Universität
Stuttgart

Fachbereich
Mathematik

Exact rate of convergence of kernel-based classification rule

Maik Döring, László Györfi, Harro Walk

Preprint 2015/002

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**

Exact rate of convergence of kernel-based classification rule

Maik Döring, László Györfi *, and Harro Walk

Institut für Angewandte Mathematik und Statistik
Universität Hohenheim
Schloss Hohenheim, 70599 Stuttgart, Germany
maik.doering@uni-hohenheim.de
Department of Computer Science and Information Theory
Budapest University of Technology and Economics
Magyar Tudósok körútja 2., H-1117 Budapest, Hungary
gyorfi@cs.bme.hu
Department of Mathematics
University of Stuttgart
Pfaffenwaldring 57, 70569 Stuttgart, Germany
walk@mathematik.uni-stuttgart.de

Abstract. A binary classification problem is considered, where the posteriori probability is estimated by the nonparametric kernel regression estimate with naive kernel. The excess error probability of the corresponding plug-in decision classification rule according to the error probability of the Bayes decision is studied such that the excess error probability is decomposed into approximation and estimation error. A general formula is derived for the approximation error. Under a weak margin condition and various smoothness conditions, tight upper bounds are presented on the approximation error. By a Berry-Esseen type central limit theorem a general expression for the estimation error is shown.

AMS Classification: 62G10.

Key words and phrases: lower bound, upper bound, classification error probability, kernel rule, margin condition

1 Introduction

We consider a binary classification problem. Using a kernel estimator for the posteriori probability, the asymptotics of the error probability is examined of the corresponding plug-in classification rule. In this paper lower and upper bounds are presented on the rate of convergence of the classification error probability.

* This work was supported in part by the National Development Agency (NFÜ, Hungary) as part of the project Introduction of Cognitive Methods for UAV Collision Avoidance Using Millimeter Wave Radar, (grant no.: KMR-12-1-2012-0008)

Let the feature vector X take values in $\mathbb{R}^d$ such that its distribution is denoted by μ and let the label Y be binary valued. If g is an arbitrary decision function, then its error probability is denoted by

$$L(g) = \mathbb{P}\{g(X) \neq Y\}.$$

The Bayes decision g^* minimizes the error probability. It follows that

$$g^*(x) = \begin{cases} 1 \text{ if } \eta(x) > 1/2 \\ 0 \ \text{ otherwise,} \end{cases}$$

where the posteriori probability η is given by

$$\eta(x) = \mathbb{P}\{Y = 1 \mid X = x\}.$$

Let the corresponding error probability be

$$L^* = \mathbb{P}\{g^*(X) \neq Y\}.$$

Put

$$D(x) = 2\eta(x) - 1,$$

then the Bayes decision has the following equivalent form:

$$g^*(x) = \begin{cases} 1 \text{ if } D(x) > 0 \\ 0 \ \text{ otherwise.} \end{cases}$$

In the standard model of pattern recognition, we are given training labeled samples, which are independent and identically copies of (X, Y): $(X_1, Y_1), \dots, (X_n, Y_n)$. Based on these labeled samples, one can estimate the regression function D by D_n, and the corresponding plug-in classification rule g_n derived from D_n is defined by

$$g_n(x) = \begin{cases} 1 \text{ if } D_n(x) > 0 \\ 0 \ \text{ otherwise.} \end{cases}$$

In the following our focus lies on the rate of convergence of the excess error probability $\mathbb{E}\{L(g_n)\} - L^*$. In Section 2 margin conditions are discussed, which measure how fast the regression function D crosses the decision boundary. A nonparametric kernel regression estimate of D is introduced and a decomposition of the excess error probability into approximation and estimation error is considered in Section 3. Tight upper bounds on the approximation error are shown in Section 4 depending on margin and smoothness conditions on D. By a Berry-Esseen type central limit theorem a general expression for the estimation error is derived in Section 5. Finally, some conclusions are given.

