British Journal of Mathematics & Computer Science
6(4): 297-306, 2015, Article no.BJMCS.2015.080
ISSN: 2231-0851

SCIENCEDOMAIN *international*
www.sciencedomain.org

SCIENCEDOMAIN

Bounds on the Expected Nearest Neighbor Distance

Mohamed M. Rizk[1*]

[1]*Mathematics and Statistics Department, Faculty of Science, Taif University, Taif, Saudi Arabia.*

Article Information

DOI: 10.9734/BJMCS/2015/15152
Editor(s):
(1) Dariusz Jacek Jakóbczak, Computer Science and Management Department, Technical University of Koszalin, Poland.
Reviewers:
(1) Anonymous, Vietnam.
(2) Anonymous, China.
(3) Anonymous, Czech Republic.
Complete Peer review History: http://www.sciencedomain.org/review-history.php?iid=734&id=6&aid=7675

Original Research Article

Received: 10 November 2014
Accepted: 09 December 2014
Published: 09 January 2015

Abstract

This paper found upper and lower bounds on the expected nearest neighbor distance for distributions having unbounded supports (-∞,∞) and induced lower and upper bounds for logistic and Laplace distributions, as typical. Then we found the risk of nearest neighbor of their distributions.

Keywords: Nearest neighbor classification, expected nearest neighbor distance, logistic distribution, Laplace's distribution.

1 Introduction

The nearest neighbor rule was first studied by Fix and Hodges [1]. Cover and Hart [2] proved that $R^* \le R_\infty \le 2R^*(1-R^*)$ under certain conditions, 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. Cover [3] 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, where R_m denotes the finite sample risk, and m is the sample size. Kulkarni and Posner [4] studied the rate of convergence for nearest neighbor estimation in terms of the covering numbers of totally bounded sets. Evans et al. [5] derived an asymptotic moments of near neighbor distance distributions. Irle and Rizk [6] 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. Liitiäinen et al. [7] studied a boundary corrected expansion of the moments of nearest neighbor distributions. Rizk and Ateya [8] found lower and upper bounds for the risk of nearest neighbor of generalized exponential distribution.

**Corresponding author: mhm96@yahoo.com;*

In this paper, we find upper and lower bounds on the expected nearest neighbor distance for distributions having unbounded support $S = (-\infty, \infty)$ for which we derive upper and lower bounds on the expected nearest neighbor distance of logistic and Laplace distributions as typical, and evaluate the bounds of the risk for their distributions.

In pattern recognition if we have a random variable (X, θ), such that $X \in R^d$ is an observed pattern from which we wish to predict the unobservable class θ. This class takes values in a finite set $M = \{1,2,\ldots, C\}$. If the joint distribution of (X, θ) is known, then the best classifier is the Bayes classifier, see [9,10]. In general the joint distribution of (X, θ) will be unknown, and we have a training sequence $D_m = \left((X^{(1)}, \theta^{(1)}), (X^{(2)}, \theta^{(2)}), \ldots, (X^{(m)}, \theta^{(m)})\right)$ at our disposal, where patterns and corresponding classes are observed and we assume that $\left((X^{(1)}, \theta^{(1)}), (X^{(2)}, \theta^{(2)}), \ldots, (X^{(m)}, \theta^{(m)})\right)$, the data, stem from a sequence of independent identically distributed random pairs with the same distribution as (X, θ). The nearest neighbor rule assigns any input feature vector to the class given by the label θ' of the nearest reference vector.

The problem to be considered is the classification of a random variable θ taking values in $M = \{1,2\}$ given a sample X in χ, with the goal of minimizing the finite-sample risk $R_m = P(\theta \neq \theta')$, where χ is a separable metric space equipped with metric ρ which we denote as the pair (χ, ρ).Define the nearest distance at time m as $d_m = \rho(X, X')$.

2 Bounds for the Expected Nearest Neighbor Distance

In this section, we find upper and lower bounds for expected nearest neighbor distance for distributions having unbounded support $S = (-\infty, \infty)$, compare [6].

