ELSEVIER

Pattern Recognition Letters 18 (1997) 1179–1186

Pattern Recognition Letters

On the use of neighbourhood-based non-parametric classifiers [1]

J.S. Sánchez [a,*], F. Pla [a], F.J. Ferri [b]

[a] *Dept. d'Informàtica, Universitat Jaume I, E-12071 Castelló, Spain*
[b] *Dept. d'Informàtica i Electrónica, Universitat de València, E-46100 Burjassot (València), Spain*

Abstract

Alternative non-parametric classification schemes, which come from the use of different definitions of neighbourhood, are introduced. In particular, the *Nearest Centroid Neighbourhood* along with the neighbourhood relation derived from the *Gabriel Graph* and the *Relative Neighbourhood Graph* are used to define the corresponding (*k*-)Nearest Neighbour-like classifiers. Experimental results are reported to compare the performance of the approaches proposed here to the one obtained with the *k*-Nearest Neighbours rule. © 1997 Elsevier Science B.V.

Keywords: Nearest neighbourhood; Gabriel graph; Relative neighbourhood graph; Nearest centroid neighbourhood

1. Introduction

Classification is a problem consisting of implicitly or explicitly generating decision boundaries which can successfully distinguish the distinct classes usually in a *d*-dimensional feature space. So, classification in pattern recognition has traditionally been tackled through two alternative approaches; namely, *parametric* and *non-parametric* (Duda and Hart, 1973). While the parametric classifiers assume a (known) functional distribution of given samples, the non-parametric approaches do not assume any particular distribution of the set of prototypes.

Although the parametric approaches have theoretically been shown to be potentially capable of yielding optimal results, in practice, they often tend to actually fail because of inappropriate assumptions of a priori distributions.

Among non-parametric methods, those which are based on sample-to-sample distances are particularly remarkable and, in particular, the family of *k-Nearest Neighbours* (*k*-NN) techniques (Duda and Hart, 1973) is often used in practice. When applied to classification, these schemes require the classes to be represented by appropriate sets of prototypes and the decision rule is generally reduced to label each given sample with the class that contains most of its *k*-nearest neighbours. It is the conceptual simplicity of such a rule, along with its asymptotical trend towards the *Bayes* rule in terms of minimum classification error, that makes the *k*-NN approach particularly appealing in many practical situations. Nevertheless, when the number of prototypes in the training set is not large enough, the *k*-NN rule is no longer optimal. This problem becomes more relevant when having few prototypes compared to the intrinsic dimensionality of the feature space, which is a very usual practical situation.

* Corresponding author. E-mail: sanchez@inf.uji.es.

[1] This work has partially been supported by projects P1B96-13 (Fundació Caixa-Castelló) and AGF95-0712-C03-01 and TIC95-676-C02-01 (Spanish CICYT).

PII S0167-8655(97)00112-8

Some alternative neighbourhood definitions are proposed to obtain new non-parametric classifiers, trying to partially overcome the practical drawbacks pointed out for the *k*-NN rule. In particular, the recently introduced concept of *Nearest Centroid Neighbourhood* (Chaudhuri, 1996) along with the neighbourhood relation derived from the *Gabriel* and the *Relative Neighbourhood* graphs (Jaromczyk and Toussaint, 1992) are used.

This paper is organised as follows. Section 2 introduces the already mentioned neighbourhood definitions. The resulting classification rules as well as conceptual differences with respect to the *k*-NN rule are provided in Section 3. Furthermore, a comparative study of the application of different non-parametric classification approaches in some critical situations (that is, few prototypes and/or high dimensionality of the feature space) is presented in Section 4. Finally, Section 5 summarises the main results of the paper and gives some concluding remarks.

2. Alternative neighbourhood definitions

Intuitively, the concept of neighbourhood should be such that the neighbours are as *close* to a sample as possible but also, the neighbours should lie as *homogeneously* around that sample as possible. The second condition is a consequence of the first in the asymptotic case but in some practical situations, the geometrical placement can become much more important than the actual distances to appropriately characterise a sample by its neighbourhood. As the *Nearest Neighbourhood* (NN) takes into account the first property only, the nearest neighbours may not be placed symmetrically around the sample if the neighbourhood in the training set is not spatially homogeneous. In fact, it has been shown that the use of local distance measures can significantly improve the behaviour of the classifier in the case of a finite sample size (Short and Fukunaga, 1980).

