The strong universal consistency of the kernel estimate & the k-NN estimate

— Seminar: Statistical Learning —

Valentin Pfisterer, Tobias Winkler

University of Tübingen

May 19, 2023

Where are we?

	Partitioning	**Kernel**	**k-NN**
weak universal consistency	✓	✓	✓
strong (universal) consistency	✓	**?**	**?**

Table of Contents

Quick reminder

Kernel & k-NN estimate

Definition (kernel estimate)

$$m_n(x) = \frac{\sum_{i=1}^{n} K(\frac{x-X_i}{h}) Y_i}{\sum_{i=1}^{n} K(\frac{x-X_i}{h})}$$

Definition (k-NN estimate)

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

Strong (universal) consistency

Definition (Strong (universal) consistency)

Strong consistency

$$\lim_{n\to\infty} \int |m_n(x) - m(x)|^2 \mu(\mathrm{d}x) = 0 \quad \text{with probability one.}$$

Strong universal consistency

$$\lim_{n\to\infty} \int |m_n(x) - m(x)|^2 \mu(\mathrm{d}x) = 0$$

for all distributions of (X, Y) with $\mathbb{E}(Y^2) < \infty$.

Kernel Estimates

Regular kernels

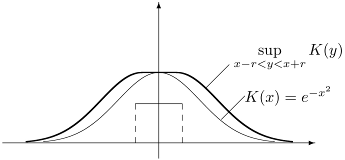
Definition (regular kernels)

Kernel K **regular**: $\iff$
K is non-negative and $\exists\ B_r(0)$ with $r > 0$ and $b > 0$, such that

$$1 \geqslant K(x) \geqslant b\mathbb{1}_{\{x \in B_r(0)\}}$$

and

$$\int \sup_{u \in x + B_r(0)} K(u)\,\mathrm{d}x < \infty.$$



Strong consistency of the kernel estimate

Theorem 2.1 (strong consistency)

Let m_n be the kernel estimator of the regression function m with a regular kernel K. Assume $\exists L < \infty$: $P(|Y| \leqslant L) = 1$. If

$$h_n \to 0 \quad and \quad nh_n^d \to \infty,$$

then the kernel estimate is **strongly consistent**.

Proof components (overview)

Theorem
(strong consistency of the kernel estimator)
$$\lim_{n\to\infty} \int |m_n(x) - m(x)|^2 d\mu(x) = 0$$

Lemma 1
$$\int \frac{K_h(x-u)}{\int K_h(x-z)d\mu(z)} \leqslant \varrho.$$

Lemma 2
$$\int_{B_R(0)} \frac{1}{\sqrt{\mu(B_h(x))}} d\mu(x) \leqslant \left(1 + \frac{R}{h}\right)^{d/2} c_d$$

Lemma 4
$$\mathbb{P}\left(\int |m(x) - m_n^*(x)| d\mu(x) > \varepsilon\right) \leqslant e^{-n\varepsilon^2/(8L\varrho)^2}$$

Lemma 3
$$\lim_{n\to\infty} \int \mathbb{E}(|m(x) - m_n^*(x)|)\, d\mu(x) = 0$$

Banach-Steinhaus
$$\lim_{n\to\infty} \int |m(x) - \int K_n(x,z) m(z) d\mu(z)| d\mu(x) = 0$$

modificated kernel
$$K_n(x,z) := \frac{K(\frac{x-z}{h_n})}{\int K(\frac{x-u}{h_n}) d\mu(u)}$$

modificated estimator
$$m_n^*(x) := \frac{\sum_{i=1}^n Y_i K_{h_n}(x - X_i)}{n\mathbb{E}(K_{h_n}(x - X))}$$

McDiarmid‘s inequality
$$\mathbb{P}(f(Z_1, ..., Z_n) - \mathbb{E}((f(Z_1, ..., Z_n))) \geqslant \varepsilon) \leqslant e^{-2\varepsilon^2/\sum_{i=1}^n c_i^2}$$

Proof Components 1

Lemma 2.1 (Covering Lemma)

