Rates of convergence for nearest neighbor estimators with the smoother regression function

Takanori Ayano[*]

Department of Mathematics, Osaka University

Abstract

Let (X, Y) be a $\mathbb{R}^d \times \mathbb{R}$-valued random vector. In regression analysis one wants to estimate the regression function $m(x) := \mathbf{E}(Y|X = x)$ from a data. In this paper we consider the rate of convergence for the k nearest neighbor estimator in case that X is uniformly distributed on $[0, 1]^d$, $\mathbf{Var}(Y|X = x)$ is bounded, and m is (p, C)-smooth. It is an open problem whether the optimal rate can be achieved by some k nearest neighbor estimator in case of $1 < p \leq 1.5$. We solve the problem affirmatively. This is the main result of this paper. Throughout this paper, we assume that the data is independent and identically distributed and as an error criterion we use the expected L_2 error.

Keywords

Regression, Nonparametric estimation, Nearest neighbor, Rate of convergence,

1 Introduction

Let (X, Y) be a $\mathbb{R}^d \times \mathbb{R}$-valued random vector. In regression analysis, one wants to predict the value of Y after having observed the value of X, i.e. to find a measurable function f such that the mean squared error $\mathbf{E}_{XY} (f(X) - Y)^2$ is minimized, where $\mathbf{E}_{XY}$ denotes the expectation with respect to (X, Y). Let $m(x) := \mathbf{E}\{Y|X = x\}$ (regression function), which is the conditional expectation of Y given $X = x$. Then $m(x)$ is the solution of the minimization problem. In fact, one can check for any measurable function f,

$$\mathbf{E}_{XY} (f(X) - Y)^2 = \mathbf{E}_{XY} (m(X) - Y)^2 + \mathbf{E}_X (f(X) - m(X))^2.$$

In statistics, only the data is available, (the distribution of (X, Y) and m are not available), and one needs to estimate the function m from the data $\{(X_i, Y_i)\}_{i=1}^n$, which are independently distributed according to the distribution of (X, Y). We wish to construct an estimator m_n of m such that the expected L_2 error

---

[*]t-ayano@cr.math.sci.osaka-u.ac.jp

$R(m_n) := \mathbf{E}_{X^n Y^n} \mathbf{E}_X \left(m_n(X) - m(X) \right)^2$ is as small as possible, where $\mathbf{E}_{X^n Y^n}$ denotes the expectation with respect to the data. In order to analyze the performance of estimators theoretically, it is very important to evaluate how fast the error $R(m_n)$ converges to zero, when the data size n tends to infinity. In this paper we consider k-NN (nearest neighbor) estimators and the rate of convergence in case that m is (p, C)-smooth (cf.Györfi et al., 2002, p.37).

The k-NN estimator is defined as follows. Given $x \in \mathbb{R}^d$, we rearrange the data $(X_1, Y_1), \ldots, (X_n, Y_n)$ in the ascending order of the values of $\|X_i - x\|$. As a tie-breaking rule, if $\|X_i - x\| = \|X_j - x\|$ and $i < j$, we declare that X_i is "closer" to x than X_j. We write the rearrange sequence by $(X_{1,x}, Y_{1,x}), \ldots, (X_{n,x}, Y_{n,x})$. Notice that $\{(X_{i,x}, Y_{i,x})\}_{i=1}^n$ is expressed by $\{(X_{\pi(i)}, Y_{\pi(i)})\}_{i=1}^n$ using a permutation $\pi : \{1, \ldots, n\} \to \{1, \ldots, n\}$ depending on $x \in \mathbb{R}^d$. Then for $1 \leq k \leq n$, the k-NN estimator m_n is defined by

$$m_n(x) = \frac{1}{k} \sum_{i=1}^k Y_{i,x}$$

For the details about k-NN estimators, for example, see Chapter 6 in Györfi et al. (2002).

Let $p, C > 0$, and express p by $p = q + r$, $q \in \mathbb{Z}_{\geq 0}$, $0 < r \leq 1$. We say that a function $m : \mathbb{R}^d \to \mathbb{R}$ is (p, C)-smooth if for all $q_1, \ldots, q_d \in \mathbb{Z}_{\geq 0}$ with $q = q_1 + \cdots + q_d$, the partial derivatives $\frac{\partial^q m}{\partial x_1^{q_1} \cdots \partial x_d^{q_d}}$ exist and for all $x, z \in \mathbb{R}^d$ the following is satisfied.

$$\left| \frac{\partial^q m}{\partial x_1^{q_1} \cdots \partial x_d^{q_d}}(x) - \frac{\partial^q m}{\partial x_1^{q_1} \cdots \partial x_d^{q_d}}(z) \right| \leq C \|x - z\|^r$$

For $p, C, \sigma > 0$, let $\mathcal{D}(p, C, \sigma)$ be the class of distributions of (X, Y) such that:

(I) X is uniformly distributed on $[0, 1]^d$;

(II) $\mathbf{Var}(Y|X = x) \leq \sigma^2$;

(III) m is (p, C)-smooth,

where $\mathbf{Var}(Y|X = x)$ denotes the variance of Y given $X = x$.

