[Information Sciences Letters](https://digitalcommons.aaru.edu.jo/isl)

Volume 11
Issue 6 *Nov. 2022*

Article 10

2022

Some Contributions to the Risk of the Nearest Neighbor Rulesv

M. M. Rizk
Mathematics & Computer Science Department, Faculty of Science, Menoufia University, Shebin El-Kom, Egypt, mhm96@yahoo.com

Information Sciences Letters
An International Journal

http://dx.doi.org/10.18576/isl/110610

Some Contributions to the Risk of the Nearest Neighbor Rules

M. M. Rizk

Mathematics & Computer Science Department, Faculty of Science, Menoufia University, Shebin El-Kom, Egypt

Received: 11 Sep. 2021, Revised: 2 Nov. 2021, Accepted: 11 Dec. 2021
Published online: 1 Nov. 2022

Abstract: Upper bounds are derived for the finite-sample risk R_m for distributions having unbounded support, for which we found upper bounds on the expected nearest neighbor distance. We look at real-valued observations and remarks on the multidimensional case. The upper bounds of some distributions are set as typical.

Keywords: Completely monotone function, expected nearest neighbor distance, moment generating function, nearest neighbor rule, pattern recognition.

1 Introduction

In statistical pattern recognition, we suppose that an independent identically distributed (i.i.d.) training sequence $\left(X^{(1)},\theta^{(1)}\right),\left(X^{(2)},\theta^{(2)}\right),\cdots,\left(X^{(m)},\theta^{(m)}\right)$ taking values in $R^d \times \{1,2,...,C\}$ is at our disposal. Given a new random variable (X,θ), such that $X \in R^d$ is an observation, our goal is to predict its corresponding unobservable class θ, this class takes values in a finite set $\{1,2,...,C\}$. Let δ be a function (classifier) from R^d to $\{1,2,...,C\}$, the probability of misclassification is $P(\delta(X) \neq \theta)$.

If the joint distribution of (X,θ) is known, we get the Bayes classifier δ^* (The minimal probability of error) which is the best classifier (see Devroye et al. [1], Györfi et al. [2]). Mostly, the joint distribution of (X,θ) will be unknown, and based on the training sequence $D_m = \left(\left(X^{(1)},\theta^{(1)}\right),\left(X^{(2)},\theta^{(2)}\right),\cdots,\left(X^{(m)},\theta^{(m)}\right)\right)$, which consists of n i.i.d. random pairs with the same distribution as (X,θ). The object is to minimize the finite-sample risk R_m (unconditional probability of error).

The supervised pattern recognition model is considered the core of many contributions to the statistical literature in recent years, the nearest neighbor rule is a popular and simple method to classify patterns in non-parametric situations; it was first originally suggested and studied by Fix and Hodges [3],[4], and several results in different directions are available, e.g., Dasarathy [5], Devroye et al. [1], Györfi et al. [2], Biau and Devroye [6], and Zhao and Lai [7].

In addition, there is a wealth of studies for the error estimation of the nearest neighbor rules, e.g., Cover and Hart [8], Cover [9], Fukunaga and Hummels [10], Psaltis et al., [11], Snapp and Venkatesh [12], Irle and Rizk [13]. Moreover, there exists a rather large amount of literature on nearest neighbor distances, e.g., Kulkarni and Posner [14], Evans et al. [15], and Liitiäinen et al. [16].

In this paper, we study the classification of a random variable θ taking values in $\{0,1\}$ given a sample X in R^d with metric ρ, which we denote the pair as (R^d,ρ). The object is to find upper bounds on the expected nearest neighbor distance for the distributions that have unbounded support, and derive bounds for the finite-sample risk R_m in terms of the expected nearest neighbor distance, we then estimate upper bounds for some distributions. We look at real-valued observations and study the multidimensional case.

2 Nearest Neighbor Classification

Given any i.i.d. training sequence D_m and (X,θ) another independent sample of the same distribution, such that X

* Corresponding author e-mail: mhm96@yahoo.com

is an observed pattern and our goal is to guess its corresponding θ. By using suitable tie-breaking, the nearest neighbor rule assigns X to a class $\theta^{(i)}$ with the property $\left\|X-X^{(i)}\right\| \leq \left\|X-X^{(j)}\right\|$ for all $i \neq j$.

In the following, let the class-conditional distributions F_l be 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 and let S be its support in R^d. By $B(\varepsilon,x) = \left\{x' \in R^d : \|x-x'\| \leq \varepsilon\right\}$ we will denote the closed ball of radius ε at x.
Assume that the conditional mean of θ given $X=x$ is defined as

$$m(x) = P(\theta = 1|X=x) = E(\theta|X=x),$$

and the conditional variance as

$$\sigma^2(x) = P(\theta=1|X=x) - [P(\theta=1|X=x)]^2 \\ = m(x) - (m(x))^2.$$

The finite-sample risk R_m can be written in integral form as following, compare [13] and [14]:

Let X' denote the nearest neighbor feature vector in the training sequence D_m, and let θ' be the class label associated with X'. Assume $R_m(X,X')$ refers to the conditional probability of error for the nearest neighbor rule which is defined as the probability of misclassification θ by θ' given X and its nearest neighbor X', then

$$\begin{aligned} & R_m\left(X,X'\right) = P\left(\theta \neq \theta'|X,X'\right) \\ & = P\left(\theta=1,\theta'=0|X,X'\right) + P\left(\theta=0,\theta'=1|X,X'\right) \\ & = P(\theta=1|X)P\left(\theta'=0|X'\right) + P(\theta=0|X)P\left(\theta'=1|X'\right) \end{aligned} \tag{1}$$

The m-sample conditional average probability of error $R_m(X)$ is given by averaging $P(\theta \neq \theta'|X,X')$ with respect to X', thus
$R_m(X) = P(\theta \neq \theta'|X) = \int_S P(\theta \neq \theta'|X,X') f_m(x'|x)\,dx'$,
hence, averaging $P(\theta \neq \theta'|X)$ with respect to X to get the unconditional probability of error (the finite-sample risk R_m), then

$$\begin{aligned} R_m &= P\left(\theta \neq \theta'\right) = \int_S P\left(\theta \neq \theta'|X\right) f(x)dx, \\ &= \int_S\int_S P\left(\theta \neq \theta'|X,X'\right) f_m\left(x'|x\right) f(x)dx'dx \\ &= m\int_S\int_S P\left(\theta \neq \theta'|X,X'\right) \times \\ & \quad \left(P\left(|X-x| > |x'-x|\right)\right)^{m-1} f(x')f(x)dx'dx, \end{aligned}$$

where $f_m(x'|x)$ refers to the conditional density of X' given $X=x$, and taken the form:

$$\begin{aligned} f_m\left(x'|x\right) &= m\left(1 - P\left(X \in B\left(|x'-x|,x\right)\right)\right)^{m-1} f(x') \\ &= m\left(P\left(|X-x| > |x'-x|\right)\right)^{m-1} f(x'). \end{aligned}$$

In the following, by define $d_m = \rho(X,X')$ as the nearest distance at time m, we derive an upper bound on R_m in terms of d_m. Firstly, we provide an upper bound on $R_m(X,X')$.