Kernel K is regular. $\Longrightarrow$
$\exists \varrho \equiv \varrho(K) < \infty$, such that $\forall u \in \mathbb{R}^d$, $h > 0$ and probability measure μ

$$\int \frac{K_h(x-u)}{\int K_h(x-z)\mu(dz)} \mu(\mathrm{d}x) \leqslant \varrho.$$

Moreover, $\forall \delta > 0$

$$\lim_{n \to \infty} \sup_{u \in \mathbb{R}^d} \int \frac{K_h(x-u)\mathbb{1}_{\{\|x-u\|>\delta\}}}{\int K_h(x-z)\mu(dz)} \mu(\mathrm{d}x) = 0.$$

Proof Components 1

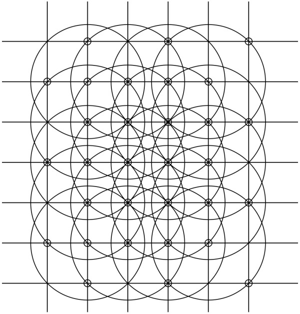


Figure: An example of a bounded overlap of $\mathbb{R}^2$.

Proof Components 2

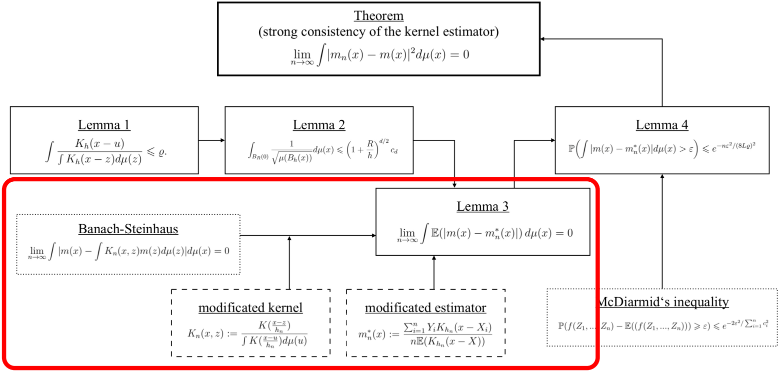
Lemma 2.2

Let $1 \leqslant R < \infty$, $0 < h \leqslant R$, $B_R(0) \subseteq \mathbb{R}^d$ ball of Radius R.
$\Longrightarrow$
For every probability measure μ,

$$\int_{B_R(0)} \frac{1}{\sqrt{\mu(B_h(x))}} \mu(\mathrm{d}x) \leqslant \left(1 + \frac{R}{h}\right)^{d/2} c_d,$$

where c_d depends upon the dimension d only.

Proof components 3



Proof Components 3

Definition (modificated estimate)

Define

$$m_n^*(x) := \frac{\sum_{i=1}^n Y_i K_{h_n}(x - X_i)}{n\mathbb{E}(K_{h_n}(x - X))}.$$

Lemma 2.3

Under the conditions of Theorem 1,

$$\lim_{n\to\infty} \int \mathbb{E}(|m(x) - m_n^*(x)|)\, \mu(\mathrm{d}x) = 0.$$

Proof Components 3

Definition (modificated kernel-function)

Define

$$K_n(x,z) := \frac{K(\frac{x-z}{h_n})}{\int K(\frac{x-u}{h_n})\mu(\mathrm{d}u)}.$$

Theorem of Banach-Steinhaus

(i) $\exists c > 0\ \forall n$: $\int |K_n(x,z)|\mu(\mathrm{d}x) \leqslant c$ for μ-almost all z

(ii) $\exists D \geqslant 1\ \forall n, x$: $\int |K_n(x,z)|\mu(\mathrm{d}z) \leqslant D$.

(iii) $\forall a > 0$, $\lim_{n\to\infty} \int \int |K_n(x,z)| \mathbb{1}_{\{\|x-z\|>a\}}\mu(\mathrm{d}z)\mu(\mathrm{d}x) = 0$.

(iv) $\lim_{n\to\infty} \operatorname{ess\,sup}_x |\int K_n(x,z)\mu(\mathrm{d}z) - 1| = 0$.

