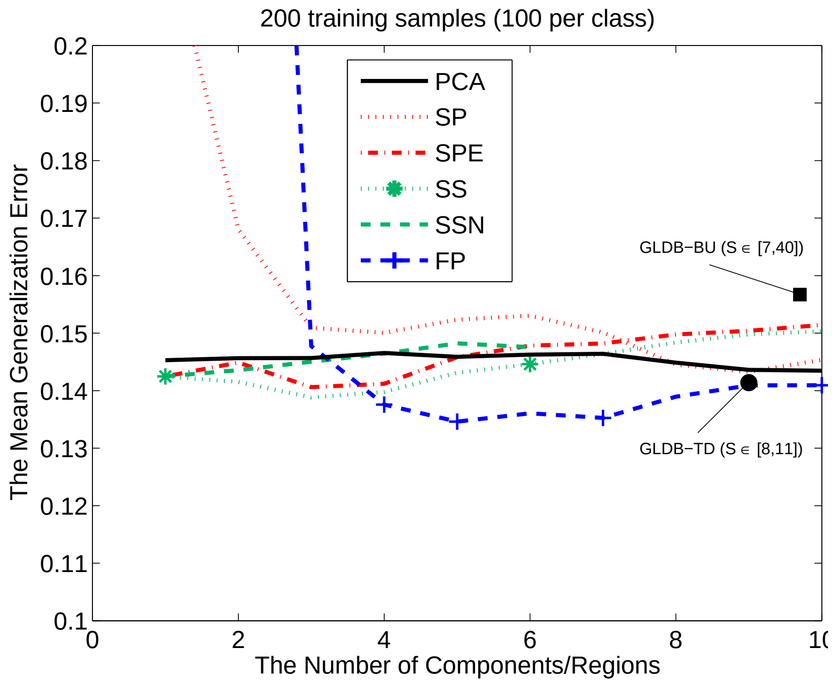
Experimental results and discussion

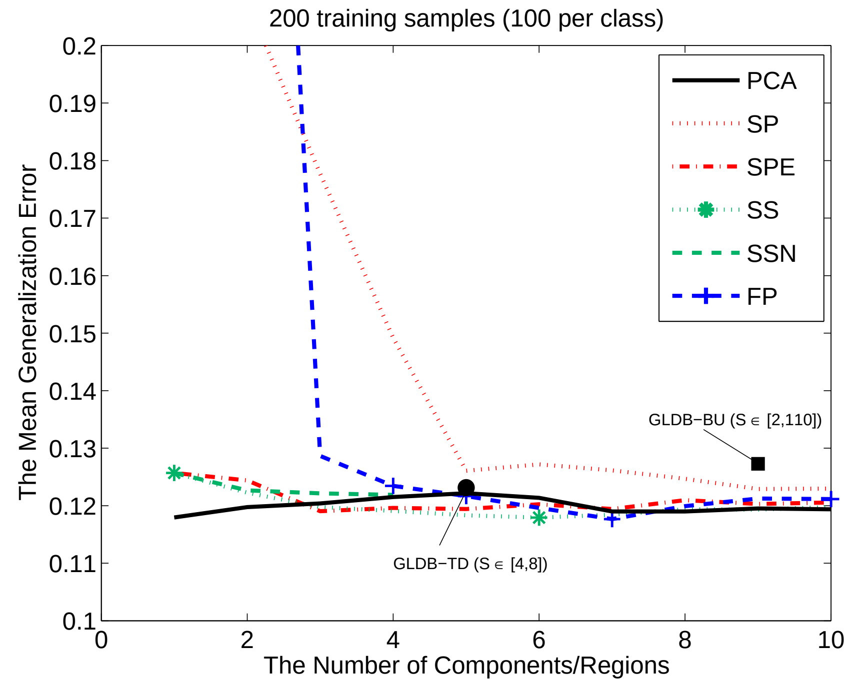
Experimental setup

- Training data sets with 100 objects per class, randomly chosen from the total set, for the two spectral representations: FFT and PSD.
- Remaining data used for testing.
- Prior class probabilities set to be equal as the data are unbalanced and the real prior class probabilities are unknown.
- Regularized Linear Classifier (RLC) for evaluating the performance of feature selection/extraction methods ($\lambda = 10^{-8}$)
- Experiments repeated 20 times on independent training sets.

