Universität Stuttgart

Nearest neighbor based conformal prediction

László Györfi
Harro Walk

Stuttgarter Mathematische Berichte

2020-002

mathematik

Fachbereich Mathematik
Fakultät Mathematik und Physik
Universität Stuttgart
Pfaffenwaldring 57
D-70 569 Stuttgart

E-Mail: preprints@mathematik.uni-stuttgart.de
WWW: http://www.mathematik.uni-stuttgart.de/preprints

ISSN **1613-8309**

Nearest neighbor based conformal prediction

László Györfi[*] Harro Walk[†]

May 24, 2020

Abstract

In this paper we introduce a nearest neighbor based estimate of the prediction interval with prescribed conditional coverage probability and with small length. In the special case, when there is no feature vector, the problem is the estimate of a confidence interval. For confidence interval estimate, we show the distribution-free strong consistency of the conditional coverage probability and of excess length of the interval, while the conditional coverage probability of prediction interval has the distribution-free strong consistency property and under weak conditions on the underlying distributions strong consistency and the fast rate of convergence of the excess length are shown. As a consequence, we construct a confidence set estimate for classification.

AMS Classification: 62G08, 62G20.

Key words and phrases: confidence interval, prediction interval, conditional coverage probability, excess length, classification, nearest neighbor estimate

[*]Corresponding author. Department of Computer Science and Information Theory, Budapest University of Technology and Economics, Magyar Tudósok krt. 2., Budapest, H-1117, Hungary. gyorfi@cs.bme.hu

[†]Institute of Stochastics and Applications, University of Stuttgart, Pfaffenwaldring 57, D-70569 Stuttgart, Germany. harro.walk@mathematik.uni-stuttgart.de

1 Introduction

Consider a training data set

$$\mathcal{D}_n = \{(X_1, Y_1), \ldots, (X_n, Y_n)\}, \tag{1}$$

and a test point (X_{n+1}, Y_{n+1}), with the training and test data all drawn i.i.d. from the same distribution. Here each $X_i \in \mathbb{R}^d$ is a feature vector, while $Y_i \in \mathbb{R}$ is a response variable. The problem of predictive inference is the following: if we observe the n training data points $\mathcal{D}_n$, and are given the feature vector X_{n+1} for a new test data point, we would like construct a prediction interval $C_n(x) = C_n(\mathcal{D}_n, x)$ for Y_{n+1} such that we believe $C_n(X_{n+1})$ is likely to contain the test point's true response value Y_{n+1}.

We are interested in the conditional coverage probability

$$\mathbb{P}\{Y_{n+1} \in C_n(X_{n+1}) \mid \mathcal{D}_n, X_{n+1} = x\} \tag{2}$$

and in the length of the interval $C_n(x)$. Most of the related literature study the (unconditional) coverage probability

$$\mathbb{P}\{Y_{n+1} \in C_n(X_{n+1})\},$$

while Vovk [16] considers the conditional coverage probability

$$\mathbb{P}\{Y_{n+1} \in C_n(X_{n+1}) \mid \mathcal{D}_n\}$$

and Barber et al. [1] investigate the conditional coverage probability

$$\mathbb{P}\{Y_{n+1} \in C_n(X_{n+1}) \mid X_{n+1} = x\}.$$

If there is no feature vector X_i, then the problem is reduced to the estimation of confidence interval.

In this paper we consider three set estimation problems:

- confidence interval of given conditional coverage probability and of small length,
- prediction interval of given conditional coverage probability (2) and of small length,
- confidence set for classification with given conditional coverage probability (2) and with small size.

Such estimates can be derived from the estimates of quantiles or conditional quantiles. For example, concerning the confidence interval one may search for the shortest interval with prescribed empirical distribution. In the sequel we modify this principle such that choose the shortest interval with a bit smaller empirical distribution. It turns out that these interval estimation problems are easier than the corresponding point estimation, which means that the rate of convergence of the length of the interval is super-polynomial.

Lei et al. [12], Papadopoulos et al. [14], Romano, Patterson and Candés [15], Vovk [16], studied estimates of prediction intervals. These authors discuss and investigate various concepts on the basis of independent and identically distributed or exchangeable samples. Concerning an overview on the validity and on efficiency, especially a negative result for the non-asymptotic situation see Vovk [16] and Barber et al. [1]. Lei and Wasserman [11] and Lei, Robins and Wasserman [13] contain results on batch mode prediction, while Vovk [16] and Lei et al. [12] are on online (or sequential) prediction.

Usually, the authors use a prediction set around a point prediction of Y, especially intervals are considered, which are symmetric around a regression function estimate, showing the consistency of the coverage probability, also refining the procedures under computational aspects (see Papadopoulos et al. [14]). Here we show the consistency of the conditional coverage probability and investigate the rate of convergence for the excess of the estimated prediction (confidence) interval length with respect to the minimal prediction (confidence) interval length.

2 Estimate the confidence interval

In this setup there is no feature vector. For the random variable Y, introduce the distribution function

$$F(y) := \mathbb{P}\{Y \leq y\}.$$

If $0 < p < 1$, then the set of possible quantiles is an interval Q such that

$$[q_{p,low}, q_{p,up}) \subset Q \subset [q_{p,low}, q_{p,up}],$$

where

$$q_{p,up} := F_{up}^{-1}(p) := \sup_{y \in \mathbb{R}: F(y) \leq p} y$$

is the upper quantile, and

$$q_{p,low} := F_{low}^{-1}(p) := \min_{y \in \mathbb{R}: F(y) \geq p} y$$

denotes the lower quantile. In general, the quantile is not unique. The quantile is unique, if

$$q_{p,low} = q_{p,up}.$$

Concerning a confidence interval

$$C = [f_{lo}, f_{hi}],$$

we need that for the *coverage probability*

$$\mathbb{P}\{Y \in C\} \geq 1 - \alpha,$$

where $0 < \alpha < 1$ is fixed, and the *size of the interval*

$$\Delta := f_{hi} - f_{lo}$$

is as small as possible.

If the lower and upper quantiles are known, then the optimal confidence interval can be constructed as follows: Put

$$D(p) := F_{low}^{-1}(p + 1 - \alpha) - F_{up}^{-1}(p),$$

$(0 < p < \alpha)$ and

$$p^* := \arg\min_{0<p<\alpha} D(p).$$

Then

$$\Delta^* = \min_{0<p<\alpha} D(p) = D(p^*),$$

$$f_{lo}^* = F_{up}^{-1}(p^*)$$

and

$$f_{hi}^* = F_{low}^{-1}(p^* + 1 - \alpha).$$

For the sake of simplicity, throughout the paper we assume that the arg min exists. In the general case, p^* can be defined such that $D(p^*) \leq \inf_{0<p<\alpha} D(p) + 1/n^2$, where n is the sample size.

If the lower and upper quantiles are unknown, then assume, that we observed data $\{Y_1, \ldots, Y_n\}$ consisting of independent and identically distributed copies of Y.

Let

$$F_n(y) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}_{Y_i \leq y},$$

be the empirical distribution function, where $\mathbb{I}$ denotes the indicator function, while

$$F_{n,low}^{-1}(p) := \min_{y \in \mathbb{R}: F_n(y) \geq p} y, \qquad 0 < p < 1$$

and

$$F_{n,up}^{-1}(p) := \sup_{y \in \mathbb{R}: F_n(y) \le p} y, \qquad 0 < p < 1$$

are the corresponding quantile function estimates. From these estimates one may derive the plug-in estimate of the optimal confidence interval. Put

$$D_n(p) := F_{n,low}^{-1}(p + 1 - \alpha - \ln n/\sqrt{n}) - F_{n,up}^{-1}(p + \ln n/\sqrt{n})$$

$(0 < p < \alpha)$ and

$$p_n := \arg\min_{0<p<\alpha} D_n(p).$$

Then

$$\Delta_n = D_n(p_n),$$
$$f_{n,lo} = F_{n,up}^{-1}(p_n + \ln n/\sqrt{n})$$

and

$$f_{n,hi} = F_{n,low}^{-1}(p_n + 1 - \alpha - \ln n/\sqrt{n}).$$

The main point here is that we underestimate the optimal confidence interval resulting in distribution-free consistency.