Lemma 2.1.

$$R_m\left(X,X'\right) = \sigma^2(X) + \sigma^2\left(X'\right) + \left(m(X) - m(X')\right)^2.$$

Proof. From (1), we have

$$\begin{aligned} R_m\left(X,X'\right) &= P(\theta=1|X)P\left(\theta'=0|X'\right) \\ &+ P(\theta=0|X)P\left(\theta'=1|X'\right) \\ &= m(X)\left(1-m(X')\right) + m\left(X'\right)\left(1-m(X)\right) \\ &= m(X) + m\left(X'\right) - 2m(X)m\left(X'\right) \\ &= \left[m(X) - m(X)^2\right] + \left[m(X') - m(X')^2\right] \\ &+ m(X)\left[m(X) - m(X')\right] \\ &- m\left(X'\right)\left[m(X) - m(X')\right] \\ &= \sigma^2(X) + \sigma^2\left(X'\right) + \left(m(X) - m(X')\right)^2. \end{aligned}$$

Assumption 1. For some $\lambda_1 > 0$ and $\alpha > 0$, we have $|m(x) - m(x')| \leq \lambda_1 \rho(x,x')^\alpha$ for all $x,x' \in R^d$.

Corollary 2.2. For some appropriate $\rho = max\{\lambda_1, \lambda_1^2\}$ independent of m and under assumption (1), we have

$$R_m\left(X,X'\right) \leq 2\sigma^2(X) + \lambda\left(d_m^\alpha + d_m^{2\alpha}\right). \tag{2}$$

Proof. From lemma 2.1, we have

$$\begin{aligned} R_m\left(X,X'\right) &= 2\sigma^2(X) + \left[\sigma^2\left(X'\right) - \sigma^2(X)\right] \\ &+ \left(m(X) - m(X')\right)^2. \end{aligned}$$

Since

$$\begin{aligned} & \left|\sigma^2\left(X'\right) - \sigma^2(X)\right| \\ & = \left|m\left(X'\right)\left(1-m(X')\right) + m(X)(1-m(X))\right| \\ & \leq \left|m\left(X'\right) - m(X)\right|. \end{aligned}$$

Thus

$$\begin{aligned} & R_m\left(X,X'\right) \\ & \leq 2\sigma^2(X) + \left|m\left(X'\right) - m(X)\right| + \left(m(X) - m(X')\right)^2 \\ & \leq 2\sigma^2(X) + \lambda_1\rho\left(x,x'\right)^\alpha + \lambda_1^2\rho\left(x,x'\right)^{2\alpha} \\ & \leq 2\sigma^2(X) + \lambda\left(d_m^\alpha + d_m^{2\alpha}\right). \end{aligned}$$

Lemma 2.3. Suppose that assumption 1 with $\alpha \leq 1$ holds, then we have

$$R_m \leq R_\infty + \lambda\left[(Ed_m)^\alpha + (Ed_m^2)^\alpha)\right], \tag{3}$$

where R_∞ denotes infinite-sample risk.
Proof. By taking expected values on (2), we obtain

$$R_m \leq R_\infty + \lambda\left[(Ed_m)^\alpha + (Ed_m^2)^\alpha)\right],$$

where $R_\infty = 2E[\sigma^2(X)]$.
Since $h(t) = t^\alpha$ is concave for $0 < \alpha \leq 1$, thus by using Jensen's inequality, we obtain

$$R_m \leq R_\infty + \lambda\left[(Ed_m)^\alpha + (Ed_m^2)^\alpha)\right].$$

3 Upper bounds for nearest neighbor distances

In this section, we find upper bounds on the expected nearest neighbor distance Ed_m, therefore, we can be derive bounds for the finite-sample risk R_m in terms of Ed_m and Ed_m^2 for $x \in R^d$, and estimate upper bounds for some distributions.

Assumption 2. (a) Suppose that $\|X-x\|$ has a finite moment generating function $\psi(t,x) = Ee^{t\|X-x\|}$ for $x \in R^d$ and $0 < t < 1$.
(b) There exists $c > 0$ such that for all x in the support of X and for all $\varepsilon > 0$, we have

$$H(x,\varepsilon) \geq c\varepsilon^d f(x), \tag{4}$$

where $H(x,\varepsilon) = -\log P(\|X-x\| > \varepsilon)$.

Now we derive upper bounds of Ed_m and Ed_m^2 for $x \in R^d$ and $d = 1$. The following Corollary provides a sufficient condition for the validity of the inequality (4) for $d = 1$.

Corollary 3.1. A sufficient condition for $H(x,\varepsilon) \geq c\varepsilon f(x)$ is given by $P(|X-x| \leq \varepsilon) \geq c\varepsilon f(x)$, where $H(x,\varepsilon) = -\log P(|X-x| > \varepsilon)$.
Proof. Since

$$-\log P(|X-x| > \varepsilon) = -\log P(1-|X-x| \leq \varepsilon),$$

and, for all $0 \leq y \leq 1$ we have $-\log P(1-y) \geq y$. Hence

$$-\log P(|X-x| > \varepsilon) \geq P(|X-x| \leq \varepsilon).$$

Consequently, a sufficient condition for $H(x,\varepsilon) \geq c\varepsilon f(x)$ is given by $P(|X-x| \leq \varepsilon) \geq c\varepsilon f(x)$.

Note that, by letting ε tend to ∞, the second condition is violated for unbounded support. Moreover, we have:
1- $P(|X-x| \leq \varepsilon) \geq \varepsilon f(x)$ if $[x,x+\varepsilon]$ ($[x-\varepsilon,x]$) is included in the support of X and f is increasing (decreasing) on $[x,x+\varepsilon]$ ($[x-\varepsilon,x]$), respectively.
2- $P(|X-x| \leq \varepsilon) \geq 2\varepsilon f(x)$ if $[x-\varepsilon,x+\varepsilon]$ is included in the support of X and f is convex on $[x-\varepsilon,x+\varepsilon]$.
3- $P(|X-x| \leq \varepsilon) \geq 2\varepsilon f(x)$ if $S = (0,\infty)$ and the probability density function $f(x)$ has a completely monotonic function, we provide a proof of this statement in Appendix A.

Lemma 3.2. By assumption 2 for $d = 1$, and a constant $\tau \geq 1$, we have

$$Ed_m \leq \tau \int_{-\infty}^{K_1} \psi\left(\frac{1}{\tau m},x\right)^m P^X(dx) + \tau \int_{K_2}^{\infty} \psi\left(\frac{1}{\tau m},x\right)^m P^X(dx) + \frac{(K_2-K_1)}{cm}. \tag{5}$$

where $K_1 = K_1(m)$, $K_2 = K_2(m)$ are constants depending on m such that $-\infty < K_1 \leq 0 \leq K_2 < \infty$.