Experimental results and discussion

Average generalization error of RLC using different methods to select discriminative spectral regions for the *Ruiz-FFT* (left plot) and the *Ruiz-PSD* (right plot) datasets:



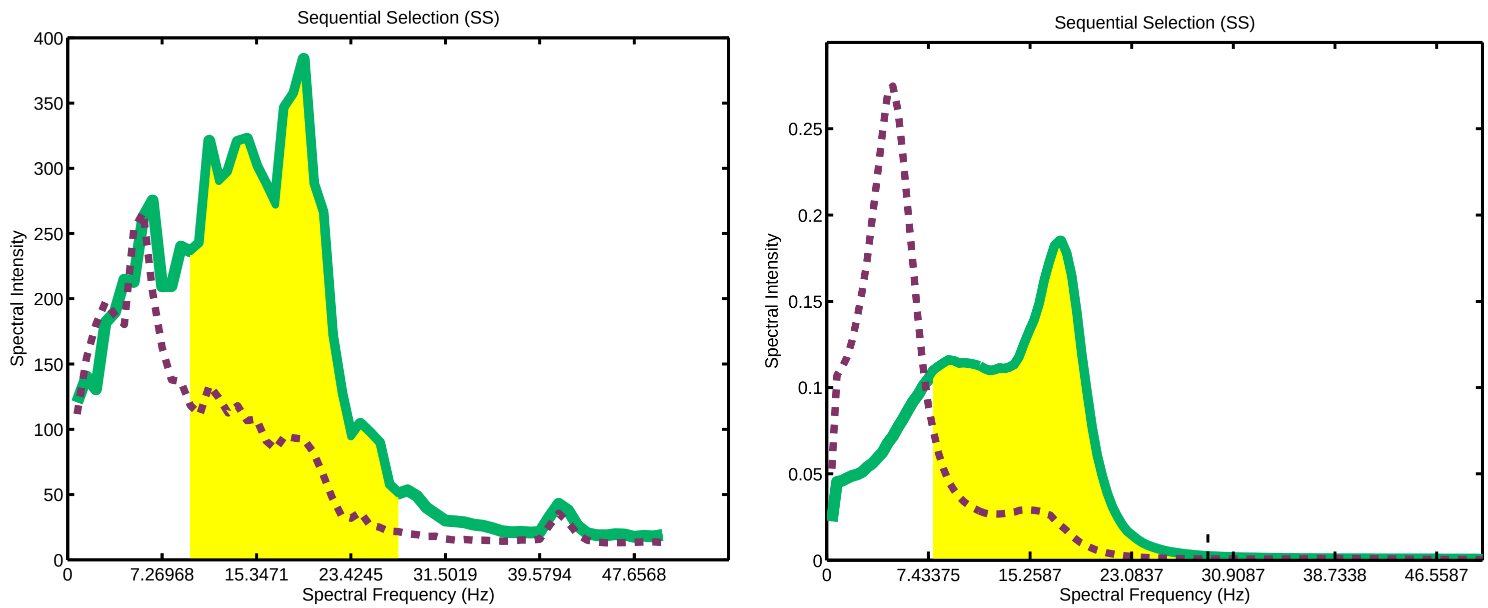


- 20 trials, maximum of 10 extracted spectral bands, $\sigma \cong 0.01$
- GLDB-TD and GLDB-BU terminate automatically according to a data-driven criterion → a single point is given in each plot

Experimental results and discussion

Selection of a single spectral region

- PCA (a single component), SS and SPE strategies are the best performing ones; in particular, PCA is the best one for the *Ruiz-PSD* dataset.
- There is no difference between SS, SSN and SPE, as they converge to the same solution.