$\Longrightarrow \forall m \in L_1(\mu)$:

$$\lim_{n\to\infty} \int |m(x) - \int K_n(x,z)m(z)\mu(\mathrm{d}z)|\mu(\mathrm{d}x) = 0.$$

Proof Components 4

Lemma 2.4

For n large enough:

$$\mathbb{P}\left(\int |m(x) - m_n^*(x)|\mu(\mathrm{d}x) > \varepsilon\right) \leqslant e^{-n\varepsilon^2/(8L\varrho)^2}.$$

Theorem (McDiarmid's Inequality)

Let $Z_1, \ldots, Z_n \in A$ be indep. RV and assume for $f : A^n \to \mathbb{R}$:

$$\sup_{z_1,\ldots,z_n,\hat{z}_i \in A} |f(z_1, \ldots, z_n) - f(z_1, \ldots, z_{i-1}, \hat{z}_i, z_{i-1}, \ldots, z_n)| \leqslant c_i.$$

Then, $\forall \varepsilon > 0$,

$$\mathbb{P}(f(Z_1, \ldots, Z_n) - \mathbb{E}(f(Z_1, \ldots, Z_n)) \geqslant \varepsilon) \leqslant e^{-2\varepsilon^2/\sum_{i=1}^n c_i^2}.$$

Proof components 5 ✓

$$\int |m_n(x) - m(x)|\mu(\mathrm{d}x) \leqslant \underbrace{\int |m_n(x) - \mathbf{m_n^*(x)}|\mu(\mathrm{d}x)}_{\Downarrow} + \underbrace{\int |\mathbf{m_n^*(x)} - m(x)|\mu(\mathrm{d}x)}_{\textit{Lemma 4 } \checkmark}$$

$$\boxed{\begin{aligned}
&|m_n^*(x) - m_n(x)| \\
&= \left| \frac{\sum Y_i K_{h_n}(x-X_i)}{n\mathbb{E}(K_{h_n}(x-X))} - \frac{\sum Y_i K_{h_n}(x-X_i)}{\sum K_{h_n}(x-X_i)} \right| \quad \text{(by definition)} \\
&= \left|\sum Y_i K_{h_n}(x - X_i)\right| \left| \frac{1}{n\mathbb{E}(K_{h_n}(x-X))} - \frac{1}{\sum K_{h_n}(x-X_i)} \right| \\
&\leqslant L\left|\sum K_{h_n}(x - X_i)\right| \left| \frac{1}{n\mathbb{E}(K_{h_n}(x-X))} - \frac{1}{\sum K_{h_n}(x-X_i)} \right| \quad (\text{by } |Y| \leqslant L) \\
&= L\left| \frac{\sum Y_i K_{h_n}(x-X_i)}{n\mathbb{E}(K_{h_n}(x-X))} - 1 \right| = L|M_n^*(x) - 1|,
\end{aligned}}$$

where M_n^* is a special form of $m_n^*(x)$ for $(X, 1)$.
(in this case: $M(x) = \mathbb{E}(1|X = x) = 1 \ \forall x \in \mathbb{R}^d$ with $Y \equiv 1$) □

Strong universal consistency of the kernel estimate

Theorem 2.2 (strong universal consistency)

Let $K(x) = \mathbb{1}_{\{\|x\| \leqslant 1\}}$ and let h_n satisfy

$$h_{n-1} \neq h_n \quad \textit{at most for the indices } n = n_1, n_2, \ldots,$$

where $n_{k+1} \geqslant D n_k$ for fixed $D > 1$. Additionally let

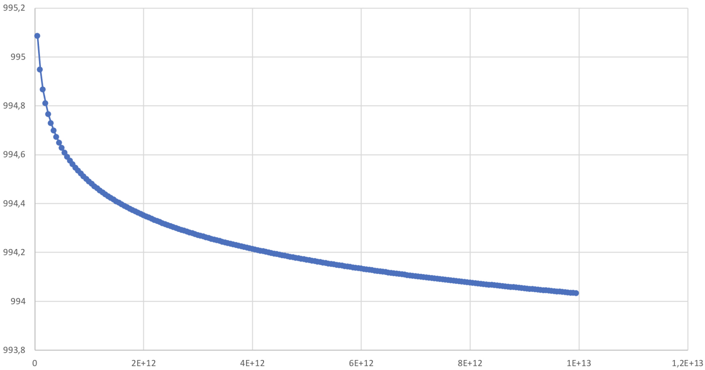
$$h_n \to 0 \quad \text{and} \quad n h_n^d \to \infty,$$

e.g., $h_n = c e^{-\gamma \lfloor q \log n \rfloor / q}$ with $c > 0$, $0 < d\gamma < 1$ and $q > 0$.
Then m_n is **strongly universally consistent**.

