Analytica Chimica Acta, 136 (1982) 15—27
Elsevier Scientific Publishing Company, Amsterdam — Printed in The Netherlands

ALTERNATIVE *k*-NEAREST NEIGHBOUR RULES IN SUPERVISED PATTERN RECOGNITION

Part 1. *k*-Nearest Neighbour Classification by using Alternative Voting Rules

D. COOMANS and D. L. MASSART*

Vrije Universiteit Brussel, Farmaceutisch Instituut, Laarbeeklaan 103 — B-1090 Brussels (Belgium)

(Received 28th September 1981)

SUMMARY

This paper discusses an extension of the well known *k*-nearest neighbour method. The majority voting procedure is replaced by an alternative voting method. This alternative *k*NN method is approached from both probabilistic and non-probabilistic points of view. On the basis of an example of differentiation between EU thyroid function and HYPER thyroid function, it is shown that alternative votes can give rise to better classification results than the majority vote.

The *k*-nearest neighbour rule (*k*NN) [1—5] is one of the best known methods for supervised pattern recognition in analytical chemistry and, more generally, the method has been proposed by Cover [6] as a reference method for the evaluation of the performance of more sophisticated techniques. The theoretical properties of this method have been studied by Fix and Hodges [7], Cover and Hart [6, 8] and Peterson [9].

In this paper one of the shortcomings of the *k*-nearest neighbour method is discussed. The method does not perform well if the number of objects of the learning classes is very different. A modified *k*-nearest neighbour method is proposed which permits better results in situations of this kind.

METHODS

kNN method

The simplest *k*NN method (with $k = 1$) classifies a test object into the learning class to which the object belongs and which is closest to the test object. When $k > 1$, the k closest members of the learning set are selected and the test object is classified in the class to which the majority of the k objects belong. This is called the majority vote procedure [7]. Figure 1 illustrates the majority vote of the 3-NN method. A test object is classified on the basis of 3 variables (x, y, z) in class ω_1 because 2 of the 3 closest learning objects belong to ω_1. The appropriate k value is usually selected by means of the leave-one-out classification of the learning set. Kowalski and Bender [1]

0003-2670/82/0000-0000/$02.75

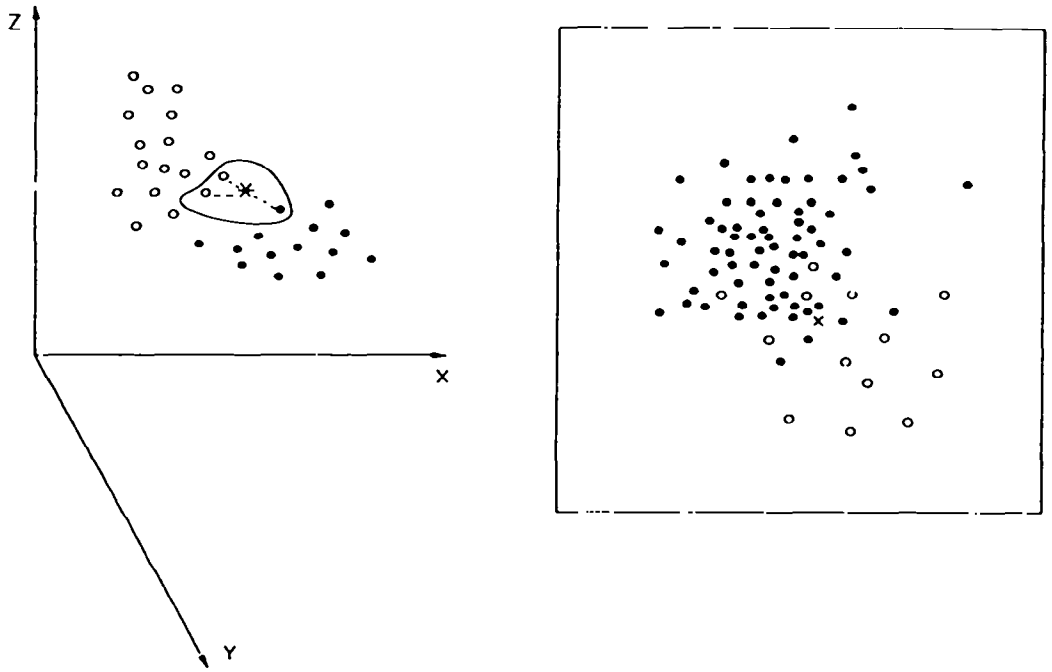

Fig. 1. Illustration of the 3-NN rule in a three-dimensional pattern space: (o) training object belonging to class ω_1; (•) training object belonging to class ω_2; (*) test object.