Proof. We divide the integral form of Ed_m into three parts as follows:

$$\begin{aligned} Ed_m &= \int_{-\infty}^{\infty}\int_0^{\infty} P(|X-x|>\varepsilon)^m d\varepsilon P^X(dx) \\ &= \int_{-\infty}^{K_1}\int_0^{\infty} P(|X-x|>\varepsilon)^m d\varepsilon P^X(dx) \\ &+ \int_{K_2}^{\infty}\int_0^{\infty} P(|X-x|>\varepsilon)^m d\varepsilon P^X(dx) \\ &+ \int_{K_1}^{K_2}\int_0^{\infty} P(|X-x|>\varepsilon)^m d\varepsilon P^X(dx) \\ &= S_1(m)+S_2(m)+S_3(m), say. \end{aligned} \tag{6}$$

Firstly we find upper bounds for $S_1(m)$ and $S_2(m)$: For any $0 < t < 1$, use Markov's inequality and assumption 2 (a) such that $|X-x|$ has a finite moment generating function $\psi(t,x) = Ee^{t|X-x|}$, $x \in R$. Then

$$\begin{aligned} \int_0^{\infty} P(|X-x|>\varepsilon)^m d\varepsilon &= \int_0^{\infty} P\left(e^{t|X-x|} > e^{t\varepsilon}\right)^m d\varepsilon \\ &\leq \int_0^{\infty} \psi(t,x)^m e^{-mt\varepsilon} d\varepsilon \\ &= \frac{1}{mt}\psi(t,x)^m, \end{aligned}$$

thus for $t = \frac{1}{\tau m}$, $\tau \geq 1$ we have

$$\int_0^{\infty} P(|X-x|>\varepsilon)^m d\varepsilon \leq \tau\psi\left(\frac{1}{\tau m},x\right)^m.$$

Therefore

$$S_1(m) \leq \tau \int_{-\infty}^{K_1} \psi\left(\frac{1}{\tau m},x\right)^m P^X(dx), \tag{7}$$

$$S_2(m) \leq \tau \int_{K_2}^{\infty} \psi\left(\frac{1}{\tau m},x\right)^m P^X(dx). \tag{8}$$

For bounding $S_3(m)$, suppose that X has a density $f(x)$, then

$$\begin{aligned} S_3(m) &= \int_{K_1}^{K_2}\int_0^{\infty} P(|X-x|>\varepsilon)^m d\varepsilon P^X(dx) \\ &= \int_{K_1}^{K_2}\int_0^{\infty} e^{-mH(x,\varepsilon)} f(x) d\varepsilon dx, \end{aligned}$$

where $H(x,\varepsilon) = -\log P(|X-x| > \varepsilon)$.
Using assumption 2 (b) for $d = 1$ we obtain

$$\begin{aligned} S_3(m) &\leq \int_{K_1}^{K_2}\int_0^{\infty} e^{-mc\varepsilon f(x)} f(x) d\varepsilon dx \\ &= \int_{K_1}^{K_2} \frac{1}{cm} dx = \frac{(K_2-K_1)}{cm} \end{aligned} \tag{9}$$

Hence, substituting (7)-(9) in (6) we obtain (5).

Example 1: Suppose X has a density function $e^{-x}, x > 0$. So, we take $K_1 = 0$ and $S_1(m)$ vanishes. We have

$$\psi(t,x) = Ee^{t|X-x|} \leq Ee^{tX+tx} = e^{tx}\int_0^{\infty} e^{ty}e^{-y}dy = e^{tx}\frac{1}{1-t},$$

thus for $t=\frac{1}{2m}$ we obtain
$\psi\left(\frac{1}{2m},x\right)^m \le e^{\frac{x}{2}}\left(\frac{1}{1-\frac{1}{2m}}\right)^m = e^{\frac{x}{2}}\left(1+\frac{1}{2m-1}\right)^m$
Consequently, using (8)

$$S_2(m) \le 2\int_{K_2}^{\infty} e^{\frac{-x}{2}}\left(1+\frac{1}{2m-1}\right)^m dx = 4\left(1+\frac{1}{2m-1}\right)^m e^{\frac{-K_2}{2}}.$$

For $K_2 = 2\log m$, it follows

$$S_2(m) \le \frac{4}{m}\left(1+\frac{1}{2m-1}\right)^m = O\left(\frac{1}{m}\right), \quad (10)$$

since $\left(1+\frac{1}{2m-1}\right)^m \to e^{\frac{1}{2}}\,(m\to\infty)$.

For $c=2$, $K_1=0$ and using $K_2=2\log m$ in (9), hence

$$S_3(m) \le \left(\frac{\log m}{m}\right). \quad (11)$$

Substituting (10) and (11) in (5), we obtain

$$Ed_m \le \left(\frac{\log m}{m}\right) + O\left(\frac{1}{m}\right). \quad (12)$$

The validity (4) for $d=1$ is provided for exponential distribution in Appendix B.

Example 2: Suppose X has a density function $\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{\frac{-x^2}{2}}$, $x\in R$, and we take $-K_1=K_2>0$. For $x\in R$, $t>0$, we obtain

$$\psi(t,x) = Ee^{t|X-x|} \le Ee^{|tx|+|tX|} = e^{|tx|}Ee^{|tX|}$$

thus, by using Jensen's inequality for $t=\frac{1}{m}$, we obtain

$$\psi\left(\frac{1}{m},x\right)^m \le e^{|x|}\left(Ee^{\frac{|X|}{m}}\right)^m \le e^{|x|}E\left(e^{\frac{|X|}{m}}\right)^m = e^{|x|}Ee^{|X|} = 2e^{|x|}\int_0^{\infty}\frac{1}{\sqrt{2\pi}}e^{y}e^{\frac{-y^2}{2}}dy = 2e^{|x|}e^{\frac{1}{2}}\int_0^{\infty}\frac{1}{\sqrt{2\pi}}e^{\frac{-(y-1)^2}{2}}dy \le 2e^{|x|}e^{\frac{1}{2}}.$$

Consequently, for $K_2>1$

$$S_1(m)+S_2(m) \le 4e^{\frac{1}{2}}\int_{K_2}^{\infty}e^{x}\phi(x)dx = 4e\int_{K_2}^{\infty}\frac{1}{\sqrt{2\pi}}e^{\frac{-(x-1)^2}{2}}dx = 4eP(X+1\ge K_2) \le \frac{4e}{(K_2-1)}e^{\frac{-(K_2-1)^2}{2}}.$$

Assume $K_2=\sqrt{2\log m}+1$, it follows

$$S_1(m)+S_2(m) \le o\left(\frac{1}{m}\right). \quad (13)$$

For a constant $c^*>0$ (see Appendix C), and using $-K_1=K_2=\sqrt{2\log m}+1$ in (9), hence

$$S_3(m) \le c^*\frac{\sqrt{2\log m}+1}{m}. \quad (14)$$

Substituting (13) and (14) in (5), we obtain

$$Ed_m \le c^*\frac{\sqrt{2\log m}+1}{m} + o\left(\frac{1}{m}\right). \quad (15)$$

The validity (4) for $d=1$ is provided for normal distribution in Appendix C.