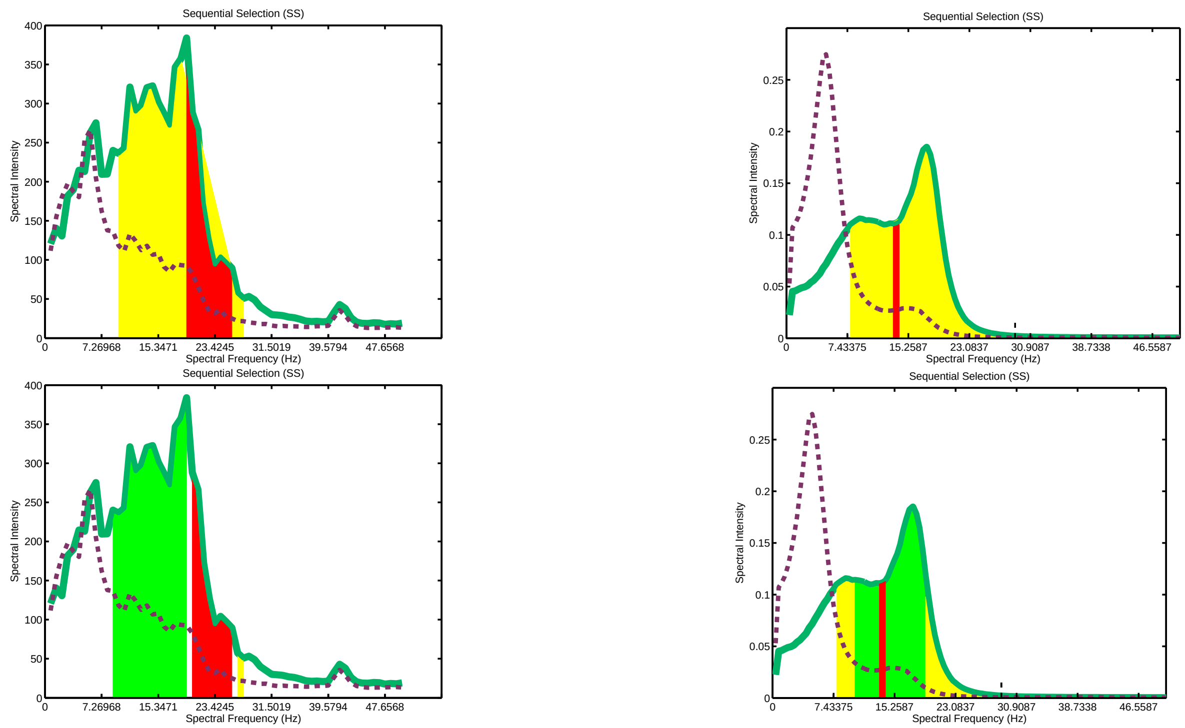


Consistent results: selected spectral region roughly from 8 Hz up to 27.5 Hz, where a significant part of energy associated to the VT events is placed.

Experimental results and discussion

The SS strategy for 2–3 regions

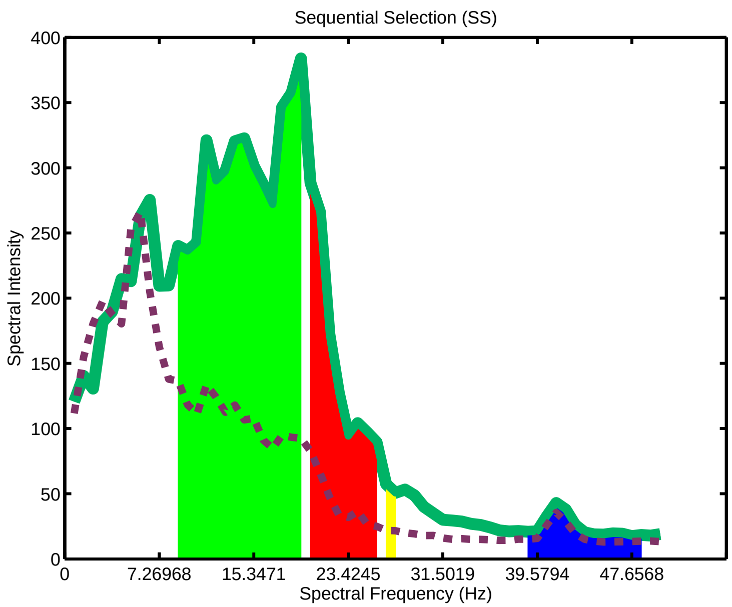
- One of the best performing spectral band selection methods. It benefits the most for small number of spectral regions
- Its performance gradually degrades because each new selected spectral region depends on the sequence of the regions selected before

