On Selecting Middle-Length Feature Lines for Dissimilarity-based Classification

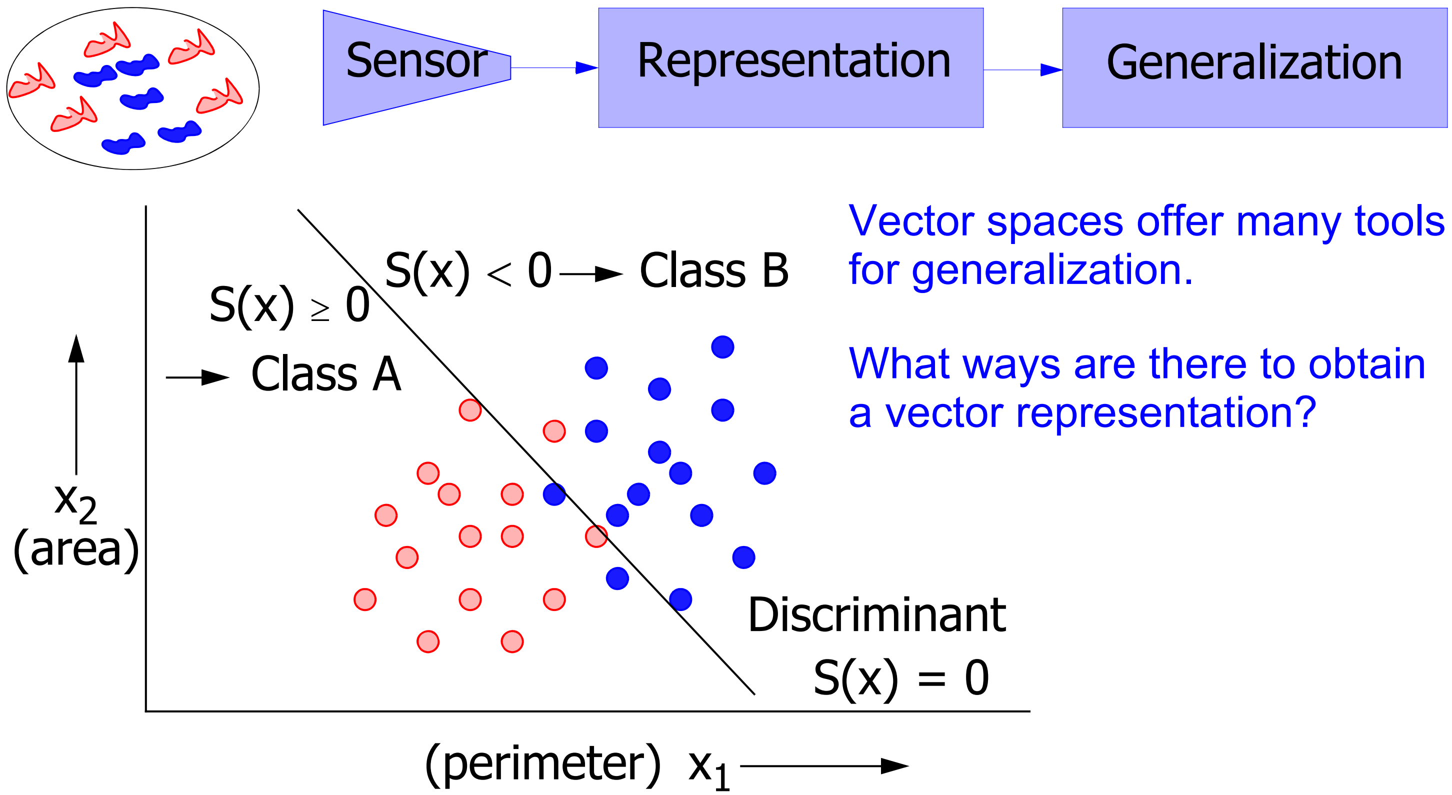
Mauricio Orozco-Alzate, Robert P. W. Duin,
Germán Castellanos Domínguez

Abstract

In dissimilarity-based classification, the use of feature lines resulting from the linear combination of pairs of prototypes is an option to overcome representational limitations. The choice of such feature lines is important issue, not just to obtain a good description but also to reduce the dimensionality. We consider the selection of the middle-length feature lines. Results show that they are more appropriate to represent moderately curved subspaces than the the longest ones.