2.1 An Upper Bound on the Expected Nearest Neighbor Distance

We use constants, $-\infty < K_1(m) \leq 0 \leq K_2(m) < \infty$ depending on m, to write

$$Ed_m = \int_{-\infty}^{\infty}\int_0^{\infty} P(|X - x| > \varepsilon)^m \, d\varepsilon f(x) dx$$

$$= \int_{-\infty}^{K_1(m)}\int_0^{\infty} P(|X - x| > \varepsilon)^m \, d\varepsilon f(x) dx$$

$$+ \int_{K_2(m)}^{\infty}\int_0^{\infty} P(|X - x| > \varepsilon)^m \, d\varepsilon f(x) dx$$

$$+ \int_{K_1(m)}^{K_2(m)}\int_0^{\infty} P(|X - x| > \varepsilon)^m d\varepsilon f(x) dx.$$

$$Ed_m = L_1(m) + L_2(m) + L_3(m). \tag{2.1}$$

where,

$$L_1(m) = \int_{-\infty}^{K_1(m)}\int_0^{\infty} P(|X - x| > \varepsilon)^m d\varepsilon f(x)\, dx, \tag{2.2}$$

$$L_2(m) = \int_{K_2(m)}^{\infty}\int_0^{\infty} P(|X - x| > \varepsilon)^m d\varepsilon f(x) dx, \tag{2.3}$$

$$L_3(m) = \int_{K_1(m)}^{K_2(m)} \int_0^\infty P(|X - x| > \varepsilon)^m d\varepsilon f(x)dx. \quad (2.4)$$

2.1.1 Bounding $L_1(m)$ and $L_2(m)$

We assume for the following that $|X - x|$ has a finite moment generating function $\varphi(t, x) = E\left(e^{t|X-x|}\right)$, $x \in R$, $0 < t < 1$. By Markov's inequality for any $0 < t < 1$, we obtain:

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

For $t = \frac{1}{\tau m}$, and$\tau \geq 1$, we have $\int_0^\infty P(|X - x| > \varepsilon)^m d\varepsilon \leq \tau\varphi\left(\frac{1}{\tau m}, x\right)^m$.It follows

$$L_1(m) \leq \tau \int_{-\infty}^{K_1(m)} \varphi\left(\frac{1}{\tau m}, x\right)^m f(x)dx, \quad (2.5)$$

$$L_2(m) \leq \tau \int_{K_2(m)}^\infty \varphi\left(\frac{1}{\tau m}, x\right)^m f(x)dx. \quad (2.6)$$

2.1.2 Bounding $L_3(m)$

From (2.4) we have

$$L_3(m) = \int_{K_1(m)}^{K_2(m)} \int_0^\infty P(|X - x| > \varepsilon)^m d\varepsilon f(x)dx$$

$$= \int_{K_1(m)}^{K_2(m)} \int_0^\infty e^{-mG(x,\varepsilon)} f(x)d\varepsilon dx, \quad (2.7)$$

where $G(x, \varepsilon) = -\log P(|X - x| > \varepsilon)$.Assume that the following inequality holds:
There exists $c > 0$ such that for all x in the support of X and for all $\varepsilon > 0$

$$G(x, \varepsilon) \geq c\varepsilon f(x). \quad (2.8)$$

From this inequality, we obtain

$$\int_{K_1(m)}^{K_2(m)} \int_0^\infty e^{-mG(x,\varepsilon)} f(x)d\varepsilon dx \leq \int_{K_1(m)}^{K_2(m)} \int_0^\infty e^{-mc\varepsilon f(x)} f(x)d\varepsilon dx$$

$$= \int_{K_1(m)}^{K_2(m)} \frac{1}{cm} dx = \frac{1}{cm}\left(K_2(m) - K_1(m)\right). \quad (2.9)$$

We can show that a sufficient condition for $G(x, \varepsilon) \geq c\varepsilon f(x)$ is given by, see [6]:

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

Note that this second condition will always be violated for unbounded Support letting ε tend to∞.

2.2 Deriving a Lower Bound

$$Ed_m = \int_0^\infty P(d_m > \varepsilon)d\varepsilon = \int_{-\infty}^\infty \int_0^\infty P(d_m > \varepsilon | X = x)\, d\varepsilon f(x)dx$$
$$= \int_{-\infty}^\infty \int_0^\infty P(|X - x| > \varepsilon)^m\, d\varepsilon f(x)dx$$
$$\geq \int_{-\infty}^\infty \int_0^\infty P(X < x - \varepsilon)^m\, d\varepsilon\, f(x)dx$$
$$= \int_{-\infty}^\infty \int_{-\infty}^x P(X < z)^m\, dz f(x)dx$$
$$= \int_{-\infty}^\infty \int_z^\infty f(x)\, dx P(X < z)^m dz$$

In the following sections we derive the upper and lower bounds for the expected nearest neighbor distance Ed_m for logistic and Laplace's distributions.

