ISSN 2347-1921

Upper Bounds of the risk of nearest Neighbor Rules for Finite Samples

Mohamed M. Rizk
Mathematics & Statistics Department
Faculty of Science, Taif University, Taif, Saudi Arabia
Permanent Address: Mathematics Department,
Faculty of Science, Menoufia University, Shebin El-Kom, Egypt
mhm96@yahoo.com
Azhari A. Elhag
Mathematics & Statistics Department
Faculty of Science, Taif University, Taif, Saudi Arabia
azhri_elhag@hotmai.com

ABSTRACT

In this paper, based on the integral expressions for the conditional risk $R_m(x)$ and the risk R_m, we evaluate the upper bounds of the risk for some important distributions with unbounded support $S = (0, \infty)$.

Keywords:

Nearest neighbor rules; finite sample risk; mixture distributions.

Council for Innovative Research
Peer Review Research Publishing System
Journal: Journal of Advances in Mathematics
Vol 7, No. 1
editor@cirworld.com
www.cirworld.com, member.cirworld.com

1 INTRODUCTION

The nearest neighbor rules method is widely used in pattern recognition. This method is useful both for classification and for estimation of density functions and is one of the nonparametric techniques. Many researches are studied for bounded supports; see Dasarathy [3], Fukunaga and Hummels [8], Psaltis et al. [11], and Snapp and Venkatesh [12]. Cover and Hart [1] proved that $R^* \le R_\infty \le 2R^*(1-R^*)$ under certain conditions on the distribution, where R^* denotes the Bayes error (the minimum probability of error over all decision rules) and R_∞ is the nearest neighbor risk in the infinite-sample limit. If $R^* \ll 1$ then the nearest neighbor classifier is nearly in the infinite sample limit. In practice, however, the sample must be finite. Furthermore, data storage and access costs favor small samples. So we are interest to the upper bounds of the finite sample risk R_m and how rapidly does the statistical risk R_m of a nearest neighbor rule approach its infinite-sample limit R_∞ in the case of unbounded support for which we find asymptotic expansion for the finite sample risk R_m for special distributions having unbounded support $S=(0,\infty)$.

The nearest neighbor rule was first studied by Fix and Hodges [5], [6]. Cover [2] has shown that $R_m = R_\infty + O(m^{-2})$ for the nearest neighbor classifier in the case one-dimensional bounded support, mixture density $f \ge c > 0$, and under some additional conditions. Wagner [14] and Fritz [7] treated convergence of the conditional error rate for nearest neighbor. Fukunaga and Hummels [8] studied the rate of convergence of the above bias in d-dimensional feature space. Psaltis et al. [11] generalized the results of Cover [2] to general dimension, and Snapp and Venkatesh [12] further extended the results to the case of multiple classes. Irle and Rizk [10] found an asymptotic evaluation of the conditional risk $R_m(x)$ (the probability of error conditioned on the event that $X=x$) by using partial integration and Laplace's method. There is a wealth of consistency results in different directions available for nearest neighbor rules; see the collection of Dasarathy [3], the monographs by Devroye et al. [4], Györfi et al. [9], and Theodoridis, Koutroumbas [14],[15].

1.1 Nearest neighbor procedure

The neatest neighbor decision rule is one of the simplest types of nonparametric methods of interest in statistical pattern recognition that can be used with arbitrary distributions and without the assumption that the forms of the underling densities are known.

Let $(X^{(1)},\theta^{(1)}),(X^{(2)},\theta^{(2)}),\dots,(X^{(m)},\theta^{(m)})$ be independent distributed random variables taking values in $R^d \times \{1,2,\dots,C\}$, and let (X,θ) be another independent sample of the same distribution, such that X is an observed pattern and it is desired to estimate θ. The neatest neighbor rule assigns X to a class $\theta^{(i)}$ if

$$\|X-X^{(i)}\| \le \|X-X^{(j)}\|, \quad \text{for all } i \ne j,$$

That is, we will call $X' \in \{X^{(1)},X^{(2)},\dots,X^{(m)}\}$ a nearest neighbor to input pattern X if

$$\min_{1\le i\le m} \|X-X^{(i)}\| = \|X-X'\|$$

In case of a tie, the candidate with the smaller index is said to closer to X, that is if $\|X-X^{(i)}\| = \|X-X^{(j)}\|$, and $i<j$, we choose $X^{(i)}$ as a nearest neighbor of X and assign it to a class $\theta^{(i)}$.