Hence, in order to tackle the problem just described, we propose to make use of some alternative neighbourhood definitions, obtaining what hereafter is called the *Surrounding Neighbourhood* (SN) of a sample, that is, in some effective way the neighbours of a sample will be considered not only in terms of proximity but also in terms of their spatial distribution with respect to that sample.

It must be pointed out that even though there is no theoretical justification to use neighbours around (and close to) a sample instead of nearest neighbours, this can effectively help in special situations (finite sample case), in which prototypes do not fully represent the underlying statistics and/or the distance used (irrelevant in the asymptotic case) exhibits some undesirable properties.

The first SN definition comes from the already mentioned *Nearest Centroid Neighbourhood* (NCN) concept (Chaudhuri, 1996). Let p be a point whose k neighbours should be found in a set of points (training set) $X = \{x_1, \ldots, x_n\}$. These k neighbours are such that (a) they are as near to p as possible, and (b) their centroid is also as close to p as possible. Both conditions can be satisfied through an iterative procedure (Chaudhuri, 1996) in the following way:

- The first neighbour of p is its nearest neighbour, q_1.
- The ith neighbour, q_i, $i \geqslant 2$, is such that the centroid of this and previously selected neighbours, $q_1, \ldots, q_i$ is the closest to p.

This definition gives rise to a neighbourhood in which both closeness and spatial distribution of neighbours are taken into account because of the centroid criterion. The application of the NCN instead of the NN has already been proven to be very useful in some representation problems (Chaudhuri, 1996).

The second SN definition considered here is based on the information derived from two analogous geometrical structures widely used in Pattern Recognition: the *Gabriel Graph* (GG) and the *Relative Neighbourhood Graph* (RNG) (Jaromczyk and Toussaint, 1992), which are two well-known examples of *Proximity Graphs*.

Let $X = \{x_1, \ldots, x_n\}$ be a set of points in $\mathbb{R}^d$. From a general point of view, a Proximity Graph, $G = (V, E)$, is an undirected graph with a set of vertices $V = X$, and a set of edges, E, such that $(x_i, x_j) \in E$ if and only if x_i and x_j satisfy some neighbourhood relation. In this case, it is said that x_i and x_j are *graph neighbours*. Consequently, the *Graph Neighbourhood* of a sample can be defined as the union of all its graph neighbours.

Let $d(\cdot,\cdot)$ be the Euclidean distance in $\mathbb{R}^d$. The GG is defined as follows:

$$(x_i, x_j) \in E \Leftrightarrow d^2(x_i, x_j) \leq d^2(x_i, x_k) + d^2(x_j, x_k) \quad \forall x_k \in X,\ k \neq i, j. \tag{1}$$

Thus, x_i and x_j are said to be *Gabriel neighbours*. In other words, two points are Gabriel neighbours if and only if there is no other point from X lying in the hypersphere centered at their middle point and whose diameter is the distance between them; namely, *hypersphere of influence* of x_i and x_j (Jaromczyk and Toussaint, 1992).

Analogously, the set of edges in the RNG can be defined as follows:

$$(x_i, x_j) \in E \Leftrightarrow d(x_i, x_j) \leq \max d(x_i, x_k), d(x_j, x_k) \quad \forall x_k \in X,\ k \neq i, j \tag{2}$$

Its geometric interpretation is based on the concept of *lune* (Jaromczyk and Toussaint, 1992), defined as the disjoint intersection between two hyperspheres centered at x_i and x_j and whose radii are equal to the distance between them. Two points are *relative neighbours* if and only if their lune does not contain other points from X.

From the definitions just provided, taking into account that two points are graph neighbours if no other point lies inside a given zone of influence between them (that is, a hypersphere in the GG and a lune in the RNG), it is possible to completely surround a sample by means of its graph neighbourhood.

3. Surrounding classification schemes

The non-parametric approaches proposed in this paper are based on the general idea of estimating the class of a sample from its neighbours, but considering a kind of neighbourhood which allows one to inspect a sufficiently small and near area around the sample, in such a way that all prototypes surrounding the sample take part in the classification process. This is accomplished by using the SN definitions introduced in the previous section, that is, the NCN and the Graph Neighbourhood. These concepts are used here to obtain what is called the *k-Nearest Centroid Neighbours* (k-NCN) and *Graph Neighbours* (GN) rules, respectively.