So if $-K_1, K_2$ have logarithmic growth as in examples 1, 2, we obtain $S_3(m)=O\left(\frac{1}{m^{\beta}}\right)$ for all $\beta<1$.

Note that, from (12) and (15) there is an additional logarithmic term over the rates for compact support for exponentially decaying tails (e.g., exponential and normal distributions), compare [14]. Indeed, it can show that these upper bounds are fairly tight and examples 1, 2 illustrate that the expected nearest neighbor distance depends on the tails of the distributions.

Lemma 3.3. By assumption 2 for $d=1$, a constant $\tau\ge 1$ and $c_1, c_2>0$, we have

$$Ed_m^2 \le 2\tau^2\int_{-\infty}^{K_1}\psi\left(\frac{1}{\tau m},x\right)^m P^X(dx) + 2\tau^2\int_{K_2}^{\infty}\psi\left(\frac{1}{\tau m},x\right)^m P^X(dx) + \frac{2c(K_2-K_1)}{c^2m^2}. \quad (16)$$

where $K_1=K_1(m)$, $K_2=K_2(m)$ are constants depending on m such that $-\infty<K_1\le 0\le K_2<\infty$.
Proof. Similarly as in the proof of lemma 3.2, we derive upper bounds for Ed_m^2 by dividing it into three parts

$$\begin{aligned} Ed_m^2 &= \int_{-\infty}^{\infty}\int_0^{\infty} 2\varepsilon P(|X-x|>\varepsilon)^m d\varepsilon P^X(dx) \\ &= \int_{-\infty}^{K_1}\int_0^{\infty} 2\varepsilon P(|X-x|>\varepsilon)^m d\varepsilon P^X(dx) \\ &+ \int_{K_2}^{\infty}\int_0^{\infty} 2\varepsilon P(|X-x|>\varepsilon)^m d\varepsilon P^X(dx) \\ &+ \int_{K_1}^{K_2}\int_0^{\infty} 2\varepsilon P(|X-x|>\varepsilon)^m d\varepsilon P^X(dx) \\ &= S_1'(m)+S_2'(m)+S_3'(m), say. \end{aligned} \quad (17)$$

For bounding $S_1'(m)$ and $S_2'(m)$, using assumption 2 (a) such that $|X-x|$ has a finite moment generating function $\psi(t,x)=Ee^{t|X-x|}$ for $x\in R, 0<t<1$.
By Markov's inequality for any $0<t<1$

$$\int_0^{\infty}2\varepsilon P(|X-x|>\varepsilon)^m d\varepsilon = 2\int_0^{\infty}\varepsilon P\left(e^{t|X-x|}>e^{t\varepsilon}\right)^m d\varepsilon \le 2\psi(t,x)^m\int_0^{\infty}\varepsilon e^{-mt\varepsilon}d\varepsilon = \frac{2}{m^2t^2}\psi(t,x)^m,$$

hence for $t = \frac{1}{\tau m}$, $\tau \geq 1$, we have

$$\int_0^\infty 2\varepsilon P(|X-x|>\varepsilon)^m d\varepsilon \leq 2\tau^2 \psi\left(\frac{1}{\tau m},x\right)^m.$$

It follows

$$S_1'(m) \leq 2\tau^2 \int_{-\infty}^{K_1} \psi\left(\frac{1}{\tau m},x\right)^m P^X(dx), \tag{18}$$

$$S_2'(m) \leq 2\tau^2 \int_{K_2}^{\infty} \psi\left(\frac{1}{\tau m},x\right)^m P^X(dx). \tag{19}$$

For bounding $S_3'(m)$, suppose that X has a density $f(x)>0$, then

$$\begin{aligned} S_3'(m) &= \int_{K_1}^{K_2}\int_0^\infty 2\varepsilon P(|X-x|>\varepsilon)^m d\varepsilon P^X(dx) \\ &= \int_0^\infty 2\varepsilon\left(\int_{K_1}^{K_2} P(|X-x|>\varepsilon)^m P^X(dx)\right)d\varepsilon \\ &= \int_0^\infty 2\varepsilon \int_{K_1}^{K_2} e^{-mH(x,\varepsilon)} f(x)dxd\varepsilon, \end{aligned}$$

where $H(x,\varepsilon) = -\log P(|X-x|>\varepsilon)$.
Using assumption 2 (b) for $d=1$, we obtain

$$\begin{aligned} S_3'(m) &\leq \int_0^\infty 2\varepsilon \int_{K_1}^{K_2} e^{-cm\varepsilon f(x)} f(x)dxd\varepsilon \\ &\leq \int_0^\infty 2\varepsilon \int_{K_1}^{K_2} c_1 e^{-c_2 m\varepsilon} dxd\varepsilon \\ &= 2c_1(K_2-K_1)\int_0^\infty \varepsilon e^{-c_2 m\varepsilon} d\varepsilon \\ &= 2c_1(K_2-K_1)\frac{\Gamma(2)}{c_2^2 m^2} = \frac{2c_1(K_2-K_1)}{c_2^2 m^2}, \end{aligned} \tag{20}$$

where $\int f(x)e^{-\alpha f(x)}dx \leq \gamma_1 e^{-\alpha\gamma_2}$ holds for $\alpha>1$, γ_1,γ_2 are positive constants, any probability density functions have bounded support and $f(x)>0$.
Hence, by substituting (18)-(20) in (17) we obtain (16).

Note that, if $\int_{K_1}^{K_2} f(x)^{-1}dx < \infty$, we show that (see appendix D).

$$S_3'(m) \leq \frac{2}{c^2 m^2}\int_{K_1}^{K_2} f(x)^{-1}dx. \tag{21}$$

We can use (21) for exponential and normal distributions by taking suitable values from K_1, K_2 in examples 1, 2.

Theorem 3.4. Let the conditions of lemmas 2.3, 3.2 and 3.3 be satisfied. Then

$$\begin{aligned} R_m \leq R_\infty &+ \lambda\tau(1+2\tau)\int_{-\infty}^{K_1}\psi\left(\frac{1}{\tau m},x\right)^m P^X(dx) \\ &+ \lambda\tau(1+2\tau)\int_{K_2}^{\infty}\psi\left(\frac{1}{\tau m},x\right)^m P^X(dx) \\ &+ \frac{\lambda(K_2-K_1)}{m}\left(\frac{1}{c}+\frac{2c_1}{c_2^2 m}\right). \end{aligned} \tag{22}$$

Proof. This is immediate from the previous results, by taking $\alpha=1$ in lemma 2.3 and after substituting (5) and (16) in (3), we obtain (22).