Fig. 2. The majority vote decision is not the best choice if the sizes of the training classes ω_1 (o) and ω_2 (•) differ considerably.

proposed the use of $k = 1$ or 3 if no important overlap existed between the learning classes.

It can be seen empirically that the kNN method does not perform well in some situations. For instance, difficulties occur if two overlapping classes differ considerably in size. Figure 2 shows that the test patient (X) is classified in the largest class because in the overlapping zone more neighbours of the larger class are found systematically. In Fig. 2, the test object is in fact relatively close to the centre of class ω_1, although most of the closest neighbours belong to ω_2.

It is possible to modify the kNN method in such a way that more relevant decisions are obtained in the situation described previously. The majority vote approach must then be replaced. The proposed replacement may be called the "alternative voting" concept and the method will be called the "alternative kNN method". For example, an adequate alternative vote for the example in Fig. 2 might be as follows: classify the test object in class ω_2 if for $k = 10$ at least 9 neighbours belong to ω_2, otherwise classify the test object in ω_1.

Thus, in the alternative kNN method, it is important not only to choose an appropriate k value but also to select a suitable "alternative vote" criterion.

The kNN method with alternative voting rules

The case of two learning classes, ω_1 and ω_2, is considered. For each k value, k alternative decision rules are possible. These decision rules are symbolized by B_{kl}. The k nearest neighbours of the object to be classified are considered, and at least l of these must belong to class ω_1 for the object to be classified in that class. The minimum number of nearest neighbours belonging to class ω_1 needed to classify an object in that class will be called hereafter n_{1kl}. For $k = 5$ this means that the following five alternative decision rules are possible.

Classify $\vec{x}_i$ in ω_1 if

$$(\text{rule } B_{5,1}: n_{1kl} = 1)\ n_{15i} \geq 1 \tag{1}$$

$$(\text{rule } B_{5,2}: n_{1kl} = 2)\ n_{15i} \geq 2 \tag{2}$$

$$(\text{rule } B_{5,3}: n_{1kl} = 3)\ n_{15i} \geq 3 \tag{3}$$

$$(\text{rule } B_{5,4}: n_{1kl} = 4)\ n_{15i} \geq 4 \tag{4}$$

$$(\text{rule } B_{5,5}: n_{1kl} = 5)\ n_{15i} \geq 5 \tag{5}$$

otherwise classify $\vec{x}_i$ in ω_2. In these rules, n_{1ki} is the number of objects from class ω_1 in the set of k nearest neighbours of test object i. The "majority vote" decision corresponds to $B_{5,3}$. Generally, the decision rule B_{kl} can be formulated as follows: classify test object $\vec{x}_i$ in ω_1 if $n_{1ki} \geq n_{1kl}$, otherwise classify $\vec{x}_i$ in ω_2.

In the same way as is done for k in the classical kNN method, an appropriate B_{kl} can be selected on the basis of the classification rate obtained for the different alternative voting rules as determined by leave-one-out classification on the learning set. The appropriate B_{kl} corresponds to the largest percentage of correct classification.

Probabilistic classification on the basis of the kNN method (majority vote rule)

In analytical chemistry, the kNN method is considered as a non-probabilistic method, i.e., test objects are classified in one of the learning classes without a degree of certainty or a probability being assigned to the decision. However, Cover and Hart [8] showed that the kNN method is based on probabilistic considerations. Duda and Hart [10] considered the kNN method as a technique which makes it possible to estimate (rather roughly) the probability densities in the position of a test object $\vec{x}_i$ and consequently to estimate the *a posteriori* probabilities of class membership if they are entered together with the *a priori* probabilities in the Bayes equation [10]. The *a posteriori* probability of a test object $\vec{x}_i$ for class ω_1 in two-class problems (ω_1 and ω_2) is given by

$$P(\omega_1/\vec{x}_i) = P(\omega_1)P(\vec{x}_i/\omega_1)/[P(\omega_1)P(\vec{x}_i/\omega_1) + P(\omega_2)P(\vec{x}_i/\omega_2)] \tag{6}$$

where $P(\omega_1)$ and $P(\omega_2)$ are the *a priori* probabilities for classes ω_1 and ω_2, respectively, and $P(\vec{x}_i/\omega_1)$ and $P(\vec{x}_i/\omega_2)$ are the probability densities for classes ω_1 and ω_2 in position $\vec{x}_i$.