Therefore, it seems clear that the ultimate goal of the application of the Surrounding Neighbourhood in the classification schemes is to acquire a different kind of information than that obtained by means of the NN concept (and more reliable in some cases). Thus, it is proposed to make a decision about the class membership of a given sample after knowing the spatial distribution of the prototypes around that sample. This means that a sample is classified not only based on its nearest neighbours but also taking into account how prototypes are placed around it.

Bearing this in mind, the NCN and Graph Neighbourhood concepts can be used to obtain two alternative non-parametric classifiers. Both of them have in common the fact that they consider a number of prototypes *around* and *relatively close to* (instead of only *close to*) a sample to estimate its class.

Given a set of prototypes $X = \{x_1, \ldots, x_n\}$ with L different classes, and a new sample p, let us introduce the k-NCN classification rule as follows:

- Find the k nearest centroid neighbours of p, $X^p = \{x_1^p, \ldots, x_k^p\}$, where $k \leqslant n$.
- Assign to p the class with a majority of votes from its k nearest centroid neighbours in the set X^p (resolve ties randomly).

Analogously, the GN rule can be expressed in the following way:

- Find all the graph neighbours of p, $X^p = \{x_1^p, \ldots, x_m^p\}$, where $m \leqslant n$.
- Assign to the sample p the class with a majority of votes from its m graph neighbours in the set X^p (resolve ties randomly).

Note that in the case of the GN rule, considering the GG and the RNG, we have two different but analogous approaches: namely, the GGN and the RNGN rules, respectively. On the other hand, it is worth mentioning that the corresponding classifiers with a reject option and other extensions could also be defined exactly in the same way as in the k-NN rule (Devijver and Kittler, 1982).

With respect to the complexity of both *Surrounding classifiers*, while k NCN's from a training set with n prototypes in a d-dimensional feature space can be computed in $O(kn)$ time (Chaudhuri, 1996),

the averaged cost to search for the graph neighbours of a sample is close to O(dn) (Toussaint et al., 1985). From a practical point of view, it is clear therefore, that one should select the most appropriate Surrounding Neighbourhood depending on the particular characteristics of each problem, taking into account issues such as the dimensionality of the feature space.

4. Experimental results

A number of experiments on synthetic and real databases have been carried out in order to illustrate the behaviour of the classifiers proposed here and compare their performance to the one achieved by the k-NN rule. The *Holdout* method averaged over five different random partitions (half for training and half for testing purposes) of each original database has been used to obtain error rate estimates. In the case of the k-NN and k-NCN rules, typical settings for the parameter k (ranging from 1 to 11) have been tried and the ones leading to the best performance have finally been included.

Four different databases have been used: the first three have been taken from a public data repository (Murphy and Aha, 1991) for benchmarking purposes, and the last one from a particular application (Pla et al., 1993).

4.1. The synthetic database

This experiment consists of a set of seven databases corresponding to the same problem but with dimensionality ranging from 2 to 8. There are two classes consisting of multivariate normal distributions with zero mean and standard deviation 1 and 2 in all dimensions, respectively. There are 2500 patterns for each class.

The performance of the classifiers has been evaluated and compared to the theoretical *Bayes* error. In the case of the k-NN and k-NCN rules, we have represented the results corresponding to the best k in each dimensionality. From Fig. 1, it is possible to draw a number of interesting conclusions: firstly, the GG-based approach and the k-NCN rule are clearly the best for most dimensionalities, although differences between them are not statistically significant and therefore, almost nothing can be said about the relative merits of each one of them for this first experiment. Furthermore, the most important conclusion is that the error rates are not too far from the optimal *Bayes* error. Secondly, while the *Bayes* classifier as well as the k-NCN and both GN rules