The finite-sample risk R_m in one-dimensional is estimated for each exponential and normal distributions, respectively, as follows: By the results in examples 1, 2 and theorem 3.4, we obtain

$$\begin{aligned} R_m &\leq R_\infty + \frac{20\lambda}{m}\left(1+\frac{1}{2m-1}\right)^m + \frac{\lambda\log m}{m}\left(1+\frac{4c_1}{c_2^2 m}\right) \\ &= R_\infty + O\left(\frac{1}{m}\right) + \frac{\lambda\log m}{m}\left(1+\frac{4c_1}{c_2^2 m}\right), \end{aligned}$$

and

$$\begin{aligned} R_m &\leq R_\infty + \frac{12e\lambda}{m\sqrt{2\log m}} + \frac{\lambda c^*\sqrt{2\log m}+1}{m}\left(1+\frac{2c_1}{c_2^2 m}\right) \\ &= R_\infty + o\left(\frac{1}{m}\right) + \frac{\lambda c^*\sqrt{2\log m}+1}{m}\left(1+\frac{2c_1}{c_2^2 m}\right), \end{aligned}$$

where λ, c_1, c_2, c^* are positive constants.
Now we find upper bounds on Ed_m and Ed_m^2 for $x \in R^d$ and $d \geq 2$, and derive bounds for the finite-sample risk R_m.

Lemma 3.5. By assumption 2 for $d \geq 2$ and a constant $\tau \geq 1$ and, we have

$$\begin{aligned} Ed_m \leq \tau &\int_{-\infty}^{K_{11}}\psi\left(\frac{1}{\tau m},x\right)^m P^X(dx) \\ &+\tau\int_{K_{22}}^{\infty}\psi\left(\frac{1}{\tau m},x\right)^m P^X(dx) \\ &+\frac{\Gamma(1/d)}{d(mc)^{(1/d)}}\int_{K_{11}}^{K_{22}} f(x)^{1-(1/d)}dx, \end{aligned} \tag{23}$$

where $K_{11}=K_{11}(m), K_{22}=K_{22}(m)$ are constants depending on m such that $-\infty < K_{11} \leq 0 \leq K_{22} < \infty$.
Proof. Similarly as the proof of lemma 3.2, we derive upper bounds of Ed_m as follows:

$$\begin{aligned} Ed_m &= \int_{-\infty}^{\infty}\int_0^\infty P(\|X-x\|>\varepsilon)^m d\varepsilon P^X(dx) \\ &= \int_{-\infty}^{K_{11}}\int_0^\infty P(\|X-x\|>\varepsilon)^m d\varepsilon P^X(dx) \\ &+ \int_{K_{22}}^{\infty}\int_0^\infty P(\|X-x\|>\varepsilon)^m d\varepsilon P^X(dx) \\ &+ \int_{K_{11}}^{K_{22}}\int_0^\infty P(\|X-x\|>\varepsilon)^m d\varepsilon P^X(dx) \\ &= S_{11}(m)+S_{22}(m)+S_{33}(m), say. \end{aligned} \tag{24}$$

For bounding $S_{11}(m)$ and $S_{22}(m)$, an analogous upper bounds also hold for $d \geq 2$ as the proof of lemma 3.2. Hence, by assumption 2 (a) and for $t=\frac{1}{\tau m}$, $\tau \geq 1$, we have

$$\begin{aligned} S_{11}(m) &= \int_{-\infty}^{K_{11}}\int_0^\infty P(\|X-x\|>\varepsilon)^m d\varepsilon P^X(dx) \\ &\leq \tau\int_{-\infty}^{K_{11}}\psi\left(\frac{1}{\tau m},x\right)^m P^X(dx), \end{aligned} \tag{25}$$

$$S_{22}(m) = \int_{K_{22}}^{\infty}\int_0^{\infty} P(\|X-x\|>\varepsilon)^m d\varepsilon P^X(dx)$$
$$\leq \tau \int_{K_{22}}^{\infty} \psi\left(\frac{1}{\tau m},x\right)^m P^X(dx). \quad (26)$$

For bounding $S_{33}(m)$, suppose X with a density $f(x)>0$

$$S_{33}(m) = \int_{K_{11}}^{K_{22}}\int_0^{\infty} P(\|X-x\|>\varepsilon)^m d\varepsilon P^X(dx)$$
$$= \int_{K_{11}}^{K_{22}}\int_0^{\infty} e^{-mH(x,\varepsilon)} f(x)d\varepsilon dx,$$

where $H(x,\varepsilon) = -\log P(\|X-x\|>\varepsilon)$.
Using assumption 2 (b) for $d\geq 2$, we obtain

$$S_{33}(m) \leq \int_{K_{11}}^{K_{22}}\int_0^{\infty} e^{-mc\varepsilon^d f(x)} f(x)d\varepsilon dx$$
$$= \int_{K_{11}}^{K_{22}} \frac{f(x)^{1-(1/d)}}{d(mc)^{(1/d)}}\left(\int_0^{\infty} e^{-t}t^{(1/d)-1}dt\right)dx,$$
$$\left(mc\varepsilon^d f(x) = t \Rightarrow \varepsilon = \frac{t^{(1/d)}}{(mcf(x))^{(1/d)}}\right)$$
$$= \frac{\Gamma(1/d)}{d(mc)^{(1/d)}}\int_{K_{11}}^{K_{22}} f(x)^{1-(1/d)}dx, \quad (27)$$

Hence, by substituting (25)-(27) in (24), we obtain (23).

Lemma 3.6. By assumption 2 for $d\geq 2$, a constant $\tau\geq 1$ and $c_3,c_4>0$, we have

$$Ed_m^2 \leq 2\tau^2\int_{-\infty}^{K_{11}} \psi\left(\frac{1}{\tau m},x\right)^m P^X(dx)$$
$$+2\tau^2\int_{K_{22}}^{\infty} \psi\left(\frac{1}{\tau m},x\right)^m P^X(dx)$$
$$+\frac{2c_3(K_{22}-K_{11})}{c_4^2 m^2}, \quad (28)$$

where $K_{11} = K_{11}(m), K_{22} = K_{22}(m)$ are constants depending on m such that $-\infty < K_{11}\leq 0\leq K_{22}<\infty$.
Proof. Similarly as the proof of lemma 3.3, we derive upper bounds of Ed_m^2 as follows:

$$Ed_m = \int_{-\infty}^{\infty}\int_0^{\infty} 2\varepsilon P(\|X-x\|>\varepsilon)^m d\varepsilon P^X(dx)$$
$$= \int_{-\infty}^{K_{11}}\int_0^{\infty} 2\varepsilon P(\|X-x\|>\varepsilon)^m d\varepsilon P^X(dx)$$
$$+\int_{K_{22}}^{\infty}\int_0^{\infty} 2\varepsilon P(\|X-x\|>\varepsilon)^m d\varepsilon P^X(dx)$$
$$+\int_{K_{11}}^{K_{22}}\int_0^{\infty} 2\varepsilon P(\|X-x\|>\varepsilon)^m d\varepsilon P^X(dx)$$
$$= S'_{11}(m)+S'_{22}(m)+S'_{33}(m), say. \quad (29)$$