Lei, Robins and Wasserman [13] studied the extension of our problem, when Y is d' dimensional and the aim was to approximate the confidence set of minimal Lebesque measure. They assumed a smooth density of Y, and from the kernel density estimate they derived a confidence set estimate. For Lipschitz continuous density and for the expected Lebesgue measure of the symmetric difference of the optimal and estimated confidence set they got a rate of convergence $O\left(\ln n/n^{1/(d'+2)}\right)$, which is $O\left(\ln n/n^{1/3}\right)$, for $d' = 1$. Here we do not assume a density and construct a confidence interval instead of a confidence set such that the tail of the excess of the estimated confidence interval length with respect to the minimal confidence interval has the rate of convergence $O\left(1/n^{2\ln n}\right)$.

Put

$$C_n = [f_{n,lo}, f_{n,hi}].$$

The following theorem is on the tail distribution of the conditional coverage probability $\mathbb{P}\{Y_{n+1} \in C_n \mid Y_1, \dots, Y_n\}$ and of the excess length $(\Delta_n - \Delta^*)^+$:

Theorem 1. *For any distribution of Y and $\varepsilon > 0$,*

$$\begin{aligned} &\mathbb{P}\{|\mathbb{P}\{Y_{n+1} \in C_n \mid Y_1, \dots, Y_n\} - (1-\alpha)| \ge \varepsilon\} \\ &\le 8(n+1)e^{-n\varepsilon^2/128} + \mathbb{I}_{\ln n/\sqrt{n} \ge \varepsilon/4}, \end{aligned} \tag{3}$$

and

$$\mathbb{P}\{\Delta_n - \Delta^* \ge 0\} \le 2/n^{2\ln n}. \tag{4}$$

The proof of this theorem is in the last section. (3) implies that

$$\lim_{n\to\infty} \mathbb{P}\{Y_{n+1} \in C_n \mid Y_1, \dots, Y_n\} = 1 - \alpha \tag{5}$$

a.s. and

$$\mathbb{E}\left\{|\mathbb{P}\{Y_{n+1} \in C_n \mid Y_1, \dots, Y_n\} - (1-\alpha)|\right\} = O\left(\ln n/\sqrt{n}\right). \tag{6}$$

Furthermore, because of

$$\sum_{n=1}^{\infty} \mathbb{P}\{\Delta_n - \Delta^* \geq 0\} \leq \sum_{n=1}^{\infty} 2/n^{2\ln n} < \infty,$$

the Borel-Cantelli lemma implies the strong distribution-free consistency of the excess length:

$$\lim_{n\to\infty} [\Delta_n - \Delta^*]^+ = 0$$

a.s.

3 Conformal prediction

Introduce the conditional distribution function

$$F(y \mid x) := \mathbb{P}\{Y \leq y \mid X = x\}$$

and the corresponding lower and upper conditional quantile functions are defined by

$$q_{p,low}(x) := F_{x,low}^{-1}(p) := \min_{y \in \mathbb{R}: F(y|x) \geq p} y.$$

and

$$q_{p,up}(x) := F_{x,up}^{-1}(p) := \sup_{y \in \mathbb{R}: F(y|x) \leq p} y,$$

respectively.

The conformal prediction is an interval

$$C(x) = [f_{lo}(x), f_{hi}(x)].$$

Concerning a prediction interval $C(x)$, we need that for the *conditional coverage probability* (called also conditional validity in [11] and in [16])

$$\mathbb{P}\{Y \in C(X) \mid X = x\} \geq 1 - \alpha, \tag{7}$$

where $0 < \alpha < 1$ is fixed, and the *length of the interval*

$$\Delta(x) := f_{hi}(x) - f_{lo}(x)$$

is as small as possible (efficiency, compare [11] and [16]).

An extension of our problem can be found in Lei et al. [12] (with a generalization to multi-dimensional Y in Lei and Wasserman [11]) such that $C(x)$ is an arbitrary (measurable) set, and under (7) the aim is to minimize $\lambda(C(x))$, where λ stands for the Lebesgue measure. From a kernel density estimate they derived a consistent estimate of the optimal prediction set, and for Lipschitz continuous conditional density showed the rate of convergence of order $O\left(\ln n / n^{1/(d+3)}\right)$. Our setup is much simpler such that there is no need to assume the existence of the density of (X, Y). Here, we do not assume a density of (X, Y). Interestingly, under mild conditions, especially Lipschitz continuity of the function $F(y \mid \cdot)$, we get a super-polynomial rate $O\left(1/n^{2 \ln n}\right)$ in the case of prediction interval, instead of a prediction set.

If the conditional distribution function and the conditional quantile functions are known, then the optimal prediction interval can be constructed. Put

$$D(p) := F_{x,low}^{-1}(p + 1 - \alpha) - F_{x,up}^{-1}(p),$$

$(0 < p < \alpha)$ and

$$p^* := p^*(x) := \arg\min_{0<p<\alpha} D(p).$$

Then

$$\Delta^*(x) = \min_{p<\alpha} D(p) = D(p^*),$$

$$f_{lo}^*(x) = F_{x,up}^{-1}(p^*)$$

and

$$f_{hi}^*(x) = F_{x,low}^{-1}(p^* + 1 - \alpha).$$

If the conditional distribution function and the conditional quantile function are unknown, then assume, that we observed data (1). The obvious way is to estimate the conditional distribution function and the conditional quantile function. Allowing additional measurement errors, Hansmann and Kohler [10] studied local averaging regression based, in detail kernel regression based, conditional quantile function estimates. Here we introduce a k-nearest-neighbor (k-NN) estimate, which is easier to analyze. For a fixed $x \in \mathbb{R}^d$, reorder the data $(X_1, Y_1), \ldots, (X_n, Y_n)$ according to increasing values of $\|X_i - x\|$. The reordered data sequence is denoted by

$$(X_{(n,1)}(x), Y_{(n,1)}(x)), \ldots, (X_{(n,n)}(x), Y_{(n,n)}(x)).$$

$X_{(n,k)}(x)$ is the k-th nearest neighbor of x. The tie breaking is done by randomization. If the distribution of feature vector X is denoted by μ, then for the sake of simplicity, we assume that μ has a density, therefore tie happens with probability 0.