UNIVERSIDAD NACIONAL DE COLOMBIA
SEDE MANIZALES

Vector Representations

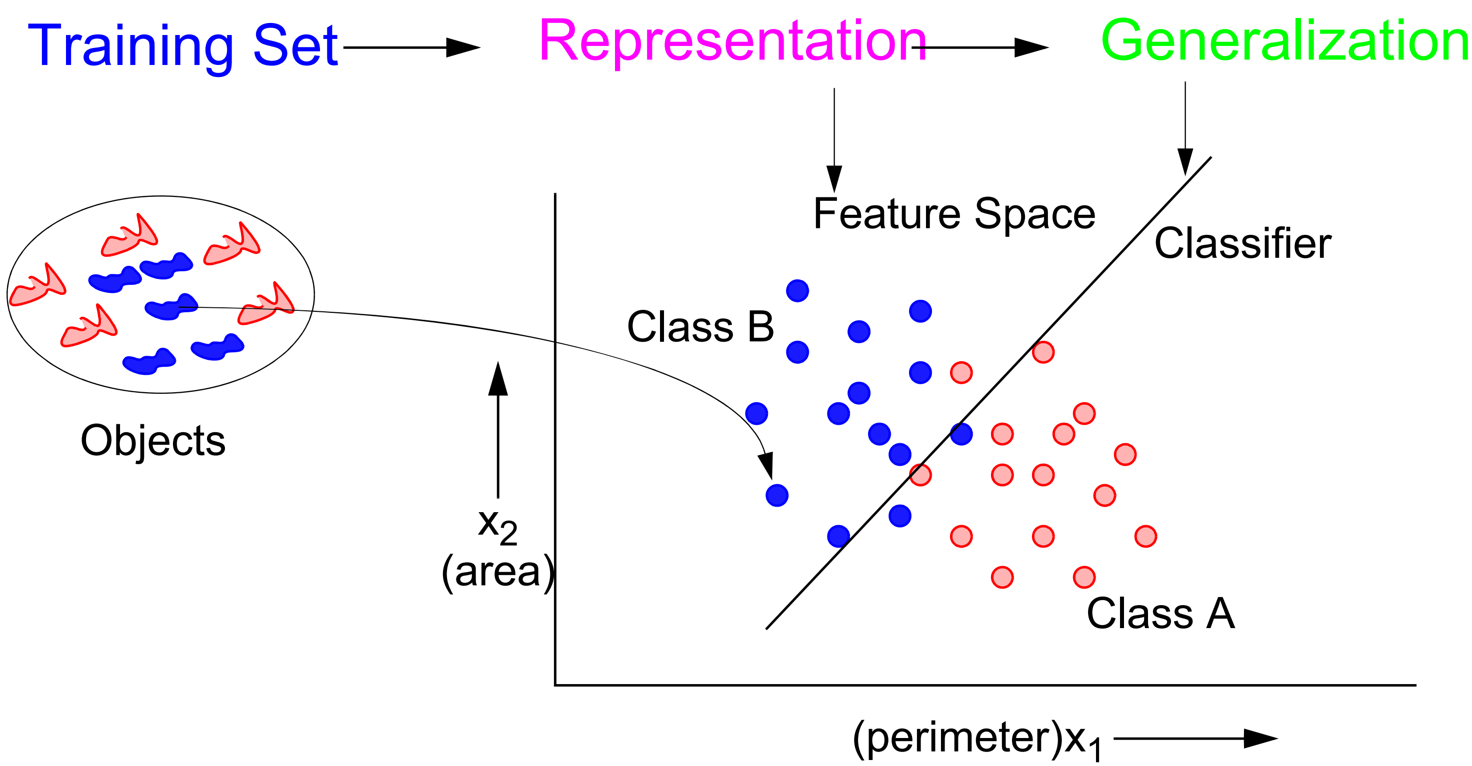


Vector spaces offer many tools for generalization.

What ways are there to obtain a vector representation?

Feature based representation: traditional approach in Pattern Recognition and based on knowledge

