Pattern Recognition Letters 1 (1983) 287–289
North-Holland

July 1983

A learning scheme for a fuzzy k-NN rule

Adam JÓŹWIK
Institute of Biocybernetics and Biomedical Engineering, Polish Academy of Sciences, 00-818 Warsaw, KRN 55, Poland

Received 1 February 1983

Abstract: The performance of a fuzzy k-NN rule depends on the number k and a fuzzy membership-array $W[l, m_R]$, where l and m_R denote the number of classes and the number of elements in the reference set X_R respectively. The proposed learning procedure consists in iterative finding such k and W which minimize the error rate estimated by the 'leaving one out' method.

Key words: NN rules, learning procedure, fuzzy decisions, probability of misclassification.

1. The fuzzy k-NN rule

The fuzzy k-NN rule operates as follows. Let

$$X_R = \{x_i\}_{i=1}^{m_R}$$

be the reference set and

$$W = \{w_i\}_{i=1}^{m_R}$$

be a set of l-dimensional vectors, where l is the number of classes,

$$w_i = (w_{i,1}, w_{i,2}, \ldots, w_{i,l}),$$

$$\sum_{j=1}^{l} w_{i,j} = 1 \quad \text{and} \quad 0 \le w_{i,j} \le 1.$$

Each $w_{i,j}$, $1 \le i \le m_R$, $1 \le j \le l$, is a membership-value of the i-th object (represented in the n-dimensional feature space by $\boldsymbol{x}_i$ and identified with $\boldsymbol{x}_i$) to the class j. The set W is identified with a membership-array $W[l, m_R]$. It is assumed that X_R and W are given.

For each $\boldsymbol{x}$ to be classified, the set K of indices that correspond to the k nearest neighbors of $\boldsymbol{x}$ in X_R is found and the fuzzy decision-vector

$$v = \left(\sum_{s \in K} w_s\right) \Big/ k$$

is assigned. If one is interested in a nonfuzzy decision then the vector $\boldsymbol{x}$ is assigned to the class associated with the largest number v_j, $1 \le j \le 1$, where

$$(v_1, v_2, \ldots, v_1) = v.$$

Ties are broken randomly or by the single NN rule. If all $w_{i,j}$, $1 \le i \le m_R$, $1 \le j \le 1$, are equal to 0 or 1 then the fuzzy k-NN rule is equivalent to the common k-NN rule.

2. The learning procedure

The following input data are given: the training set X_T (m_T elements), the set X_C (m_C elements if $X_C \ne \emptyset$) of objects to be classified and the membership-sequence

$$u_T = (u_{T,i})_{i=1}^{m_T},$$

where $u_{T,i} = j$ if $\boldsymbol{x}_i$ belongs to the class j.

The procedure described below consists of two stages. In the first stage the following sequence is generated:

$$(W_0, k_0, p_0), (W_1, k_1, p_1), \ldots, (W_h, k_h, p_h), \ldots, \tag{1}$$

where W_h, k_h and p_h denote the membership-array, the number of nearest neighbours and the expected probability of misclassification for the

k_h-NN rule characterized by the array W_h and the reference set $X_R = X_T$. The array W_0 is binary. It is derived directly from the sequence u_T, i.e. $w_{i,j}=1$ if $u_{T,i}=j$. The sequence (1) is generated by an application of the 'leaving one out' method (introduced by Lachenbruch (1965)). W_{h+1}, k_h and p_h are derived from the array W_h and the set X_T. Each x_i, $i=1,2,\dots,m_T$, is simultaneously classified by the fuzzy $1,2,\dots,m_T-2$ and m_T-1 NN rules with respect to $X_R = X_T \setminus \{x_i\}$. Thus, to each x_i and k, with k the number of neighbours, the fuzzy decision-vector $v_{k,i}$ and, at the same time, the nonfuzzy decision $u_{k,i}$ are assigned. Comparing the decision-sequences

$$u_k = (u_{k,i})_{i=1}^{m_T}, \quad k=1,2,\dots,m_T,$$

with the membership-sequence u_T, we may find the error rates $q_k = e_k/m_T$, where e_k denotes the number of objects x_i, $i=1,2,\dots,m_T$, misclassified by the fuzzy k-NN rule. As k_h such k is taken that offers the minimum error rate

$$p_h = \min_k q_k.$$

The vector

$$w_{h+1,i} = (v_{k_h,i}k_h + w_{h,i})/(k_h+1) \tag{2}$$

forms a row of the array W_{h+1}. The relation (2) gives the same effect as if the vector x_i were classified by the fuzzy (k_h+1)-NN rule subject to $X_R = X_T$ (not $X_R = X_T \setminus \{x_i\}$).