The probabilistic properties of the kNN method and the corresponding selection of the optimal k may be illustrated by the following simple example. If a test object i in Fig. 3 is classified by using the simple neighbour (k = 1) or 1-NN method, one assumes that the nearest neighbour (NB) is representative for i. This means that around i a region with radius $d(\vec{x}_i, \vec{x}_{NB})$ is used for the estimation of the probability densities in $\vec{x}_i$. For the 1-NN method, the probability density depends on one reference object. In this way, there are only two possibilities: either $P(\vec{x}_i/\omega_1) = 1/n_1$ (n_1 is the number of objects in ω_1) together with $P(\vec{x}_i/\omega_2) = 0$ if the nearest neighbour belongs to ω_1, or $P(\vec{x}_i/\omega_1) = 0$ together with $P(\vec{x}_i/\omega_2) = 1/n_2$ (n_2 is the number of objects in ω_2) if the nearest neighbour belongs to ω_2. From Eqn. (6), it follows that $P(\omega_1/\vec{x}_i)$ is 1 or 0 which means that x_i is classified in ω_1 or excluded from ω_1 with absolute certainty. Consequently, a too "black or white" picture is given of the real certainty with which the classification may be made.

It might seem that a larger k value would give rise to the selection of more probability levels and consequently to a more differentiated and therefore better estimation of the probability densities.

This is not the case because a larger number of nearest neighbours gives rise to an enlargement of the decision region around $\vec{x}_i$ which is defined by the distance between $\vec{x}_i$ and the further neighbour (i.e., the kth). In Fig. 3, k = 41 for instance gives rise to a region which cannot be considered as representative of the position $\vec{x}_i$, and the relevance of the probability densities determined on the basis of k = 41 can be considered to be decreased. Thus the determination of the appropriate k value is obviously an optimization process.

Probabilistic classification using an alternative voting rule

In order to assign an *a posteriori* probability to a test object when a particular alternative voting rule B_{kl} is used, a decision matrix is needed. It is a probabilistic decision matrix which collects for each k and l considered

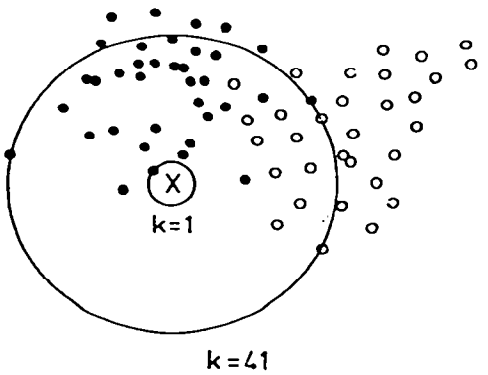

Fig. 3. Probability density estimation by means of the kNN method: (○) training object belonging to class ω_1; (●) training object belonging to class ω_2; (X) test object i. The circles are the decision regions around $\vec{x}_i$ for k = 1 and k = 41.

in the matrix, the *a posteriori* probabilities according to a given class (say ω_1); essentially, each n_{1kl} corresponds to the *a posteriori* probability $P(\omega_1/n_{1kl})$. The *a posteriori* probabilities are determined in such a way that they take explicitly the difference between the number of objects between two learning classes into account, and are given by

$$P(\omega_1/n_{1kl}) = n_{1kl}/n_1/[n_{1kl}/n_1 + (k - n_{1kl})/n_2] \tag{7}$$

The estimated *a posteriori* probability $P(\omega_1/\vec{x}_i)$ of a test object i with measurement vector $\vec{x}_i$ for class ω_1, and for which k nearest neighbours were selected comprising n_{1ki} objects belonging to ω_1, is given by

$$P(\omega_1/\vec{x}_i) = P(\omega_1/n_{1ki}) \tag{8}$$

From the relation $n_{1ki} \geqslant n_{1kl}$ (see above), n_{1kl} is the minimum number of nearest neighbours belonging to ω_1 necessary for classification in ω_1 with decision rule B_{kl} of a test object i. Thus $P(\omega_1/n_{1kl})$ can also be seen as the minimum *a posteriori* probability for the classification of a test object with B_{kl} into class ω_1. Table 1 shows such a probability matrix for a class ω_1 ($n_1 = 150$) and ω_2 ($n_2 = 35$). It can be seen that in this situation the "majority vote" decision rules (indicated by dotted lines) do not correspond to the minimum probability-of-error decision boundary: $P(\omega_1/\vec{x}_i) = P(\omega_2/\vec{x}_i) = 0.5$. For instance, for $B_{5,3}$, $P(\omega_1/n_{1,5,3})$ is 0.26 which is far from 0.5. The boundaries for the different decision rules B_{kl} can be approximated by

$$\bar{P}(\omega_1/n_{1kl}) = \tfrac{1}{2}[P(\omega_1/n_{1kl}) + P(\omega_1/n_{1kl-1})] \tag{9}$$

and again arranged in a decision matrix.