The k-NN estimate of the conditional distribution function is defined by

$$F_{k,n}(y \mid x) = \frac{1}{k} \sum_{i=1}^{k} \mathbb{I}_{Y_{(n,i)}(x) \le y},$$

while

$$F^{-1}_{x,k,n,low}(p) := \min_{y \in \mathbb{R}: F_{k,n}(y|x) \ge p} y, \qquad 0 < p < 1$$

and

$$F^{-1}_{x,k,n,up}(p) := \sup_{y \in \mathbb{R}: F_{k,n}(y|x) \le p} y, \qquad 0 < p < 1$$

are the corresponding conditional quantile function estimates. From these estimates one may derive the plug-in estimate of the optimal prediction interval. Put

$$D_n(p) := F^{-1}_{x,k,n,low}(p + 1 - \alpha - t_{k,n}) - F^{-1}_{x,k,n,up}(p + t_{k,n})$$

$(0 < p < \alpha)$ with

$$t_{k,n} := \ln n(1/\sqrt{k} + (k/n)^{1/d}),$$

and

$$p_n := p_n(x) := \operatorname*{arg\,min}_{0<p<\alpha} D_n(p).$$

Then

$$\Delta_n(x) = D_n(p_n),$$

$$f_{k,n,lo}(x) = F^{-1}_{x,k,n,up}(p_n + t_{k,n})$$

and

$$f_{k,n,hi}(x) = F^{-1}_{x,k,n,low}(p_n + 1 - \alpha - t_{k,n}).$$

This estimate has a simple interpretation. If $y_1 < \cdots < y_k$ denotes the ordered samples of $Y_{(n,1)}(x), \ldots, Y_{(n,k)}(x)$, then

$$F_{k,n}(y \mid x) = \frac{1}{k} \sum_{i=1}^{k} \mathbb{I}_{y_i \le y}.$$

For

$$\begin{aligned} y_{g(n)} &= \arg\min_{y_i;\, i/k\le\alpha} [F^{-1}_{x,k,n,low}(i/k+1-\alpha-2t_{k,n}) - y_i] \\ &= \arg\min_{y_i;\, i\le k\alpha} [y_{\lfloor k(i/k+1-\alpha-2t_{k,n})\rfloor} - y_i] \\ &= \arg\min_{y_i;\, i\le k\alpha} [y_{\lfloor i+k(1-\alpha-2t_{k,n})\rfloor} - y_i], \end{aligned}$$

we have that

$$f_{k,n,lo}(x) = y_{g(n)+\lfloor kt_{k,n}\rfloor}$$

and

$$f_{k,n,hi}(x) = y_{g(n)+\lfloor k(1-\alpha)-kt_{k,n}\rfloor}.$$

Put

$$C_{k,n}(x) = [f_{k,n,lo}(x), f_{k,n,hi}(x)].$$

The next theorem is on the strong universal consistency of the conditional coverage probability and on the rate of convergence under some smoothness condition. In order to have non-trivial rate of convergence of the conditional coverage probability, one has to assume tail and smoothness conditions, otherwise the rate of convergence can be arbitrarily slow, see Chapter 3 in [9]. For most of the related results, the feature vector X is assumed to be bounded, which excludes the classical parametric problem, where the conditional distributions of X given Y are multidimensional Gaussian distributions. Next, we introduce a mild combined tail and smoothness condition, under which we get fast rate of convergence.

Definition 1. *For each y, the function $F(y \mid \cdot)$ satisfies the modified Lipschitz condition, if there is a constant C^* such that for any $y \in \mathbb{R}$ and $x, z \in \mathbb{R}^d$*

$$|F(y \mid x) - F(y \mid z)| \le C^* \mu(S_{x,\|x-z\|})^{1/d}, \tag{8}$$

where $S_{x,r}$ denotes the sphere centered at x and having the radius r (cf. [3], [8]).

If the density f of μ is continuous, then the right hand side of (8) is approximately equal to $C^* f(x)^{1/d}\|x-z\|$. For $d = 1$, the standard exponential distribution is an example, where the right hand side of (8) is approximately equal to $C^* e^{-x}\max\{x-z, z-x\}$. Interestingly, one can show densities for modified Lipschitz condition, where $\mathbb{E}\{\|X\|\} = \infty$, for example, f is the Cauchy density.

Theorem 2. *Assume that μ has a density. If $k = k_n$ such that $k_n/(\ln n)^2 \to \infty$ and $(\ln n)^d k_n/n \to 0$, then for any distribution of (X, Y),*

$$\lim_{n\to\infty} \mathbb{P}\{Y_{n+1} \in C_{k_n,n}(X_{n+1}) \mid \mathcal{D}_n, X_{n+1} = x\} = 1 - \alpha \tag{9}$$

a.s. for μ-almost-all x. If, in addition, for each y, the function $F(y \mid \cdot)$ satisfies the modified Lipschitz condition, then for any fixed x,

$$\begin{aligned}&\mathbb{E}\left\{\left|\mathbb{P}\{Y_{n+1} \in C_{k_n,n}(X_{n+1}) \mid \mathcal{D}_n, X_{n+1} = x\} - (1-\alpha)\right|\right\}\\ &= O\left(\ln n/\sqrt{k_n}\right) + O\left(\ln n (k_n/n)^{1/d}\right).\end{aligned} \tag{10}$$

For the choice

$$k_n = \lfloor const \cdot n^{2/(d+2)} \rfloor, \tag{11}$$

the rate (10) is of order

$$O\left(\frac{\ln n}{n^{1/(d+2)}}\right). \tag{12}$$

Under mild condition, the following theorem states strong consistency and the rate of convergence of the excess length $[\Delta_n(x) - \Delta^*(x)]^+$:

Theorem 3. *Assume that μ has a density and $k_n/(\ln n)^2 \to \infty$ and $(\ln n)^d k_n/n \to 0$.*

(i) If either the conditional quantiles at p^* and at $p^* + 1 - \alpha$ are unique, or the function $F(y \mid \cdot)$ satisfies the modified Lipschitz condition, then

$$\lim_{n\to\infty} [\Delta_n(x) - \Delta^*(x)]^+ = 0 \tag{13}$$

a.s. for μ-almost-all x.

(ii) Under the modified Lipschitz condition we have that

$$\mathbb{P}\{\Delta_n(x) - \Delta^*(x) \geq 0\} \leq 2n^{-2\ln n} + 2e^{-3k_n/14\mathbb{I}_{\ln n \geq 2C^*}}. \tag{14}$$

We get that $\mathbb{I}_{\ln n \geq 2C^*} = 1$, if n is large enough. Thus, the choice (11) and (14) imply the super-polynomial rate of convergence:

$$\mathbb{P}\{\Delta_n(x) - \Delta^*(x) \geq 0\} \leq 2n^{-2\ln n} + 2e^{-3k_n/14} = O(1/n^{2\ln n}).$$

It is an open problem whether (13) is satisfied without assuming anything on the conditional distribution function, or whether there is an estimate of the optimal prediction interval with distribution-free consistency of the excess length of the prediction interval.

4 Estimate the confidence set for classification

The previous results have some consequences in classification. For the classification, the label Y takes values in the finite set $\{1, \ldots, K\}$ and the Bayes decision g^* minimizes the error probability:

$$g^*(x) = \arg\max_j P_j(x),$$

where

$$P_j(x) = \mathbb{P}\{Y = j \mid X = x\}, \quad j = 1, \ldots, K.$$

A possibility for improving the confidence of the classification is the Bayes decision with rejection option, which means that for small values of $\max_j P_j(x)$ one does not make any decision, see Problem 2.5 in [7].

Another possibility is, that instead of deciding on a single label, we give a list of labels. Such list is called confidence set for classification. The problem is to construct a confidence set $C_n(x) = C_n(\mathcal{D}_n, x)$ such that for the *conditional coverage probability*

$$\mathbb{P}\{Y \in C(X) \mid X = x\} \geq 1 - \alpha,$$

where $0 < \alpha < 1$ is fixed, and the *cardinality* $|C(x)|$ *of the set* $C(x)$ is as small as possible. (Cf. [4] and [5].)