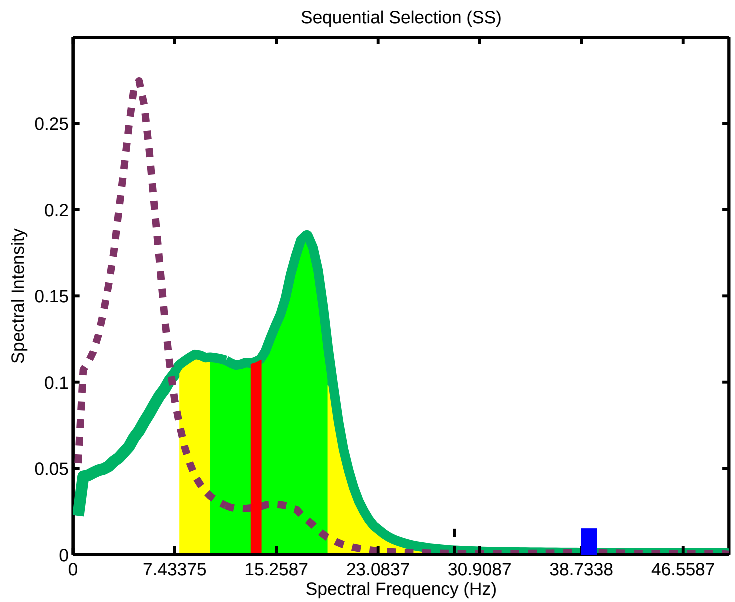


Experimental results and discussion

The SS strategy for 4 regions

- Interesting case!
- The first three bands indicate that most of the discriminative information is contained between 7.5 Hz and 25 Hz approximately.
- A narrow band around 40 Hz, associated to the peak observed in both classes, is also selected.