Indeed, for decision rule B_{kl}, n_{1kl} is the minimum number of the neighbours from class ω_1 to classify a test object in ω_1, or $P(\omega_1/n_{1kl})$ is the minimum

TABLE 1

$P(\omega_1/n_{1kl})$ for different k and l values if $n_1 = 150$ and $n_2 = 35$

k \ l	(0)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	0.00	1.00															
2	0.00	0.19	1.00														
3	0.00	0.10	0.32	1.00													
4	0.00	0.07	0.19	0.41	1.00												
5	0.00	0.06	0.13	0.26	0.48	1.00											
6	0.00	0.04	0.10	0.19	0.32	0.54	1.00										
7	0.00	0.04	0.09	0.15	0.24	0.37	0.58	1.00									
8	0.00	0.03	0.07	0.12	0.19	0.24	0.41	0.62	1.00								
9	0.00	0.03	0.06	0.10	0.16	0.23	0.32	0.45	0.65	1.00							
10	0.00	0.03	0.06	0.09	0.13	0.19	0.26	0.35	0.48	0.68	1.00						
11	0.00	0.02	0.05	0.08	0.12	0.15	0.22	0.29	0.38	0.51	0.70	1.00					
12	0.00	0.02	0.04	0.07	0.10	0.14	0.19	0.25	0.32	0.41	0.54	0.72	1.00				
13	0.00	0.02	0.04	0.07	0.09	0.13	0.17	0.21	0.27	0.34	0.44	0.56	0.74	1.00			
14	0.00	0.02	0.04	0.06	0.09	0.11	0.15	0.19	0.24	0.30	0.37	0.46	0.58	0.75	1.00		
15	0.00	0.02	0.03	0.06	0.08	0.10	0.13	0.17	0.19	0.26	0.32	0.39	0.48	0.61	0.77	1.00	
16	0.00	0.02	0.03	0.05	0.07	0.10	0.12	0.15	0.19	0.23	0.28	0.34	0.41	0.50	0.62	0.78	1.00

a posteriori probability for ω_1 which can be obtained for a test object classified in ω_1. In contrast, n_{1kl-1} is the maximum number of neighbours from class ω_1 needed to classify a test object in ω_2, or $P(\omega_1/n_{1kl-1})$ is the maximum *a posteriori* probability for ω_1 which can be obtained for a test object classified in ω_2. The decision boundary is situated between $P(\omega_1/n_{1kl})$ and $P(\omega_1/n_{1kl-1})$ and is thus approximated by Eqn. (9).

The probabilistic decision rule B_{kl} can be formulated as follows:
classify test object i with $\bar{x}_i$ in ω_1 if $P(\omega_1/\bar{x}_i) > \bar{P}(\omega_1/n_{1kl})$, otherwise classify $\bar{x}_i$ in ω_2 and $P(\omega_1/\bar{x}_i)$ can be found in Table 1 taking $n_{1kl} = n_{1ki}$. Table 2 shows the matrix of the $\bar{P}(\omega_1/n_{1kl})$ values for the different alternative decision rules B_{kl} for $n_1 = 150$ and $n_2 = 35$.

It is assumed that the appropriate alternative decision rule for each k corresponds to a $\bar{P}(\omega_1/n_{1kl})$ closest to 0.5. The optimal decision rules are indicated by full lines in Table 2, and the "majority vote" decisions by dotted lines. The discrepancy between the majority votes and the optimal rules is more meaningful for large k values. The optimal rules are situated in this example at the right-hand side of the majority votes because more objects are required from ω_1 in the set of k nearest neighbours to compensate for the discrepancy between n_1 and n_2 ($n_1 \gg n_2$).

Only two-class problems have been considered so far. Although an extension of the *k*NN matrix concept is possible, it becomes too complex. It is easier to split multi-class problems into several binary problems and to make a decision on the basis of the results of the binary classifications.