Vector Representations

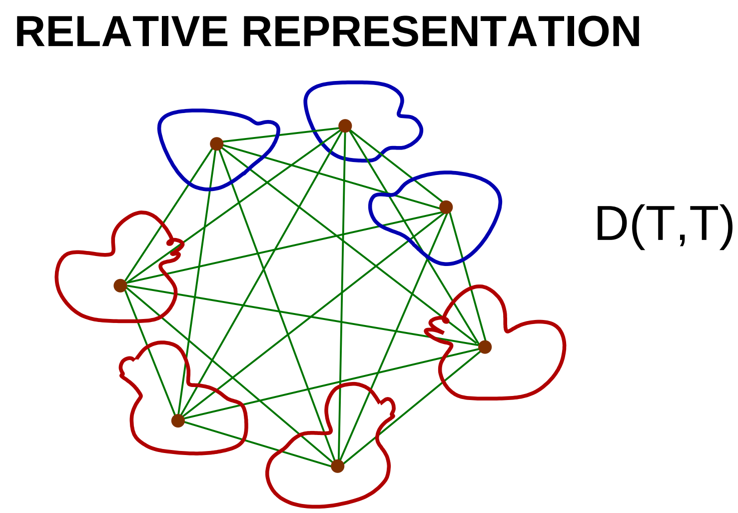


Knowledge ⇒ good features ⇒ (almost) separable classes
Lack of knowledge ⇒ (too many) bad features ⇒ hardly separable classes
Many features ∼ Lack of knowledge

Vector Representations

Other (alternative) representations:

- Pixel (sample) based representations: Raw data
- **Dissimilarity based representation:** Recent, expert based



- Conceptual representation: Combining classifiers

Dissimilarity representations

- A number of studies showed advantages of learning from dissimilarity representations instead of learning from feature-based representations: [Duin et al., 1998, Pękalska et al., 2001, Pękalska and Duin, 2002, Paclík and Duin, 2003, Pękalska and Duin, 2005].

- Based on pairwise comparisons and expressed as an $N \times n$ dissimilarity matrix $D(T,R)$, where each entry corresponds to a dissimilarity between pairs of objects:

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

Dissimilarity representations

- 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.

Dissimilarity-based classification

- The nearest neighbor rule (1-NN): traditional, simple but computationally expensive and sensitive to noise
- The dissimilarity space approach:

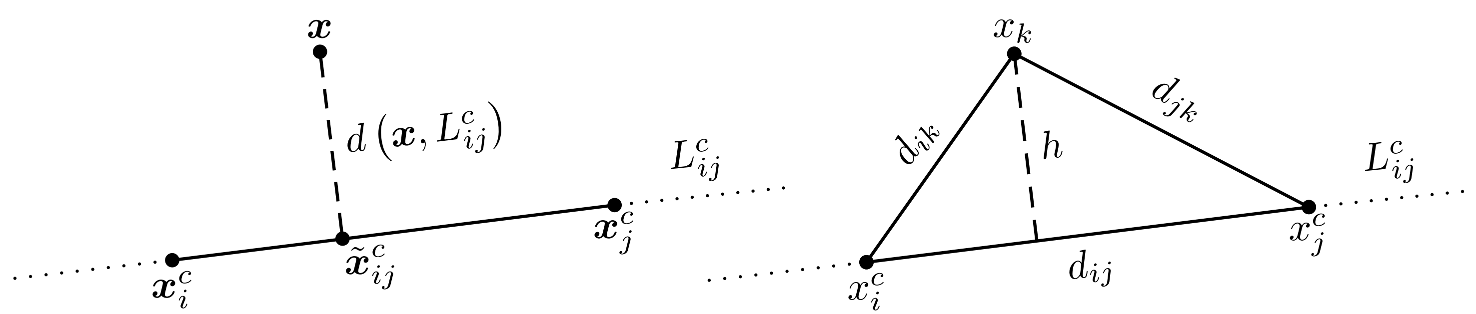
$$D(T,T) = \begin{matrix} & \Rightarrow \\ \Downarrow & \end{matrix}\begin{pmatrix} d_{11} & d_{12} & d_{13} & d_{14} & d_{15} & d_{16} & d_{17} \\ d_{21} & d_{22} & d_{23} & d_{24} & d_{25} & d_{26} & d_{27} \\ d_{31} & d_{32} & d_{33} & d_{34} & d_{35} & d_{36} & d_{37} \\ d_{41} & d_{42} & d_{43}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

  - Consider dissimilarities as “features”
  - Select “features”, i.e. representation objects
  - Build a trained classifier, e.g. a Bayesian one, on top of the representation

Generalization of dissimilarity representations using feature lines

$\Longrightarrow$ Dissimilarity representation computed from a very small set of prototypes: representational limitation

- The principle of the nearest feature line [Li and Lu, 1999]
- Deriving the generalized dissimilarity representation: distances to feature lines in terms of dissimilarities [Orozco-Alzate and Castellanos-Domínguez, 2007]



Generalization of dissimilarity representations using feature lines

Solve for h:

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

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

$$\Downarrow$$

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

where $d_{jk} = D_L(x_j, L_k)$.

