Pattern Recognition Vol. 13, No. 3, pp. 251–255, 1981.
Printed in Great Britain.

0031-3203/81/030251-05 $02.00/0
Pergamon Press Ltd.

ON THE RELATION OF PERFORMANCE TO EDITING IN NEAREST NEIGHBOR RULES

JACK KOPLOWITZ
Clarkson College of Technology, Potsdam, NY 13676, U.S.A.
and
THOMAS A. BROWN
General Electric Company, Syracuse, NY 13221, U.S.A.

(*Received* 29 *May* 1979; *in revised form* 10 *April* 1980; *received for publication* 5 *August* 1980)

Abstract—A general class of editing schemes is examined which allows for the relabeling as well as the deletion of samples. It is shown that there is a trade-off between asymptotic performance and sample deletion which can adversely affect the finite sample performance. A kk' rule is proposed to minimize the proportion of deleted samples. A slight modification of the rule is introduced which allows for an exact analysis in any dimension.

Nearest neighbors Editing Probability of misclassification NN rules Asymptotic performance

INTRODUCTION

Nearest neighbor (NN) rules are well-known procedures for classifying unknown observations by examining the nearest samples from the data set. In the asymptotic case as the number of samples becomes arbitrarily large, the k nearest neighbors will be infinitesimally close to the sample being classified.[1,2] Clearly for a finite number of samples the value of k should be small enough so that all k neighbors are close to the unknown sample. For any k the asymptotic error rates for k-NN rules can be bounded in terms of the Bayes risk, the error rate if the Bayes or optimum classifier was used.

The single NN rule, although its asymptotic performance is not as good as with higher values of k, is simpler to implement and has available algorithms for condensing its data set.

To retain these advantages as well as achieve performance closer to that of the Bayes classifier, editing schemes have been proposed.[3–7] The data set is edited according to a prescribed rule and then a single NN rule is used for the final decision.

To examine the motivation of editing, the Bayes classifier which assigns an observation to the class having the greatest *a posteriori* probability at that point in the feature space is considered. These probabilities are approximately equal to the fraction of samples from each class in the neighborhood of the observation. A sample is defined to be of minority class if it is misclassified by Bayes rule. With the NN rule the performance is decreased by the presence of minority class samples. Editing attempts to remove the minority samples and thus obtain performance closer to the Bayes classifier. The proportion $\rho(x)$ of the number of minority class samples to the total number of samples in a neighborhood of x gives an indication of performance. If the points are distributed uniformly then $\rho(x)$ is the probability that for an observation x the NN classifier differs from the Bayes classifier.

Wilson[5] examined an editing procedure using the k-NN rule. The procedure tests a sample using the k-NN rule with the remainder of the data. The sample is discarded if it is misclassified. The edited data set is then used for single NN classification of unknown observations. The convergence of this editing rule has been proved by Wilson[5] and Wagner.[6]

GENERALIZED EDITING RULES

Consider Wilson's editing scheme with the 5-NN rule. A sample is edited if at least three of its five nearest neighbors are not of the same class. However, if relabeling of samples is allowed one finds that a further decrease of $\rho(x)$ occurs by including the following rule. If all five neighbors are of one class then the sample is labeled as belonging to that class. Further improvement occurs if we delete samples with only three of five neighbors of the same class. This indicates that rather than deleting on the basis of incorrect classification, deleting should be based on a lack of a strong indication of the true class of a sample.

One might consider a class of editing rules which, based on the class of a sample and its nearest neighbors, either delete, relabel or leave the sample unchanged. In an attempt to minimize $\rho(x)$ which is indicative of the performance, one quickly finds that the best rule is simply to edit unless the sample and all of its neighbors are of the same class. However, the proportion of deleted samples grows considerably and for a finite sample size this may reduce rather than improve the performance. Thus with editing there exists a trade-off between asymptotic and finite sample performance. To minimize $\rho(x)$ and limit the proportion of deleted samples a kk'-NN rule is proposed. Unless a sample and its $k-1$ nearest neighbors form a majority of k' out of k it is edited. Otherwise, it is labeled according to the majority class. Penrod and Wagner[7] point out that the editing process is not independent at each sample. This makes it difficult to determine the performance of the edited NN rules since the remaining data samples in a small neighborhood are no longer necessarily uniformly distributed. By restricting the problem to one dimension and modifying the editing procedure so that nearest neighbors to a sample x from those samples greater than x are used, an exact analysis is provided.[7] In this paper, a second alternative is presented for slightly modifying the editing procedure. The modification makes possible an exact analysis of the proposed kk'-NN rule.