In this paper, we suppose that the class-conditional distributions F_l are absolutely continuous with corresponding densities f_l, for each $l \in \{1,2\}$. Let $f = p_1 f_1 + p_2 f_2$ denote the mixture density, where $p_1+p_2=1,\ 0<p_l<1$ for all l, $\rho = |x'-x|$ is the distance between two points x' and x, and let S be its support in R.

1.2 The finite sample risk

The neatest neighbor rule decides X belong to the class θ' of its nearest neighbor X'. A mistake is made if $\theta' \ne \theta$. Notice that the neatest neighbor rule utilizes only the classification of the nearest neighbor. The $m-1$ remaining classifications $\theta^{(i)}$ are ignored. The risk of the nearest neighbor procedure from a training sequence of size m is defined by $R_m = P(\theta' \ne \theta)$.

The finite-sample risk R_m can be written in integral form (see [10],[11]), The conditional probability of error for the nearest neighbor rule is defined as the probability of error in classification θ by θ' given X and its nearest neighbor X', and denoted by $P(\theta \ne \theta' | X, X')$. By averaging $P(\theta \ne \theta' | X, X')$ over X', we obtain the probability of error conditioned on the event that $X=x$ and denoted by $R_m(X)$, such that

$$\begin{aligned} R_m(X) &= P(\theta \ne \theta' | X = x), \\ &= \frac{p_1p_2f_1(x)}{f(x)}\int_S m\, f_2(x')\,[P(|X-x|>|x'-x|)]^{m-1}\,dx' + \frac{p_1p_2f_2(x)}{f(x)}\int_S m\, f_1(x')\,[P(|X-x|>|x'-x|)]^{m-1}\,dx' \\ &= \frac{p_1p_2f_1(x)}{f(x)} I + \frac{p_1p_2f_2(x)}{f(x)} J, \qquad (1.1) \end{aligned}$$

where

$$I = I(x) = \int_S m f_2(x')\,[P(|X-x|>|x'-x|)]^{m-1} dx', \qquad (1.2)$$

$$J = J(x) = \int_S m f_1(x') \, [P(|X - x| > |x' - x|)]^{m-1} dx', \tag{1.3}$$

and by averaging $P(\theta \neq \theta' | X = x)$ with respect to X, we obtain the finite-sample risk R_m (the unconditional probability of error) in the form:

$$R_m = P(\theta' \neq \theta) = \int_S P(\theta' \neq \theta | X = x) f(x) dx \tag{1.4}$$

$$= p_1 p_2 \int_S \int_S m \left(P(|X - x| > |x' - x|)\right)^{m-1} \cdot (f_1(x) f_2(x') + f_1(x') \, f_2(x)) dx' dx. \tag{1.5}$$

2. The main result

In this section, we evaluate the probability of error conditioned on the event that $X = x$ for a two-class pattern recognition problem for support $S = (0, \infty)$, and we derive the finite-sample risk R_m.

2.1 Theorem

Let $x \in S$. Assume that the densities f_l are differentiable for $l = 1,2$, $f(x) > 0$, and $f(x - \rho) + f(x + \rho) > 0$ for $\rho > 0$. Define

$$q_0(x, \rho) = \frac{f_2(x-\rho) + f_2(x+\rho)}{f(x-\rho) + f(x+\rho)}, \; q_1(x, \rho) = \frac{q_0'(x,\rho)}{f(x-\rho) + f(x+\rho)}, \; \bar{q}_0(x, \rho) = \frac{f_1(x-\rho) + f_1(x+\rho)}{f(x-\rho) + f(x+\rho)} \text{ and } \bar{q}_1(x, \rho) = \frac{\bar{q}_0{}'(x,\rho)}{f(x-\rho) + f(x+\rho)}.$$

Then

$$R_m \leq R_\infty + \frac{1}{2m}\left(1 - e^{-2m)}\right) - p_1 p_2 \int_0^\infty \left([f_1(x) q_0(x, x) + f_2(x) \bar{q}_0(x, x)] [1 - \left(F(2x)\right)]^m \right) dx$$
$$+ \frac{p_1 p_2}{(m+1)} \int_0^\infty \left([f_1(x) q_1(x, x) + f_2(x) \bar{q}_1(x, x)] \left[1 - \left[1 - \left(F(2x)\right)\right]^{m+1}\right] \right) dx$$

where $R_\infty = \int_0^\infty \frac{2 p_1 p_2 f_1(x) f_2(x)}{f(x)} dx$, that was first derived by Cover and Hart [1].