3 Bounds on the Expected Nearest Neighbor Distance for Logistic Distribution

Let X have a probability density function$f(x) = \frac{e^{\frac{-(x-a)}{b}}}{b\left(1+e^{\frac{-(x-a)}{b}}\right)^2}$, $-\infty < x < \infty$, where a the location parameter and $b > 0$ the scale parameter. We use the method in the previous section to find the upper bound for the logistic distribution. Now, without loss of the generality, we assume that $a = 0$.

3.1 An Upper Bound for the Logistic Distribution

Using the method in (2.1), we take $-K_1(m) = K_2(m) > 0$, for $x \in R, t > 0$.

3.1.1 Bounding $L_1(m)$ and $L_2(m)$

Firstly, we evaluate $\varphi(t, x)$,i.e, we find the moment generating function of $|X - x|$. For $x \in R, 0 < t < 1$, we have

$$\varphi(x,t) = E\left(e^{t|X-x|}\right) \leq E(e^{tx+tX})$$
$$= e^{tx} \int_{-\infty}^{\infty} e^{ty} \frac{e^{\frac{-x}{b}}}{b\left(1+e^{\frac{-x}{b}}\right)^2} dy$$

$$= e^{tx}\Gamma(1 - bt)\Gamma(1 + bt) = e^{tx}\frac{\pi bt}{\sin \pi bt}, \qquad bt < 1,$$

whereΓ gamma function.Hence, for $t = \frac{1}{2bm}$, we obtain $\varphi\left(x, \frac{1}{2bm}\right) \leq e^{\frac{x}{2bm}}\left(\frac{\frac{\pi}{2m}}{\sin\frac{\pi}{2m}}\right)$. Therefore

$$\varphi\left(x, \frac{1}{2bm}\right)^m \leq e^{\frac{x}{2b}}\left[\left(\frac{\frac{\pi}{2m}}{\sin\frac{\pi}{2m}}\right)\right]^m.$$

It follows

$$L_1(m)+L_2(m)\le 4b\left[\left(\frac{\frac{\pi}{2m}}{\sin\frac{\pi}{2m}}\right)\right]^m\int_{K_2(m)}^{\infty}\frac{e^{\frac{-x}{2b}}}{b\left(1+e^{\frac{-x}{b}}\right)^2}dx$$

$$=8b\left[\left(\frac{\frac{\pi}{2m}}{\sin\frac{\pi}{2m}}\right)\right]^m\int_0^{e^{\frac{-K_2(m)}{2b}}}\frac{dy}{(1+y^2)^2}$$

$$=8b\left[\left(\frac{\frac{\pi}{2m}}{\sin\frac{\pi}{2m}}\right)\right]^m\int_0^{\tan^{-1}e^{\frac{-K_2(m)}{2b}}}cos^2\theta d\theta$$

$$=4b\left[\left(\frac{\frac{\pi}{2m}}{\sin\frac{\pi}{2m}}\right)\right]^m\left[\tan^{-1}e^{\frac{-K_2(m)}{2b}}+\frac{1}{2}sin2\left(\tan^{-1}e^{\frac{-K_2(m)}{2b}}\right)\right]$$

For $K_2(m)=2b\log m$, it follows

$$L_1(m)+L_2(m)\le 4b\left[\left(\frac{\frac{\pi}{2m}}{\sin\frac{\pi}{2m}}\right)\right]^m\left[\tan^{-1}\frac{1}{m}+\frac{1}{2}sin\left(2\tan^{-1}\frac{1}{m}\right)\right], \tag{3.1}$$

we note that $\left(\frac{\frac{\pi}{2m}}{\sin\frac{\pi}{2m}}\right)\to 1$ for large m.

<u>3.1.2Bounding $L_3(m)$</u>

From (2.7) we have

$$L_3(m)=\int_{K_1(m)}^{K_2(m)}\int_0^{\infty}e^{-mG(x,\varepsilon)}f(x)d\varepsilon dx, \tag{3.2}$$

where,

$$G(x,\varepsilon)=-\log P(|X-x|>\varepsilon)\ge P(|X-x|\le\varepsilon)=F(x+\varepsilon)-F(x-\varepsilon). \tag{3.3}$$

Then we need good asymptotic estimates for $F(x+\varepsilon)-F(x-\varepsilon)$, as $\varepsilon\to 0$, By using the Taylor expansion for the functions $F(x+\varepsilon)$ and $F(x-\varepsilon)$ we obtain

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

since $f^{(n)}(x)\ge 0$ for $n=0,2,4,\ldots$, then we obtain $G(x,\varepsilon)\ge 2\varepsilon f(x)$. Hence

$$L_3(m)\le\int_{K_1(m)}^{K_2(m)}\int_0^{\infty}e^{-2m\varepsilon f(x)}f(x)d\varepsilon dx$$

$$=\int_{K_1(m)}^{K_2(m)}\frac{1}{2m}dx=\frac{2b\log m}{m}, \tag{3.4}$$

where $-K_1(m)=K_2(m)>0$, and $K_2(m)=2b\log m$.