For the single NN rule the modification consists of grouping the original data set into pairs. With the first sample a pair is formed with its nearest neighbor. With each next sample a pair is formed with the nearest neighbor remaining from the unpaired samples. Asymptotically, the probability that sample pairs are arbitrarily close approaches one. A sample is deleted if it is not of the same class as its corresponding sample in the pair. For the kk'-NN rule the modification is as follows.

(i) Samples are placed into groups of k.
(ii) If there is not a majority of k' of one class the group of k samples is deleted. Otherwise, all samples are labelled as belonging to the majority class.
(iii) The single NN rule is used on the edited data set.

It is not proposed that this grouping should necessarily be implemented in practice. In fact for a finite sample size one may prefer the ungrouped kk'-NN rule. The purpose of introducing this modification is to enable comparison of rules for various k', k and analyze the trade-off between performance and editing.

Let $p_1(x)$ be the probability that a sample at x belongs to class 1. Let $P_1(x)$ be the probability of classifying an unknown observation at x into class 1. Asymptotically, as the samples in a group become artibrarily close

$$P_1(x) = \frac{\sum_{i=k'}^{k} \binom{k}{i} p_1^i(x)[1 - p_1(x)]^{k-i}}{\sum_{i=k'}^{k} \binom{k}{i} [p_1^i(x)(1 - p_1(x))^{k-i} + (1 - p_1(x))^i p_1^{k-i}(x)]}. \quad (1)$$

Due to the relabeling of samples this is the proportion of class 1 samples after editing. Since the editing of each group is independent, this is equal to the probability of a class 1 decision.

Table 1. Table of bounds (b) for kk' group editing rule ($R^* \le R \le b \cdot R^*$).

k'	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$	$k=7$
2	1.21	1.31				
3		1.12	1.16	1.22		
4			1.08	1.10	1.12	1.17
5				1.06	1.08	1.09
6					1.05	1.06
7						1.04

The probability of error at x is given by

$$p(e|x) = p_1(x)P(2|x) + [1 - p_1(x)]P(1|x). \quad (2)$$

Rewriting (1) in terms of $r(x)$, the local Bayes risk $[0 \le r(x) \le 0.5]$, by replacing $P_1(x)$ with $r(x)$ and substituting into (2) yields

$$p(e|x) = \frac{\sum_{i=k'}^{k} \binom{k}{i} [r(x)^i(1 - r(x))^{k+1-i} + (1 - r(x))^i r^{k+1-i}(x)]}{\sum_{i=k'}^{k} \binom{k}{i} [r(x)^i (1 - r(x))^{k-i} + (1 - r(x))^i r^{k-i}(x)]}. \quad (3)$$

Due to the independence of each group this is equal to the proportion of deleted samples in a neighborhood of x.

These results depend on the assumption that after editing the group containing the NN to x is arbitrarily close to x. To show this convergence it is assumed, without loss of generality, that any neighborhood $N(x)$ around x has non-zero probability measure. It is shown by Cover and Hart[1] that x has this property with probability one. As the number of samples $n \to \infty$ the number of groups, m, within $N(x)$ becomes arbitrarily large. The probability of editing a group is

$$P(ed|x) = \sum_{i=k-k'+1}^{k'-1} \binom{k}{i} r^i(x)[1 - r(x)]^{k-i} < 1. \quad (4)$$

Since the editing of each group is independent, the probability of no groups remaining in $N(x)$ is $P^m(ed|x) \to 0$ as $m \to \infty$. Thus the probability that x is not arbitrarily close to the group containing its NN approaches zero.