For bounding $S'_{11}(m)$ and $S'_{22}(m)$, an analogous upper bounds also hold for $d\geq 2$ as the proof of lemma 3.3. Hence, by assumption 2 (a) and for $t=\frac{1}{\tau m}$, $\tau\geq 1$, we have

$$S'_{11}(m) = \int_{-\infty}^{K_{11}}\int_0^{\infty} 2\varepsilon P(\|X-x\|>\varepsilon)^m d\varepsilon P^X(dx)$$
$$\leq 2\tau^2\int_{-\infty}^{K_{11}} \psi\left(\frac{1}{\tau m},x\right)^m P^X(dx), \quad (30)$$

$$S'_{22}(m) = \int_{K_{22}}^{\infty}\int_0^{\infty} 2\varepsilon P(\|X-x\|>\varepsilon)^m d\varepsilon P^X(dx)$$
$$\leq 2\tau^2\int_{K_{22}}^{\infty} \psi\left(\frac{1}{\tau m},x\right)^m P^X(dx). \quad (31)$$

For bounding $S'_{33}(m)$, suppose X with a density $f(x)>0$

$$S'_{33}(m) = \int_{K_{11}}^{K_{22}}\int_0^{\infty} 2\varepsilon P(\|X-x\|>\varepsilon)^m d\varepsilon P^X(dx)$$
$$= \int_0^{\infty} 2\varepsilon\left(\int_{K_{11}}^{K_{22}} P(\|X-x\|>\varepsilon)^m P^X(dx)\right)d\varepsilon$$
$$\leq \int_0^{\infty} 2\varepsilon\int_{K_{11}}^{K_{22}} e^{-mH(x,\varepsilon)} f(x)dxd\varepsilon,$$

where $H(x,\varepsilon) = -\log P(\|X-x\|>\varepsilon)$.
Using assumption 2 (b) for $d\geq 2$, we obtain

$$S'_{33}(m) \leq \int_{K_{11}}^{K_{22}}\int_0^{\infty} 2\varepsilon e^{-mc\varepsilon^d f(x)} f(x)d\varepsilon dx$$
$$= 2\int_{K_{11}}^{K_{22}} \frac{f(x)^{1-(2/d)}}{d(mc)^{(2/d)}}\left(\int_0^{\infty} e^{-t}t^{(2/d)-1}dt\right)dx,$$
$$\left(mc\varepsilon^d f(x) = t \Rightarrow \varepsilon = \frac{t^{(1/d)}}{(mcf(x))^{2/d}}\right)$$
$$= \frac{2\Gamma(2/d)}{d(mc)^{(2/d)}}\int_{K_{11}}^{K_{22}} f(x)^{1-(2/d)}dx, \quad (32)$$

Hence, by substituting (30)-(32) in (29), we obtain (28).
Theorem 3.7. Let the conditions of lemmas 2.3, 3.5 and 3.6 be satisfied. Then

$$R_m \leq R_\infty + \lambda\tau(1+2\tau)\int_{-\infty}^{K_{11}} \psi\left(\frac{1}{\tau m},x\right)^m P^X(dx)$$
$$+\lambda\tau(1+2\tau)\int_{K_{22}}^{\infty} \psi\left(\frac{1}{\tau m},x\right)^m P^X(dx)$$
$$+\frac{\lambda\Gamma(1/d)}{d(mc)^{(1/d)}}\int_{K_{11}}^{K_{22}} f(x)^{1-(1/d)}dx$$
$$+\frac{2\lambda\Gamma(2/d)}{d(mc)^{(2/d)}}\int_{K_{11}}^{K_{22}} f(x)^{1-(2/d)}dx. \quad (33)$$

Proof. This is immediate from the previous results, by taking $\alpha=1$ in lemma 2.3 and after substituting (23) and (28) in (3), we obtain (33).

We now estimate the finite-sample risk R_m for exponential and normal distributions, respectively, for $x\in R^d$, $d=2$ as follows:
Example 3: Let X be an exponential distribution with a density function $e^{-(x+y)}, x,y>0$. Taking $\tau=2, K_{11}=0$ and $K_{22}=2\log m$ in (33), we obtain

$$R_m \leq R_\infty + \frac{40\lambda}{m^2}\left(1+\frac{1}{2m-1}\right)^{2m}$$
$$+\frac{2\lambda}{\sqrt{mc}}\left[\sqrt{\pi}\left(1-\frac{1}{m}\right)^2+\frac{2(\log m)^2}{\sqrt{mc}}\right]$$
$$= R_\infty + O\left(\frac{1}{m^2}\right)+\frac{2\lambda}{\sqrt{mc}}\left[\sqrt{\pi}\left(1-\frac{1}{m}\right)^2+\frac{2(\log m)^2}{\sqrt{mc}}\right]$$

where
$\int_{K_{22}}^{\infty}\psi\left(\frac{1}{\tau m},x\right)^m P^X(dx) \le \frac{4}{m^2}\left(1+\frac{1}{2m-1}\right)^{2m} = O\left(\frac{1}{m^2}\right)$,
since $\left(1+\frac{1}{2m-1}\right)^{2m} \to e \ (m\to\infty)$,

$$\int_{K_{11}}^{K_{22}} f(x)^{1-(1/d)}dx = \int_0^{2\log m} f(x)^{1/2}dx$$
$$= \int_0^{2\log m}\int_0^{2\log m} e^{\frac{-(u+v)}{2}}dudv = 4\left(1-\frac{1}{m}\right)^2,$$

and

$$\int_{K_{11}}^{K_{22}} f(x)^{1-(2/d)}dx = \int_0^{2\log m}\int_0^{2\log m} dudv = 4(\log m)^2.$$

Example 4: Let X be a normal distribution with a density function $\frac{1}{2\pi}e^{-(x^2+y^2)/2}, -\infty < x,y < \infty$. Taking $\tau = 1$ and $-K_{11} = K_{22} = \sqrt{2\log m}$ in (33), we obtain

$$R_m \le R_\infty + \frac{1.2833\lambda}{\pi m^2 \log m} + \frac{\lambda}{\sqrt{mc}}\left[\sqrt{2\pi} + \frac{8\log m}{\sqrt{mc}}\right],$$

where
$\int_{K_{22}}^{\infty}\psi\left(\frac{1}{\tau m},x\right)^m P^X(x) \le \frac{1.2833}{\pi m^2 \log m}$,
such that, for $t = \frac{1}{m}$ and using Jensen's inequality, we have:
$\psi\left(\frac{1}{\tau m},x\right)^m = \left(Ee^{t\|X-x\|}\right)^m \le E\left(e^{\|X-x\|}\right) = E\left(e^Z\right)$,
where
$Z = \|X-x\| = \sqrt{(X_1-x_1)^2+(X_2-x_2)^2} = \sqrt{X_1'^2+X_2'^2}$,
and $X_1' = X_1 - x_1, X_2' = X_2 - x_2$.
Hence Z has a Rayleigh distribution, that is, $f_Z(z) = ze^{-z^2/2}, z > 0$, see [17].