Substituting (3.1) and (3.4) in (2.1), we obtain the upper bound of Ed_m for the logistic distribution.

$$Ed_m \leq 4b\left[\left(\frac{\frac{\pi}{2m}}{\sin\frac{\pi}{2m}}\right)\right]^m\left[\tan^{-1}\frac{1}{m}+\frac{1}{2}sin\left(2\tan^{-1}\frac{1}{m}\right)\right]+\frac{2b\log m}{m}. \tag{3.5}$$

3.2Alower Bound for the Logistic Distribution

Appling the method in section 2.2, we have

$$\begin{aligned}
Ed_m &= \int_0^\infty P(d_m > \varepsilon)d\varepsilon \\
&= \int_{-\infty}^\infty \int_0^\infty P(d_m > \varepsilon | X = x)\, d\varepsilon \frac{e^{\frac{-x}{b}}}{b\left(1+e^{\frac{-x}{b}}\right)^2} dx \\
&= \int_{-\infty}^\infty \int_0^\infty P(|X-x| > \varepsilon)^m\, d\varepsilon \frac{e^{\frac{-x}{b}}}{b\left(1+e^{\frac{-x}{b}}\right)^2} dx \\
&\geq \int_{-\infty}^\infty \int_0^\infty P(X < x-\varepsilon)^m\, d\varepsilon \frac{e^{\frac{-x}{b}}}{b\left(1+e^{\frac{-x}{b}}\right)^2} dx \\
&= \int_{-\infty}^\infty \int_{-\infty}^x P(X < z)^m\, dz \frac{e^{\frac{-x}{b}}}{b\left(1+e^{\frac{-x}{b}}\right)^2} dx \\
&= \int_{-\infty}^\infty \int_z^\infty \frac{e^{\frac{-x}{b}}}{b\left(1+e^{\frac{-x}{b}}\right)^2} dx P(X < z)^m dz \\
&= \int_{-\infty}^{\infty} \frac{e^{\frac{-z}{b}}}{\left(1+e^{\frac{-z}{b}}\right)^{m+1}} dz = \frac{b}{m}.
\end{aligned} \tag{3.6}$$

Now, we can find an upper bound on the finite sample risk $R_m$in terms of the expected nearest neighbor distance for logistic distribution by using the following result:

If, for some $\omega_1 > 0$ and$0 < \gamma \leq 1$ we have $|m(x) - m(x')| \leq \omega_1 \rho(x, x')^\gamma$, for all$x, x' \in \chi$, then for some suitable $\omega > 0$independent of m,

$$R_m \leq R_\infty + \omega\left[(Ed_m)^\gamma + \left(Ed_m^{2\gamma}\right)\right], \tag{3.7}$$

where$\omega = max\{\omega_1, \omega_1^2\}$.

This result is due to Irle and Rizk [6], for which they found an upper bound on the finite sample risk $R_m$in terms of the expected nearest neighbor distance. Putting$\gamma = \frac{1}{2}$in (3.7). Hence, from (3.5) we have

$$\begin{aligned}
R_m \leq R_\infty + \omega\Bigg[&\left(4b\left[\left(\frac{\frac{\pi}{2m}}{\sin\frac{\pi}{2m}}\right)\right]^m\left[\tan^{-1}\frac{1}{m}+\frac{1}{2}sin\left(2\tan^{-1}\frac{1}{m}\right)\right]+\frac{2b\log m}{m}\right)^{\frac{1}{2}} \\
&+\left(4b\left[\left(\frac{\frac{\pi}{2m}}{\sin\frac{\pi}{2m}}\right)\right]^m\left[\tan^{-1}\frac{1}{m}+\frac{1}{2}sin\left(2\tan^{-1}\frac{1}{m}\right)\right]+\frac{2b\log m}{m}\right)\Bigg]
\end{aligned} \tag{3.8}$$

4 Bounds on the Expected Nearest Neighbor Distance for Laplace's Distribution

We look at d_m in the case of the Laplace's distribution with probability density function $f(x) = \frac{1}{2\beta}e^{-\frac{|x|}{\beta}}$, where $\beta > 0, -\infty < x < \infty$.