Proof.

Firstly, we evaluate the asymptotic expansions for I and J in (1.1).

From equation (1.2), we have

$$I = I(x) = m \int_0^\infty f_2(x') \; [P(|X - x| > |x' - x|)]^{m-1} \, dx'$$
$$= m \int_0^x f_2(x') \, [P(|X - x| > |x' - x|)]^{m-1} \, dx'$$
$$+ m \int_x^\infty f_2(x') \, [P(|X - x| > |x' - x|)]^{m-1} \, dx'$$
$$= m \int_0^x f_2(z) \left[P(X < z) + P\left(X > x + (x - z)\right)\right]^{m-1} dz$$
$$+ m \int_x^\infty f_2(z) \left[P(X > z) + P\left(X < x - (z - x)\right)\right]^{m-1} dz$$
$$= m \int_0^x f_2(x - \rho) \, [P(X < x - \rho) + P(X > x + \rho)]^{m-1} \, d\rho$$
$$+ m \int_0^\infty f_2(x + \rho) \, [P(X > x + \rho) + P(X < x - \rho)]^{m-1} \, d\rho$$
$$= m \int_0^x \left(f_2(x - \rho) + f_2(x + \rho)\right). [P(X < x - \rho) + P(X > x + \rho)]^{m-1} d\rho$$
$$+ m \int_x^\infty f_2(x + \rho) \left(P(X > x + \rho)\right)^{m-1} d\rho = I' + I'', \tag{2.1}$$

where

$$I' = m \int_0^x \left(f_2(x - \rho) + f_2(x + \rho)\right). [P(X < x - \rho) + P(X > x + \rho)]^{m-1} d\rho \tag{2.2}$$

$$I'' = m \int_x^\infty f_2(x + \rho) \left(P(X > x + \rho)\right)^{m-1} \tag{2.3}$$

Similarly,

$$J = J(x) = \int_S m f_1(x') \, [P(|X - x| > |x' - x|)]^{m-1} dx'$$
$$= m \int_0^x \left(f_1(x - \rho) + f_1(x + \rho)\right). [P(X < x - \rho) + P(X > x + \rho)]^{m-1} d\rho$$

$$+ m\int_x^\infty f_1(x+\rho)\big(P(X>x+\rho)\big)^{m-1}\,d\rho = J' + J'', \qquad (2.4)$$

where

$$J' = m\int_0^x \big(f_1(x-\rho)+f_1(x+\rho)\big).\,[P(X<x-\rho)+P(X>x+\rho)]^{m-1}d\rho, \qquad (2.5)$$

$$J'' = m\int_x^\infty f_1(x+\rho)\big(P(X>x+\rho)\big)^{m-1}\,d\rho. \qquad (2.6)$$

Now, we evaluate I' and I'',

$$I' = m\int_0^x \big(f_2(x-\rho)+f_2(x+\rho)\big).\,[P(X<x-\rho)+P(X>x+\rho)]^{m-1}d\rho$$

$$= -\int_0^x \frac{f_2(x-\rho)+f_2(x+\rho)}{f(x-\rho)+f(x+\rho)}\,\frac{d}{d\rho}[P(X<x-\rho)+P(X>x+\rho)]^{m}d\rho$$

$$= -\int_0^x q_0(x,\rho)\,\frac{d}{d\rho}[P(X<x-\rho)+P(X>x+\rho)]^{m}d\rho,$$

where $q_0(x,\rho) = \frac{f_2(x-\rho)+f_2(x+\rho)}{f(x-\rho)+f(x+\rho)}$ (2.7)