EXAMPLE

As an example of the application and performance of the alternative voting rule approach, the EU/HYPER differentiation in the THYROID

TABLE 2

$\bar{P}(\omega_1/n_{1kl})$ for the different alternative decision rules if $n_1 = 150$ and $n_2 = 35$

l =	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
k = 1	0.50															
2	0.10	0.60														
3	0.05	0.21	0.66													
4	0.04	0.13	0.30	0.71												
5	0.03	0.10	0.20	0.37	0.74											
6	0.02	0.07	0.15	0.26	0.43	0.77										
7	0.02	0.07	0.12	0.20	0.31	0.48	0.79									
8	0.02	0.05	0.10	0.16	0.22	0.33	0.52	0.81								
9	0.02	0.05	0.08	0.13	0.20	0.28	0.39	0.55	0.83							
10	0.02	0.05	0.08	0.11	0.16	0.23	0.31	0.42	0.58	0.84						
11	0.01	0.04	0.07	0.10	0.14	0.19	0.26	0.34	0.45	0.61	0.85					
12	0.01	0.03	0.06	0.09	0.12	0.17	0.22	0.29	0.37	0.48	0.63	0.86				
13	0.01	0.03	0.06	0.08	0.11	0.15	0.19	0.24	0.31	0.39	0.50	0.65	0.87			
14	0.01	0.03	0.05	0.08	0.10	0.13	0.17	0.22	0.27	0.34	0.42	0.52	0.67	0.88		
15	0.01	0.03	0.05	0.07	0.09	0.12	0.15	0.18	0.23	0.29	0.36	0.44	0.54	0.69	0.89	
16	0.01	0.03	0.04	0.06	0.09	0.11	0.14	0.17	0.21	0.26	0.31	0.38	0.46	0.56	0.70	0.89

function on the basis of five laboratory tests is used. This data set has been used for pattern recognition in several papers [11–13]. The EU/HYPER differentiation is a typical example of two learning classes that differ in size. Such a situation is very common in clinical chemistry. Another common difficulty is that the dispersion of the training samples differs greatly in the different classes. The normal class is usually denser than pathological classes. The "alternative voting" concept may be useful in this kind of situation. For the normal class (EU), large numbers of objects (patients, samples) are often available. The pathological cases (HYPER, in this instance) are not so numerous. The EU class consists of 150 cases and the HYPER class of 35 cases. For each patient, five laboratory results are available. In Table 3 the matrix of the classification results is given for k = 1 to 20. The results are expressed as the Youden Index (YI %): YI % = (% correct classified HYPER + % correct classified EU) — 100.

In the matrix of Table 3, the results are arranged according to the votes indicated in Tables 1 and 2 with ω_1 = EU and ω_2 = HYPER. Table 3 shows that with the classical majority vote kNN method, the Youden Index obtained is 93.6% for k = 1 and 3. An increase of the k values corresponds to a significant decrease of the classification rate obtained by the majority vote rule. The empirical rule proposed by Kowalski and Bender [1] (choose k = 1 or 3) is clearly valid here for the classical kNN method. However, if alternative voting rules are considered, it can be seen that larger k values give rise to larger YI % values.

The largest YI % for each k is underlined with a full line. It can be seen that the largest correct classification rates are situated to the right of the majority votes (dotted underlines). This agrees with expectations based on the difference in size of the two learning classes. The theoretically expected optimal decision rules are indicated in Table 2. There is relatively good agreement between the assumed optimal decision such as indicated in Table 2 and the maximal classification rates in Table 3. A maximal YI of 99.3% was found for the decision rules $B_{13,10}$, $B_{14,11}$, $B_{15,12}$ and $B_{16,13}$.

Screening of EU cases or HYPER cases can be done mainly by choosing a decision rule B_{kl} suitable for the particular purpose. Obviously, for screening of EU cases, the classification rule must miss as few EU cases as possible. In the decision matrix used, the classification rates are expressed in percent EU correctly classified (EU %). Table 4 shows the matrix for the example studied; it can be seen that for a number of decision rules all the EU cases are correctly classified. In such cases, it is preferable to choose the decision rule which gives 100% correct classification of EU and also the optimal YI % (Table 3). In Table 4 the best choice for each k is indicated by a dotted line and the individual best choice is indicated by a full line, i.e., decision rules $B_{17,11}$ and $B_{18,12}$. The decision rule for the screening of HYPER cases can be selected on the basis of a matrix in which the classification rates relate to the correct classification of HYPER cases, i.e., the percentage of HYPER cases correctly classified (HYPER %) (Table 5). Here again, many classification

