Journal of Statistical Research *ISSN 0256 - 422 X*
2007, Vol. 41, No. 1, pp. 69–79
Bangladesh

COMPARISON BETWEEN k–NN AND k–RNN CLASSIFICATION RULES : A MONTE CARLO SIMULATION STUDY

S. S. Bagui

Department of Computer Science, The University of West Florida, USA

Email: bagui@uwf.edu

S. C. Bagui

Department of Mathematics and Statistics, The University of West Florida, USA

Email: sbagui@uwf.edu

N. R. Pal

Electronics and Communication Sciences Unit, Indian Statistical Institute, India

Email: Nikhil@isical.ac.in

M. A. Matin

Visiting Scientist, Applied Statistics Unit, Indian Statistical Institute, India

Email: matin@maths.leeds.ac.uk

SUMMARY

In this paper we conduct a comparison between the k–NN and k–RNN classification rules using an extensive monte carlo simulation study. We observe that the k–RNN rule performs better than the k–NN rule both in small and large sample cases.

Keywords and phrases: Classification, Discrimination, k–NN, k–RNN, Simulation, Error rates.

AMS Classification: 62G99, 62H30, 68T10.

1 Introduction

Classification and pattern recognition is about assigning labels (or objects) to one of two or more predefined classes (or categories). Various parametric and nonparametric procedures are available in the literature for classification and pattern recognition. We will focus our attention to only nearest neighbor (NN) type classification rules. Nearest neighbor classification rules are widely used in statistical pattern recognition and also popular in computer science, marine science, biological sciences applications, and other areas. Two types of NN classification rules are (i) distance based nearest neighbor rules and (ii) rank based nearest neighbor rules. Distance based nearest neighbor rules will be denoted by NN rules and the rank based nearest neighbor rules will be denoted by RNN rules. Below we describe the k–NN and k–RNN classification rules.

The k—NN Methods: The k–NN classification rule was first introduced by Fix and Hodges (1951). Their k–NN rule is based on the density estimates using distance nearest neighbors. The rule may be described as follows: Let $\{X_1, X_2, \ldots, X_{n_1}\}$ and $\{Y_1, Y_2, \ldots, Y_{n_2}\}$ be training samples from two given populations, π_1 and π_2, respectively. Let Z be an unknown observation known to be either from π_1 or π_2 to be correctly classified between π_1 and π_2. Using a distance function d, rank order the distances of all the observations from Z. For a fixed integer k, the k–NN rule assigns Z to π_i if $k_1/n_1 > k_2/n_2$, where k_i is the number of observations from $\pi_i, (i = 1, 2)$, among the $k = k_1 + k_2$ observations nearest to Z. Cover and Hart (1967) proposed a slightly modified version of the above rule and is described as follows:

The Conventional k–NN Algorithm:
This rule assigns Z to the population $\pi_i, (i = 1, 2)$, if $k_i = \overset{max}{j}\{k_j\}$. That is, this rule assigns Z to π_i if the majority of the k nearest neighbors (in a distance sense), of Z come from $\pi_i, (i = 1, 2)$. We call this rule the conventional k–NN classification rule.

Cover and Hart (1967) studied the 1–NN rule and found that the limiting risk, $R_D(1)$, of this rule, has the following lower and upper bounds, $R^* \leq R_D(1) \leq 2R^*(1-R^*)$, where R^* is the minimum Bayes risk obtained as $R^* = \int \min(\xi_1 f_1(z), \xi_2 f_2(z))dz$, where ξ_i and f_i are the prior probability and the probability density function of the population $\pi_i, (i = 1, 2)$, respectively. Devroye (1981a) investigated the k–NN rule and derived the following upper bound on the asymptotic risk, $R_D(k)$, of the rule, $R_D(1) < (1+\alpha_i)R^*$, where $\alpha_k = \frac{\alpha\sqrt{k}}{k-3.25}(1 + \frac{\beta}{\sqrt{k-3}})$, k odd, $k \geq 5$, and $\alpha = 0.3399$ and $\beta = 0.9749$ are universal constants. This bound is the best possible in a certain sense. For other properties and aspects of the conventional k–NN rule, we refer the reader to Devijver and Kittler (1982), Wagner (1971), Fritz (1975), Devroye (1981b), and Xiru (1985). It should be pointed out that, as the important reference by Fix and Hodges (1951) is rather inaccessible, the paper has been reprinted at the end of the commentary on it by Silverman and Jones (1989).