2 Margin conditions

Given the plug-in classification rule g_n derived from D_n it follows

$$\mathbb{E}\{L(g_n)\} - L^* \leq \mathbb{E}\{|D(X) - D_n(X)|\}$$

(cf. Theorem 2.2 in Devroye, Györfi, Lugosi [3]). Therefore we may get an upper bound on the rate of convergence of the excess error probability $\mathbb{E}\{L(g_n)\} - L^*$ via the L_1 rate of convergence of the corresponding regression estimation.

However, according to Section 6.7 in Devroye, Györfi, Lugosi [3], the classification is easier than L_1 regression function estimation, since the rate of convergence of the error probability depends on the behavior of the function D in the neighborhood of the decision boundary

$$B_0 = \{x; D(x) = 0\}. \tag{1}$$

This phenomenon has been discovered by Mammen and Tsybakov [10], Tsybakov [13], who formulated the (strong) margin condition:

– *The strong margin condition.* Assume that for all $0 < t \leq 1$,

$$\mathbb{P}\{|D(X)| \leq t\} \leq ct^{\alpha}, \tag{2}$$

where $\alpha > 0$ and $c > 0$.

Kohler and Krzyżak [7] introduced the weak margin condition:

– *The weak margin condition.* Assume that for all $0 < t \leq 1$,

$$\mathbb{E}\left\{\mathbb{I}_{\{|D(X)|\leq t\}}|D(X)|\right\} \leq ct^{1+\alpha}, \tag{3}$$

where $\mathbb{I}$ denotes the indicator function.

Obviously, the strong margin condition implies the weak margin condition:

$$\mathbb{E}\left\{\mathbb{I}_{\{|D(X)|\leq t\}}|D(X)|\right\} \leq \mathbb{E}\left\{\mathbb{I}_{\{|D(X)|\leq t\}}t\right\} = t\mathbb{P}\left\{|D(X)| \leq t\right\} \leq ct \cdot t^{\alpha}.$$

The difference between the strong and weak margin condition is that, for the strong margin condition, the event

$$\{D(X) = 0\}$$

counts. One can weaken the strong margin condition (2) such that we require

$$\mathbb{P}\{0 < |D(X)| \leq t\} \leq ct^{\alpha}. \tag{4}$$

Obviously, (4) implies (3). Under some mild conditions we have that $\alpha = 1$. (Cf. Lemma 2.) The margin conditions measure how fast the probability of a t-neighborhood of the decision boundary increases with t. A large value of α corresponds to a small probability of the neighborhood of the decision boundary, which means that the probability for events far away of the decision boundary is high. Therefore, a classifier can make the right decision more easily, hence one can expect smaller errors for larger values of α.

Recently, Audibert and Tsybakov [1] proved that if the plug-in classifier g has been derived from the regression estimate $\tilde{D}$ and if D satisfies the strong margin condition, then

$$L(g) - L^* \leq \left(\int (\tilde{D}(x) - D(x))^2 \mu(dx)\right)^{\frac{1+\alpha}{2+\alpha}}. \tag{5}$$

It is easy to see that (5) holds even under weak margin condition: we have that

$$L(g) - L^* = \mathbb{E}\left\{\mathbb{I}_{\{g(X)\neq g^*(X)\}}|D(X)|\right\} \tag{6}$$

(cf. Theorem 2.2 in Devroye, Györfi, Lugosi [3]). Let $sign(x) = 1$ for $x > 0$ and $sign(x) = -1$ for $x \leq 0$. For fixed $t_n > 0$,