TABLE 3

Classification results for the EU/HYPER differentiation according to different alternative votes B_{ld}
(The results are expressed in YI %)

l =	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
k = 1	93.62																			
2	82.19	93.62																		
3	70.76	93.62	97.33																	
4	60.00	85.05	92.29	97.33																
5	54.29	74.29	87.90	95.14	97.33															
6	51.43	68.57	80.00	87.90	95.14	96.67														
7	45.71	60.00	74.29	85.71	87.90	94.48	96.00													
8	45.71	51.43	65.71	74.29	85.71	93.62	94.48	96.00												
9	45.71	51.43	54.29	68.57	82.86	87.90	92.95	97.33	95.33											
10	45.71	51.43	51.43	65.71	77.14	85.05	87.90	98.67	97.33	95.33										
11	45.71	45.71	51.43	57.14	68.57	77.14	87.90	90.76	98.67	97.33	95.33									
12	42.86	45.71	48.57	57.14	60.00	74.29	80.00	87.90	93.62	98.67	97.33	94.00								
13	42.86	42.86	48.57	54.29	57.14	62.86	74.29	82.86	87.90	99.33	98.00	96.67	94.00							
14	42.86	42.86	45.71	51.43	54.29	60.00	65.71	77.14	85.71	93.62	99.33	98.00	96.67	94.00						
15	40.00	42.86	45.71	48.57	54.29	57.14	65.71	71.43	82.86	87.90	93.62	99.33	98.00	96.67	92.67					
16	40.00	42.86	45.71	48.57	48.57	54.29	60.00	68.57	77.14	85.71	90.76	93.62	99.33	98.00	96.67	92.00				
17	37.14	42.86	42.86	48.57	48.57	48.57	54.29	60.00	74.29	80.00	91.43	93.62	93.62	98.67	98.00	96.67	92.00			
18	37.14	40.00	42.86	48.57	48.57	48.57	51.43	57.14	62.86	74.29	80.00	91.43	93.62	93.62	98.67	98.00	96.67	92.00		
19	37.14	40.00	40.00	42.86	48.57	48.57	51.43	51.43	57.14	68.57	74.29	82.86	90.76	93.62	96.48	98.67	97.33	96.67	91.33	
20	37.14	40.00	40.00	42.86	45.71	48.57	51.43	51.43	51.43	60.00	71.43	74.29	85.71	90.76	93.62	96.48	98.00	96.67	96.67	90.67

TABLE 4

Classification results for the EU/HYPER differentiation according to different alternative votes B_{hl}
(The results are expressed in EU %)

l =	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
h = 1	99.33																			
2	99.33	99.33																		
3	99.33	99.33	97.33																	
4	100.00	99.33	98.00	97.33																
5	100.00	100.00	99.33	98.00	97.33															
6	100.00	100.00	100.00	99.33	98.00	96.67														
7	100.00	100.00	100.00	100.00	99.33	97.33	96.00													
8	100.00	100.00	100.00	100.00	100.00	99.33	97.33	96.00												
9	100.00	100.00	100.00	100.00	100.00	99.33	98.67	97.33	95.33											
10	100.00	100.00	100.00	100.00	100.00	99.33	99.33	98.67	97.33	95.33										
11	100.00	100.00	100.00	100.00	100.00	100.00	99.33	99.33	98.67	97.33	95.33									
12	100.00	100.00	100.00	100.00	100.00	100.00	100.00	99.33	99.33	98.67	97.33	94.00								
13	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	99.33	99.33	98.00	96.67	94.00							
14	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	99.33	99.33	98.00	96.67	94.00						
15	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	99.33	99.33	99.33	98.00	96.67	92.67					
16	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	99.33	99.33	99.33	98.00	96.67	92.00				
17	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	99.33	99.33	98.67	98.00	96.67	92.00			
18	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	99.33	99.33	98.67	98.00	96.67	92.00		
19	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	99.33	99.33	99.33	98.67	97.33	96.67	91.33	
20	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	99.33	99.33	99.33	98.00	96.67	96.67	90.67	