Then, by partial integration

$$I' = \int_0^x q_0'(x,\rho)\,[P(X<x-\rho)+P(X>x+\rho)]^{m}d\rho - \Big[q_0(x,\rho)\big(P(X<x-\rho)+P(X>x+\rho)\big)^m\Big]_0^x$$

$$= q_0(x,0) + \int_0^x q_0'(x,\rho)\,[P(X<x-\rho)+P(X>x+\rho)]^{m}d\rho$$

$$= q_0(x,0) - q_0(x,x)\cdot\big(P(X>2x)\big)^m + I_1', \qquad (2.8)$$

where $I_1' = I_1'(x) = \int_0^x q_0'(x,\rho)\,[P(X<x-\rho)+P(X>x+\rho)]^{m}d\rho$.

Now we evaluate I_1', and write I_1' in the form

$$I_1' = \frac{-1}{m+1}\int_0^x \frac{q_0'(x,\rho)}{f(x-\rho)+f(x+\rho)}\,\frac{d}{d\rho}[P(X<x-\rho)+P(X>x+\rho)]^{m+1}d\rho$$

$$= \frac{-1}{m+1}\int_0^x q_1(x,\rho)\,\frac{d}{d\rho}[P(X<x-\rho)+P(X>x+\rho)]^{m+1}d\rho,$$

where $q_1(x,\rho) = \frac{q_0'(x,\rho)}{f(x-\rho)+f(x+\rho)}$. (2.9)

We integrate by parts with

$u = q_1(x,\rho),\quad dv = \frac{d}{d\rho}[P(X<x-\rho)+P(X>x+\rho)]^{m+1}d\rho$, then

$$I_1' = \frac{1}{m+1}\int_0^x q_1'(x,\rho)\,[P(X<x-\rho)+P(X>x+\rho)]^{m+1}d\rho - \frac{1}{m+1}\Big[q_1(x,\rho)\big(P(X<x-\rho)+P(X>x+\rho)\big)^{m+1}\Big]_0^x$$

$$= \frac{I_2'}{m+1} + \frac{q_1(x,0)}{m+1} - \frac{q_1(x,x)}{m+1}\big(P(X>2x)\big)^{m+1}, \qquad (2.10)$$

where $I_2' = \int_0^x q_1'(x,\rho)\,[P(X<x-\rho)+P(X>x+\rho)]^{m+1}d\rho$.

From (2.8) and (2.10) we obtain I' in the following form

$$I' = q_0(x,0) - q_0(x,x)\cdot\big(P(X>2x)\big)^m + \frac{q_1(x,0)}{m+1} - \frac{q_1(x,x)}{m+1}\big(P(X>2x)\big)^{m+1} + \frac{I_2'}{m+1} \qquad (2.11)$$

Evaluating I_2':

$$I_2' = \int_0^x q_1'(x,\rho)\,[P(X<x-\rho)+P(X>x+\rho)]^{m+1}d\rho$$

$\le \int_0^x q_1'(x,\rho)d\rho = [q_1(x,\rho)]_0^x = q_1(x,x) - q_1(x,0)$, then

$$I' \le q_0(x,0) - q_0(x,x)\cdot\big(P(X>2x)\big)^m + \frac{q_1(x,x)}{m+1}\Big[1-\big(P(X>2x)\big)^{m+1}\Big] \qquad (2.12)$$

Now, we evaluate I'',

$$I'' = m\int_x^\infty f_2(x+\rho)\big(P(X > x+\rho)\big)^{m-1}\,d\rho$$

$$\le m\int_x^\infty \frac{1}{p_2} f(x+\rho)\big[1-\big(F(x+\rho)\big)\big]^{m-1}\,d\rho$$