The lower bound for the class $\mathcal{D}(p, C, \sigma)$ is known (cf.Györfi et al., 2002, p.38):

$$\liminf_{n \to \infty} \inf_{m_n} \sup_{(X,Y) \in \mathcal{D}(p,C,\sigma)} n^{2p/(2p+d)} R(m_n) \geq \text{const.} > 0 \tag{1}$$

where $\inf_{m_n}$ denotes the infimum over all the estimators.

For $0 < p \leq 1$, the rate $n^{-2p/(2p+d)}$ is achieved by the k-NN estimator (cf.Györfi et al., 2002, pp.93,99):

$$\sup_{(X,Y) \in \mathcal{D}(p,C,\sigma)} R(m_n) \leq \text{const.} \; n^{-2p/(2p+d)}$$

For $p > 1.5$, it is shown that the rate $n^{-2p/(2p+d)}$ is unachievable by any k-NN estimator and it is presented as a conjecture that even for $1 < p \leq 1.5$, the rate $n^{-2p/(2p+d)}$ will be achieved by some k-NN estimator (cf.Györfi et al., 2002, p.96). In this paper, we show that the conjecture is right (Theorem). Regression analysis is used in many fields for example economics, medicine, pattern recognition etc. (cf.Györfi et al., 2002, pp.4-9). Nearest neighbor estimators are very important in regression analysis. We have shown the performance of the nearest neighbor estimator theoretically.

Throughout this paper we will use the following notations : $\mathbb{R}, \mathbb{R}_{>0}, \mathbb{Z}_{\geq 0}, \mathbb{N}$ are the sets of reals, positive reals, nonnegative integers and positive intgers. For a measurable set $D \subset \mathbb{R}^d$, $\mathrm{vol}(D)$ denotes the Lebesgue measure of D. For $x \in \mathbb{R}^d$, $\|x\|$ denotes the Euclidean norm of x. For $u, v \in \mathbb{R}^d$, we define $H(u, v) := \{w \in \mathbb{R}^d \mid \|w - u\| \leq \|v - u\|\}$ and $G(u, v) := H(u, v) \cap [0, 1]^d$. For $a > 0$, $\lfloor a \rfloor$ denotes b such that $b \leq a < b + 1$.

2 Related Work

In this section, we overview the related work about consistency and the rate of convergence. For consistency, it was shown in Stone (1977) that the k-NN estimators are universally consistent. Since then it was shown that many estimators share this property (cf.Devroye et al.,1994, Greblicki et al.,1984, Györfi and Walk, 1997, Kohler, 1999, Kohler and Krzyżak, 2001, Kohler, 2002, Lugosi and Zeger, 1995, Nobel, 1996, Walk, 2002, Walk, 2005, Walk, 2008). For the rate of convergence, we know several results as follow:

- Stone (1982) proved the lower bound (1);

- for the distributions satisfying (II)(III) with $0 < p \leq 1$ and the partitioning, kernel, and k-NN estimator, the rate $n^{-2p/(2p+d)}$ is achievable if X is bounded (for the k-NN estimator, the condition $d > 2p$ is required as well) (cf.Györfi, 1981, Györfi et al., 2002, Kulkarni and Posner, 1995, Spiegelman and Sacks, 1980);

- Kohler et al.(2006, 2009) proved the same statement later without assuming that X should be bounded;

- for the partitioning estimators and the class $\mathcal{D}(p, C, \sigma)$ with $p > 1$, the rate $n^{-2p/(2p+d)}$ is unachievable (cf.Györfi et al., 2002);

- for the kernel estimators, the rate $n^{-2p/(2p+d)}$ is achievable for $\mathcal{D}(p, C, \sigma)$ with $0 < p \leq 1.5$ and is unachievable for that with $p > 1.5$ (cf.Györfi et al., 2002);

If we summarize the above results in Table 1, only the following problem remains: Does the k-NN estimator achieve the rate $n^{-2p/(2p+d)}$ under (II)(III) even for $1 < p \leq 1.5$? The problem is still hard, but we solve the statement affirmatively under (I)(II)(III).