If the a posteriori probabilities $P_j(x)$, $j = 1, \ldots, K$ are known, then the optimal confidence set can be constructed. Let $P_{i_1(x)}(x) \geq \cdots \geq P_{i_K(x)}(x)$ be the ordered values of $P_1(x), \ldots, P_K(x)$. Define $L^*(x)$ by the inequalities

$$\sum_{j=1}^{L^*(x)} P_{i_j(x)}(x) \geq 1 - \alpha > \sum_{j=1}^{L^*(x)-1} P_{i_j(x)}(x).$$

Then, the optimal confidence set is

$$C^*(x) = \{i_1(x), \ldots, i_{L^*(x)}(x)\}.$$

If the a posteriori probabilities $P_j(x)$ are unknown, then an approximation of P_j is denoted by $\tilde{P}_j$, $j = 1, \ldots, K$, from which the corresponding plug-in confidence set is as follows: Let $\tilde{P}_{\tilde{i}_1(x)}(x) \geq \cdots \geq \tilde{P}_{\tilde{i}_K(x)}(x)$ be the ordered values of $\tilde{P}_1(x), \ldots, \tilde{P}_K(x)$. Define $\tilde{L}(x)$ by the inequalities

$$\sum_{j=1}^{\tilde{L}(x)} \tilde{P}_{\tilde{i}_j(x)}(x) \geq 1 - \alpha - \tilde{t} > \sum_{j=1}^{\tilde{L}(x)-1} \tilde{P}_{\tilde{i}_j(x)}(x)$$

with a $\tilde{t} > 0$. Then, the plug-in confidence set is

$$\tilde{C}(x) = \{\tilde{i}_1(x), \dots, \tilde{i}_{\tilde{L}(x)}(x)\}.$$

In this way we underestimate the optimal confidence set.

Choose an integer k less than n, then the k-nearest-neighbor estimate of P_j is

$$P_{n,j}(x) = \frac{1}{k} \sum_{i=1}^{k} \mathbb{I}_{\{Y_{(n,i)}(x)=j\}}. \tag{15}$$

As before, from these estimates we derive the plug-in confidence set $C_n(x)$ such that C_n and $P_{n,j}$ correspond to $\tilde{C}$ and $\tilde{P}_j$, respectively. Furthermore,

$$\tilde{t} = t_{k,n} := \ln n(1/\sqrt{k} + (k/n)^{1/d}).$$

Theorem 4. *Assume that μ has a density. If $k = k_n$ such that $k_n/(\ln n)^2 \to \infty$ and $(\ln n)^d k_n/n \to 0$, then for any distribution of (X, Y),*

$$\liminf_{n\to\infty} \mathbb{P}\{Y_{n+1} \in C_n(X_{n+1}) \mid \mathcal{D}_n, X_{n+1} = x\} \geq 1 - \alpha \tag{16}$$

a.s. for μ-almost-all x. If, in addition, the modified Lipschitz condition is satisfied, then for any fixed x,

$$\begin{aligned} &\mathbb{E}\left\{((1-\alpha) - \mathbb{P}\{Y_{n+1} \in C_n(X_{n+1}) \mid \mathcal{D}_n, X_{n+1} = x\})^+\right\} \\ &= O\left(\ln n/\sqrt{k_n}\right) + O\left(\ln n(k_n/n)^{1/d}\right). \end{aligned} \tag{17}$$

Theorem 5. *Assume that μ has a density and $k_n/(\ln n)^2 \to \infty$ and $(\ln n)^d k_n/n \to 0$. Under the modified Lipschitz condition we have that*

$$\mathbb{P}\{|C_n(x)| - |C^*(x)| > 0\} \leq 2n^{-2\ln n} + 2e^{-3k_n/14\mathbb{I}_{\ln n \geq 2C^*}}. \tag{18}$$

5 Proofs

Proof Theorem 1. We have that

$$\mathbb{P}\{Y_{n+1} \in C_n \mid Y_1, \dots, Y_n\} = F(f_{n,hi}) - F(f_{n,lo}).$$

and

$$\begin{aligned} &\left|[F(f_{n,hi}) - F(f_{n,lo})] - (1-\alpha)\right| \\ &\leq \left|[F(f_{n,hi}) - F(f_{n,lo})] - (1-\alpha - 2\ln n/\sqrt{n})\right| + 2\ln n/\sqrt{n} \\ &= \left|[F(f_{n,hi}) - F(f_{n,lo})] - [F_n(f_{n,hi}) - F_n(f_{n,lo})]\right| + 2\ln n/\sqrt{n} \\ &\leq 2\sup_y \left|F(y) - F_n(y)\right| + 2\ln n/\sqrt{n}. \end{aligned}$$

Thus, Theorem 12.4 in Devroye, Györfi and Lugosi [7] implies that

$$\begin{aligned}
&\mathbb{P}\{|[F(f_{n,hi}) - F(f_{n,lo})] - (1-\alpha)| \geq \varepsilon\} \\
&\leq \mathbb{P}\{2 \sup_y |F(y) - F_n(y)| \geq \varepsilon/2\} + \mathbb{I}_{2\ln n/\sqrt{n} \geq \varepsilon/2} \\
&\leq 8(n+1)e^{-n\varepsilon^2/128} + \mathbb{I}_{\ln n/\sqrt{n} \geq \varepsilon/4}.
\end{aligned}$$

Furthermore,

$$\begin{aligned}
\Delta_n - \Delta^* &= D_n(p_n) - D(p^*) \\
&\leq D_n(p^*) - D(p^*) \\
&\leq \left(F_{up}^{-1}(p^*) - F_{n,up}^{-1}(p^* + \ln n/\sqrt{n})\right)^+ \\
&\qquad + \left(F_{n,low}^{-1}(p^* + 1 - \alpha - \ln n/\sqrt{n}) - F_{low}^{-1}(p^* + 1 - \alpha)\right)^+.
\end{aligned} \tag{19}$$

One has

$$\begin{aligned}
&\mathbb{P}\{F_{up}^{-1}(p^*) - F_{n,up}^{-1}(p^* + \ln n/\sqrt{n}) \geq 0\} \\
&\leq \mathbb{P}\{F_n(F_{up}^{-1}(p^*)) \geq p^* + \ln n/\sqrt{n}\} \\
&= \mathbb{P}\{F_n(F_{up}^{-1}(p^*)) - F(F_{up}^{-1}(p^*)) \geq \ln n/\sqrt{n}\}.
\end{aligned}$$

The Hoeffding inequality implies that

$$\mathbb{P}\{F_{up}^{-1}(p^*) - F_{n,up}^{-1}(p^* + \ln n/\sqrt{n}) \geq 0\} \leq e^{-2n(\ln n/\sqrt{n})^2} = 1/n^{2\ln n}.$$

Similarly, we get that

$$\mathbb{P}\{F_{n,low}^{-1}(p^* + 1 - \alpha - \ln n/\sqrt{n}) - F_{low}^{-1}(p^* + 1 - \alpha) \geq 0\} \leq 1/n^{2\ln n}.$$

□

Proof of Theorem 2. One has that

$$\mathbb{P}\{Y_{n+1} \in C_{k,n}(X_{n+1}) \mid \mathcal{D}_n, X_{n+1} = x\} = F(f_{k,n,hi}(x) \mid x) - F(f_{k,n,lo}(x) \mid x).$$

Devroye [6] proved that $k_n/\ln n \to \infty$ and $k_n/n \to 0$ imply that

$$\left|F(y \mid x) - F_{k_n,n}(y \mid x)\right| \to 0 \tag{20}$$

a.s. for all y and for μ-almost all x, from which we get

$$\sup_y \left|F(y \mid x) - F_{k_n,n}(y \mid x)\right| \to 0 \tag{21}$$

a.s. for μ-almost all x, where we used the argument of the standard proof of Glivenko-Cantelli theorem showing that the pointwise convergence of empirical distribution functions implies the uniform convergence. We have that

$$
\begin{aligned}
&\big|[F(f_{k,n,hi}(x) \mid x) - F(f_{k,n,lo}(x) \mid x)] - (1-\alpha)\big| \\
&\leq \big|[F(f_{k,n,hi}(x) \mid x) - F(f_{k,n,lo}(x) \mid x)] - (1-\alpha-2t_{k,n})\big| + 2t_{k,n} \\
&= \big|[F(f_{k,n,hi}(x) \mid x) - F(f_{k,n,lo}(x) \mid x)] \\
&\qquad - [F_{k,n}(f_{k,n,hi}(x) \mid x) - F_{k,n}(f_{k,n,lo}(x) \mid x)]\big| + 2t_{k,n} \\
&\leq 2\sup_y \big|F(y \mid x) - F_{k,n}(y \mid x)\big| + 2t_{k,n}.
\end{aligned}
\tag{22}
$$