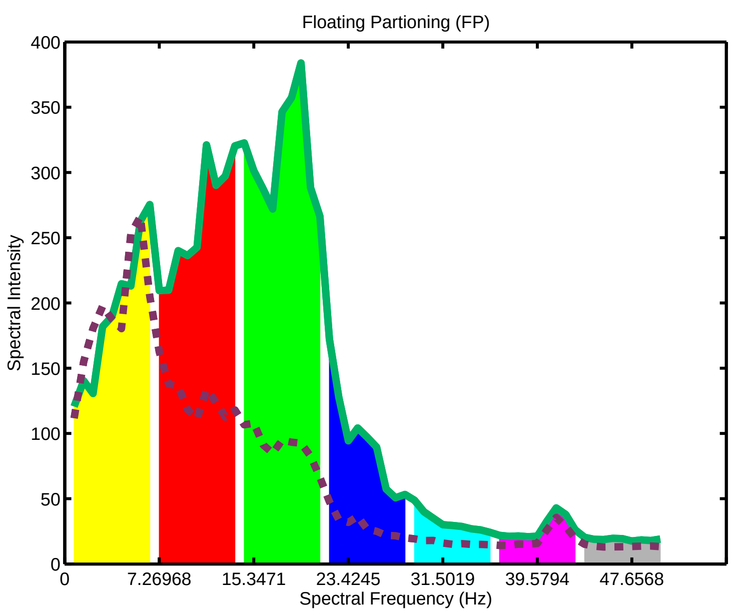


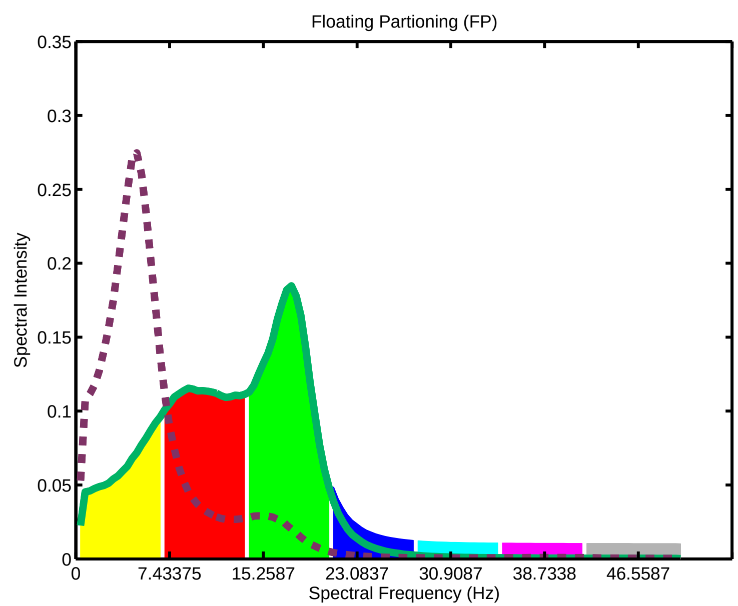


Experimental results and discussion

The FP strategy

- It depends very much on the initial partition
- It outperforms other techniques when spectra are split into five or more spectral regions
- Result is very close to the initial uniform partition.
- It is not very useful for interpretation and dimensionality reduction





Concluding remarks

- Spectral band selection techniques allow us to find a small number of discriminative spectral regions as well as to discard spectral peaks, apparently significant, which might be attributed to path effects and near surface resonance.

- Differences between selected bands and average spectra shapes suggest that the influence of the spectral estimation methods should be analyzed carefully.

- The two types of volcanic events share some dominant and discriminative regions, e.g. the narrow band near 40 Hz, suggesting a common source of volcanic process.

- Further discussion, based on the opinion of experts (geologists, volcanologists), might give some insight about the physical phenomena associated to VT and LP events. For instance, *magma viscosity*, *rock hardness/composition* among others properties.

References

[Bicego et al., 2004] Bicego, M., Murino, V., and Figueiredo, M. A. T. (2004). Similarity-based classification of sequences using hidden Markov models. *Pattern Recognition*, 37(12):2281–2291.

[Chien and Wu, 2002] Chien, J.-T. and Wu, C.-C. (2002). Discriminant waveletfaces and nearest feature classifiers for face recognition. *IEEE Trans. Pattern Anal. Machine Intell.*, 24(12):1644–1649.

[Duin and Pękalska, 2005] Duin, R. P. W. and Pękalska, E. (2005). Open issues in pattern recognition. In Kurzynski, M., Puchala, E., Wozniak, M., and Zolnierek, A., editors, *Computer Recognition Systems (Proc. of 4th Int. Conf. on Computer Recognition Systems CORES'05)*, volume 30 of *Advances in Soft Computing*, pages 27–42, Rydzyna Castle, Poland. Springer.

[Goldfarb, 2004a] Goldfarb, L. (2004a). Pattern representation and the future of pattern recognition: Introductory talk. Satellite Workshop of 17th International Conference on Pattern Recognition.

[Goldfarb, 2004b] Goldfarb, L. (2004b). Representational formalisms: why we haven't had one. In Goldfarb, L., editor, *Pattern Representation and the Future of Pattern Recognition (Proc. Satellite Workshop of 17th International Conference on Pattern Recognition)*, Cambridge, UK.