$\le \frac{-1}{p_2}\big[1-\big(F(x+\rho)\big)\big]^m\Big|_x^\infty = \frac{1}{p_2}\big[1-\big(F(2x)\big)\big]^m$, then

$$I'' \le \frac{1}{p_2}\big[1-\big(F(2x)\big)\big]^m. \tag{2.13}$$

Substituting (2.12) and (2.13) in to (2.1) we obtain

$$I = I' + I'' \le q_0(x,0) - q_0(x,x)\cdot\big(P(X > 2x)\big)^m + \frac{q_1(x,x)}{m+1}\Big[1-\big(P(X>2x)\big)^{m+1}\Big] + \frac{1}{p_2}\big[1-\big(F(2x)\big)\big]^m.$$

$$= q_0(x,0) + \left[\frac{1}{p_2} - q_0(x,x)\right]\big[1-\big(F(2x)\big)\big]^m + \frac{q_1(x,x)}{m+1}\Big[1-\big[1-\big(F(2x)\big)\big]^{m+1}\Big], \tag{2.14}$$

where $q_0(x,0) = \frac{f_2(x)}{f(x)}$, $q_0(x,x) = \frac{f_2(0)+f_2(2x)}{f(0)+f(2x)}$, and $q_1(x,x) = \frac{q_0'(x,x)}{f(0)+f(2x)}$.

Similarly, we can show that

$$J = J' + J'' \le \bar q_0(x,0) + \left[\frac{1}{p_1} - \bar q_0(x,x)\right]\big[1-\big(F(2x)\big)\big]^m + \frac{\bar q_1(x,x)}{m+1}\Big[1-\big[1-\big(F(2x)\big)\big]^{m+1}\Big], \tag{2.15}$$

where $\bar q_0(x,\rho) = \frac{f_1(x-\rho)+f_1(x+\rho)}{f(x-\rho)+f(x+\rho)}$, $\bar q_1(x,\rho) = \frac{\bar q_0{}'(x,\rho)}{f(x-\rho)+f(x+\rho)}$,

$\bar q_0(x,0) = \frac{f_1(x)}{f(x)}$, $\bar q_0(x,x) = \frac{f_1(0)+f_1(2x)}{f(0)+f(2x)}$, and $\bar q_1(x,x) = \frac{\bar q_0{}'(x,x)}{f(0)+f(2x)}$.

Substituting (2.14) and (2.15) in to (1.1) we obtain

$$R_m(x) \le \frac{2p_1p_2f_1(x)f_2(x)}{f^2(x)} + \left[1 - \frac{p_1p_2f_1(x)}{f(x)}q_0(x,x) - \frac{p_1p_2f_2(x)}{f(x)}\bar q_0(x,x)\right]\big[1-\big(F(2x)\big)\big]^m$$
$$+ \frac{p_1p_2\Big[1-\big[1-\big(F(2x)\big)\big]^{m+1}\Big]}{(m+1)f(x)}\Big[f_1(x)q_1(x,x)+f_2(x)\bar q_1(x,x)\Big] \tag{2.16}$$

Now, we evaluate R_m, by substituting (2.16) in (1.4) we obtain

$$R_m = P(\theta' \ne \theta) \le \int_0^\infty \left(\frac{2p_1p_2f_1(x)f_2(x)}{f^2(x)}\right) f(x)\,dx$$
$$+ \int_0^\infty \left(\left[1 - \frac{p_1p_2f_1(x)}{f(x)}q_0(x,x) - \frac{p_1p_2f_2(x)}{f(x)}\bar q_0(x,x)\right]\big[1-\big(F(2x)\big)\big]^m\right) f(x)\,dx$$
$$+ \int_0^\infty \left(\frac{p_1p_2\Big[1-\big[1-\big(F(2x)\big)\big]^{m+1}\Big]}{(m+1)f(x)}\,[f_1(x)q_1(x,x)+f_2(x)\bar q_1(x,x)]\right) f(x)\,dx$$
$$= \int_0^\infty \frac{2p_1p_2f_1(x)f_2(x)}{f(x)}\,dx + \int_0^\infty \big[1-\big(F(2x)\big)\big]^m f(x)\,dx$$
$$-p_1p_2\int_0^\infty \Big([f_1(x)q_0(x,x) + f_2(x)\bar q_0(x,x)]\big[1-\big(F(2x)\big)\big]^m\Big)\,dx$$
$$+\frac{p_1p_2}{(m+1)}\int_0^\infty \Big([f_1(x)q_1(x,x)+f_2(x)\bar q_1(x,x)]\Big[1-\big[1-\big(F(2x)\big)\big]^{m+1}\Big]\Big)\,dx$$
$$\le \int_0^\infty \frac{2p_1p_2f_1(x)f_2(x)}{f(x)}\,dx + \int_0^\infty f(x).e^{-2mF(2x)}\,dx$$
$$-p_1p_2\int_0^\infty \Big([f_1(x)q_0(x,x) + f_2(x)\bar q_0(x,x)]\big[1-\big(F(2x)\big)\big]^m\Big)\,dx$$
$$+\frac{p_1p_2}{(m+1)}\int_0^\infty \Big([f_1(x)q_1(x,x)+f_2(x)\bar q_1(x,x)]\Big[1-\big[1-\big(F(2x)\big)\big]^{m+1}\Big]\Big)\,dx,$$