Sequence of bandwidths

Let $q = 1$, $c = d = 1000$ and $\gamma = 1/5000$.

$$\Rightarrow h_n = 1000 \cdot e^{-\frac{1}{5000}\lfloor \log n \rfloor}.$$



Proof component

Theorem A

(i) m_n local averaging estimate with subprobability weights

(ii) m_n strongly universal consistent with bounded Y

(iii) $\exists c < \infty : \forall Y$ with $\mathbb{E}(Y^2) < \infty$:

$$\limsup_{n \to \infty} \sum_{i=1}^{n} Y_i^2 \int \alpha_{n,i}(x)\mu(\mathrm{d}x) \leqslant c\mathbb{E}\left(Y^2\right) \ \mathbb{P}\text{-a.s.}$$

Then m_n is strongly universally consistent.

k-NN Estimates

Strong consistency of the k-NN estimate

Theorem 3.1 (strong consistency)

(1) $\exists L < \infty$ s.t. $|Y| \leqslant L$ $\mathbb{P}$-a.s.
(2) $\forall x \in \mathbb{R}^d$: $\|X - x\|$ absolutely continuous.
If

$$k_n \to \infty \quad \text{and} \quad k_n/n \to 0,$$

$\implies$ k_n-NN regression function estimate is strongly consistent.

Proof components (overview)

Theorem

(strong consistency of the k-NN estimator)

$$\lim_{n\to\infty}\int |m_n(x)-m(x)|^2 d\mu(x) = 0$$

McDiarmid‘s inequality

$$\mathbb{P}(f(Z_1,\dots,Z_n)-\mathbb{E}((f(Z_1,\dots,Z_n)))\geqslant\varepsilon)\leqslant e^{-2\varepsilon^2/\sum_{i=1}^n c_i^2}$$

weak universal consistency of k-NN

$$\lim_{n\to\infty}\mathbb{E}\Big(\int|m(x)-m_n(x)|\mu(\mathrm{d}x)\Big)=0$$

$$\mathbb{E}\Big(\int|m_n(x)-m(x)|\mu(\mathrm{d}x)\Big)\leqslant L\mathbb{E}\Big(\int|M_n(x)-M_n^*(x)|\mu(\mathrm{d}x)\Big)+\mathbb{E}\Big(\int|m_n^*(x)-m(x)|\mu(\mathrm{d}x)\Big)$$

$$|m_n(x)-m_n^*(x)|\leqslant L|M_n(x)-M_n^*(x)|$$

modificated estimator

$$M_n^*(x)$$

Decomposition

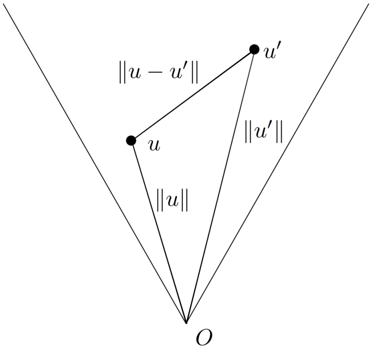
$$|m(x)-m_n(x)|\leqslant|m_n(x)-m_n^*(x)|+|m_n^*(x)-m(x)|$$

modificated estimator

$$m_n^*(x)=\frac{1}{k_n}\sum_{j=1}^n Y_j \mathbb{1}_{\{\|X_j-x\|<\rho_n(x)\}}$$

Reminder

Remembering the cone property



$$\|u\| < \|u'\| \implies \|u - u'\| < \|u'\|$$

Some useful sets

Definition (A_i)

Let A_i be the collection of all $x \in \mathbb{R}^d$, s.t. X_i is one of its k_n nearest neighbors of x in $\{X_1, \ldots, X_n\}$.