The k-R-NN Methods: The rank nearest neighbor classification rule was first introduced by Anderson (1966). In recent days, R-NN classification rule is gaining more popularity

in the area of classification and statistical pattern recognition because of its simplicity and robustness. The k-R-NN classification rule for univariate populations may be described as follows:

The k-R-NN Algorithm:
Pool the sample observations X_i's, Y_j's, and Z, and rank them in increasing order; then count up k observations to the right-hand side of Z and count down k observations to the left-hand side of Z; (i) if there are more X-neighbors than Y-neighbors among rank nearest neighbors, classify Z into the X-population π_1; (ii) if there are more Y-neighbors than X-neighbors among $2k$ rank nearest neighbors, classify Z into Y-population π_1; (iii) if there are exactly k X-neighbors and k Y-neighbors, classify Z into either of the two populations with probability $\frac{1}{2}$ each (to break the tie); and if on any side of Z k observations are not available then use as many as available.

Dasgupta and Lin (1980) investigated the 1-R-NN rule. They derived the asymptotic risk, $R_R(1)$, of the k-R-NN rule and showed that $R^* \leq R_R(1) \leq 2R^*(1 - R^*)$, where R^* is the minimum Bayes error rate defined earlier. In fact, this asymptotic risk is exactly the same as that of the 1–NN rule. Bagui and Pal (1995) extended Dasgupta and Lins work to more than two populations and also suggested a 1-R-NN rule for multivariate data. Bagui and Vaughn (1998) examined the k-R-NN rule for univariate populations and derived an upper bound on the asymptotic risk, $R_R(k)$, of this rule which is parallel to the upper bound obtained by Devroye (1981a). Bagui and Vaughn (1998) also noted that this risk converges to the Bayes risk twice as fast as the conventional k–NN rule. Bagui et al. (2003) proposed and studied a k-R-NN for multivariate data. In this article we present a wide-ranging monte carlo simulation comparison between k–NN and k-R-NN classification rules under univariate populations.

2 Asymptotic Relationship Between k–NN Rule and k-R-NN Rule

Let X_1 possess a density function f_1 and Y_1 possess a density function f_2. We know that Z has density either f_1 or f_2. Let $N = \min(n_1, n_2)$. The finite sample probabilities of misclassification (PMC) of the k-R-NN classification rule are given as follows:

$$\alpha_{12}(k; n_1, n_2) = P(\text{classify } Z \in \pi_2 | Z \in \pi_1) \tag{2.1}$$

$$\alpha_{21}(k; n_1, n_2) = P(\text{classify } Z \in \pi_1 | Z \in \pi_2). \tag{2.2}$$

Thus, the total probability of misclassification (TPMC) or the risk of the k-R-NN rule is given by

$$R_R(k; n_1, n_2) = \xi_1 \alpha_{12}(k; n_1, n_2) + \xi_2 \alpha_{21}(k; n_1, n_2). \tag{2.3}$$

Thus, the asymptotic risk of the k-R-NN rule may be written as

$$R_k = \xi_1 \alpha_{12}(k) + \xi_2 \alpha_{21}(k). \tag{2.4}$$

Let $\pi(z)$ denote the asymptotic conditional probability of classifying Z into π_1, given $Z = z$. Viewing Bagui and Vaughn (1998), $\pi(z)$ can be written as

$$\pi(z) = \sum_{j=k+1}^{2k} \left(\begin{array}{c} 2k \\ j \end{array} \right) \eta_1^j(z)\eta_2^{2k-j}(z) + \frac{1}{2} \left(\begin{array}{c} 2k \\ k \end{array} \right) \eta_1^k(z)\eta_2^k(z), \tag{2.5}$$

where, $\eta_i(z) = \frac{\xi_i f_i(z)}{\xi_1 f_1(z)+\xi_2 f_2(z)}, i = 1, 2.$

Thus, the asymptotic PMC's of the k-R-NN rule is given by

$$\alpha_{12}(k) = \lim_{n\to\infty} \alpha_{12}(k; n_1, n_2) = \lim_{n\to\infty} P(\text{classify} Z \in \pi_2 | Z \in \pi_1) = \int (1 - \pi(z)) f_1(z) dz \tag{2.6}$$

$$\alpha_{21}(k) = \lim_{n\to\infty} \alpha_{21}(k; n_1, n_2) = \lim_{n\to\infty} P(\text{classify} Z \in \pi_1 | Z \in \pi_2) = \int \pi(z) f_2(z) dz. \tag{2.7}$$