$$E(e^Z) = \int_0^\infty e^z z e^{-z^2/2}dz = e^{\frac{1}{2}}\int_0^\infty z e^{-(z-1)^2/2}dz$$
$$= -e^{\frac{1}{2}}\left[\int_0^\infty -(z-1)e^{-(z-1)^2/2}dz - \int_0^\infty e^{-(z-1)^2/2}dz\right]$$
$$\le 1 + \sqrt{2\pi}e^{\frac{1}{2}} \cong 5.133,$$

$$\int_{K_{22}}^\infty \psi\left(\frac{1}{\tau m},x\right)^m P^X(dx)$$
$$\le \frac{5.133}{2\pi}\int_{K_{22}}^\infty\int_{K_{22}}^\infty e^{-(u^2+v^2)/2}dudv$$
$$\le \frac{5.133}{2\pi(K_{22})^2}e^{-(K_{22})^2} = \frac{5.133e^{-(2\log m)}}{2\pi(2\log m)} \cong \frac{1.2833}{\pi m^2 \log m},$$

since $\int_{K_{22}}^\infty e^{-x^2} \le \frac{1}{K_{22}}e^{-(K_{22})^2}$,

$$\int_{K_{11}}^{K_{22}} f(x)^{1-(1/d)}dx = 4\int_0^{\sqrt{2\log m}} f(x)^{1/2}dx$$
$$= 4\int_0^{\sqrt{2\log m}}\int_0^{\sqrt{2\log m}} \frac{1}{\sqrt{2\pi}}e^{-(u^2+v^2)/4}dudv \le 2\sqrt{2\pi},$$

such that

$\int_0^{\sqrt{2\log m}} \frac{1}{\sqrt{2\pi}}e^{-u^2/4}du \le \sqrt{2}\int_0^\infty \frac{1}{2\sqrt{2\pi}}e^{-u^2/4}du \le \frac{\sqrt{2}}{2}$,
and
$\int_{K_{11}}^{K_{22}} f(x)^{1-(2/d)}dx = 4\int_0^{\sqrt{2\log m}}\int_0^{\sqrt{2\log m}} dudv = 8\log 8.$

Conclusion. We have found upper bounds on the expected nearest neighbor distance for the distributions that have unbounded support and estimated bounds for the finite-sample risk R_m in terms of the expected nearest neighbor distance. We provided rate of convergence for expected nearest neighbor distance Ed_m in the one-dimensional unbounded support which we showed that Ed_m converges at rate of $o\left(\frac{1}{m^\beta}\right)$ for all $\beta < 1$, and it depends on the tails of the distributions for which there is an additional logarithmic term compared with the rates for compact support for exponentially decaying tails. We looked at real-valued observations and given some contributions for $x \in R^d$ and $d \ge 2$. We found upper bounds for the exponential and normal distributions as typical.

Appendix.

A. Proof the sufficient condition $P(|X-x| \le \varepsilon) \ge 2\varepsilon f(x)$ of the inequality $H(x,\varepsilon) \ge c\varepsilon f(x)$ when $S = (0,\infty)$ and the probability density function $f(x)$ has a completely monotonic function.

Recall that, a function (probability density function) $f(x)$ with domain $(0,\infty)$ is said to be completely monotone function if all derivatives of the f exist and $(-1)^n f^{(n)}(x) \ge 0$ for all $x > 0$ and $n > 0$, see Feller [18].

By corollary 3.1 we have

$$\begin{aligned} H(x,\varepsilon) &= -\log P(|X-x| > \varepsilon) \\ &\ge P(|X-x| \le \varepsilon) \\ &= F(X+\varepsilon) - F(X-\varepsilon). \end{aligned} \quad (34)$$

Therefore, we can get a good asymptotic estimates for $F(X+\varepsilon) - F(X-\varepsilon)$, by using the Taylor expansion for the functions $F(X+\varepsilon)$ and $F(X-\varepsilon)$ as $\varepsilon \to 0$, respectably, we have

$$F(X+\varepsilon) = F(X) + \frac{f(x)\varepsilon}{1!} + \frac{f'(x)\varepsilon^2}{2!} + \frac{f''(x)\varepsilon^3}{3!} + \frac{f'''(x)\varepsilon^4}{4!} + \frac{f^{(4)}(x)\varepsilon^5}{5!} + \cdots \quad (35)$$

$$F(X-\varepsilon) = F(X) - \frac{f(x)\varepsilon}{1!} + \frac{f'(x)\varepsilon^2}{2!} - \frac{f''(x)\varepsilon^3}{3!} + \frac{f'''(x)\varepsilon^4}{4!} - \frac{f^{(4)}(x)\varepsilon^5}{5!} + \cdots \quad (36)$$

Substituting (35) and (36) in (34) yields.

$$\begin{aligned} H(x,\varepsilon) &\ge F(X+\varepsilon) - F(X-\varepsilon) \\ &= \frac{2f(x)\varepsilon}{1!} + \frac{2f''(x)\varepsilon^3}{3!} + \frac{2f^{(4)}(x)\varepsilon^5}{5!} + \cdots \\ &\ge 2\varepsilon f(x) > 0, \end{aligned} \quad (37)$$

where, $f^{(n)}(x) \ge 0$ for $n = 0,2,4,\cdots$.

Then, the inequality $H(x,\varepsilon) \ge c\varepsilon f(x)$ holds with $c = 2$ if a probability density function $f(x)$ is a completely monotone (e.g., the densities proportional to the functions $(1+x)^{-k}, x^{-2}e^{x^{-1}}, x^{(\alpha-1)}e^{-x}$ and $e^{-x^\alpha}(0 < \alpha \le 1)$ are satisfying this criterion).

B. Proof the validity of the inequality (4) for $d = 1$ for exponential distribution.
Since the probability density function of the exponential distribution is a completely monotone, so the inequality (4) for $d = 1$ validity with $c = 2$. Then $S_3(m) \leq \frac{\log m}{m}$ and we used it.
Or, applying a different way: If $[x-\varepsilon, x+\varepsilon] \subseteq [0,\infty)$, then by convexity $P(|X-x| \leq \varepsilon) \geq 2\varepsilon f(x)$.
If $(x-\varepsilon) \leq 0$, then

$$\begin{aligned}-\log P(|X-x| > \varepsilon) &= -\log P(X > x+\varepsilon)\\ &= -\log e^{-(x+\varepsilon)} = x+\varepsilon \geq \varepsilon f(x).\end{aligned}$$

This shows validity with $c = 1$. Substituting by $K_1 = 0$ and $K_2 = 2\log m$ in (8) we obtain $S_3(m) \leq \frac{2\log m}{m}$.