Table 1 : the achievability of $n^{-2p/(2p+d)}$ for the estimators and $\mathcal{D}(p, C, \sigma)$

	achievable	unachievable
partitioning	$0 < p \le 1$	$p > 1$
kernel	$0 < p \le 1.5$	$p > 1.5$
k-NN	$0 < p \le 1$	$p > 1.5$

3 Main Result

For $p, C, \sigma > 0$, let $\mathcal{D}(p, C, \sigma)$ be the class of distributions of (X, Y) such that:

(I) X is uniformly distributed on $[0, 1]^d$;

(II) $\mathbf{Var}(Y|X = x) \le \sigma^2$;

(III) m is (p, C)-smooth,

where $\mathbf{Var}(Y|X = x)$ denotes the variance of Y given $X = x$.

Then we get the following theorem:

Theorem
Let $1 < p \le 1.5$ and let m_n be the k-NN estimator with $k = \lfloor n^{2p/(2p+d)} \rfloor$. Then there exists $C_1 > 0$ (which does not depend on n) such that

$$\sup_{(X,Y)\in\mathcal{D}(p,C,\sigma)} \mathbf{E}_{X^n Y^n} \mathbf{E}_X \left(m_n(X) - m(X) \right)^2 \le C_1 n^{-2p/(2p+d)}.$$

4 Proof of Theorem

Suppose we are given $X = x, X_1 = x_1, \ldots, X_n = x_n$. We take the expectation with respect to $Y_1, \ldots, Y_n$. Then the following bias-variance decomposition is well-known (cf.Györfi et al., 2002, p.94):

$$\mathbf{E}_{Y^n} \left(m_n(x) - m(x) \right)^2 = \mathbf{E}_{Y^n} \left(\frac{1}{k} \sum_{i=1}^{k} (Y_{i,x} - m(x)) \right)^2$$

$$= \mathbf{E}_{Y^n} \left(\frac{1}{k} \sum_{i=1}^{k} (Y_{i,x} - m(x_{i,x})) \right)^2 + \left\{ \frac{1}{k} \sum_{i=1}^{k} (m(x_{i,x}) - m(x)) \right\}^2$$

$$\le \frac{\sigma^2}{k} + \frac{1}{k^2} \left\{ \sum_{i=1}^{k} (m(x_{i,x}) - m(x)) \right\}^2 . \quad (\because \text{(II)}) \tag{2}$$

We evaluate the second term of (2). Let $x_{i,x} = (x_{i,x}^{(1)}, \ldots, x_{i,x}^{(d)})$ and $x = (x^{(1)}, \ldots, x^{(d)})$. Let m_s be the partial derivative of m with respect to the s-th component. Then by the mean-value theorem, there exists $u_i \in \mathbb{R}^d$ such that $\|u_i - x\| \le \|x_{i,x} - x\|$ and

$$\left\{ \sum_{i=1}^{k} (m(x_{i,x}) - m(x)) \right\}^2 = \left\{ \sum_{i=1}^{k} \sum_{s=1}^{d} m_s(u_i)(x_{i,x}^{(s)} - x^{(s)}) \right\}^2,$$

4

(the idea using the mean-value theorem is due to Györfi et al., 2002, p.84) by Cauchy-Schwarz's inequality

$$\leq 2\left\{\sum_{i=1}^{k}\sum_{s=1}^{d}(m_s(u_i)-m_s(x))(x_{i,x}^{(s)}-x^{(s)})\right\}^2 + 2\left\{\sum_{s=1}^{d}m_s(x)\sum_{i=1}^{k}(x_{i,x}^{(s)}-x^{(s)})\right\}^2$$

$$\leq 2kd\sum_{i=1}^{k}\sum_{s=1}^{d}(m_s(u_i)-m_s(x))^2(x_{i,x}^{(s)}-x^{(s)})^2 + 2d\sum_{s=1}^{d}m_s(x)^2\left\{\sum_{i=1}^{k}(x_{i,x}^{(s)}-x^{(s)})\right\}^2,$$

let $L > 0$ such that $\max_{1\leq s\leq d, x\in[0,1]^d}|m_s(x)| \leq L$, because m is (p,C)-smooth and $\|u_i - x\| \leq \|x_{i,x} - x\|$,

$$\leq 2kdC^2\sum_{i=1}^{k}\sum_{s=1}^{d}\|x_{i,x}-x\|^{2p-2}(x_{i,x}^{(s)}-x^{(s)})^2 + 2dL^2\sum_{s=1}^{d}\left\{\sum_{i=1}^{k}(x_{i,x}^{(s)}-x^{(s)})\right\}^2$$

$$= 2kdC^2\sum_{i=1}^{k}\|x_{i,x}-x\|^{2p} + 2dL^2\sum_{i=1}^{k}\|x_{i,x}-x\|^2$$

$$+2dL^2\sum_{s=1}^{d}\sum_{1\leq i\neq j\leq k}(x_{i,x}^{(s)}-x^{(s)})(x_{j,x}^{(s)}-x^{(s)})$$

We regard $x, x_1, \ldots, x_n$ as the random variables $X, X_1, \ldots, X_n$ and take the expectation with respect to $X, X_1, \ldots, X_n$.