Under the conditions of the theorem

$$
t_{k_n,n} \to 0.
$$

Therefore (21) and (22) imply (9). Because of (22), (10) is proved if

$$
\mathbb{E}\left\{\sup_y |F(y \mid x) - F_{k,n}(y \mid x)|\right\} = O\left(\ln k/\sqrt{k}\right) + O\left((k/n)^{1/d}\right). \tag{23}
$$

Put

$$
\bar{F}_{k,n}(y \mid x) = \mathbb{E}\{F_{k,n}(y \mid x) \mid X_1, \dots, X_n\} = \frac{1}{k}\sum_{i=1}^{k} F(y \mid X_{(n,i)}(x)).
$$

Concerning (23), we show that

$$
\mathbb{E}\left\{\sup_y |F_{k,n}(y \mid x) - \bar{F}_{k,n}(y \mid x)|\right\} = O\left(\ln k/\sqrt{k}\right) \tag{24}
$$

and

$$
\mathbb{E}\left\{\sup_y |\bar{F}_{k,n}(y \mid x) - F(y \mid x)|\right\} = O\left((k/n)^{1/d}\right). \tag{25}
$$

Proof of (24): We have

$$
\begin{aligned}
&\mathbb{E}\left\{\sup_y |F_{k,n}(y \mid x) - \bar{F}_{k,n}(y \mid x)|\right\} \\
&= \mathbb{E}\left\{\mathbb{E}\left\{\sup_y |F_{k,n}(y \mid x) - \bar{F}_{k,n}(y \mid x)| \mid X_1, \dots, X_n\right\}\right\} \\
&= \mathbb{E}\left\{\int_0^1 \mathbb{P}\left\{\sup_y |F_{k,n}(y \mid x) - \bar{F}_{k,n}(y \mid x)| \geq s \mid X_1, \dots, X_n\right\} ds\right\}.
\end{aligned}
$$

Conditioning with respect to $X_1, \ldots, X_n$ we note that $F_{k,n}(y \mid x)$ is an arithmetic mean of independent $\{0,1\}$ valued random variables. Although these random variables are not identically distributed, we can extend the proof of Theorem 12.4 in Devroye, Györfi and Lugosi [7], which is a sharpened version of the Glivenko-Cantelli theorem. Thus, we obtain that

$$\mathbb{P}\left\{\sup_y |F_{k,n}(y \mid x) - \bar{F}_{k,n}(y \mid x)| \geq s \mid X_1, \ldots, X_n\right\} \leq 2(k+1)e^{-ks^2/32},$$

$0 < s < 1$, which implies

$$\begin{aligned}
\mathbb{E}\left\{\sup_y |F_{k,n}(y \mid x) - \bar{F}_{k,n}(y \mid x)|\right\} &\leq \int_0^1 \min\{1, 2(k+1)e^{-ks^2/32}\} ds \\
&\leq \varepsilon + 2(k+1)\int_\varepsilon^1 e^{-ks^2/32} ds \\
&\leq \varepsilon + 2(k+1)\frac{2\sqrt{2}}{\sqrt{k}}\int_{k\varepsilon^2/32}^\infty e^{-t}t^{-1/2} dt \\
&\leq \varepsilon + \frac{64}{\varepsilon}e^{-k\varepsilon^2/32},
\end{aligned}$$

which yields (24) by the choice $\varepsilon = \ln k/\sqrt{k}$.
Proof of (25): The modified Lipschitz condition (8) implies that for each y,

$$\begin{aligned}
|\bar{F}_{k,n}(y \mid x) - F(y \mid x)| &\leq \frac{1}{k}\sum_{i=1}^k |F(y \mid X_{(n,i)}(x)) - F(y \mid x)| \\
&\leq C^* \frac{1}{k}\sum_{i=1}^k \mu(S_{x,\|x-X_{(n,i)}(x)\|})^{1/d} \\
&\leq C^* \mu(S_{x,\|x-X_{(n,k)}(x)\|})^{1/d}.
\end{aligned}$$

For i.i.d. uniformly distributed $U_1, \ldots, U_n$, let $U_{(1,n)}, \ldots, U_{(n,n)}$ denote the corresponding order statistic. If μ has a density, then from Section 1.2 in Biau and Devroye [2] we have that

$$\mu(S_{x,\|x-X_{(n,k)}(x)\|}) \stackrel{\mathcal{D}}{=} U_{(k,n)}. \tag{26}$$

Thus, for any fixed y and $s > 0$ with $(s/C^*)^d > k/n$, the Bernstein inequality implies that

$$
\begin{aligned}
&\mathbb{P}\left\{\sup_y |\bar{F}_{k,n}(y \mid x) - F(y \mid x)| \geq s\right\} \\
&\leq \mathbb{P}\left\{C^* U_{(k,n)}^{1/d} \geq s\right\} \\
&= \mathbb{P}\left\{U_{(k,n)} \geq (s/C^*)^d\right\} \\
&= \mathbb{P}\left\{\frac{1}{n}\sum_{i=1}^{n}\left(\mathbb{I}_{U_i \leq (s/C^*)^d} - (s/C^*)^d\right) < -[(s/C^*)^d - k/n]\right\} \\
&\leq e^{-\frac{n[(s/C^*)^d - k/n]^2}{2(s/C^*)^d + 2[(s/C^*)^d - k/n]/3}}.
\end{aligned}
$$

Therefore

$$
\begin{aligned}
&\mathbb{E}\left\{\sup_y |\bar{F}_{k,n}(y \mid x) - F(y \mid x)|\right\} \\
&= \int_0^1 \mathbb{P}\left\{\sup_y |\bar{F}_{k,n}(y \mid x) - F(y \mid x)| \geq s\right\} ds \\
&\leq 2C^*(k/n)^{1/d} + \int_{2C^*(k/n)^{1/d}}^{\infty} e^{-\frac{n[(s/C^*)^d - k/n]^2}{2(s/C^*)^d + 2[(s/C^*)^d - k/n]/3}} ds \\
&\leq 2C^*(k/n)^{1/d} + \int_{2C^*(k/n)^{1/d}}^{\infty} e^{-3n(s/C^*)^d/32} ds \\
&\leq 2C^*(k/n)^{1/d} + O((1/n)^{1/d}) \\
&= O((k/n)^{1/d}),
\end{aligned}
$$

and so we get (25). □

Proof of Theorem 3. Notice that

$$
\begin{aligned}
&\Delta_n(x) - \Delta^*(x) \\
&= D_n(p_n) - D(p^*) \\
&\leq D_n(p^*) - D(p^*) \\
&\leq \left(F_{x,up}^{-1}(p^*) - F_{x,k,n,up}^{-1}(p^* + t_{k,n})\right)^+ \\
&\quad + \left(F_{x,k,n,low}^{-1}(p^* + 1 - \alpha - t_{k,n}) - F_{x,low}^{-1}(p^* + 1 - \alpha)\right)^+. \qquad (27)
\end{aligned}
$$