Some useful sets

Definition ($C_{i,j}$)

Let C_j be a cone of radius $\pi/3$, s.t. $C_1, \dots, C_{\gamma_d}$ covers $\mathbb{R}^d$.
We define $C_{i,j} := X_i + C_j \quad \forall i \in \{1, \dots, n\}\ \forall j \in \{1, \dots, \gamma_d\}$.

Definition ($B_{i,j}$)

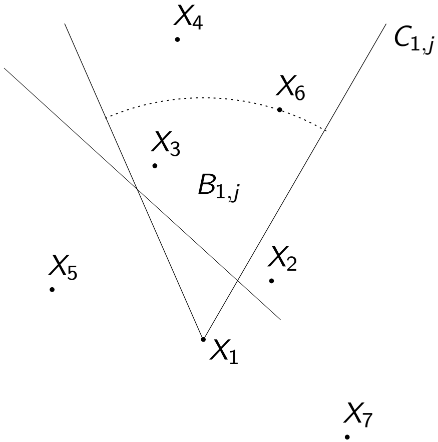
Let $B_{i,j}$ be the subset of $C_{i,j}$ consisting of all $x \in C_{i,j}$ that are among the k_n nearest neighbors of X_i in the set

$$\{X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n, x\} \cap C_{i,j}.$$

Equivalently $B_{i,j}$ is the subset of $C_{i,j}$ consisting of all x that are closer to X_i than the k_n-th nearest neighbor of X_i in $\{X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n\} \cap C_{i,j}$.

Some useful sets

Example of $B_{i,j}$ and $C_{i,j}$ with $k_n = 2, i = 1$.



Strong universal consistency of the k-NN estimate

Lemma 3.2 (k-NN covering)

If $x \in A_i$, then $x \in \bigcup_{j=1}^{\gamma_d} B_{i,j}$, and thus

$$\mu(A_i) \leqslant \sum_{j=1}^{\gamma_d} \mu(B_{i,j}).$$

Proof idea: Cone property!

Take $X_l \in C_{i,j}$ that is k_n-NN of X_i
Then $\|X_l - X_i\| < \|x - X_i\| \implies \|X_l - x\| < \|X_i - x\|$
$x \in A_i \implies X_l$ k_n-NN of $x \implies$ there can exist at most $k_n - 1$ points $X_l \in C_{i,j}$ closer to X_i than x. $\implies x \in B_{i,j}$

Strong universal consistency of the k-NN estimate

Lemma 3.3 (k-NN upper bound for $B_{i,j}$)

If $k_n/\log(n) \to \infty$ and $k_n/n \to 0$, $\forall x \in \mathbb{R}^d$: $\|X - x\|$ absolutely continuous

$$\implies \limsup_{n\to\infty} \frac{n}{k_n} \max_{1\leqslant i\leqslant n} \mu(B_{i,j}) \leqslant 2 \quad \mathbb{P}\text{-a.s.} \quad \forall j \in \{1, \ldots, \gamma_d\}$$

Proof idea: Borel-Cantelli Lemma

Show that $\sum_{n=1}^{\infty} \mathbb{P}\left(\frac{n}{k_n} \max_{1\leqslant i\leqslant n} \mu(B_{i,j}) > 2\right) < \infty$

Strong universal consistency of the k-NN estimate

Theorem 3.2 (strong universal consistency)

$\forall x \in \mathbb{R}^d$: $\|X - x\|$ absolutely continuous.
If

$$k_n / \log n \to \infty \quad \text{and} \quad k_n / n \to 0$$

$\implies$ k_n-NN regression function estimate is strongly universally consistent

Proof Sketch SUC

Only thing to show:

Theorem A

(i) m_n local averaging estimate with subprobability weights

(ii) m_n strongly universal consistent with bounded Y

(iii) $\exists c < \infty : \forall Y$ with $\mathbb{E}(Y^2) < \infty :$

$$\limsup_{n\to\infty} \sum_{i=1}^{n} Y_i^2 \int \alpha_{n,i}(x)\mu(\mathrm{d}x) \leqslant c\mathbb{E}\left(Y^2\right) \ \mathbb{P}\text{-a.s.}$$

Then m_n is strongly universally consistent.