$$\mathbf{E}_X\mathbf{E}_{X^nY^n}\left(m_n(X)-m(X)\right)^2$$

$$\leq \frac{\sigma^2}{k} + \frac{2dC^2}{k}\mathbf{E}_X\mathbf{E}_{X^n}\sum_{i=1}^{k}\|X_{i,X}-X\|^{2p} + \frac{2dL^2}{k^2}\mathbf{E}_X\mathbf{E}_{X^n}\sum_{i=1}^{k}\|X_{i,X}-X\|^2 \quad (3)$$

$$+\frac{2dL^2}{k^2}\mathbf{E}_X\mathbf{E}_{X^n}\sum_{s=1}^{d}\sum_{1\leq i\neq j\leq k}\left(X_{i,X}^{(s)}-X^{(s)}\right)\left(X_{j,X}^{(s)}-X^{(s)}\right) \quad (4)$$

In order to evaluate the second and third terms in (3), the following proposition is available.

Proposition (Györfi et al., 2002, pp.95,99)
For any $\gamma > 0$, there exists $c_1 > 0$ (depending on γ and d) such that,

$$\frac{1}{k}\mathbf{E}_X\mathbf{E}_{X^n}\sum_{i=1}^{k}\|X_{i,X}-X\|^{2\gamma} \leq c_1\left(\frac{k}{n}\right)^{2\gamma/d}.$$

The proposition is proved originally for $\gamma = 1$ in Györfi et al., 2002, but we have extended it to the general $\gamma > 0$. We proceed to evaluate (4).

Let $D = \{(x, x_1, \ldots, x_n) \mid \|x_i - x\| < \|x_{k+1} - x\|, i = 1, \ldots, k, \|x_j - x\| > \|x_{k+1} - x\|, j = k+2, \ldots, n\}$.

5

Claim 1

$$\mathbf{E}_X \mathbf{E}_{X^n} \sum_{1 \le i \ne j \le k} \left(X_{i,X}^{(s)} - X^{(s)} \right) \left(X_{j,X}^{(s)} - X^{(s)} \right)$$

$$= \frac{n \cdots (n-k)}{k!} \int_D \sum_{1 \le i \ne j \le k} \left(x_i^{(s)} - x^{(s)} \right) \left(x_j^{(s)} - x^{(s)} \right) dx_1 \cdots dx_n dx$$

(See Appendix for proof)

From Claim 1,
since $x_1, \ldots, x_k \in G(x, x_{k+1})$ and $x_{k+2}, \ldots, x_n \in [0,1]^d \backslash G(x, x_{k+1})$ on D,

$$T := \frac{1}{k^2} \mathbf{E}_X \mathbf{E}_{X^n} \sum_{s=1}^d \sum_{1 \le i \ne j \le k} \left(X_{i,X}^{(s)} - X^{(s)} \right) \left(X_{j,X}^{(s)} - X^{(s)} \right)$$

$$= \frac{n \cdots (n-k)}{k^2 k!} \sum_{s=1}^d \sum_{1 \le i \ne j \le k} \int_D \left(x_i^{(s)} - x^{(s)} \right) \left(x_j^{(s)} - x^{(s)} \right) dx_1 \cdots dx_n dx,$$

$$= \frac{n \cdots (n-k)}{k^2 k!} \sum_{s=1}^d \sum_{1 \le i \ne j \le k} \int_{[0,1]^d} dx \int_{[0,1]^d} dx_{k+1} \int_{G(x,x_{k+1})} (x_i^{(s)} - x^{(s)}) dx_i$$

$$\int_{G(x,x_{k+1})} (x_j^{(s)} - x^{(s)}) dx_j \cdot \mathrm{vol}[G(x,x_{k+1})]^{k-2} (1 - \mathrm{vol}[G(x,x_{k+1})])^{n-k-1}.$$

Let $W := \{ (x, x_{k+1}) \mid G(x, x_{k+1}) \ne H(x, x_{k+1}) \}$. Since for $(x, x_{k+1}) \notin W$, $\int_{G(x,x_{k+1})} (x_i^{(s)} - x^{(s)}) dx_i = \int_{H(x,x_{k+1})} (x_i^{(s)} - x^{(s)}) dx_i = 0$, we obtain

$$T = \sum_{s=1}^d \frac{n \cdots (n-k)}{k^2 (k-2)!} \int_W dx_{k+1} dx \left(\int_{G(x,x_{k+1})} (x_1^{(s)} - x^{(s)}) dx_1 \right)^2$$

$$\mathrm{vol}[G(x,x_{k+1})]^{k-2} (1 - \mathrm{vol}[G(x,x_{k+1})])^{n-k-1}.$$

Claim 2   There exists $c_2 > 0$ (depending only on d) such that

$$\left| \int_{G(x,x_{k+1})} (x_1^{(s)} - x^{(s)}) \, dx_1 \right| \le c_2 \cdot \mathrm{vol}[G(x,x_{k+1})]^{(d+1)/d}$$