Proof of (ii): We have that

$$
\begin{aligned}
&\mathbb{P}\left\{F_{x,up}^{-1}(p^*) - F_{x,k,n,up}^{-1}(p^* + t_{k,n}) \geq 0\right\}\\
&\leq \mathbb{P}\left\{F_{k,n}(F_{x,up}^{-1}(p^*) \mid x) \geq p^* + t_{k,n}\right\}\\
&= \mathbb{P}\Big\{F_{k,n}(F_{x,up}^{-1}(p^*) \mid x) - F(F_{x,up}^{-1}(p^*) \mid x) \geq \ln n(1/\sqrt{k} + (k/n)^{1/d})\Big\}\\
&\leq \mathbb{P}\Big\{F_{k,n}(F_{x,up}^{-1}(p^*) \mid x) - \bar{F}_{k,n}(F_{x,up}^{-1}(p^*) \mid x) \geq \ln n/\sqrt{k}\Big\}\\
&+ \mathbb{P}\Big\{\bar{F}_{k,n}(F_{x,up}^{-1}(p^*) \mid x) - F(F_{x,up}^{-1}(p^*) \mid x) \geq \ln n(k/n)^{1/d}\Big\}.
\end{aligned}
$$

As in the previous proofs, the Hoeffding inequality and the Bernstein inequality imply that

$$
\begin{aligned}
\mathbb{P}\Big\{F_{k,n}(F_{x,up}^{-1}(p^*) \mid x) - \bar{F}_{k,n}(F_{x,up}^{-1}(p^*) \mid x) \geq \ln n/\sqrt{k}\Big\} &\leq e^{-2k(\ln n/\sqrt{k})^2}\\
&= n^{-2\ln n},
\end{aligned}
$$

and for $\ln n \geq 2C^*$,

$$
\begin{aligned}
&\mathbb{P}\left\{\bar{F}_{k,n}(F_{x,up}^{-1}(p^*) \mid x) - F(F_{x,up}^{-1}(p^*) \mid x) \geq \ln n(k/n)^{1/d}\right\}\\
&\leq \mathbb{P}\left\{C^* U_{(k,n)}^{1/d} \geq \ln n(k/n)^{1/d}\right\}\\
&\leq \mathbb{P}\left\{U_{(k,n)} \geq 2k/n\right\}\\
&= \mathbb{P}\left\{\frac{1}{n}\sum_{i=1}^{n}\left(\mathbb{I}_{U_i \leq 2k/n} - \mathbb{E}\{\mathbb{I}_{U_i \leq 2k/n}\}\right) \leq -k/n\right\}\\
&\leq e^{-\frac{n(k/n)^2}{4(k/n)+2(k/n)/3}}\\
&= e^{-3k/14}.
\end{aligned}
$$

Thus,

$$
\mathbb{P}\left\{F_{x,up}^{-1}(p^*) - F_{x,k,n,up}^{-1}(p^* + t_{k,n}) \geq 0\right\} \leq n^{-2\ln n} + e^{-3k/14\mathbb{I}_{\ln n \geq 2C^*}}.
$$

The proof of

$$
\begin{aligned}
&\mathbb{P}\left\{F_{x,k,n,low}^{-1}(p^* + 1 - \alpha - t_{k,n}) - F_{x,low}^{-1}(p^* + 1 - \alpha) \geq 0\right\}\\
&\leq n^{-2\ln n} + e^{-3k/14\mathbb{I}_{\ln n \geq 2C^*}}
\end{aligned}
$$

is similar.

Proof of (i): Because of (27) we have to prove that

$$
\left(F_{x,up}^{-1}(p^*) - F_{x,k,n,up}^{-1}(p^* + t_{k,n})\right)^+ \to 0
$$

and

$$\left(F^{-1}_{x,k,n,low}(p^*+1-\alpha-t_{k,n})-F^{-1}_{x,low}(p^*+1-\alpha)\right)^+\to 0$$

a.s. for μ-almost-all x.
If the conditional quantiles at p^* and at $p^*+1-\alpha$ are unique, then Theorem 1 in Hansmann and Kohler [10] and (20) imply

$$\begin{aligned}\left(F^{-1}_{x,up}(p^*)-F^{-1}_{x,k,n,up}(p^*+t_{k,n})\right)^+ &\leq \left(F^{-1}_{x,up}(p^*)-F^{-1}_{x,k,n,up}(p^*)\right)^+\\ &\leq \left|F^{-1}_{x,up}(p^*)-F^{-1}_{x,k,n,up}(p^*)\right|\\ &\to 0\end{aligned}$$

and

$$\begin{aligned}&\left(F^{-1}_{x,k,n,low}(p^*+1-\alpha-t_{k,n})-F^{-1}_{x,low}(p^*+1-\alpha)\right)^+\\ &\leq \left(F^{-1}_{x,k,n,low}(p^*+1-\alpha)-F^{-1}_{x,low}(p^*+1-\alpha)\right)^+\\ &\leq \left|F^{-1}_{x,k,n,low}(p^*+1-\alpha)-F^{-1}_{x,low}(p^*+1-\alpha)\right|\\ &\to 0\end{aligned}$$

a.s. for μ-almost-all x.
If the function $F(y \mid \cdot)$ satisfies the modified Lipschitz condition, and $\mathbb{I}_{\ln n\geq 2C^*}=1$ for n large enough, then (14) implies that

$$\mathbb{P}\{\Delta_n(x)-\Delta^*(x)\geq 0\}\leq 2n^{-2\ln n}+2e^{-3k_n/14}$$

and so the condition $k_n/\ln n\to\infty$ together with the Borel-Cantelli lemma implies the strong consistency (13). □

Proof of Theorem 4. For the plug-in confidence set, we can bound the conditional coverage probability and the excess size as follows.

$$\begin{aligned}\mathbb{P}\{Y\in\tilde{C}(X)\mid X=x\} &= \sum_{j\in\tilde{C}(x)}P_j(x)\\ &\geq \sum_{j\in\tilde{C}(x)}\tilde{P}_j(x)-\sum_{j=1}^{K}|P_j(x)-\tilde{P}_j(x)|\\ &\geq 1-\alpha-\tilde{t}-\sum_{j=1}^{K}|P_j(x)-\tilde{P}_j(x)|. \qquad (28)\end{aligned}$$

Furthermore,

$$\begin{aligned}
\mathbb{I}_{|\tilde{C}(x)|-|C^*(x)|>0} &= \mathbb{I}_{\tilde{L}(x)>L^*(x)} \\
&= \mathbb{I}_{\tilde{L}(x)-1\geq L^*(x)} \\
&\leq \mathbb{I}_{1-\alpha-\tilde{t}>\sum_{j=1}^{L^*(x)} \tilde{P}_{i_j(x)}(x)} \\
&\leq \mathbb{I}_{1-\alpha-\tilde{t}>\sum_{j=1}^{L^*(x)} P_{i_j(x)}(x)-\sum_{j=1}^{K}|P_j(x)-\tilde{P}_j(x)|} \\
&\leq \mathbb{I}_{1-\alpha-\tilde{t}>1-\alpha-\sum_{j=1}^{K}|P_j(x)-\tilde{P}_j(x)|} \\
&= \mathbb{I}_{\tilde{t}<\sum_{j=1}^{K}|P_j(x)-\tilde{P}_j(x)|}. \qquad (29)
\end{aligned}$$

(28) implies that

$$\begin{aligned}
&\mathbb{P}\{Y_{n+1}\in C_n(X_{n+1}) \mid \mathcal{D}_n, X_{n+1}=x\} \\
&\geq 1-\alpha-t_{k,n}-\sum_{j=1}^{K}|P_j(x)-P_{n,j}(x)|. \qquad (30)
\end{aligned}$$