4.1 An Upper Bound for the Laplace's Distribution

Using the method in 2.1, we take $-K_1(m) = K_2(m) > 0$, for $x \in R, t > 0$.

4.1.1 Bounding $L_1(m)$ and $L_2(m)$

Now, we evaluate $\varphi(t,x)$. For $x \in R, 0 < t < 1$, we have

$$\varphi(t,x) = Ee^{t|X-x|} \le Ee^{t|X|+t|x|}$$

$$= e^{t|x|}\int_{-\infty}^{\infty}\frac{1}{2\beta}e^{-\frac{|y|}{\beta}}e^{ty}dy$$

$$= \frac{e^{t|x|}}{2\beta}\left[\int_{-\infty}^{0}e^{y\left(\frac{1}{\beta}+t\right)}dy + \int_{0}^{\infty}e^{-y\left(\frac{1}{\beta}-t\right)}dy\right]$$

$$= \frac{e^{t|x|}}{2\beta}\left[\frac{\beta}{(1+\beta t)} + \frac{\beta}{(1-\beta t)}\right] = \frac{e^{t|x|}}{(1-\beta^2 t^2)}.$$

Hence for$t = \frac{1}{2m} < \frac{1}{\beta}$, we obtain

$$\varphi\left(\frac{1}{2m}, x\right)^m \le \frac{e^{\frac{|x|}{2}}}{\left(1-\frac{\beta^2}{4m^2}\right)^m} = e^{\frac{|x|}{2}}\left(1+\frac{\beta^2}{4m^2-\beta^2}\right)^m. \text{ Then}$$

$$L_1(m) + L_2(m) \le 4\ \left(1+\frac{\beta^2}{4m^2-\beta^2}\right)^m \int_{K_2(m)}^{\infty} e^{\frac{x}{2}}\frac{1}{2\beta}e^{-\frac{x}{\beta}}dx$$

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

$$= 4\left(1+\frac{\beta^2}{4m^2-\beta^2}\right)^m \frac{e^{-K_2(m)\left(\frac{2-\beta}{2\beta}\right)}}{(2-\beta)}$$

For $K_2(m) = \log m^{\frac{2\beta}{2-\beta}}$, it follows

$$L_1(m) + L_2(m) \le \frac{4}{2-\beta}\left(1+\frac{\beta^2}{4m^2-\beta^2}\right)^m \frac{1}{m} = O\left(\frac{1}{m}\right), \tag{4.1}$$

since $\left(1+\frac{\beta^2}{4m^2-\beta^2}\right)^m \to 1$ as $m \to \infty$.

4.1.2 Bounding $L_3(m)$

From (2.7) we have

$$L_3(m) = \int_{K_1(m)}^{K_2(m)} \int_0^\infty e^{-mG(x,\varepsilon)} f(x) d\varepsilon dx$$
$$= 2\int_0^{K_2(m)} \int_0^\infty e^{-mG(x,\varepsilon)} \frac{1}{2\beta} e^{-\frac{x}{\beta}} d\varepsilon dx,$$

where $G(x,\varepsilon) = -\log P(|X-x| > \varepsilon)$, and $f(x)$ even function. Similarly, as in the previous section we have $F(x+\varepsilon) - F(x-\varepsilon) \geq 2\varepsilon f(x)$, as $\varepsilon \to 0$. Then we obtain $G(x,\varepsilon) \geq 2\varepsilon f(x)$. Hence

$$L_3(m) \leq \int_0^{K_2(m)} \int_0^\infty e^{-2m\varepsilon f(x)} f(x) d\varepsilon dx$$
$$= \int_0^{K_2(m)} \frac{1}{2m} dx = \frac{\log m^{\frac{2\beta}{2-\beta}}}{2m}, \quad (4.2)$$

where $K_2(m) = \log m^{\frac{2\beta}{2-\beta}}$.

From (4.1), (4.3), we obtain the upper bound of Ed_m for Laplace distribution in the form:

$$Ed_m \leq \frac{4}{(2-\beta)m}\left(1 + \frac{\beta^2}{4m^2 - \beta^2}\right)^m + \frac{\log m^{\frac{2\beta}{2-\beta}}}{2m}. \quad (4.3)$$

Thus

$$R_m \leq R_\infty + \lambda\left[\sqrt{\frac{C_3}{m} + \frac{\log m^{\frac{2\beta}{2-\beta}}}{2m}} + \frac{C_3}{m} + \frac{\log m^{\frac{2\beta}{2-\beta}}}{2m}\right], \quad (4.4)$$

where, $C_3 = \frac{4}{(2-\beta)}\left(1 + \frac{\beta^2}{4m^2-\beta^2}\right)^m$. Note that C_3 dependent on m and $\left(1 + \frac{\beta^2}{4m^2-\beta^2}\right)^m \to 1$, as $m \to \infty$.