(See Appendix for proof)

From Claim 2,

$$T \le c_2^2 d \frac{n \cdots (n-k)}{k^2 (k-2)!} \int_W \mathrm{vol}[G(x,x_{k+1})]^{k+\frac{2}{d}} \{ 1 - \mathrm{vol}[G(x,x_{k+1})] \}^{n-k-1} dx_{k+1} dx$$

$$= c_2^2 d \frac{n \cdots (n-k)}{k^2 (k-2)!} \int_0^1 u^{k+\frac{2}{d}} (1-u)^{n-k-1} dF(u) \tag{5}$$

where $F(u)$ is the Lebesgue measure of $S(u) := \{(x, x_{k+1}) \in W \mid 0 \leq \text{vol}[G(x, x_{k+1})] \leq u\}$ for $0 \leq u \leq 1$.

Claim 3

$$F(u) = \begin{cases} 2d\sum_{i=0}^{d-1}\left\{\dfrac{{}_{d-1}C_i(-2)^{d-1-i}}{(d-i)e_2^{(d-i)/d}} - \dfrac{{}_{d-1}C_i(-2)^{d-1-i}}{(2d-i)e_2^{(d-i)/d}}\right\}u^{(2d-i)/d} & \left(0 \leq u \leq \dfrac{e_2}{2^d}\right) \\ u - 2de_2\displaystyle\int_0^{1/2}(x^{(1)})^d(1 - 2x^{(1)})^{d-1}dx^{(1)} & \left(\dfrac{e_2}{2^d} \leq u \leq 1\right) \end{cases}$$

where $e_2 = \text{vol}[\{y \in \mathbb{R}^d \mid \|y\| \leq 1\}]$, and $e_2 \leq 2^d$.

(See Appendix for proof)

For $0 < u < e_2/2^d$ and $e_2/2^d < u < 1$, let $f(u) := F'(u)$. ($f(u) \geq 0$)

$$f(u) = \begin{cases} \displaystyle\sum_{i=0}^{d-1}(4d-2i)\left\{\dfrac{{}_{d-1}C_i(-2)^{d-1-i}}{(d-i)e_2^{(d-i)/d}} - \dfrac{{}_{d-1}C_i\,(-2)^{d-1-i}}{(2d-i)e_2^{(d-i)/d}}\right\}u^{(d-i)/d} & \left(0 < u < \dfrac{e_2}{2^d}\right) \\ 1 & \left(\dfrac{e_2}{2^d} < u < 1\right) \end{cases}$$

Let $f(0) = f(e_2/2^d) = f(1) = 0$. There exists $c_3 > 0$ (depending only on d) such that $f(u) \leq c_3 u^{1/d}$, because, for $e_2/2^d < u < 1$, $f(u) = 1 \leq (2/e_2^{1/d})u^{1/d}$ and for the other u it is trivial.

For $\alpha \in \mathbb{R}_{>0}$ and $\beta \in \mathbb{N}$, let $B(\alpha, \beta) := \displaystyle\int_0^1 u^{\alpha-1}(1 - u)^{\beta-1}du$ (Beta function). Then the following formula is well-known:

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)} = \frac{\Gamma(\alpha)\,(\beta - 1)!}{(\alpha + \beta - 1)\cdots\alpha\Gamma(\alpha)} = \frac{(\beta - 1)!}{\alpha\cdots(\alpha + \beta - 1)},$$

where Γ is Gamma function.

On the other hand,

$$\lim_{n\to\infty}\frac{n!}{(1 + \frac{3}{d})\cdots(n + \frac{3}{d})}\cdot n^{\frac{3}{d}} = \lim_{n\to\infty}\frac{\Gamma(n + 1)\cdot\Gamma(1 + \frac{3}{d})}{\Gamma(n + 1 + \frac{3}{d})}\cdot n^{\frac{3}{d}}$$

By Stirling's formula,

$$= \lim_{n\to\infty}\Gamma(1 + \frac{3}{d})\frac{\sqrt{2\pi n}\left(\frac{n}{e}\right)^n}{\sqrt{2\pi(n + \frac{3}{d})}\left(\frac{n + \frac{3}{d}}{e}\right)^{n + \frac{3}{d}}}\cdot n^{\frac{3}{d}} = \Gamma(1 + \frac{3}{d})$$

Therefore, there exist $c_4, c_5 > 0$ (depending only on d) such that

$$c_4 n^{-\frac{3}{d}} \leq \frac{n!}{(1 + \frac{3}{d})\cdots(n + \frac{3}{d})} \leq c_5 n^{-\frac{3}{d}}$$