Table 2. Table of bounds (b) for the kNN rule ($R^* \le R \le b \cdot R^*$).

k	b
1	2
3	1.31
5	1.22
7	1.17

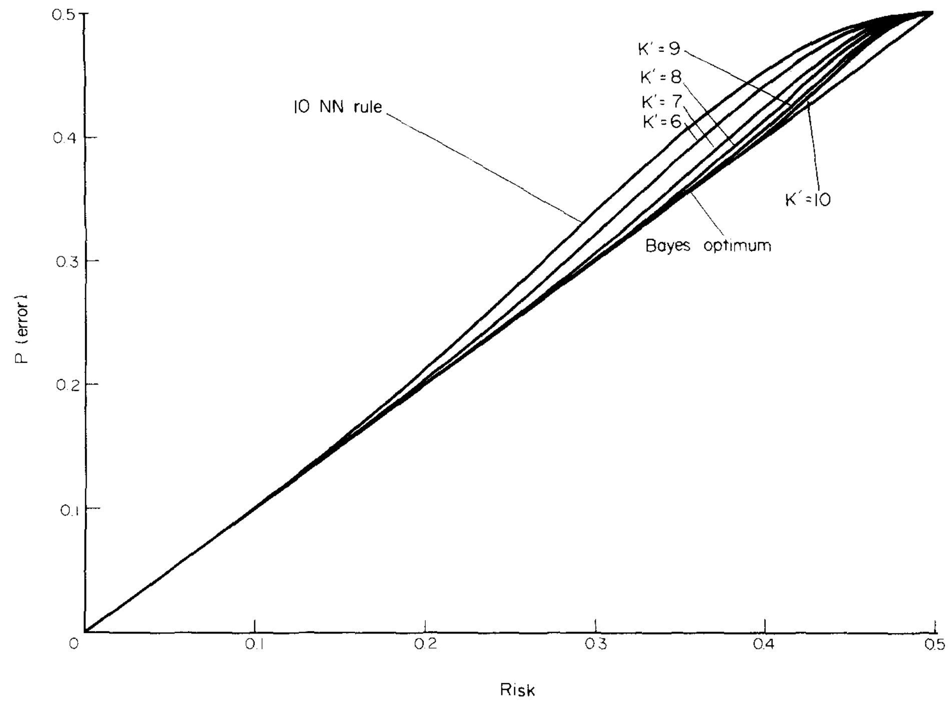

Fig. 1. Probability of error for kk' editing of groups, $k = 10$.

Bounds on the error rate in terms of the Bayes risk (R^*) for this grouping rule are given in Table 1 for various k and k'. Comparisons are made to the bounds for the ordinary k-NN rule given in Table 2. The results in Penrod and Wagner[7] indicate that without the grouping modification performance may be slightly worse. We conjecture that as the dimensionality becomes arbitrarily large the editing of each sample becomes independent, making the results in Table 1 exact for the unmodified rule.

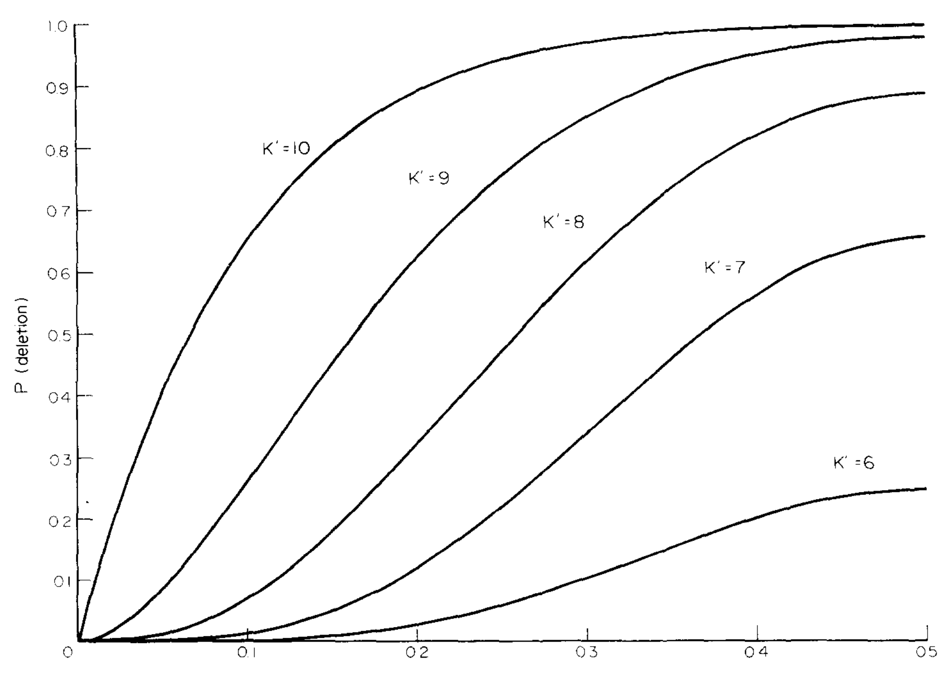

Fig. 2. Probability of deletion for kk' editing of groups, $k = 10$.