Using (2.4) to (2.7), the asymptotic TPMC (or risk) of the k-R-NN rule can be expressed as

$$\begin{aligned} R_R(k) &= \xi_1 \int (1 - \pi(z)) f_1(z) dz + \xi_2 \int \pi(z) f_2(z) dz \\ &= \int \xi_1 f_1(z) [\sum_{j=k+1}^{2k} \left(\begin{array}{c} 2k \\ j \end{array} \right) \eta_1^j(z)\eta_2^{2k-j}(z) + \frac{1}{2} \left(\begin{array}{c} 2k \\ k \end{array} \right) \eta_1^k(z)\eta_2^k(z)] dz \\ &+ \int \xi_2 f_2(z) [\sum_{j=k+1}^{2k} \left(\begin{array}{c} 2k \\ j \end{array} \right) \eta_1^j(z)\eta_2^{2k-j}(z) + \frac{1}{2} \left(\begin{array}{c} 2k \\ k \end{array} \right) \eta_1^k(z)\eta_2^k(z)] dz. \end{aligned} \tag{2.8}$$

For simplicity of notation, writing η_i for $\eta_i(z)$, $R_R(k)$ can be re-expressed as

$$\begin{aligned} R_R(k) = E\eta_1\eta_2(\eta_1^{2k-1} + \eta_2^{2k-1}) + \left(\begin{array}{c} 2k \\ 1 \end{array} \right) E\eta_1^2\eta_2^2(\eta_1^{2k-3} + \eta_2^{2k-3}) \\ + \ldots + \left(\begin{array}{c} 2k \\ k-1 \end{array} \right) E\eta_1^k\eta_2^k + \frac{1}{2} \left(\begin{array}{c} 2k \\ k \end{array} \right) E\eta_1^k\eta_2^k. \end{aligned} \tag{2.9}$$

Theorem 2.1. (Bagui and Vaughn (1998)). The asymptotic risk has the following property:

$$R_R(k) \leq R_R(k-1), \quad k = 2, 3, \ldots .$$

Let $R_D(k)$ denote the asymptotic risk k–NN rule. From Cover and Hart (1967), $R_D(k)$

can be expressed as

$$R_D(k) = \int \xi_1 f_1(z)[\sum_{j=0}^{m-1} \binom{k}{j} \eta_1^j(z)\eta_2^{k-j}(z) + \frac{1}{2}\binom{k}{m}\eta_1^m(z)\eta_2^m(z)]dz + \int \xi_2 f_2(z)[\sum_{j=m+1}^{k} \binom{k}{j} \eta_1^j(z)\eta_2^{k-j}(z) + \frac{1}{2}\binom{k}{m}\eta_1^m(z)\eta_2^m(z)]dz \tag{2.10}$$

where $k = 2m$.

Clearly, from (2.8) and (2.10) we note that $R_D(k) = R_R(\frac{k}{2})$ for k even. Now from Theorem 2.1, we may conclude that $R_R(k) \leq R_R(\frac{k}{2}) = R_D(k)$. Thus, asymptotic risk of k-R-NN rule is less than that of asymptotic risk of k–NN rule is less than that of asymptotic risk of k–NN rule.

In this article we present an extensive monte carlo simulation comparison between k–NN and k-R-NN classification rules.

3 Simulation Methods and Results

The asymptotic properties are not necessarily valid for small sample cases. In this section we examine the performance of k-R-NN and k–NN rules using monte carlo simulation study in small, moderate, and large sample cases. In order to compare the k-R-NN rule with the conventional k–NN rule, random samples of equal size ($n = n_1 = n_2$) were simulated from pairs of each of the following univariate distribution, namely, normal, lognormal, gamma, exponential, logistic. Given these samples, (treated as training samples), 1000 random observations from π_1 and another 1000 observations from π_2 were simulated and were classified according to k-R-NN and k–NN classification rules. The proportion among the 20000 Zs that were misclassified by the k-R-NN and k–NN rules were computed and displayed in the following tables. We varied from 10 to 1000 and k from 1 to 6.