$$\begin{aligned}
L(g) - L^* &= \mathbb{E}\left\{\mathbb{I}_{\{sign\,\tilde{D}(X)\neq sign\,D(X),|D(X)|\leq t_n\}}|D(X)|\right\}\\
&\quad+\mathbb{E}\left\{\mathbb{I}_{\{sign\,\tilde{D}(X)\neq sign\,D(X),|D(X)|> t_n\}}|D(X)|\right\}\\
&\leq \mathbb{E}\left\{\mathbb{I}_{\{|D(X)|\leq t_n\}}|D(X)|\right\}\\
&\quad+\mathbb{E}\left\{\mathbb{I}_{\{sign\,\tilde{D}(X)\neq sign\,D(X),|\tilde{D}(X)-D(X)|> t_n\}}|\tilde{D}(X)-D(X)|\right\},
\end{aligned}$$

therefore the weak margin condition implies that

$$\begin{aligned}
L(g) - L^* &\leq ct_n^{1+\alpha} + t_n\mathbb{E}\left\{\mathbb{I}_{\{|\tilde{D}(X)-D(X)|>t_n\}}\frac{|\tilde{D}(X)-D(X)|}{t_n}\right\}\\
&\leq ct_n^{1+\alpha} + t_n\mathbb{E}\left\{\frac{|\tilde{D}(X)-D(X)|^2}{t_n^2}\right\}.
\end{aligned}$$

For the choice

$$t_n = \left(\mathbb{E}\left\{|\tilde{D}(X)-D(X)|^2\right\}\right)^{\frac{1}{2+\alpha}}$$

we get (5).

For bounding the error probability, assume, for example, that D satisfies the *Lipschitz condition*: for any $x, z \in \mathbb{R}^d$

$$|D(x) - D(z)| \leq C\|x - z\|.$$

If D is Lipschitz continuous and X is bounded then there are regression estimates such that

$$\int (D_n(x) - D(x))^2\mu(dx) \leq c_1^2 n^{-\frac{2}{d+2}},$$

therefore (5) means that

$$L(g) - L^* \leq \left(c_1^2 n^{-\frac{2}{d+2}}\right)^{\frac{1+\alpha}{2+\alpha}} = \left(c_1^{1+\alpha} n^{-\frac{1+\alpha}{d+2}}\right)^{\frac{2}{2+\alpha}}.$$

Kohler and Krzyżak [7] proved that for the standard plug-in classification rules (partitioning, kernel, nearest neighbor) and for weak margin condition we get that the order of the upper bound can be smaller:

$$n^{-\frac{1+\alpha}{d+2}}.$$

The main aim of this paper is to show tight upper bounds on the excess error probability $\mathbb{E}\{L(g_n)\} - L^*$ of the kernel classification rule g_n.

3 Kernel classification

We fix $x \in \mathbb{R}^d$, and, for an $h > 0$, let the (naive) kernel estimate of $D(x)$ be

$$D_{n,h}(x) = \frac{1}{n}\sum_{i=1}^{n}(2Y_i - 1)\mathbb{I}_{\{X_i \in S_{x,h}\}}/\mu(S_{x,h}),$$

where $S_{x,h}$ denotes the sphere centered at x with radius h. Notice that $D_{n,h}$ is not a true estimate, because its denominator contains the unknown distribution μ. However, the corresponding plug-in classification rule defined below depends only on the sign of $D_{n,h}(x)$, and so μ doesn't count. The (naive) kernel classification rule is

$$g_{n,h}(x) = \begin{cases} 1 \text{ if } D_{n,h}(x) > 0 \\ 0 \ \text{ otherwise,} \end{cases}$$

(cf. Devroye [2], Devroye and Wagner [4], Krzyżak [8], Krzyżak and Pawlak [9]).