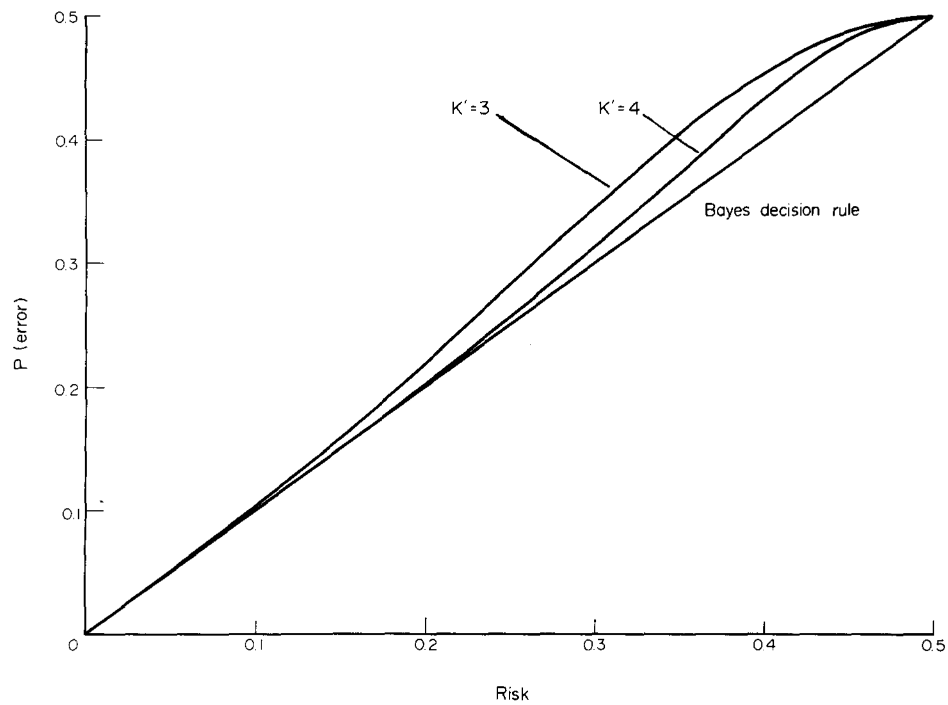

Fig. 3. Probability of error for kk' editing of groups, $k = 4$.

The effect of varying k' with k fixed is shown in Figs 1 and 2 for $k = 10$ and in Figs 3 and 4 for $k = 4$. As k' approaches k the error rate decreases toward the Bayes risk but the sample deletion rate increases greatly. In the finite sample case for classifiers with edited data sets it is necessary to compromise between the asymptotic or infinite sample performance and the degree of editing.