7

From (5),

$$T \le c_2^2 d \frac{n \cdots (n-k)}{k^2(k-2)!} \int_0^1 u^{k+\frac{2}{d}} (1-u)^{n-k-1} f(u) du$$

$$\le c_2^2 c_3 d \frac{n \cdots (n-k)}{k^2(k-2)!} \int_0^1 u^{k+\frac{3}{d}} (1-u)^{n-k-1} du$$

$$= c_2^2 c_3 d \frac{n \cdots (n-k)}{k^2(k-2)!} B\left(k+1+\frac{3}{d}, \; n-k\right)$$

$$= c_2^2 c_3 d \frac{n \cdots (n-k)}{k^2(k-2)!} \frac{(n-k-1)!}{(k+1+\frac{3}{d}) \cdots (n+\frac{3}{d})} \le c_2^2 c_3 d \frac{(k+1) \cdots n}{(k+1+\frac{3}{d}) \cdots (n+\frac{3}{d})}$$

$$= c_2^2 c_3 d \frac{n!}{(1+\frac{3}{d}) \cdots (n+\frac{3}{d})} / \frac{k!}{(1+\frac{3}{d}) \cdots (k+\frac{3}{d})} \le \frac{c_2^2 c_3 c_5 d}{c_4} \left(\frac{k}{n}\right)^{3/d} \tag{6}$$

Therefore, from (3), (4), (6), and Proposition, there exist $C_2, C_3, C_4 > 0$ (which do not depend on n) such that

$$\mathbf{E}_X \mathbf{E}_{X^n Y^n} (m_n(X) - m(X))^2 \le \frac{\sigma^2}{k} + C_2 \left(\frac{k}{n}\right)^{2p/d} + \frac{C_3}{k} \left(\frac{k}{n}\right)^{2/d} + C_4 \left(\frac{k}{n}\right)^{3/d}$$

Assuming $p \le 1.5$, if we set $k = \lfloor n^{2p/(2p+d)} \rfloor$, there exists $C_1 > 0$ (which does not depend on n) such that

$$\mathbf{E}_X \mathbf{E}_{X^n Y^n} (m_n(X) - m(X))^2 \le C_1 n^{-2p/(2p+d)}$$

We have got Theorem.

$\square$

Appendix

A   Proof of Claim 1

Let $h \in N := \{1, \ldots, n\}$ and $I, J \subset N \backslash \{h\}$ such that $\sharp I = k, I \cap J = \{\}(empty), I \cup J = N \backslash \{h\}$, where, $\sharp$ denotes the number of the elements. Let $D(I, J, h) := \{(x, x_1, \ldots, x_n) \mid \|x_i - x\| < \|x_h - x\|, i \in I, \|x_j - x\| > \|x_h - x\|, j \in J\}$. Since $\mathrm{vol}\left\{[0,1]^{d(n+1)} \backslash \cup_{I,J,h} D(I,J,h)\right\} = 0$ and for $(I, J, h) \ne (I', J', h'), D(I, J, h) \cap D(I', J', h') = \{\}$, we have

$$\mathbf{E}_X \mathbf{E}_{X^n} \sum_{1 \le i \ne j \le k} \left(X_{i,X}^{(s)} - X^{(s)}\right) \left(X_{j,X}^{(s)} - X^{(s)}\right)$$

$$= \sum_{I,J,h} \int_{D(I,J,h)} \sum_{i,j \in I, i \ne j} \left(x_i^{(s)} - x^{(s)}\right) \left(x_j^{(s)} - x^{(s)}\right) dx_1 \cdots dx_n dx$$

Since for each (I, J, h) the above integral has the same value and the number of (I, J, h) is $_nC_k \cdot (n-k)$, we get Claim 1.

$\square$

B Proof of Claim 2

Let $e_1 := \int_{\|y\|\leq 1,\, y^{(s)}\geq 0} y^{(s)} dy$ and $e_2 := \int_{\|y\|\leq 1} dy$, then for any $R \geq 0$,

$$\int_{\|y\|\leq R,\, y^{(s)}\geq 0} y^{(s)} dy = e_1 R^{d+1},$$

and

$$\int_{\|y\|\leq R} dy = e_2 R^d, \tag{7}$$

thus we have

$$\left| \int_{G(x,x_{k+1})} (x_1^{(s)} - x^{(s)}) dx_1 \right| \leq \int_{H(x,x_{k+1})} |x_1^{(s)} - x^{(s)}| dx_1 = 2 \int_{\|y\|\leq \|x-x_{k+1}\|,\, y^{(s)}\geq 0} y^{(s)} dy$$

$$= 2e_1 \|x - x_{k+1}\|^{d+1} = \left\{ \frac{(2e_1)^{d/(d+1)}}{e_2} \mathrm{vol}[H(x,x_{k+1})] \right\}^{(d+1)/d}$$