If D is Lipschitz continuous and X is bounded then, for the L_1 error, one has that

$$\mathbb{E}\{|D(X) - D_{n,h}(X)|\} \le c_2 h + \frac{c_3}{\sqrt{nh^d}},$$

(cf. Györfi et al. [5]), so for the choice

$$h = n^{-\frac{1}{d+2}}, \tag{7}$$

the L_1 upper bound implies that

$$\mathbb{E}\{L(g_{n,h})\} - L^* \le c_4 n^{-\frac{1}{d+2}}.$$

Because of (6), we have that the excess error probability of any plug-in classification rule has the following decomposition:

$$\mathbb{E}\{L(g_{n,h})\} - L^* = \mathbb{E}\left\{\int_{\{sign\ D_{n,h}(x) \neq sign\ D(x)\}} |D(x)|\mu(dx)\right\} = I_{n,h} + J_{n,h},$$

where

$$I_{n,h} = \mathbb{E}\left\{\int_{\{sign\ \bar{D}_h(x) = sign\ D_{n,h}(x) \neq sign\ D(x)\}} |D(x)|\mu(dx)\right\}$$

and

$$J_{n,h} = \mathbb{E}\left\{\int_{\{sign\ D_{n,h}(x) \neq sign\ D(x) = sign\ \bar{D}_h(x)\}} |D(x)|\mu(dx)\right\}$$

with $\bar{D}_h(x) = \mathbb{E}\{D_{n,h}(x)\}$. $I_{n,h}$ is called approximation error, while $J_{n,h}$ is the estimation error.

4 Approximation error

First we consider the approximation error. The following proposition means that the lower bound of the approximation error is approximately the half of the upper bound. Further, it shows that the bounds of the approximation error are mainly determined by the bandwidth h.

Proposition 1.

$$\left(\frac{1}{2}+o(1)\right)\bar{I}_h \le I_{n,h} \le \bar{I}_h,$$

where

$$\bar{I}_h = \int_{\{sign\,\bar{D}_h(x)\neq sign\,D(x)\}} |D(x)|\mu(dx).$$

Proof. The upper bound is obvious. The lower bound follows from the central limit theorem, since

$$\begin{aligned}
&\mathbb{E}\left\{\int_{\{0\ge\bar{D}_h(x),D_{n,h}(x)\le 0<D(x)\}} |D(x)|\mu(dx)\right\}\\
&= \int_{\{\bar{D}_h(x)\le 0<D(x)\}} |D(x)|\mathbb{P}\{D_{n,h}(x)\le 0\}\mu(dx)\\
&= \int_{\{\bar{D}_h(x)\le 0<D(x)\}} |D(x)|\mathbb{P}\{D_{n,h}(x)-\bar{D}_h(x)\le -\bar{D}_h(x)\}\mu(dx)\\
&\ge \int_{\{\bar{D}_h(x)\le 0<D(x)\}} |D(x)|\mathbb{P}\{D_{n,h}(x)-\bar{D}_h(x)\le 0\}\mu(dx)\\
&\ge \int_{\{\bar{D}_h(x)\le 0<D(x)\}} |D(x)|\left(\frac{1}{2}+o(1)\right)\mu(dx),
\end{aligned}$$

where $o(1)$ is uniform in x. This can be seen, for example, using the Berry-Esseen inequality as in the proof of Proposition 2 below. The handling of remaining integral is analogous. ∎

Kohler and Krzyżak [7] bounded the rate of convergence of the excess error probability assuming that D satisfies the weak margin condition and the Lipschitz condition. Further they assume that X has a density which is bounded away from zero:

$$f(x)\ge c'>0 \tag{8}$$

They proved that

$$\mathbb{E}\{L(g_{n,h})\}-L^*\le c_5h^{1+\alpha}+\frac{c_6}{nh^d} \tag{9}$$

such that, for the choice (7),

$$\mathbb{E}\{L(g_{n,h})\}-L^*\le c_7n^{-\frac{1+\alpha}{d+2}}.$$

In (9) the approximation error is upper bounded by $c_5 h^{1+\alpha}$. Next we show how it follows from Proposition 1. Denote by

$$B_{0,h} = \left\{ x; \min_{z \in B_0} \|x - z\| \le h \right\}$$

the h-neighborhood of the decision boundary B_0 defined by (1). Let λ be the Lebesgue measure and let $M^*(B_0)$ be the outer surface (Minkowski content) of the decision boundary B_0 defined by