The number of lines increases combinatorially!: selection of feature lines based on a simple criterion

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 :

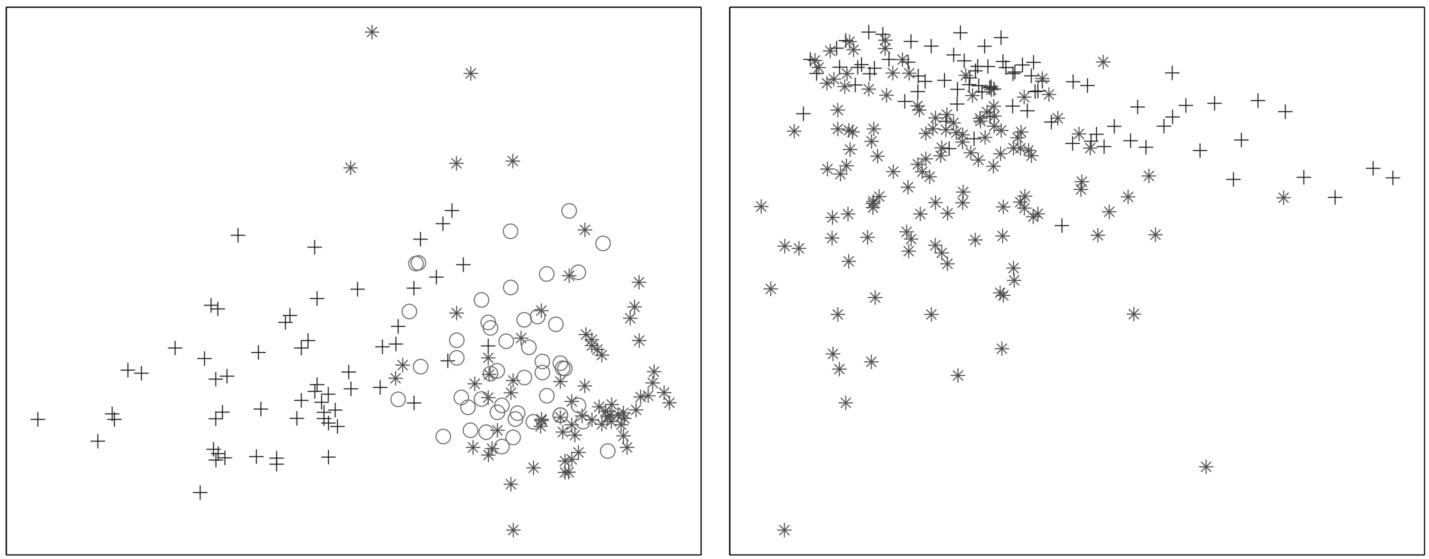
- Start the selection in the middle of the ranking.
- Include in R_L the feature line exactly placed in in the middle of the sorted list.
- take feature lines placed at its left and right sides alternately.

$$\Downarrow$$

Piece-wise description!

Experimental setup

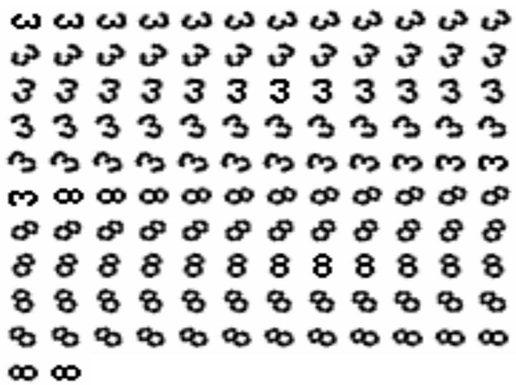
Data sets



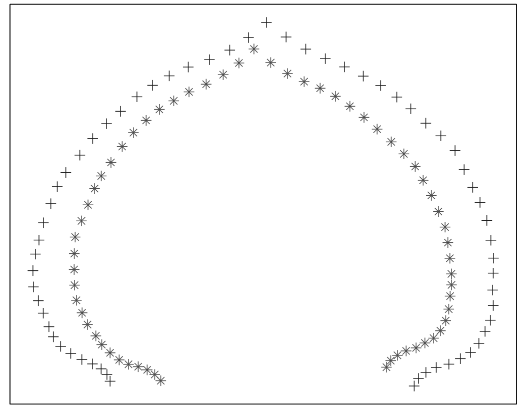
(a) *Wine* data set   (b) *Laryngeal* data set