The mean separations in Table 1 to 3 increase from 1 to 2 to 3 while the variance in all cases remains fixed at 1. As mean separation $|\mu_i - \mu_j|$ increases in Table 1 to Table 3, the error rate drops; as k increases, the error rate drops. Both of these trends agree with the theory. From Table 1, we find that the average error rate of the k-R-NN rule is 0.3647 with a standard deviation (s.d.) of 0.02510 and the average error rate of the k–NN rule is 0.3801 with a standard deviation (s.d.) of 0.02366. Thus the Z value of $2.679 > 2.326$ supports the hypothesis that the error rate for the k–NN rule is significantly larger than the k-R-NN rule at 1% level of significance. More accurately, since the P-value = 0.00369, thus the exact level of significance is 0.369%.

From Table 2, we observe that the average error rate of the k-R-NN rule is 0.1898 with a standard deviation (s.d.) of 0.03271 and the average error rate of the k–NN rule is 0.2046 with a standard deviation (s.d.) of 0.03. Thus the Z value of $2.001 > 1.96$ supports the hypothesis that error rate for the k–NN rule is significantly larger than the k-R-NN rule

Table 1: Average proportion of misclassification by k-R-NN and k–NN rules for the pair N(0, 1) vs. N(1, 1)

k	1	2	3	4	5	6
n	R-NN	R-NN	R-NN	R-NN	R-NN	R-NN
	-NN	-NN	-NN	-NN	-NN	-NN
10	0.4145	0.3907	0.3825	0.3825	0.3825	0.3682
	0.4210	0.4292	0.3665	0.3615	0.3735	0.3552
20	0.4150	0.3517	0.3397	0.3545	0.3380	0.3417
	0.4140	0.4225	0.3430	0.3692	0.3480	0.3472
50	0.4077	0.3880	0.3570	0.3547	0.3372	0.3367
	0.4060	0.4130	0.3805	0.3850	0.3810	0.3730
100	0.3965	0.3747	0.3562	0.3520	0.3432	0.3365
100	0.3945	0.3967	0.3690	0.3765	0.3725	0.3727
500	0.4042	0.3727	0.3555	0.3585	0.3500	0.3470
	0.4085	0.3992	0.3780	0.3727	0.3615	0.3592
1000	0.4070	0.3675	0.3482	0.3487	0.3387	0.3300
	0.4000	0.4000	0.3725	0.3682	0.3455	0.3480

N(a, b) Normal distribution with mean a and variance b.

Table 2: Average proportion of misclassification by k-R-NN and k–NN rules for the pair N(0, 1) vs. N(2, 1)

k	1	2	3	4	5	6
n	R-NN	R-NN	R-NN	R-NN	R-NN	R-NN
	-NN	-NN	-NN	-NN	-NN	-NN
10	0.2992	0.2672	0.2505	0.1767	0.1767	0.1775
	0.2825	0.2825	0.2250	0.2310	0.2360	0.2152
20	0.2515	0.2037	0.1842	0.1787	0.1595	0.1607
	0.2490	0.2500	0.2060	0.2185	0.1980	0.1905
50	0.2097	0.1707	0.1725	0.1670	0.1642	0.1645
	0.2145	0.2092	0.1740	0.1720	0.1655	0.1670
100	0.2210	0.1867	0.1792	0.1682	0.1630	0.1635
	0.2120	0.2145	0.1745	0.1785	0.1795	0.1770
500	0.2160	0.1790	0.1750	0.1747	0.1705	0.1677
	0.2180	0.2092	0.1770	0.1797	0.1800	0.1790
1000	0.2197	0.1935	0.1852	0.1780	0.1800	0.1777
	0.2185	0.2230	0.1955	0.1922	0.1865	0.1847