[Jain et al., 2000] Jain, A. K., Duin, R. P. W., and Mao, J. (2000). Statistical pattern recognition: A review. *IEEE Trans. Pattern Anal. Machine Intell.*, 22(1):4–37.

[Kim, 2006] Kim, S.-W. (2006). On using a dissimilarity representation method to solve the small sample size problem for face recognition. In Blanc-Talon, J., Philips, W., Popescu, D., and Scheunders, P., editors, *Advanced Concepts for Intelligent Vision Systems, Proceedings of the 8th International Conference, ACIVS 2006*, volume 4179 of *Lecture Notes in Computer Science*, pages 1174–1185, Antwerp, Belgium. Springer.

[Langer et al., 2006] Langer, H., Falsaperla, S., Powell, T., and Thompson, G. (2006). Automatic classification and a-posteriori analysis of seismic event identification at Soufrière Hills volcano, Montserrat. *Journal of Volcanology and Geothermal Research*, 153:1–10.

[Li and Lu, 1999] Li, S. Z. and Lu, J. (1999). Face recognition using the nearest feature line method. *IEEE Trans. Neural Networks*, 10(2):439–443.

[Orozco-Alzate and Castellanos-Domínguez, 2006] Orozco-Alzate, M. and Castellanos-Domínguez, C. G. (2006). Comparison of the nearest feature classifiers for face recognition. *Machine Vision and Applications (Springer)*, 17(5):279–285.

[Orozco-Alzate and Castellanos-Domínguez, 2007a] Orozco-Alzate, M. and Castellanos-Domínguez, C. G. (2007a). Clustering on dissimilarity representations for detecting mislabelled seismic signals at Nevado del Ruiz volcano. *Earth Sciences Research Journal*, 11(2):131–138.

[Orozco-Alzate and Castellanos-Domínguez, 2007b] Orozco-Alzate, M. and Castellanos-Domínguez, C. G.

(2007b). Nearest feature rules and dissimilarity representations for face recognition problems. In Delac, K. and Grgic, M., editors, *Face Recognition*, Artificial Intelligence, chapter 18, pages 337–356. I-TECH Education and Publishing, Vienna, Austria.

[Orozco-Alzate et al., 2007] Orozco-Alzate, M., Duin, R. P. W., and Castellanos-Domínguez, C. G. (2007). Generalizing dissimilarity representations using feature lines. In Rueda, L., Mery, D., and Kittler, J., editors, *Progress in Pattern Recognition, Image Analysis and Applications. Proceedings of the 12th Iberoamerican Congress on Pattern Recognition CIARP 2007*, volume 4756 of *Lecture Notes in Computer Science*, pages 370–379, Viña del Mar-Valparaiso, Chile. IAPR, Springer.

[Orozco-Alzate et al., 2008a] Orozco-Alzate, M., Duin, R.

P. W., and Castellanos-Domínguez, C. G. (2008a). A generalization of dissimilarity representations using feature lines and feature planes. To appear in Pattern Recognition Letters.

[Orozco-Alzate et al., 2008b] Orozco-Alzate, M., Skurichina, M., and Duin, R. P. W. (2008b). Spectral characterization of volcanic earthquakes at Nevado del Ruiz Volcano using spectral band selection/extraction techniques. In Ruiz-Shulcloper, J. and Kropatsch, W. G., editors, *Progress in Pattern Recognition, Image Analysis and Applications. Proceedings of the 13th Iberoamerican Congress on Pattern Recognition CIARP 2008*, volume 5197 of *Lecture Notes in Computer Science*, pages 708–715, Havana, Cuba. IAPR, Springer.

[Pękalska and Duin, 2002] Pękalska, E. and Duin, R. P. W.

(2002). Dissimilarity representations allow for building good classifiers. *Pattern Recognition Lett.*, 23:943–956.

[Pękalska and Duin, 2005] Pękalska, E. and Duin, R. P. W. (2005). *The Dissimilarity Representation for Pattern Recognition: Foundations and Applications*, volume 64 of *Machine Perception and Artificial Intelligence*. World Scientific, Singapore.