TABLE 5

Classification results for the EU/HYPER differentiation according to different alternative votes B_{hl}
(The results are expressed in HYPER %)

l =	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
h = 1	94.29																			
2	82.86	94.29																		
3	71.43	94.29	100.00																	
4	60.00	85.71	94.29	100.00																
5	54.29	74.29	88.57	97.14	100.00															
6	51.43	68.57	80.00	88.57	97.14	100.00														
7	45.71	60.00	74.29	85.71	88.57	97.14	100.00													
8	45.71	51.43	65.71	74.29	85.71	94.29	97.14	100.00												
9	45.71	51.43	54.29	68.57	82.86	88.57	94.29	100.00	100.00											
10	45.71	51.43	51.43	65.71	77.14	85.71	88.57	100.00	100.00	100.00										
11	45.71	45.71	51.43	57.14	68.57	77.14	88.57	91.43	100.00	100.00	100.00									
12	42.86	45.71	48.57	57.14	60.00	74.29	80.00	88.57	94.29	100.00	100.00	100.00								
13	42.86	42.86	48.57	54.29	57.14	62.86	74.29	82.86	88.57	100.00	100.00	100.00	100.00							
14	42.86	42.86	45.71	51.43	54.29	60.00	65.71	77.14	85.71	94.29	100.00	100.00	100.00	100.00						
15	40.00	42.86	45.71	48.57	54.29	57.14	65.71	71.43	82.86	88.57	94.29	100.00	100.00	100.00	100.00					
16	40.00	42.86	45.71	48.57	48.57	54.29	60.00	68.57	77.14	85.71	91.43	94.29	100.00	100.00	100.00	100.00				
17	37.14	42.86	42.86	48.57	48.57	48.57	54.29	60.00	74.29	80.00	91.43	94.29	94.29	100.00	100.00	100.00	100.00			
18	37.14	40.00	42.86	48.57	48.57	48.57	51.43	57.14	62.86	74.29	80.00	91.43	94.29	94.29	100.00	100.00	100.00	100.00		
19	37.14	40.00	40.00	42.86	48.57	48.57	51.43	51.43	57.14	68.57	74.29	82.86	91.43	94.29	97.14	100.00	100.00	100.00	100.00	
20	37.14	40.00	40.00	42.86	45.71	48.57	51.43	51.43	51.43	60.00	71.43	74.29	85.71	91.43	94.29	97.14	100.00	100.00	100.00	100.00

rules give 100% correct classification. The decision rule selected is that for which the YI % is largest. In Table 5, the preferred rules are underlined with full lines, i.e., $B_{13,10}$, $B_{14,11}$, $B_{15,12}$ and $B_{16,13}$. The dotted lines indicate the best choice for each other k.

In Fig. 4, the *a posteriori* classification probabilities obtained for the patients (EU/HYPER) with the probabilistic approach of the alternative kNN method are compared with those from the potential method of ALLOC [14]. The horizontal axis of the diagram indicates the *a posteriori* probabilities for the alternative kNN method and the vertical axis the *a posteriori* probabilities for the ALLOC method.

In the alternative kNN method, k was taken equal to 16 because a maximum YI % was found for this value (see Table 3). Consequently, the range of the *a posteriori* probability for ω_1 = EU (from 0 to 1) was divided into 17 regions according to the boundaries $\bar{P}(\omega_1/n_{1kl})$ for the 17 possible l values (l = 0—16). The $\bar{P}(\omega_1/n_{1kl})$ values are derived from Table 2. Furthermore, for each test object i (according to the leave-one-out procedure, each case from the EU/HYPER differentiation is considered once as a test object) the *a posteriori* probability $P(\omega_1/\vec{x}_i)$ was estimated according to the number of training cases from the EU class (n_{1kl}) which belongs to the set of 16 nearest neighbours. The estimated *a posteriori* probability is given by $P(\omega_1/n_{1kl})$ and determined on the basis of Eqn. (7). According to this value, the test object is placed in one of the 17 regions on the horizontal axis. The same test object can be positioned on the vertical axis on the basis of its $P(\omega_1/\vec{x}_i)$ value obtained with ALLOC. However, ALLOC produces continuous probabilities from 0 to 1. To permit an easy comparison of the two methods,