Table 3: Average proportion of misclassification by k-R-NN and k–NN rules for the pair N(0, 1) vs. N(3, 1)

k	1	2	3	4	5	6
n	R-NN	R-NN	R-NN	R-NN	R-NN	R-NN
	-NN	-NN	-NN	-NN	-NN	-NN
10	0.1175	0.0842	0.0875	0.0747	0.0747	0.0747
	0.1035	0.1310	0.1250	0.1270	0.1305	0.1130
20	0.1127	0.0715	0.0715	0.0730	0.0735	0.0730
	0.1070	0.1225	0.0715	0.0707	0.0715	0.0725
50	0.0872	0.0735	0.0727	0.0705	0.0715	0.0720
	0.0785	0.0825	0.0755	0.0720	0.0715	0.0710
100	0.0807	0.0745	0.0730	0.0727	0.0715	0.0695
	0.0790	0.0817	0.0790	0.0737	0.0715	0.0750
500	0.0882	0.0732	0.0752	0.0730	0.0737	0.0730
	0.0885	0.0875	0.0700	0.0735	0.0730	0.0742
1000	0.0938	0.0795	0.0805	0.0762	0.0755	0.0747
	0.0975	0.0980	0.0760	0.0805	0.0760	0.0777

at 2.5% level of significance. Since the P-value = 0.0225, the exact level of significance is 2.25%.

From Table 3, we find that the average error rate of the k-R-NN rule is 0.0782 with a standard deviation (s.d.) of 0.01072 and the average error rate of the k–NN rule is 0.0851 with a standard deviation (s.d.) of 0.02364. Thus the Z value of $1.595 > 1.282$ supports the hypothesis that error rate for the k–NN rule is significantly larger than the k-R-NN rule at 10% level of significance. Since the P-value = 0.05536, thus exact level of significance is 5.536%.

From Table 4, we find that the average error rate of the k-R-NN rule is 0.1898 with a standard deviation (s.d.) of 0.03265 and the average error rate of the k–NN rule is 0.1999 with a standard deviation (s.d.) of 0.02731. Thus the Z value of $1.42 > 1.282$ supports the hypothesis that the error rate for the k–NN rule is significantly larger than the k-R-NN rule at 10% level of significance. Since the P-value = 0.0778, the exact level of significance is 7.78%.

From Table 5, we find that the average error rate of the k-R-NN rule is 0.0781 with a standard deviation (s.d.) of 0.01072 and the average error rate of the k–NN rule is 0.0811 with a standard deviation (s.d.) of 0.01315. Thus the $Z = 1.060$ with a P-value = 0.14456 (0.10) does not support the hypothesis that error rate for the k–NN rule is significantly larger than the k-R-NN rule at 10% level of significance. The exact level of significance is 14.456%. In this case, on the average, the k-R-NN rule performed better than the k–NN rule but not significantly better.

Table 4: Average proportion of misclassification by k-R-NN and k–NN rules for the pair LG(0, 1) vs. LG(2, 1)

k	1	2	3	4	5	6
n	R-NN	R-NN	R-NN	R-NN	R-NN	R-NN
	-NN	-NN	-NN	-NN	-NN	-NN
10	0.2992	0.2672	0.2505	0.1767	0.1767	0.1775
	0.2765	0.2650	0.2005	0.1990	0.2005	0.1782
20	0.2515	0.2037	0.1842	0.1787	0.1595	0.1607
	0.2485	0.2505	0.2080	0.2155	0.1955	0.1835
50	0.2097	0.1707	0.1725	0.1670	0.1642	0.1645
	0.2130	0.2092	0.1745	0.1722	0.1645	0.1662
100	0.2210	0.1867	0.1792	0.1682	0.1630	0.1635
	0.2115	0.2147	0.1740	0.1780	0.1780	0.1785
500	0.2160	0.1790	0.1750	0.1747	0.1705	0.1677
	0.2180	0.2082	0.1770	0.1797	0.1800	0.1792
1000	0.2197	0.1935	0.1852	0.1780	0.1800	0.1777
	0.2185	0.2227	0.1955	0.1922	0.1865	0.1847

LG(a, b) Logistic distribution with location parameter a and scale parameter b.