4.2Alower Bound for the Laplace's Distribution

$$Ed_m = \int_0^\infty P(d_m > \varepsilon) d\varepsilon$$
$$= \int_{-\infty}^\infty \int_0^\infty P(d_m > \varepsilon | X = x)\, d\varepsilon \frac{1}{2\beta} e^{-\frac{|x|}{\beta}} dx$$
$$= \int_{-\infty}^\infty \int_0^\infty P(X < x - \varepsilon)^m\, d\varepsilon \frac{1}{2\beta} e^{-\frac{|x|}{\beta}} dx$$
$$= \int_0^\infty \int_0^\infty P(X < x - \varepsilon)^m\, d\varepsilon \frac{1}{\beta} e^{-\frac{x}{\beta}} dx$$
$$\geq \int_0^\infty \int_0^\infty P(X < x - \varepsilon)^m\, d\varepsilon \frac{1}{\beta} e^{-\frac{x}{\beta}} dx$$
$$= \int_0^\infty \int_0^x P(X < z)^m\, dz \frac{1}{\beta} e^{-\frac{x}{\beta}} dx$$

$$= \int_0^\infty \int_z^\infty \frac{1}{\beta} e^{-\frac{x}{\beta}} dx \left(1 - e^{-\frac{z}{\beta}}\right)^m dz$$
$$= \int_0^\infty e^{-\frac{z}{\beta}} \left(1 - e^{-\frac{z}{\beta}}\right)^m dz$$
$$= \frac{\beta}{(m+1)}. \qquad (4.5)$$

5 Conclusion

We found upper and lower bounds on the expected nearest neighbor distance for distributions having unbounded supports$S = (-\infty, \infty)$, and induced lower and upper bounds for logistic and Laplace distributions, as typical. Then we found the risk of nearest neighbor of their distributions. Note that, from $(3.5),(3.6)$ and $(4.3),(4.5)$ the lower and upper bounds of the expected nearest neighbor distance are different in constants, extra term depend on m and very small for large m, and the term $\log m$ and $\log m^{\frac{2\beta}{2-\beta}}$ respectively. That is, for the distributions have exponentially decaying tails there is an additional logarithmic term over the rates for compact support. This example illustrates that the expected nearest neighbor distance depends on the tails of the distribution.

Acknowledgements

The author would like to thank the anonymous referees for their useful comments, suggestions and corrections for improving the paper.

Competing Interests

Author has declared that no competing interests exist.

References

[1] Dasarathy B. Nearest neighbor classification techniques. IEEE, Los Alamitos, CA; 1991.

[2] Cover TM, Hart PE. Nearest neighbor pattern classification. IEEE Transactions on Information Theory. 1967;13:21-27.

[3] Cover TM. Rates of convergence for nearest neighbor procedures. In Proceedings of the Hawaii International Conference on Systems Sciences. 1968;413-415. Honolulu, HI.

[4] Kulkarni SR, Posner SE. Rates of convergence of nearest neighbor estimation under arbitrary sampling. IEEE Transactions on Information Theory. 1995;41:1028-1039.

[5] Evans D, Jones AJ, Schmidt WM. Asymptotic moments of near neighbor distance distributions. Proc. R. Soc. Lond. A. 2002;458:2839-2849.

[6] Irle A, Rizk M. On the risk of nearest neighbor rules. Dsserationzur Erlangung des Doktorgrades; 2004.

[7] Liitiäinen E, Lendasse A, Corona F. A boundary corrected expansion of the moments of nearest neighbor distributions. Random Structures and Algorithms. 2010;37:223-247.

[8] Rizk MM, Ateya SF. On lower and upper bounds for the risk in pattern recognition. Applied Mathematical Sciences. 2013;7:6417-6428.

[9] Devroye L, Györfi L, Lugosi G. Probabilistic theory of pattern recognition. Springer-Verlag, New York; 1996.

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

Peer-review history:
The peer review history for this paper can be accessed here (Please copy paste the total link in your browser address bar)
www.sciencedomain.org/review-history.php?iid=734&id=6&aid=7675