If we prove the following lemma, the proof of Claim 2 is complete:

Lemma
There exists e_3 (depending only on d) such that for any $u, v \in [0,1]^d$,

$$\mathrm{vol}[G(u,v)] \geq e_3 \mathrm{vol}[H(u,v)]$$

(Proof of Lemma)

Suppose $\|u - v\| \leq 1/2$. Let $I := \{i \mid 0 \leq u^{(i)} \leq 1/2\}$ and $M := \{w \mid \|w - u\| \leq \|u - v\|,\ w^{(i)} \geq u^{(i)},\ i \in I,\ w^{(j)} \leq u^{(j)},\ j \notin I\}$. Then, $M \subset G(u,v)$ and $\mathrm{vol}[M] = 2^{-d}\mathrm{vol}[H(u,v)]$, thus,

$$\mathrm{vol}[G(u,v)] \geq 2^{-d} \mathrm{vol}[H(u,v)] \tag{8}$$

Suppose $\|u - v\| > 1/2$. Since $\|u - v\| \leq \sqrt{d}$, we have $\mathrm{vol}[H(u,v)] \leq e_2 d^{d/2}$. From (8), for $z \in \mathbb{R}^d$ such that $\|u - z\| = 1/2$,

$$\mathrm{vol}[G(u,v)] \geq \mathrm{vol}[G(u,z)] \geq 2^{-d} \mathrm{vol}[H(u,z)] = 2^{-d} \left(e_2 2^{-d} \right).$$

$$\therefore \quad \frac{\mathrm{vol}[G(u,v)]}{\mathrm{vol}[H(u,v)]} \geq 2^{-2d} d^{-d/2}$$

Let $e_3 := \min\{2^{-d}, 2^{-2d} d^{-d/2}\}$, then we get Lemma.

$\square$

C Proof of Claim 3

Let
$$V_i := \left\{ x \in [0,1]^d \mid x^{(i)} \leq \min\{x^{(j)}, 1 - x^{(j)}\}, j = 1, \ldots d \right\}$$

$$V_{i+d} := \left\{ x \in [0,1]^d \mid 1 - x^{(i)} \leq \min\{x^{(j)}, 1 - x^{(j)}\}, j = 1, \ldots d \right\}.$$

Since $\cup_{i=1}^{2d} V_i = [0,1]^d$ and for $i \neq j$, $\mathrm{vol}[V_i \cap V_j] = 0$, by Fubini's theorem,

$$F(u) = \int_{S(u)} dx_{k+1} \, dx = \int_{[0,1]^d} \left\{ \int_{S(u)} dx_{k+1} \right\} dx = \sum_{i=1}^{2d} \int_{V_i} \left\{ \int_{S(u)} dx_{k+1} \right\} dx$$

Without loss of generality, we assume $x \in V_1$.
Let $y := (0, x^{(2)}, \ldots, x^{(d)})$. Since $H(x, y) \subset [0,1]^d$, for $u < \mathrm{vol}[H(x,y)]$,

$$0 \leq \mathrm{vol}[G(x, x_{k+1})] \leq u \implies G(x, x_{k+1}) = H(x, x_{k+1}) \implies (x, x_{k+1}) \notin W$$

$$\implies \{x_{k+1} \mid (x, x_{k+1}) \in S(u)\} = \{\}$$

For $\mathrm{vol}[H(x,y)] \leq u$ and $z \in \mathbb{R}^d$ such that $\mathrm{vol}[G(x,z)] = u$, we have

$$\{x_{k+1} \mid (x, x_{k+1}) \in S(u)\} = G(x, z) \backslash H(x, y)$$

and from (7), $\mathrm{vol}[G(x, z) \backslash H(x, y)] = u - e_2(x^{(1)})^d$.

$$U := \int_{V_1} \left\{ \int_{S(u)} dx_{k+1} \right\} dx = \int_{V_1} \max\{u - e_2(x^{(1)})^d, 0\} \, dx$$

$$= \int_0^{\min\left\{(u/e_2)^{1/d}, 1/2\right\}} \left\{ \left(u - e_2(x^{(1)})^d\right) \int_{x^{(1)}}^{1-x^{(1)}} dx^{(2)} \cdots \int_{x^{(1)}}^{1-x^{(1)}} dx^{(d)} \right\} dx^{(1)}$$

$$= \int_0^{\min\left\{(u/e_2)^{1/d}, 1/2\right\}} \left(u - e_2(x^{(1)})^d\right) (1 - 2x^{(1)})^{d-1} dx^{(1)}$$