Figure 1: Scatter plots using a classical multidimensional scaling

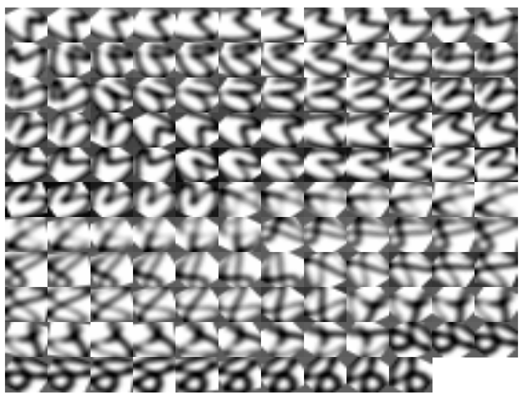
Experimental setup



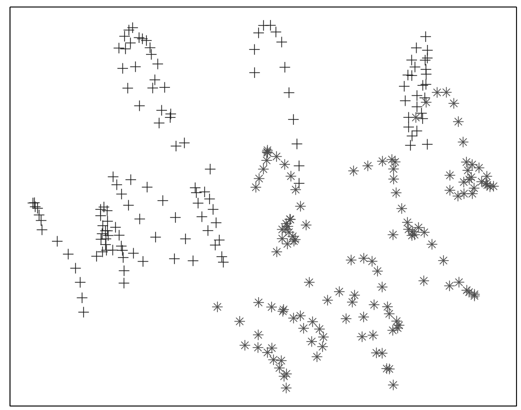
(a) Rotated machine-printed digits



(b) Scatter plot using classical MDS



(c) Rotated handwritten digits



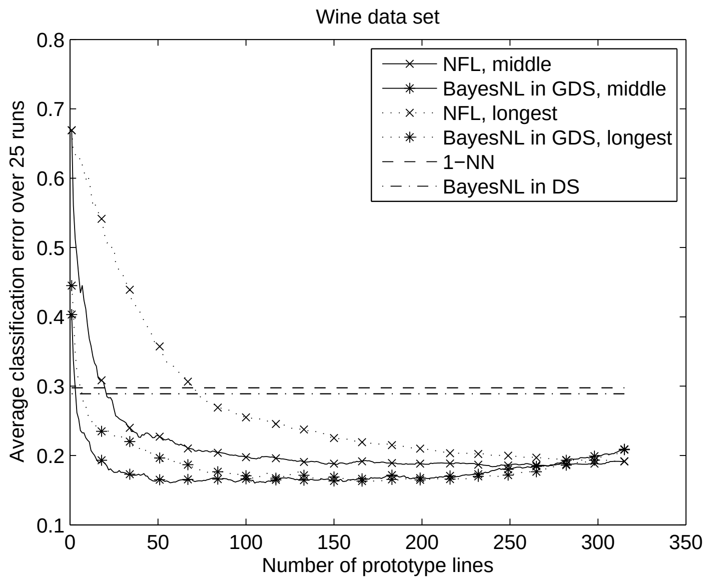
(d) Scatter plot using classical MDS

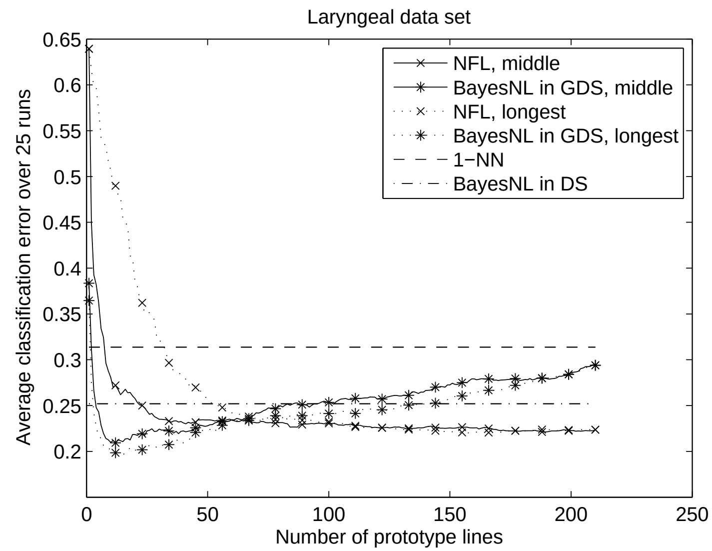
Figure 2: Digits ‘3’ and ‘8’ rotated (a)-(b) between -90° and 90° with steps of 3° and (c)-(b) between -60° and 60° with steps of 10°.