The triplet (W_*, k_*, p_*) that corresponds to the smallest index h such that $p_h \le p_{h+1}$ is treated as a result of the first stage of learning procedure.

In the second stage also a sequence of the form (1) is generated. But this time the set $W_0 = W_* \cup V_*$, where W_* is the set of the membership-vectors that form the array W_* and V_* is the set of the decision-vectors obtained by classification of all objects from the set

$$X_C = \{x_i\}_{i=m_T+1}^{m_C+m_T}$$

by the fuzzy k_*-NN rule with the membership-array W_* and the reference set $X_R = X_T$. The arrays

$$W_{h+1}[1, m_T+m_C], \quad h=0,1,\dots,$$

are derived from the arrays W_h respectively, exactly in the same way as they were derived in the first stage, i.e. classifying all objects from the set $X_T \cup X_C$ by the 'leaving one out' method and applying formula (2). However, the optimum numbers k_h and p_h are found by comparing the decision-sequences

$$u_k = (u_{k,i})_{i=1}^{m_T}, \quad k=1,2,\dots,m_T+m_C-1,$$

with the membership-sequence u_T.

The triplet (W_{**}, k_{**}, p_{**}) corresponding to the smallest index h_0 such that $p_{h_0} \le p_{h_0+1}$ is a result of the second stage of the learning procedure.

If $p_* \le p_{**}$ then the triplet (W_*, k_*, p_*) is taken as a final result. If $p_* > p_{**}$ then the triplet (W_{**}, k_{**}, p_{**}) is taken. This final result will be denoted by (W^f, k^f, p^f). Similarly $X_R^f = X_T$ if $p_* \le p_{**}$ and $X_R^f = X_T \cup X_C$ if $p_* > p_{**}$.

3. The final classification

The final decisions for the objects from the set X_C are gathered in the set $V^f = V_*$ if $p_* \le p_{**}$ and $V^f = W_{h_0+1} \setminus W'_*$ if $p_* > p_{**}$, i.e. $v_i^f \in V^f$ corresponds to $x_i \in X_C$, where W_{h_0+1} is the set of the fuzzy decisions derived from W_{h_0} ($W_{h_0} = W_{**}$) during the realization of the second stage of the learning procedure and W'_* consists of the first m_T rows of W_{h_0+1}.

Additional new objects (gathered in a set ΔX_C) can be classified by

(1) the classifier characterized by W^f, k^f and X_R^f,

(2) the new classifier obtained by the realization of the second stage of the procedure, started with the membership-array W_0 that corresponds to the set $W_0 = W^f \cup V_{**}$, where V_{**} is a set of fuzzy decision-vectors received for the set ΔX_C and the fuzzy k^f-NN rule with respect to W^f and X_R^f,

(3) the new classifier obtained by the realization of the second stage of the procedure, started with the result of the first stage, i.e. $W_0 = W_* \cup V_*$, where V_* is a set of the fuzzy decision-vectors found for set $X_C \cup \Delta X_C$ and the fuzzy k_*-NN rule with respect to W_* and $X_R = X_T$.

4. Concluding remarks

The efectiveness of the proposed learning procedure has been checked experimentally on some real small-size data. The results were better than for the k_0k'-NN rule with edited data proposed by Koplowitz and Brown (1980), where k_0 is the same as in the sequence (1) and k' is also chosen by the application of the 'leaving one out' method.

In the case when the nonfuzzy membership-sequence u_T is missing but the fuzzy membership-array W_0 is given (see Duin (1982)), error rates are found according to the following relation:

$$p = \sum_{i=1}^{m_T} \left(\sum_{j=1}^{l} \tfrac{1}{2} |v_{i,j} - w_{i,j}| \right) \Big/ m_T \tag{3}$$

where $v_{i,j}$ and $w_{i,j}$ are elements of the fuzzy decision-array and the fuzzy membership-array respectively. The relation (3) allows to generalize the case considered in Sections 2 and 3.

References

Lachenbruch, P.A. (1965), Estimation of error rates in discriminant analysis, Ph.D. dissertation, Univ. of California, Los Angeles (Chapter 5).

Koplowitz, J. and T.A. Brown (1981), On the relation of performance to editing in nearest neighbor rules, *Pattern Recognition* 13, 251-255

Duin, R.P.W. (1982), The use of continuous variables for labeling objects, *Pattern Recognition Letters* 1, 15-20.