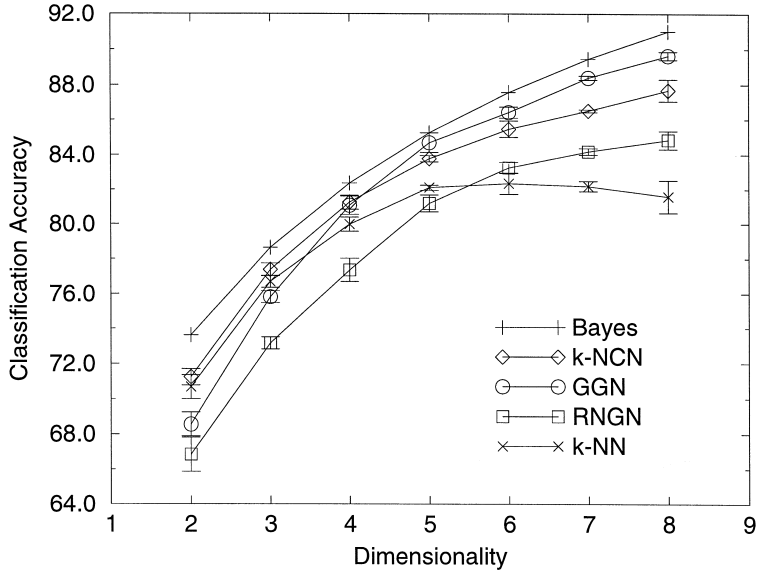

Fig. 1. Classification accuracy with varying dimensionalities.

yield better results as the dimensionality increases, the performance for the *k*-NN rule clearly degrades from dimensionality 5. Finally, the RNGN classification scheme has a behaviour worse than the other approaches introduced here, although it shows an improvement in performance similar to the GGN and the *k*-NCN rules.

4.2. The phoneme database

The second database was already used in the ROARS Project ESPRIT II (Alinat, 1993) about the development and the implementation of a real time analytical system for French and Spanish speech recognition.

There are two different classes: the nasal and the oral vowels. It contains vowels from 1809 different isolated syllables. Five different attributes were chosen to characterise each vowel: the amplitudes of the five first harmonics, normalised by the total energy. There are a total of 5404 patterns available: 3818 for nasal vowels and 1586 for oral ones.

Fig. 2 illustrates the classification accuracy obtained by the Surrounding classifiers as well as the *k*-NN rule on the phoneme database. Even though this is not a well-behaved problem (1-NN and 1-NCN achieve the highest accuracy), it clearly shows the better results of the new classifiers (except the case of the GGN approach) compared to the *k*-NN rule.

4.3. The Landsat database

The aim of this experiment is the classification of the multi-spectral values of a real image from the Landsat satellite with 2340×3380 pixels. This database is a sub-area of a scene, consisting of 82×100 pixels (Murphy and Aha, 1991).

There are 6435 patterns (each one corresponds to a 3×3 square neighbourhood of pixels completely contained within the 82×100 sub-area) with 36 characteristics (4 spectral bands $\times$ 9 pixels in the neighbourhood). There are six distinct classes: red soil, cotton crop, gray soil, damp gray soil, soil with vegetation stubble, and very damp gray soil.

Fig. 3 shows a better performance of the RNGN and *k*-NCN rules with respect to the other classification schemes. On the other hand, it is worth mentioning that the GGN approach clearly yields the worst result once again; this effect has already been observed in the previous experiment, although differences are especially significant here.

4.4. The image database

This database was already used in (Pla et al., 1993) in order to study colour segmentation to locate citrus fruits in outdoor scenes under daylight condi-

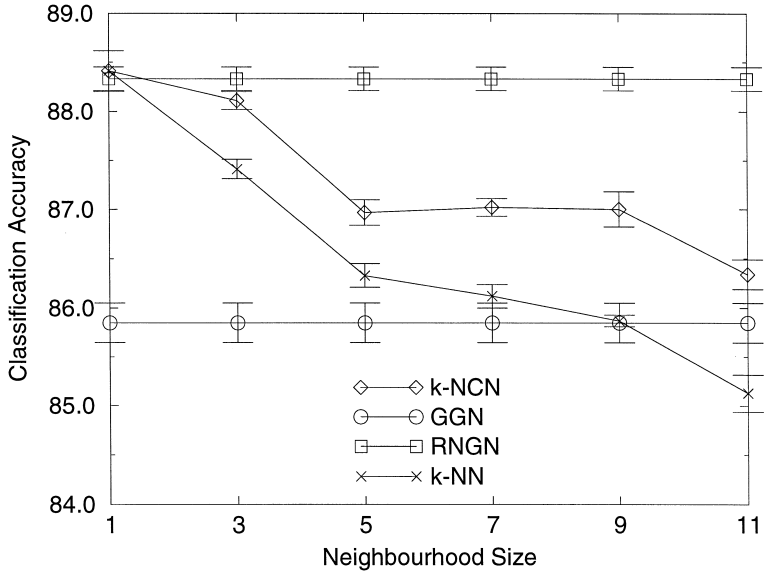

Fig. 2. Performance for data from an uttered vowel recognition problem.