Table 5: Average proportion of misclassification by k-R-NN and k–NN rules for the pair LG(0, 1) vs. LG(3, 1)

k	1	2	3	4	5	6
n	R-NN	R-NN	R-NN	R-NN	R-NN	R-NN
	-NN	-NN	-NN	-NN	-NN	-NN
10	0.1175	0.0842	0.0875	0.0747	0.0747	0.0747
	0.1065	0.1060	0.0750	0.0737	0.0735	0.0760
20	0.1127	0.0715	0.0715	0.0730	0.0735	0.0730
	0.1110	0.1242	0.0705	0.0707	0.0725	0.0707
50	0.0872	0.0735	0.0727	0.0705	0.0715	0.0702
	0.0790	0.0827	0.0760	0.0727	0.0715	0.0712
100	0.0807	0.0745	0.0730	0.0727	0.0715	0.0695
	0.0790	0.0815	0.0795	0.0740	0.0735	0.0745
500	0.0882	0.0732	0.0752	0.0730	0.0737	0.0730
	0.0885	0.0877	0.0700	0.0737	0.0730	0.0742
1000	0.0938	0.0795	0.0805	0.0762	0.0755	0.0747
	0.0975	0.0982	0.0760	0.0802	0.0760	0.0780

Table 6: Average proportion of misclassification by k-R-NN and k–NN rules for the pair LN(0, 1) vs. LN(2, 1)

k	1	2	3	4	5	6
n	R-NN	R-NN	R-NN	R-NN	R-NN	R-NN
	-NN	-NN	-NN	-NN	-NN	-NN
10	0.1877	0.1735	0.1735	0.1675	0.1675	0.1675
	0.1940	0.1882	0.1660	0.1682	0.1725	0.1670
20	0.2085	0.1775	0.1675	0.1677	0.1647	0.1657
	0.2085	0.2032	0.1690	0.1690	0.1660	0.1662
50	0.1882	0.1732	0.1672	0.1715	0.1672	0.1650
	0.1910	0.2005	0.1660	0.1702	0.1675	0.1690
100	0.2135	0.1875	0.1702	0.1692	0.1687	0.1665
	0.2130	0.2172	0.1810	0.1820	0.1670	0.1692
500	0.2150	0.1847	0.1790	0.1762	0.1765	0.1692
	0.2240	0.2202	0.1875	0.1885	0.1770	0.1765
1000	0.2115	0.1832	0.1767	0.1752	0.1725	0.1715
	0.2185	0.2205	0.1815	0.1860	0.1815	0.1770

LN(a, b) Lognormal with mean a and variance b.

From Table 6, we find that the average error rate of the k–RNN rule is 0.1774 with a standard deviation (s.d.) of 0.01407 and the average error rate of the k–NN rule is 0.1853 with a standard deviation (s.d.) of 0.01892. Thus the Z value of $2.010 > 1.96$ supports the hypothesis that the error rate for the k–NN rule is significantly larger than the k–RNN rule at 2.5% level of significance. Since the P-value = 0.02222, the exact level of significance is 2.222%.

From Table 7, we find that the average error rate of the k–RNN rule is 0.3447 with a standard deviation (s.d.) of 0.02412 and the average error rate of the k–NN rule is 0.3597 with a standard deviation (s.d.) of 0.02289. Thus the Z value of $2.706 > 2.576$ supports the hypothesis that error rate for the k–NN rule is significantly larger than the k–RNN rule at 0.5% level of significance. Since the P-value = 0.00341, the exact level of significance is 0.341%.

In majority of the cases, k–RNN rule performed significantly better than k–NN rule. From Table 1, 2, 6, and 7, we see that when there is a greater overlap between two populations the k–RNN rule tends to perform significantly better then the k–NN rule. However, when the overlap between two populations decreases, the significance level increases. This means both will tend to perform statistically equivalent way. The fact is that there is no significant difference between the two rules which is more apparent in Table V.

Computational Complexities of the k–RNN and the k–NN Rules:

Table 7: Average proportion of misclassification by k–RNN and k–NN rules for the pair EX(0, 1) vs. GA(2, 1)

k	1	2	3	4	5	6
n	-RNN	-RNN	-RNN	-RNN	-RNN	-RNN
	-NN	-NN	-NN	-NN	-NN	-NN
10	0.3540	0.3527	0.3512	0.3535	0.3567	0.4302
	0.3600	0.3552	0.3575	0.3637	0.3610	0.4505
20	0.3702	0.3282	0.3312	0.3312	0.3262	0.3422
	0.3770	0.3467	0.3365	0.3415	0.3350	0.3315
50	0.3920	0.3438	0.3347	0.3232	0.3147	0.3182
	0.4030	0.3867	0.3485	0.3627	0.3380	0.3345
100	0.3860	0.3575	0.3410	0.3310	0.3255	0.3250
	0.3845	0.3837	0.3725	0.3665	0.3525	0.3445
500	0.3740	0.3497	0.3397	0.3292	0.3277	0.3252
	0.3690	0.3717	0.3530	0.3470	0.3370	0.3425
1000	0.3760	0.3475	0.3372	0.3335	0.3287	0.3212
	0.3605	0.3710	0.3525	0.3582	0.3440	0.3490

EX(a, b) Exponential distribution with location parameter a and scale parameter b.
GA(a, b) Gamma distribution with shape parameter a and scale parameter b.