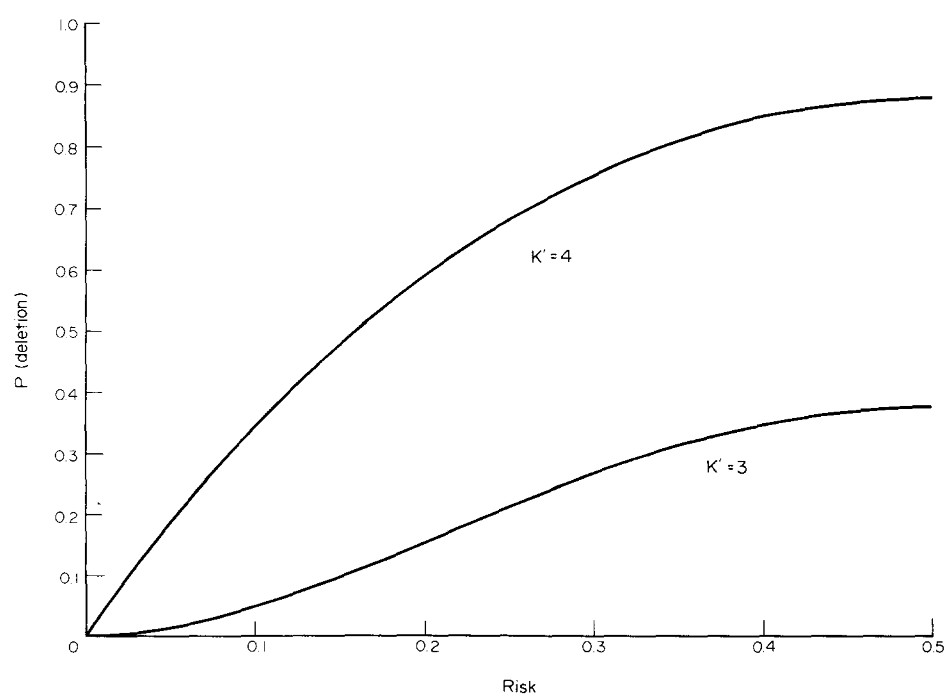

Fig. 4. Probability of deletion for kk' editing of groups, $k = 4$.

SUMMARY

An edited nearest neighbor (NN) rule has been proposed by Wilson.[5] The procedure tests a sample using the k-NN rule with the remaining data. If it is misclassified the sample is discarded. The classifier then uses the single NN rule with the edited data set. Penrod and Wagner[7] have pointed out the difficulty in analyzing the performance with edited data. By modifying the procedure slightly they have obtained exact results for one-dimensional patterns.

In this paper a general class of editing and relabeling schemes based on k-nearest neighbors is examined. It is shown that there is a trade-off between performance and editing. A kk' rule is proposed to minimize the editing for a fixed level of performance. A modification of the kk' rule is introduced which makes possible an exact analysis of the proposed rule for any dimension. Asymptotic results on the performance of kk' rules are compared to the k-nearest neighbor rules showing that editing can significantly improve the level of performance. The results point out the inverse relationship between asymptotic performance and the proportion of non-edited data.

REFERENCES

1. T. M. Cover and P. E. Hart, Nearest neighbor pattern classification, *IEEE Trans. Inf. Theory* **IT-13,** 21–27 (1967).
2. T. J. Wagner, Convergence of the nearest neighbor rule, *IEEE Trans. Inf. Theory,* **IT-17,** 566–572 (1971).
3. I. Tomek, A generalization of the k-NN rule, *IEEE Trans. Syst., Man Cybernet.* **SMC-6,** 121–127 (1976).
4. I. Tomek, An experiment with the edited nearest neighbor rule, *IEEE Trans. Syst., Man Cybernet.* **SMC-6,** 448–452 (1976).
5. D. L. Wilson, Asymptotic properties of NN rules using edited data, *IEEE Trans. Syst., Man Cybernet.* **SMC-2,** 408–421 (1972).
6. T. J. Wagner, Convergence of the edited nearest neighbor, *IEEE Trans. Inf. Theory* **IT-19,** 696–697 (1973).
7. C. S. Penrod and T. J. Wagner, Another look at the edited nearest neighbor rule, *IEEE Trans. Syst., Man Cybernet.,* **SMC-7,** 92–94 (1977).

About the Author – JACK KOPLOWITZ was born in Lenger, Kazakstan, USSR, on March 12, 1944. He received the B.E.E. degree from the City College of New York, N.Y., in 1967; the M.E.E. degree from Stanford University, Stanford CA, in 1968; and the Ph.D. degree from the University of Colorado, Boulder, CO, in 1973.

From 1968 to 1970 he was a member of the technical staff at Bell Laboratories, Holmdel, NJ, working in the area of data communications. He is presently Associate Professor of Electrical and Computer Engineering, Clarkson College of Technology, Potsdam NY. During the summers of 1975 and 1978 he was a visiting scholar at the Information Systems Laboratory, Department of Electrical Engineering, Stanford University. He spent the 1979–1980 academic year as a Research Fellow with the Department of Electrical Engineering, Technion, Israel Institute of Technology, Haifa, Israel.

Dr. Koplowitz is a member of Tau Beta Pi, Eta Kappa, Nu and Sigma Xi.

About the Author – THOMAS A. BROWN received a BSEE in 1977 and a MEEE in 1978 from Clarkson College of Technology in Potsdam, New York. Currently he is employed by General Electric Co. at the Electronics Laboratory in Syracuse, New York. He is a member of Tau Beta Pi, Eta Kappa Nu, and Phi Kappa Phi.