Under the conditions of the theorem $t_{k_n,n}\to 0$. Devroye [6] proved that $k_n/\ln n\to\infty$ and $k_n/n\to 0$ imply that

$$|P_j(x)-P_{n,j}(x)|\to 0 \qquad (31)$$

a.s. for all j and for μ-almost all x, from which we get the first half of the theorem. Because of (30), (17) is proved if

$$\mathbb{E}\left\{|P_j(x)-P_{n,j}(x)|\right\} = O\left(\ln k/\sqrt{k}\right)+O\left((k/n)^{1/d}\right). \qquad (32)$$

This last step can be verified in the same way as in the proof of Theorem 2. □

Proof of Theorem 5. (29) implies that

$$\begin{aligned}
\mathbb{P}\{|C_n(x)|-|C^*(x)|>0\} &\leq \mathbb{P}\left\{t_{k_n,n}<\sum_{j=1}^{K}|P_j(x)-P_{n,j}(x)|\right\} \\
&\leq 2n^{-2\ln n}+2e^{-3k_n/14\mathbb{I}_{\ln n\geq 2C^*}},
\end{aligned}$$

where the last step can be made as in the proof of Theorem 3. □

References

[1] Barber, R. F., Candés, E. J., Ramdas, A. and Tibshirani, R. J.: The limits of distribution-free conditional predictive inference. (submitted), 2020.

[2] Biau, G. and Devroye, L.: *Lectures on the Nearest Neighbor Method*, Springer–Verlag, New York, 2015.

[3] Chaudhuri K. and Dasgupta, S.: Rates of convergence for nearest neighbor classification. In *Advances in Neural Information Processing Systems 27*, Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence and K. Q. Weinberger, editors, 3437–3445, 2014.

[4] Chzhen, E., Denis, C. and Hebiri, M.: Minimax semi-supervised set-valued approach to multi-class classification. (submitted), 2020.

[5] Denis, C. and Hebiri, M.: Confidence sets with expected sizes for multiclass classification. *Journal of Machine Learning Research*, 18:1–28, 2017.

[6] Devroye, L.: Necessary and sufficient conditions for the pointwise convergence of nearest neighbor regression function estimates, *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 61:467–481, 1982.

[7] Devroye, L., Györfi, L. and Lugosi, G.: *A Probabilistic Theory of Pattern Recognition*, Springer–Verlag, New York, 1996.

[8] Döring, M., Györfi, L. and Walk, H.: Exact rate of convergence of k-nearest-neighbor classification rule, *Journal of Machine Learning Research*, 18:1–16, 2018.

[9] Györfi, L., Kohler, M., Krzyżak, A. and Walk, H.: *A Distribution-Free Theory of Nonparametric Regression*, Springer–Verlag, New York, 2002.

[10] Hansmann, M. and Kohler, M.: Estimation of conditional quantiles from data with additional measurement errors, *J. Statistical Planning and Inference*, 200:176–195, 2019.

[11] Lei, J. and Wasserman, L.: Distribution-free prediction bands for non-parametric regression. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76:71–96, 2014.

[12] Lei, J., G'Sell, M., Rinaldo, A., Tibshirani, R. J. and Wasserman, L.: Distribution-free predictive inference for regression. *Journal of the American Statistical Association*, 113:1094–1111, 2018.

[13] Lei, J., Robins, J. and Wasserman, L.: Distribution-free prediction sets. *Journal of the American Statistical Association*, 108:278–287, 2013.

[14] Papadopoulos, H., Proedrou, K., Vovk, V. and Gammerman, A.: Inductive confidence machines for regression. In European Conference on Machine Learning, pages 345–356. Springer-Verlag, New York, 2002.

[15] Romano, Y., Patterson, E. and Candés, E. J.: Conformalized quantile regression, *Advances in Neural Information Processing Systems 32 (NIPS 2019)*, pp. 3538–3548, 2019.

[16] Vovk, V.: Conditional validity of inductive conformal predictors. *Machine Learning*, 92:349–376, 2013.

László Györfi
Department of Computer Science and Information Theory, Budapest University of Technology and Economics, Hungary
E-Mail: gyorfi@cs.bme.hu

Harro Walk
Fachbereich Mathematik, Universität Stuttgart, Stuttgart, Germany
E-Mail: harro.walk@mathematik.uni-stuttgart.de

Erschienene Preprints ab Nummer 2013-001

Komplette Liste:
www.f08.uni-stuttgart.de/mathematik/forschung/publikationen/mathematische_berichte

2020-002 *Györfi, L.; Walk, H.:* Nearest neighbor based conformal prediction

2020-001 *Sändig, A.-M.:* Regularity results to semilinear Parabolic Initial Boundary Value Problems with Nonlinear Newton Boundary Conditions in a polygonal space-time cylinder

2019-005 *Knarr, N.; Stroppel, M.:* Embeddings and ambient automorphisms of the Pappus configuration

2019-004 *Fischer, S.:* Some New Bounds on the Entropy Numbers of Diagonal Operators

2019-003 *Braun, A.; Kohler, M.; Walk, H.:* On the rate of convergence of a neural network regression estimate learned by gradient descent

2019-002 *Györfi, L.; Henze, N.; Walk, H.:* The limit distribution of the maximum probability nearest neighbor ball

2019-001 *Gaspoz, F.D.; Heine, C.-J.; Siebert, K.G.:* An Alternative Proof of H^1-Stability of the L_2-Projection on Graded Meshes

2018-003 *Kollross, A.:* Octonions, triality, the exceptional Lie algebra F4, and polar actions on the Cayley hyperbolic plane

2018-002 *Diaź-Ramos, J.C.; Domínguez-Vázquez, M.; Kollross, A.:* On homogeneous manifolds whose isotropy actions are polar

2018-001 *Grundhöfer, T.; Stroppel, M.; Van Maldeghem, H.:* Embeddings of hermitian unitals into pappian projective planes

2017-011 *Hansmann, M.; Kohler, M.; Walk, H.:* On the strong universal consistency of local averaging regression estimates

2017-010 *Devroye, L.; Györfi, L.; Lugosi, G.; Walk, H.:* A nearest neighbor estimate of a regression functional

2017-009 *Steinke, G.; Stroppel, M.:* On elation Laguerre planes with a two-transitive orbit on the set of generators

2017-008 *Steinke, G.; Stroppel, M.:* Laguerre planes and shift planes

2017-007 *Blunck, A.; Knarr, N.; Stroppel, B.; Stroppel, M.:* Transitive groups of similitudes generated by octonions

2017-006 *Blunck, A.; Knarr, N.; Stroppel, B.; Stroppel, M.:* Clifford parallelisms defined by octonions

2017-005 *Knarr, N.; Stroppel, M.:* Subforms of Norm Forms of Octonion Fields

2017-004 *Apprich, C.; Dieterich, A.; Höllig, K.; Nava-Yazdani, E.:* Cubic Spline Approximation of a Circle with Maximal Smoothness and Accuracy

2017-003 *Fischer, S.; Steinwart, I.:* Sobolev Norm Learning Rates for Regularized Least-Squares Algorithm

2017-002 *Farooq, M.; Steinwart, I.:* Learning Rates for Kernel-Based Expectile Regression

2017-001 *Bauer, B.; Devroye, L; Kohler, M.; Krzyzak, A.; Walk, H.:* Nonparametric Estimation of a Function From Noiseless Observations at Random Points

2016-006 *Devroye, L.; Györfi, L.; Lugosi, G.; Walk, H.:* On the measure of Voronoi cells