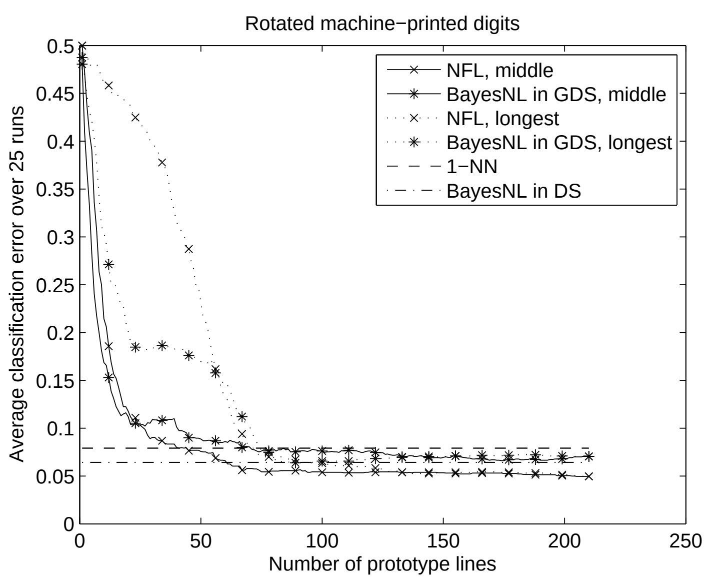
Experimental setup

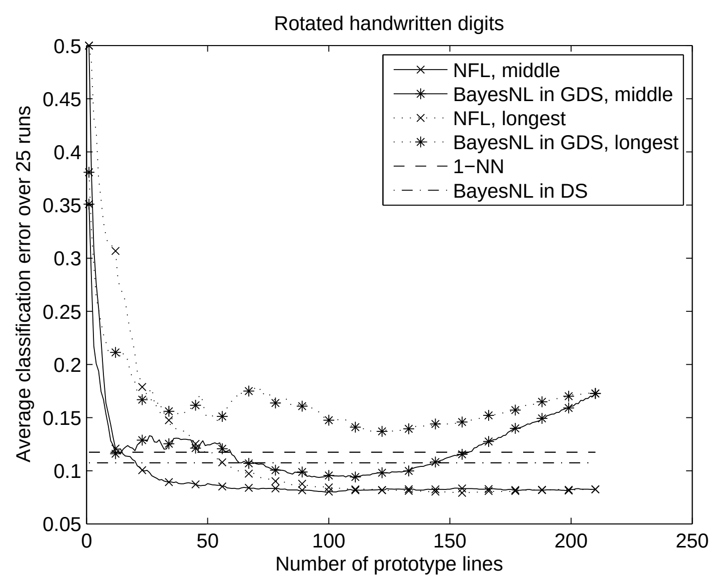
- Maximum number of prototypes considered: $r_c = 15$.
- Total number of feature lines: 315 (*Wine*, three-class problem) and 210 (other two-class data sets).
- For each repetition, a new representation set R is randomly chosen.
- The best results, that we used as a reference (horizontal line), do not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

Experimental results









Discussion and conclusions

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

Discussion and conclusions

- 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.

Contact:

M. Orozco-Alzate `http://orozco.co.nr/`
R.P.W. Duin `http://www-ict.ewi.tudelft.nl/~duin/`

References

[Duin et al., 1998] Duin, R. P. W., de Ridder, D., and Tax, D. M. J. (1998). Featureless pattern classification. *Kybernetika*, 34(4):399–404.

[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, 2007] Orozco-Alzate, M. and Castellanos-Domínguez, C. G. (2007). Nearest feature rules and dissimilarity representations for face recognition problems. In Delac, K. and Grgic, M., editors, *Face Recognition*, chapter 18, pages 337–356. I-TECH Education and Publishing, Vienna, Austria. To appear.

[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 *accepted to be presented at the 12th Iberoamerican Congress on Pattern Recognition*.

[Paclík and Duin, 2003] Paclík, P. and Duin, R. P. W. (2003). Dissimilarity-based classification of spectra: computational issues. *Real Time Imaging*, 9:237–244.

[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. World Scientific, Singapore.

[Pękalska et al., 2001] Pękalska, E., Paclík, P., and Duin, R. P. W. (2001). A generalized kernel approach to dissimilarity based classification. *J. Mach. Learn. Res.*, 2(2):175–211.