Proof Sketch SUC

In our case:

$$
\begin{aligned}
&\sum_{i=1}^{n} Y_i^2 \int \alpha_{n,i}(x)\mu(\mathrm{d}x) \\
&= \frac{1}{k_n}\sum_{i=1}^{n} Y_i^2 \int \mathbb{1}_{\{X_i \text{ is among the } k_n \text{ NNs of } x\}}\mu(\mathrm{d}x) \\
&= \frac{1}{k_n}\sum_{i=1}^{n} Y_i^2 \int \mathbb{1}_{\{x\in A_i\}}\mu(\mathrm{d}x) \\
&= \frac{1}{k_n}\sum_{i=1}^{n} Y_i^2 \mu(A_i)
\end{aligned}
$$

Proof Sketch SUC

$$\frac{1}{k_n}\sum_{i=1}^{n} Y_i^2 \mu(A_i) \leqslant \left(\frac{n}{k_n}\max_{1\leqslant i\leqslant n}\mu(A_i)\right)\frac{1}{n}\sum_{i=1}^{n} Y_i^2$$

Proof Sketch SUC

Utilizing the k-NN covering Lemma

$$
\begin{aligned}
&\limsup_{n\to\infty} \frac{n}{k_n} \max_{1\leqslant i\leqslant n} \mu(A_i) \\
&\leqslant \limsup_{n\to\infty} \frac{n}{k_n} \max_{1\leqslant i\leqslant n} \sum_{j=1}^{\gamma_d} \mu(B_{i,j}) \\
&\leqslant \limsup_{n\to\infty} \frac{n}{k_n} \sum_{j=1}^{\gamma_d} \max_{1\leqslant i\leqslant n} \mu(B_{i,j}) \\
&= \sum_{j=1}^{\gamma_d} \limsup_{n\to\infty} \frac{n}{k_n} \max_{1\leqslant i\leqslant n} \mu(B_{i,j})
\end{aligned}
$$

Proof Sketch SUC

With our second Lemma:

$$\sum_{j=1}^{\gamma_d} \limsup_{n\to\infty} \frac{n}{k_n} \max_{1\leqslant i\leqslant n} \mu(B_{i,j}) \leqslant \sum_{j=1}^{\gamma_d} 2 = 2\gamma_d$$

Proof Sketch SUC

And finally by the strong law of large numbers

$$\limsup_{n\to\infty}\left(\frac{n}{k_n}\max_{1\leqslant i\leqslant n}\mu(A_i)\right)\frac{1}{n}\sum_{i=1}^{n}Y_i^2\leqslant 2\gamma\mathbb{E}\left(Y^2\right)\ \mathbb{P}\text{-a.s.}$$

and we are done.

Comparison

Where are we?

	Partitioning	**Kernel**	**k-NN**
weak universal consistency	✓	✓	✓
strong (universal) consistency	✓	✓	✓

Comparison: Strong vs. weak (universal) consistency

KERNEL ESTIMATE		
boxed	regular	naive
	$h_n \to 0$ $nh_n^d \to \infty$	
–	$\lvert Y\rvert \leqslant L$ in $\mathbb{P}$	$h_n \neq h_{n+1}$
$\Downarrow$	$\Downarrow$	$\Downarrow$
weak universal consistency	**strong** consistency	**strong universal** consistency

Comparison:
Strong vs. weak (universal) consistency

<table>
<tr><th colspan="3">K-NN ESTIMATE</th></tr>
<tr><td colspan="2">$k_n \to \infty$</td><td>$k_n/log(n) \to \infty$</td></tr>
<tr><td colspan="3">$k_n/n \to 0$</td></tr>
<tr><td colspan="3">$\mathbb{P}(ties) = 0$</td></tr>
<tr><td>–</td><td>$|Y| \leqslant L$ a.s.</td><td>–</td></tr>
<tr><td>–</td><td colspan="2">$\|X - x\|$ abs. continuous</td></tr>
<tr><td>$\Downarrow$</td><td>$\Downarrow$</td><td>$\Downarrow$</td></tr>
<tr><td>weak universal consistency</td><td>strong consistency</td><td>strong universal consistency</td></tr>
</table>

Programming part