CC/NUMBER 13
MARCH 29, 1982

# This Week's Citation Classic

**Cover T M & Hart P E.** Nearest neighbor pattern classification.
*IEEE Trans. Inform. Theory* **IT-13**:21-7, 1967.
[Dept. Electrical Engineering, Stanford Univ., Stanford, and Stanford Res. Inst., Menlo Park, CA]

**The nearest neighbor decision rule assigns to an unclassified sample the classification of the nearest of a set of previously classified samples. This paper proves that the probability of error of the nearest neighbor rule is bounded above by twice the Bayes minimum probability of error. In this sense, it may be said that half the classification information in an infinite sample set is contained in the nearest neighbor. [The *SCI*® indicates that this paper has been cited over 190 times since 1967.]**

Thomas M. Cover
Departments of Statistics and
Electrical Engineering
Stanford University
Stanford, CA 94305

March 5, 1982

"Early in 1966 when I first began teaching at Stanford, a student, Peter Hart, walked into my office with an interesting problem. He said that Charles Cole and he were using a pattern classification scheme which, for lack of a better word, they described as the nearest neighbor procedure. This scheme assigned to an as yet unclassified observation the classification of the nearest neighbor. Were there any good theoretical properties of this procedure? Of course the motivation for such a classification rule must go back to prehistoric times. The idea is that 'things that look alike must be alike.'

"The problem seemed extremely inviting from a theoretical point of view. We began meeting for two or three hours every afternoon in an attempt to find some distribution-free properties of this classification rule. By distribution-free, I mean properties that are true regardless of the underlying joint distribution of the categories and observations. Obviously, we could not hope to prove that a procedure always has, for example, a zero probability of error, because there are many cases where the observations yield no information about the underlying category. In those problems, the goal would be more modest. Apparently, the proper goal would be to relate the probability of error of this procedure to the minimal probability of error given complete statistical information, namely, the Bayes risk.

"After some effort, we were able to prove that the nearest neighbor risk is always less than the Bayes risk plus one-sixth (if I remember correctly). This was the sort of result we were looking for, but it seemed quite unnatural. Also, it was not a very ambitious bound when the Bayes risk is near zero. We would have preferred to relate risks by a factor rather than by an additive constant. Soon thereafter we found what we were looking for. The nearest neighbor risk is less than twice the Bayes risk for all reasonable distributions and for any number of categories. Thus ancient man was proved right—'things that look alike are alike'—with a probability of error that is no worse than twice the probability of error of the most sophisticated modern day statistician using the same information. Moreover, we were soon able to prove that our bound was the best possible. So the search was over.

"The simplicity of the bound and the sweeping generality of the statement, combined with the obvious simplicity of the nearest neighbor rule itself, have caused this result to be used by others, thus accounting for the high number of citations. Since the properties of the nearest neighbor rule can be easily remembered, the bound yields a benchmark for other more sophisticated data analysis procedures, which sometimes actually perform worse than the nearest neighbor rule. This is probably due to the fact that more ambitious rules have too many parameters for the data set.

"It should be mentioned that we had to exclude a certain technical set of joint distributions from the proof of our theorem. The attendant measure-theoretic difficulties in eliminating the so-called singular distributions almost delayed the publication of our paper. It was wise that we did not hold up publication, because the theorem was not proved in total generality until ten years later in Charles Stone's 1977 paper in the *Annals of Statistics*.[1] The result remains the same, but now it applies to all possible probability distributions."

1. **Stone C J.** Consistent nonparametric regression. *Ann. Statist.* **5**:595-645, 1977.