2016-005 *Kohls, C.; Kreuzer, C.; Rösch, A.; Siebert, K.G.:* Convergence of Adaptive Finite Elements for Optimal Control Problems with Control Constraints

2016-004 *Blaschzyk, I.; Steinwart, I.:* Improved Classification Rates under Refined Margin Conditions

2016-003 *Feistauer, M.; Roskovec, F.; Sändig, AM.:* Discontinuous Galerkin Method for an Elliptic Problem with Nonlinear Newton Boundary Conditions in a Polygon

2016-002 *Steinwart, I.:* A Short Note on the Comparison of Interpolation Widths, Entropy Numbers, and Kolmogorov Widths

2016-001 *Köster, I.:* Sylow Numbers in Spectral Tables

2015-016 *Hang, H.; Steinwart, I.:* A Bernstein-type Inequality for Some Mixing Processes and Dynamical Systems with an Application to Learning

2015-015 *Steinwart, I.:* Representation of Quasi-Monotone Functionals by Families of Separating Hyperplanes

2015-014 *Muhammad, F.; Steinwart, I.:* An SVM-like Approach for Expectile Regression

2015-013 *Nava-Yazdani, E.:* Splines and geometric mean for data in geodesic spaces

2015-012 *Kimmerle, W.; Köster, I.:* Sylow Numbers from Character Tables and Group Rings

2015-011 *Györfi, L.; Walk, H.:* On the asymptotic normality of an estimate of a regression functional

2015-010 *Gorodski, C, Kollross, A.:* Some remarks on polar actions

2015-009 *Apprich, C.; Höllig, K.; Hörner, J.; Reif, U.:* Collocation with WEB–Splines

2015-008 *Kabil, B.; Rodrigues, M.:* Spectral Validation of the Whitham Equations for Periodic Waves of Lattice Dynamical Systems

2015-007 *Kollross, A.:* Hyperpolar actions on reducible symmetric spaces

2015-006 *Schmid, J.; Griesemer, M.:* Well-posedness of Non-autonomous Linear Evolution Equations in Uniformly Convex Spaces

2015-005 *Hinrichs, A.; Markhasin, L.; Oettershagen, J.; Ullrich, T.:* Optimal quasi-Monte Carlo rules on higher order digital nets for the numerical integration of multivariate periodic functions

2015-004 *Kutter, M.; Rohde, C.; Sändig, A.-M.:* Well-Posedness of a Two Scale Model for Liquid Phase Epitaxy with Elasticity

2015-003 *Rossi, E.; Schleper, V.:* Convergence of a numerical scheme for a mixed hyperbolic-parabolic system in two space dimensions

2015-002 *Döring, M.; Györfi, L.; Walk, H.:* Exact rate of convergence of kernel-based classification rule

2015-001 *Kohler, M.; Müller, F.; Walk, H.:* Estimation of a regression function corresponding to latent variables

2014-021 *Neusser, J.; Rohde, C.; Schleper, V.:* Relaxed Navier-Stokes-Korteweg Equations for Compressible Two-Phase Flow with Phase Transition

2014-020 *Kabil, B.; Rohde, C.:* Persistence of undercompressive phase boundaries for isothermal Euler equations including configurational forces and surface tension

2014-019 *Bilyk, D.; Markhasin, L.:* BMO and exponential Orlicz space estimates of the discrepancy function in arbitrary dimension

2014-018 *Schmid, J.:* Well-posedness of non-autonomous linear evolution equations for generators whose commutators are scalar

2014-017 *Margolis, L.:* A Sylow theorem for the integral group ring of $PSL(2, q)$

2014-016 *Rybak, I.; Magiera, J.; Helmig, R.; Rohde, C.:* Multirate time integration for coupled saturated/unsaturated porous medium and free flow systems

2014-015 *Gaspoz, F.D.; Heine, C.-J.; Siebert, K.G.:* Optimal Grading of the Newest Vertex Bisection and H^1-Stability of the L_2-Projection

2014-014 *Kohler, M.; Krzyżak, A.; Walk, H.:* Nonparametric recursive quantile estimation

2014-013 *Kohler, M.; Krzyżak, A.; Tent, R.; Walk, H.:* Nonparametric quantile estimation using importance sampling

2014-012 *Györfi, L.; Ottucsák, G.; Walk, H.:* The growth optimal investment strategy is secure, too.

2014-011 *Györfi, L.; Walk, H.:* Strongly consistent detection for nonparametric hypotheses

2014-010 *Köster, I.:* Finite Groups with Sylow numbers $\{q^x, a, b\}$

2014-009 *Kahnert, D.:* Hausdorff Dimension of Rings

2014-008 *Steinwart, I.:* Measuring the Capacity of Sets of Functions in the Analysis of ERM

2014-007 *Steinwart, I.:* Convergence Types and Rates in Generic Karhunen-Loève Expansions with Applications to Sample Path Properties

2014-006 *Steinwart, I.; Pasin, C.; Williamson, R.; Zhang, S.:* Elicitation and Identification of Properties

2014-005 *Schmid, J.; Griesemer, M.:* Integration of Non-Autonomous Linear Evolution Equations

2014-004 *Markhasin, L.:* L_2- and $S^r_{p,q}B$-discrepancy of (order 2) digital nets

2014-003 *Markhasin, L.:* Discrepancy and integration in function spaces with dominating mixed smoothness

2014-002 *Eberts, M.; Steinwart, I.:* Optimal Learning Rates for Localized SVMs

2014-001 *Giesselmann, J.:* A relative entropy approach to convergence of a low order approximation to a nonlinear elasticity model with viscosity and capillarity

2013-016 *Steinwart, I.:* Fully Adaptive Density-Based Clustering

2013-015 *Steinwart, I.:* Some Remarks on the Statistical Analysis of SVMs and Related Methods

2013-014 *Rohde, C.; Zeiler, C.:* A Relaxation Riemann Solver for Compressible Two-Phase Flow with Phase Transition and Surface Tension

2013-013 *Moroianu, A.; Semmelmann, U.:* Generalized Killling spinors on Einstein manifolds

2013-012 *Moroianu, A.; Semmelmann, U.:* Generalized Killing Spinors on Spheres

2013-011 *Kohls, K; Rösch, A.; Siebert, K.G.:* Convergence of Adaptive Finite Elements for Control Constrained Optimal Control Problems

2013-010 *Corli, A.; Rohde, C.; Schleper, V.:* Parabolic Approximations of Diffusive-Dispersive Equations

2013-009 *Nava-Yazdani, E.; Polthier, K.:* De Casteljau's Algorithm on Manifolds

2013-008 *Bächle, A.; Margolis, L.:* Rational conjugacy of torsion units in integral group rings of non-solvable groups

2013-007 *Knarr, N.; Stroppel, M.J.:* Heisenberg groups over composition algebras

2013-006 *Knarr, N.; Stroppel, M.J.:* Heisenberg groups, semifields, and translation planes

2013-005 *Eck, C.; Kutter, M.; Sändig, A.-M.; Rohde, C.:* A Two Scale Model for Liquid Phase Epitaxy with Elasticity: An Iterative Procedure

2013-004 *Griesemer, M.; Wellig, D.:* The Strong-Coupling Polaron in Electromagnetic Fields

2013-003 *Kabil, B.; Rohde, C.:* The Influence of Surface Tension and Configurational Forces on the Stability of Liquid-Vapor Interfaces

2013-002 *Devroye, L.; Ferrario, P.G.; Györfi, L.; Walk, H.:* Strong universal consistent estimate of the minimum mean squared error

2013-001 *Kohls, K.; Rösch, A.; Siebert, K.G.:* A Posteriori Error Analysis of Optimal Control Problems with Control Constraints