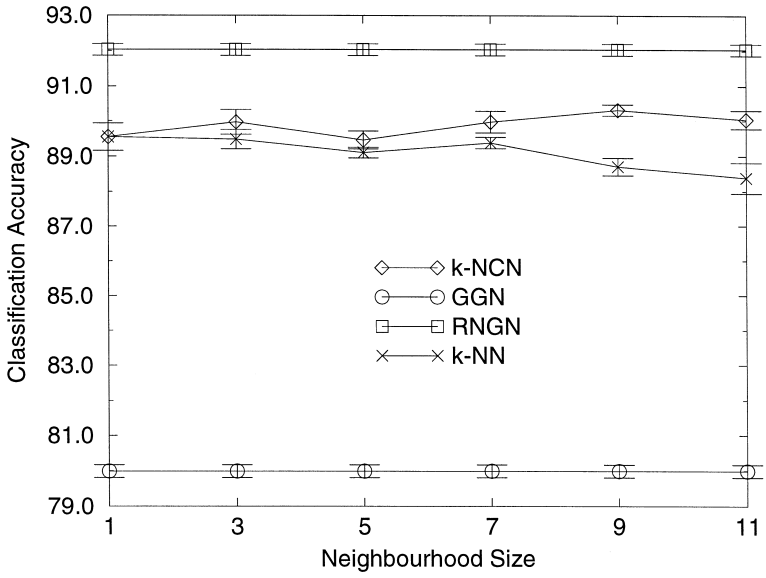

Fig. 3. Error rate estimates for the Landsat database.

tions. It consists of values from RGB colour images with a resolution of 256×256 pixels.

There are a total of 19 164 patterns selected from a 256×256 colour image with two attributes (coordinates Φ and Θ of the colour vectors in the RGB space) and three classes.

For this experiment, only one unbalanced partition of the data set (6386 for training and 12 778 for testing) was considered in order to compare the results obtained here to those achieved in previous related works (Pla et al., 1993), whose performance using an ad hoc classifier was close to 99%.

As can be seen in Fig. 4, the schemes proposed have a better behaviour than the *k*-NN classifier. The

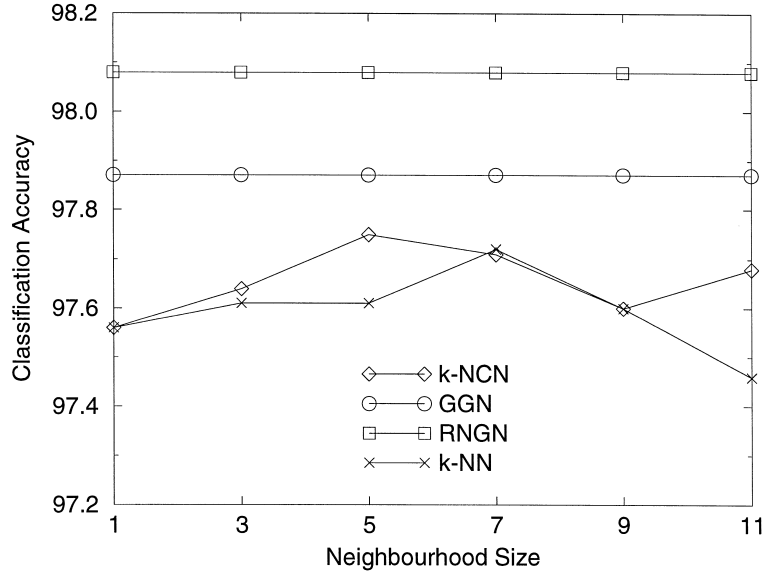

Fig. 4. Accuracy level for a real image recognition problem.

RNGN and *k*-NCN approaches achieve the highest rates for this problem, as occurred in all experiments.

5. Concluding remarks and further work

It is well known that the *k*-NN rule exhibits a very good performance with a sufficiently large number of prototypes in the training set. Nevertheless, when the size of the training set is not large enough compared to intrinsic dimensionality of the feature space, the behaviour of the *k*-NN rule is far from its optimal in terms of *Bayes* error.