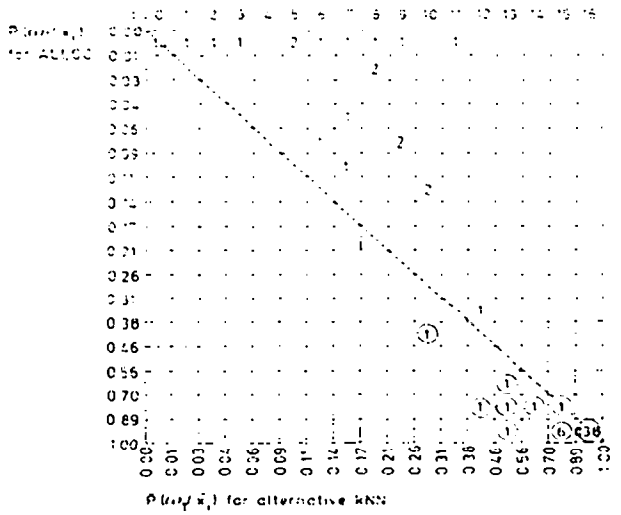

Fig. 4. Comparison between the *a posteriori* probabilities obtained with the alternative kNN method with k = 16 and those obtained with the potential method ALLOC for the EU/HYPER differentiation.

the range of the probabilities of ALLOC was divided into 17 regions identical to those of the alternative kNN method. In this way the objects could be positioned in one of the 172 cells of the diagram. The numbers in the cells of Fig. 4 indicate the number of HYPER cases positioned in the particular cell and the numbers represented by a circle indicate the number of EU cases.

The diagonal dotted line represents the complete correspondence between the alternative kNN and ALLOC methods. It can be seen that the agreement is good but not excellent. Good agreement was found for the EU class. However, in the HYPO class, there is considerable disagreement in some cases; e.g., in one case $P(\omega_1/\vec{x}_i)$ lies between 0.00 and 0.01 for the ALLOC method and between 0.31 and 0.38 for the alternative kNN method.

This is not surprising because of the rough way in which the probabilities are estimated by the kNN method. The kNN probabilities are useful for "quick look" purposes or in cases where the probabilities need not be known very accurately. An extended and more formal evaluation of the quality of the probabilities will be given in a subsequent paper, and compared with other techniques for probability density estimation.

Computer programs in Fortran IV have been developed in this laboratory for a CDC CYBER 170/750 computer. The probabilistic and non-probabilistic approaches are available.

Conclusion

For the first-level pattern recognition [15], the use of the kNN matrix based on alternative votes is more flexible than the classical kNN method. Moreover, it is possible to calculate *a posteriori* probabilities which enable some degree of certainty about the decision to be taken into account. However, it is essential to be aware of the very approximate character of the probabilistic kNN procedure.

The authors thank I. Broeckaert for making the data available and F.G.W.O. for financial support.

REFERENCES

1 B. R. Kowalski and C. F. Bender, Anal. Chem., 44 (1972) 1405.
2 B. E. H. Saxberg, D. L. Duewer, J. L. Booker and B. R. Kowalski, Anal. Chim. Acta, 103 (1978) 201.
3 D. L. Duewer, B. R. Kowalski, K. J. Clayson and R. J. Roby, Comp. Biomed. Res., 11 (1978) 567.
4 D. L. Massart, A. Dijkstra and L. Kaufman, Evaluation and Optimization of Laboratory Methods and Analytical Procedures, Elsevier, Amsterdam, 1978.
5 M. Sjöström and B. R. Kowalski, Anal. Chim. Acta, 112 (1979) 11.
6 T. M. Cover, IEEE Trans. Inf. Theory, IT-14 (1968) 50.
7 E. Fix and J. L. Hodges, Discriminatory analysis. USAF School of Aviation Medicine, Randolph Field, TX, Project 21-49-004, Contract AF41 (148)-31, Rep. 4, February 1951, and Rep. 11, August 1952.

8 T. M. Cover and P. E. Hart, IEEE Trans. Inf. Theory, IT-13 (1967) 21.
9 D. W. Peterson, IEEE Trans. Inf. Theory, IT-16 (1970) 26.
10 R. O. Duda and P. E. Hart, Pattern Classification and Scene Analysis, Wiley, New York, 1973.
11 D. Coomans, I. Broeckaert, M. Jonckheer, P. Blockx and D. L. Massart, Anal. Chim. Acta, 103 (1978) 409.
12 D. Coomans, I. Broeckaert, A. Tassin and D. L. Massart, Anal. Chim. Acta, 133 (1981) 215.
13 D. Coomans, I. Broeckaert and D. L. Massart, Anal. Chim. Acta, 134 (1982) 139, and references therein.
14 J. Hermans and J. D. F. Habbema, Manual for the ALLOC-discriminant analysis program, Department of Medical Statistics, University of Leiden, P.O. Box 2060 Leiden, The Netherlands, 1976.
15 C. Albano, W. Dunn, U. Edlund, E. Johansson, B. Norden, M. Sjöström and S. Wold, Anal. Chim. Acta, 103 (1978) 429.