since $F(2x) \ge F(x) \Rightarrow e^{-2mF(2x)} \le e^{-2mF(x)}$, then we have

$$\int_0^\infty f(x).e^{-2mF(2x)}\,dx \le \int_0^\infty f(x).e^{-2mF(x)}\,dx = \frac{-1}{2m}e^{-2mF(x)}\Big|_0^\infty = \frac{1}{2m}\big(1-e^{-2m}\big).$$

Then

$$R_m \le R_\infty + \frac{1}{2m}\left(1 - e^{-2m)}\right) - p_1 p_2 \int_0^\infty \left([f_1(x) q_0(x,x) + f_2(x)\bar{q}_0(x,x)]\left[1 - \left(F(2x)\right)\right]^m\right) dx$$

$$+ \frac{p_1 p_2}{(m+1)} \int_0^\infty \left([f_1(x) q_1(x,x) + f_2(x)\bar{q}_1(x,x)] \left[1 - \left[1 - \left(F(2x)\right)\right]^{m+1}\right]\right) dx$$

We can apply this theorem for some important distributions having unbounded support $S = (0, \infty)$.

References

[1] Cover, T. M. and Hart, P. E. (1967). Nearest neighbor pattern classification. IEEE Transactions on Information Theory, 13: 21-27.

[2] Cover, T. M. (1968b). Rates of convergence for nearest neighbor procedures. In Proceedings of the Hawaii International Conference on Systems Sciences, pages 413-415. Honolulu, HI.

[3] Dasarathy, B. (1991). Nearest neighbor classification techniques. IEEE, Los Alamitos, CA.

[4] Devroye, L., Györfi, L., and Lugosi, G. (1996). Probabilistic theory of pattern recognition. Springer-Verlag, New York.

[5] Fix, E., Hodges, J. (1991). Discriminatory analysis, nonparametric discrimination, consistency properties. in nearest neighbor classification techniques, Dasarathy, B., editor, pages 32-39, IEEE Computer Society Press, Los Alamitos, CA.

[6] Fix, E., Hodges, J. (1991). Discriminatory analysis, small sample performance. in nearest neighbor classification techniques, Dasarathy, B., editor, pages 40-56, IEEE Computer Society Press, Los Alamitos, CA.

[7] FRITZ, J. (1975). Distribution-free exponential error bound for nearest neighbor pattern classi_cation. IEEE Trans. Inform. Theory 21, 552-557.

[8] Fukunaga, K. and Hummels, D. M. (1987). Bias of nearest neighbor estimates. IEEE Trans. Pattern Anal. Mach. Intell., 9: 103-112.

[9] Györfi, L., Kohler, M., Krzyżak, A. and Walk, H. (2002). A distribution-free theory of nonparametric regression, Springer-Verlag, New York.

[10] Irle, A. Rizk, M. (2004). On the risk of nearest neighbor rules. Dsseration zur Erlangung des Doktorgrades.

[11] Psaltis, D., Snapp, R, R. and Venkatesh, S. S. (1994). On the finite sample performance of the nearest neighbor classifier. IEEE Transaction on Information Theory, 40: 820-837.

[12] Snapp, R. R. and Venkatesh, S. S. (1998). Asymptotic expansion of the k nearest neighbor risk. Ann. Statist. 26: 850-878.

[13] WAGNER, T.J. (1971). Convergence of the nearest neighbor rule. IEEE Trans. Inform. Theory 17, 566-571.

[14] Theodoridis S., Koutroumbas K. (2009) *Pattern Recognition*, 4th ed., Academic Press.

[15] Theodoridis S., Koutroumbas K. (2010). An introduction to pattern recognition: A mat lab approach. Academic Press.