In this paper, two non-parametric classification schemes have been proposed in an attempt to minimise the practical drawbacks present in the *k*-NN rule. These approaches come from using alternative definitions of a neighbourhood, which try to consider neighbours of a given sample not only in terms of distance-based proximity but also taking into account the spatial distribution of the prototypes in the training set. In particular, the NCN concept (Chaudhuri, 1996) as well as the *Graph Neighbourhood* (by means of the GG and the RNG) have been used to define alternative decision rules, here called *k*-NCN and GN, respectively (jointly referred to as *Surrounding classifiers*).

The performance of these approaches has been evaluated by means of some experiments. The comparative study with respect to the *k*-NN rule confirms that *Surrounding classifiers* can be useful in a number of practical situations. Furthermore, it is worth emphasizing that the *k*-NCN and RNGN rules show a relatively better behaviour than the GGN approach. On the other hand, differences between the performance of the GGN and RNGN rules can be explained by the large number of Gabriel neighbours considered for a given sample while, in practice, the neighbourhoods should be relatively small. The most relevant fact is that the proposed SN definitions have successfully been applied to classification, obtaining error rate estimates comparable or even better than those achieved by means of the *k*-NN approach.

A theoretical study of the proposed methods seems very difficult due to their heuristic nature. Nevertheless, the results obtained are encouraging enough to pursue further in this direction. In particular, a number of extensions proposed for the plain *k*-NN as, for example, using weights (Dudani, 1976) would be possible under these proposed schemes.

Discussion

Sklansky: Can you provide an intuitive explanation for these results. How do you explain the success of your results?

Sánchez: We think that this better behaviour of the surrounding approaches is because the classifiers take into account not only the proximity but also geometrical information of the neighbourhood of a sample.

Kamel: How do you explain the results of the first experiment, in which the Gabriel neighbours rule did best?

Sánchez: This can be explained by the distribution of the samples in this particular problem. And also by the characteristics of the overlap among classes. However, in all other real databases it is systematically the worst option. The reason why the Gabriel neighbours rule is systematically the worst is because there are too many neighbours in this kind of graph. But in the case of the artificial database in the first experiment, we have a distribution that is not so affected by the large number of Gabriel neighbours.

Raghavan: The rule that you use to decide which neighbours are close according to the Gabriel graph and the relative neighbourhood, was it based on the Euclidean distance, and have you tried to work with other ways of defining distance? The cityblock distance would be an alternative.

Sánchez: We have used only the Euclidean distance.

Raghavan: I think that when you change the way distances are defined, this neighbourhood will change in very dramatic ways. And I think that you have to

look at the interaction between your definitions and the distance function that is used.

References

Alinat, P., 1993. Periodic Progress Report 4: ROARS Project ESPRIT II 5516.

Chaudhuri, B.B., 1996. A new definition of neighbourhood of a point in multi-dimensional space. Pattern Recognition Lett. 17, 11–17.

Devijver, P.A., Kittler, J., 1982. Pattern Recognition: A Statistical Approach. Prentice-Hall, Englewood Cliffs, NJ.

Duda, R., Hart, P.E., 1973. Pattern Classification and Scene Analysis. Wiley, New York.

Dudani, S.A., 1976. The distance-weighted *k*-nearest neighbour rule. IEEE Trans. Systems Man Cybernet. 6, 325–327.

Jaromczyk, J.W., Toussaint, G.T., 1992. Relative neighbourhood graphs and their relatives. Proc. IEEE 80, 1502–1517.

Murphy, P.M., Aha, D.W., 1991. UCI Repository of Machine Learning Databases. Dept. of Information and Computer Science, University of California, Irvine.

Pla, F., Juste, F., Ferri, F.J., Vicens, M., 1993. Colour segmentation based on a light reflection model to locate citrus fruits for robotic harvesting. Computers and Electronics in Agriculture 9, 53–70.

Short, R.D., Fukunaga, K., 1980. A new nearest neighbour distance measure. In: Proc. 5th Internat. Conf. on Pattern Recognition, pp. 81–86.

Toussaint, G.T., Bhattacharya, B.K., Poulsen, R.S., 1985. The application of Voronoi diagrams to nonparametric decision rules. In: Computer Science and Statistics. Elsevier, Amsterdam, pp. 97–108.