Suppose there are N, $N = n_1 + n_2$, observations in the training data, then the cost of finding the rank of Z is equal to $logN$. On the other hand, for the conventional k–NN rule, the cost of finding k-neighbors is $(N - K)(N - 2)\dots(N - k)$ comparisons and distance calculations involving a cost of operations as each square distance requires 1 subtraction and 1 multiplication, which is much higher than that of the k–RNN rule.

4 Concluding Remarks

In this investigation we have done an extensive simulation study to compare the k–RNN and the k–NN rules in univariate cases. We observe that the k–RNN rule performs significantly better than the k–NN rule in majority of the cases. We also note superior performance of the k–RNN rule when there is larger overlap between two populations. The k–RNN rule is based on ranks of X's, Y's, and Z's, so it has better robustness properties. Also, the k–RNN lessens the burden of calculating the distances of all observations X's and Y's from Z as they are needed for the conventional the k–NN rule. Thus, the computational complexity for the k–RNN rule is much less than the k–NN rule. Users of the k–NN rule may find the k–RNN rule as a computationally simpler alternative to the k–NN rule. The k–RNN rule is particularly useful when the observations are available in terms of their ranks. The comparison of the k–RNN rule and k–NN rule under multivariate data is under investigation and findings of this will be reported later.

References

[1] Anderson, T.W. (1966). Some nonparametric multivariate procedures based on statistical equivalent blocks. In *Proceedings of the 1st International Symposium on Multivariate Analysis, Ed., P.R. Krishnaiah*, Academic Press, New York.

[2] Bagui, S.C. and Pal, N.R. (1995). A multistage generalization of the rank nearest neighbor classification rule, *Pattern Recognition Letter,* **16***, 601-614.*

[3] Bagui, S.C. and Vaughn, B. (1998). Statistical classification based on k-rank nearest neighbor rule, *Statistics and Decisions,* **16***, 181-189.*

[4] Bagui, S.C., Bagui, S.S., Pal, K., and Pal, N.R. (2003). Breast cancer detection using rank nearest neighbor classification rules, *Pattern recognition,* **36***, 25-34.*

[5] Cover, T.M. and Hart, P.E. (1967). Nearest neighbor pattern classification, *IEEE Transformation on Information Theory, 13, 21-26.*

[6] Dasgupta, S. and Lin, H.E. (1980). Nearest neighbor rules for statistical classification based on ranks, *Sankhya A,* **42***, 219-230.*

[7] Devijver, P.A. and Kittler, J. (1982). *Pattern Recognition: A Statistical Approach*, Prentice-Hall, Englewood Cliffs, NJ.

[8] Devroye, L. (1981a). On the asymptotic probability of error in nonparametric discrimination, *Annals of Statistics,* **9***, 1320-1327.*

[9] Devroye, L. (1981b). On the almost everywhere convergence of nonparametric regression function estimates, *Annals of Statistics,* **9***, 1310-1319.*

[10] Fix, E. and Hodges, J.L. (1951). Nonparametric discrimination: Consistency Properties. *U.S. Air Force School of Aviation Medicine,* **Report No. 4***, Randolph Field, Texas, USA.*

[11] Fritz, J. (1975). Distribution-free exponential error bound for nearest neighbor pattern classification, *IEEE Transformation on Information Theory,* **21***, 552-557.*

[12] Silverman, B.W. and Jones, M.C. (1989). E. Fix and Hodges (1951): an important contribution to nonparametric discriminant analysis and density estimation, *International Statistical Review,* **57***, 233-247.*

[13] Wagner, T.J. (1971). Convergence of the nearest neighbor rule, *IEEE Transformation on Information Theory,* **17***, 566-571.*

[14] Xiru, C. (1985). Exponential posterior error bound for the k–NN discrimination rule, *Scientia Sinicia A,* **XXVIII***, 673-682.*