C. Proof the validity of the inequality (4) for $d = 1$ for normal distribution Due to symmetry it is enough to treat $x > 0$. Let $\varepsilon > 0$.
(a) If $x+\varepsilon \geq -x$, then

$$P(|X-x| \leq \varepsilon) \geq P(X - x \leq X \leq x) \geq \varepsilon\varphi(x).$$

(b) So assume $x-\varepsilon < -x$, i.e. $\varepsilon > 2x$. We can show

$$\begin{aligned}&-\log P(|X-x| > \varepsilon)\\ &= -\log(P(X < -(\varepsilon-x)) + P(X > x+\varepsilon))\\ &\geq -\log 2P(X < -(\varepsilon-x)) = -\log 2 - \log P(X > \varepsilon - x)\\ &\geq -\log 2 - \log\left(\frac{1}{\varepsilon-x}\frac{1}{\sqrt{2\pi}}e^{\frac{-(\varepsilon-x)^2}{2}}\right)\\ &= \log\left(\sqrt{2\pi}/2\right) + \log(\varepsilon-x) + (\varepsilon-x)^2.\end{aligned}$$

(c) For $x > 1$, then $\log(\varepsilon - x) \geq 0$ and from $\varepsilon > 2x$ we have $(\varepsilon-x)^2 \geq \frac{\varepsilon^2}{4} \geq \frac{1}{2}\varepsilon\varphi(x)$.
And, for $x < 1$ if $\varepsilon \geq 2$, then $\log(\varepsilon - x) \geq 0$ and we proceed as above, and if $\varepsilon < 2$ we have $P(|X-x| \leq \varepsilon) \geq c\varepsilon$, where $c = \inf_{|y|<3}\varphi(y)$, hence $P(|X-x| \leq \varepsilon) \geq c\varepsilon\varphi(x)$.
Therefore, we can find a constant $c^* > 0$ such that $-\log P(|X-x| > \varepsilon) \geq c^*\varepsilon\varphi(x)$, for all $x, \varepsilon > 0$.

D. Proof (21)

$$\begin{aligned}S'_3(m) &= \int_{K_1}^{K_2}\int_0^\infty P\left(|X-x| > \sqrt{\varepsilon}\right)^m d\varepsilon P^X(dx)\\ &= \int_{K_1}^{K_2}\int_0^\infty e^{-mH(x,\sqrt{\varepsilon})} f(x)d\varepsilon dx\\ &\leq \int_{K_1}^{K_2}\int_0^\infty e^{-cm\sqrt{\varepsilon}f(x)} f(x)d\varepsilon dx\\ &\quad\left(cm\sqrt{\varepsilon}f(x) = t \Rightarrow d\varepsilon = \frac{2tdt}{c^2m^2f(x)^2}\right)\\ &= \int_{K_1}^{K_2}\frac{2f(x)}{c^2m^2f(x)^2}\left(\int_0^\infty te^{-t}dt\right)dx\\ &= \frac{2}{c^2m^2}\int_{K_1}^{K_2} f(x)^{-1}dx.\end{aligned}$$

Conflicts of Interests
The authors declare that they have no conflicts of interests

References

[1] L. Devroye, L. Györfi, G. Lugosi, A probabilistic theory of pattern recognition, Springer-Verlag, New York (1996).
[2] L. Györfi, M. Kohler, A. Krzyak, and H. Walk, A distribution-free theory of nonparametric regression, Springer-Verlag, New York, (2002).
[3] E. Fix, J. Hodges, Discriminatory analysis, nonparametric discrimination, consistency properties, In nearest neighbor classification techniques, B. Dasarathy, editor, IEEE Computer Society Press, Los Alamitos, CA, 32-39 (1991)
[4] E. Fix, J. Hodges, Discriminatory analysis, small samples editor, performance, In nearest neighbor classification techniques, B. Dasarathy, editor, IEEE Computer Society Press, Los Alamitos, CA, 40-56 (1991).
[5] B. Dasarathy, Nearest neighbor classification techniques. IEEE, Los Alamitos, CA. (1991).
[6] G. Biau and L. Devroye, Lectures on the nearest neighbor method, Springer, (2015).
[7] P. Zhao, and L. Lai, Analysis of KNN information estimators for smooth distributions, IEEE Trans. Inform. Theory, 66 (6), 3798-3826, (2019).
[8] T. M. Cover, and P. E. Hart, Nearest neighbor pattern classification, IEEE Transactions on Information Theory, 13, 21-27 (1967).
[9] T. M. Cover, Rates of convergence for neatest neighbor procedures, In Proceedings of the Hawaii International Conference on System Sciences, Honolulu, HI, 413-415 (1968).
[10] K. Fukunaga, D. M. Hummel, Bias of nearest neighbor estimates, IEEE Trans, Pattern Anal. Mach. Intel., 9, 103-112 (1987).
[11] D. Psaltis, R. Snapp, S. S. Venkatesh, On the finite sample performance ofthe nearest neighbor classifier, IEEE Transactions on Information Theory, 40, 820-837 (1994).
[12] R. R. Snapp, S. S. Venkatesh, Asymptotic expansion of the k nearest neighbor risk, Ann. Statist., 26, 850-878 (1998).
[13] A. Irle, M. Rizk, On the risk of nearest neighbor rules, Ph.D. thesis, Kiel University, Germany, (2004).
[14] S. R. Kulkarni, and S. E. Posner, Rates of convergence of nearest neighbor estimation under arbitrary sampling, IEEE Transactions on Information Theory, 41, 1028-1039, (1995).
[15] D. Evans, A. J. Jones, and W. M. Schmidt, Asymptotic moments of near neighbor distance distributions, Proc. R. Soc. Lond. A 458, 2839-2849, (2002).
[16] E. Liitiäinen, A. Lendasse, and F. Corona, A boundary corrected expdansion of the moments of nearest neighbor distributions, Random Structures and Algorithms 37, 223-247, (2010).
[17] A. Papoulis, and S. U. Pillai, Probability, Random Variables, and Stochastic Processes, 4th ed., McGraw-Hill, (2002).
[18] M. Döring, L. Györfi, and H. Walk, Rate of convergence of k-nearest-neighbor classification rule, journal of machine learning research, 18, 1-16 (2018).
[19] P. Zhao and L. Lai, Analysis of KNN information estimators for smooth distributions, IEEE Trans. Inform. Theory, 66 (6), 3798-3826 (2019).
[20] N. G. De Bruijn, Asymptotic methods in analysis, Wiley, New York (1958).