For $(u/e_2)^{1/d} \leq 1/2$, i.e. $0 \leq u \leq e_2/2^d$, @

$$U = \int_0^{(u/e_2)^{1/d}} \left(u - e_2(x^{(1)})^d\right) (1 - 2x^{(1)})^{d-1} dx^{(1)}$$

$$= \int_0^{(u/e_2)^{1/d}} \left(u - e_2(x^{(1)})^d\right) \left\{ \sum_{i=0}^{d-1} {}_{d-1}C_i \, 1^i \, (-2x^{(1)})^{d-1-i} \right\} dx^{(1)}$$

$$= \sum_{i=0}^{d-1} \left\{ \frac{{}_{d-1}C_i (-2)^{d-1-i}}{(d-i)e_2^{(d-i)/d}} - \frac{{}_{d-1}C_i \, (-2)^{d-1-i}}{(2d-i) \, e_2^{(d-i)/d}} \right\} u^{(2d-i)/d}$$

For $1/2 \leq (u/e_2)^{1/d}$, i.e. $e_2/2^d \leq u \leq 1$,

$$U = \int_0^{1/2} \left(u - e_2(x^{(1)})^d \right) (1 - 2x^{(1)})^{d-1} dx^{(1)}$$

$$= \frac{u}{2d} - e_2 \int_0^{1/2} (x^{(1)})^d (1 - 2x^{(1)})^{d-1} dx^{(1)}$$

Now we have got Claim 3.

$\square$

Acknowledgement

References

[Devroye,L.,Györfi,L.,Krzyżak,A.,Lugosi,G.,(1994)] On the strong universal consistency of nearest neighbor regression function estimates. Ann.Statist.22,No.3,1371-1385.

[Greblicki,W.,Krzyżak,A.,Pawlak,M.,(1984)] Distribution-free pointwise consistency of kernel regression estimate. Ann.Statist.12,No.4,1570-1575.

[Györfi,L.,(1981)] The rate of convergence of k_n-NN regression estimates and classification rules. IEEE Trans.Inform.Theory 27,No.3,362-364.

[Györfi,L.,Walk,H.,(1997)] On the strong universal consistency of a recursive regression estimate by Pál Révész. Statist.Probab.Lett.31,177-183.

[Györfi,L.,Kohler,M.,Krzyżak,A.,Walk,H.,(2002)] A Distribution-Free Theory of Nonparametric Regression. Springer Series in Statistics. Springer,New York.

[Kohler,M.,(1999)] Universally consistent regression function estimation using hierarchial B-splines. J.Multivariate Anal.68,138-164.

[Kohler,M.,Krzyżak,A.,(2001)] Nonparametric regression estimation using penalized least squares. IEEE Trans.Inform.Theory 47,No.7,3054-3058.

[Kohler,M.,(2002)] Universal consistency of local polynomial kernel regression estimates. Ann.Inst.Statist.Math.54,No.4,879-899.

[Kohler,M.,Krzyżak,A.,Walk,H.,(2006)] Rates of convergence for partitioning and nearest neighbor regression estimates with unbounded data. J.Multivariate Anal.97,311-323.

[Kohler,M.,Krzyżak,A.,Walk,H.,(2009)] Optimal global rates of convergence for nonparametric regression with unbounded data. J.Statist.Plann.Inference 139,1286-1296.

[Kulkarni,S.R.,Posner,S.E.,(1995)] Rates of convergence of nearest neighbor estimation under arbitrary sampling. IEEE Trans.Inform.Theory 41,No.4,1028-1039.

[Lugosi,G.,Zeger,K.,(1995)] Nonparametric estimation via empirical risk minimization. IEEE Trans.Inform.Theory 41,No.3,677-687.

[Nobel,A.,(1996)] Histogram regression estimation using data-dependent partitions. Ann.Statist.24,No.3,1084-1105.

[Spiegelman,C.,Sacks,J.,(1980)] Consistent window estimation in nonparametric regression. Ann.Statist.8,No.2,240-246.

[Stone, C.J.,(1977)] Consistent nonparametric regression. Ann.Statist.5, No.4,595-645.

[Stone, C.J.,(1982)] Optimal global rates of convergence for nonparametric regression. Ann.Statist.10,No.4,1040-1053.

[Walk,H.,(2002)] Almost sure convergence properties of Nadaraya-Watson regression estimates, in:M.Dror,P.L'Ecuyer,F.Szidarovszky (Eds.),Modeling Uncertainty:An Examination of its Theory, Methods and Applications, Kluwer Academic Publishers, Dordrecht, pp.201-223.

[Walk,H.,(2005)] Strong universal consistency of smooth kernel regression estimates. Ann.Inst.Statist.Math.57,No.4,665-685.

[Walk,H.,(2008)] A universal strong law of large numbers for conditional expectations via nearest neighbors. J.Multivariate.Anal.99,1035-1050.