$$M^*(B_0) = \lim_{h \downarrow 0} \frac{\lambda(B_{0,h} \setminus B_0)}{h}.$$

Lemma 1. *If D satisfies the weak margin condition and the Lipschitz condition, then*

$$\bar{I}_h \le c_8 h^{1+\alpha}.$$

If D satisfies the Lipschitz condition, the density f of X exists, it is bounded by f_{max} and $M^*(B_0)$ is finite, then

$$\bar{I}_h \le c_9 h^2.$$

Proof. If $x \notin B_{0,h}$ then

$$sign\, \bar{D}_h(x) = sign\, D(x).$$

Therefore

$$\begin{aligned} \bar{I}_h &= \int_{\{sign\, \bar{D}_h(x) \neq sign\, D(x)\}} |D(x)| \mu(dx) \\ &= \int_{\{sign\, \bar{D}_h(x) \neq sign\, D(x), x \in B_{0,h}\}} |D(x)| \mu(dx). \end{aligned}$$

For any fixed $x \in B_{0,h}$, there is a $z_x \in B_0$ such that $\|x - z_x\| \le h$, which together with the Lipschitz condition implies that

$$|D(x)| = |D(x) - D(z_x)| \le Ch.$$

Thus, by the weak margin condition

$$\begin{aligned} \bar{I}_h &\le \int_{\{|D(x)| \le Ch, x \in B_{0,h}\}} |D(x)| \mu(dx) \\ &\le \int_{\{|D(x)| \le Ch\}} |D(x)| \mu(dx) \\ &\le c(Ch)^{1+\alpha}. \end{aligned}$$

Concerning the second half of the lemma, we have that

$$\begin{aligned}\bar{I}_h &\leq \int_{\{|D(x)|\leq Ch, x\in B_{0,h}\}} |D(x)|\mu(dx)\\ &\leq Ch \int_{\{0<|D(x)|, x\in B_{0,h}\}} 1\mu(dx)\\ &= Ch\mu\{B_{0,h}\setminus B_0\}\\ &\leq Chf_{max}\lambda\{B_{0,h}\setminus B_0\}\\ &\leq Cc_{10}h^2.\end{aligned}$$

■

The technique of the second half of the previous proof implies that $\alpha = 1$.

Lemma 2. *Let D satisfies the lower Lipschitz inequality at B_0, which means a $c^* > 0$ exists, such that for all $t \in [0,1]$ and*

$$x \notin B_{0,c^*t}$$

it follows

$$|D(x)| > t.$$

If the density f of X exists, it is bounded by f_{max}, and the outer surface $M^*(B_0)$ is finite, then the weak margin condition holds with $\alpha = 1$.

Proof. We verify (4) with $\alpha = 1$.

$$\begin{aligned}&\mathbb{P}\{0 < |D(X)| \leq t\}\\ &= \mathbb{P}\{0 < |D(X)| \leq t, X \in B_{0,c^*t}\setminus B_0\} + \mathbb{P}\{0 < |D(X)| \leq t, X \notin B_{0,c^*t}\}\\ &\leq \mathbb{P}\{X \in B_{0,c^*t}\setminus B_0\} + \mathbb{P}\{0 < |D(X)| \leq t, t < |D(X)|\}\\ &\leq c_{10}c^*t.\end{aligned}$$

■

Hall and Kang [6] investigated the bandwidth choice. They assumed that conditional densities exist, which are bounded away from zero. Under twice differentiable conditional densities, they proved that

$$\mathbb{E}\{L(g_{n,h})\} - L^* \leq c_{11}h^4 + o\left(\frac{1}{nh^d}\right). \tag{10}$$

In (10) the approximation error is upper bounded by $c_{11}h^4$. Next we show how it follows from Proposition 1.

Let us introduce some notations:

$$p_+ := \mathbb{P}\{Y = 1\},\ p_- := \mathbb{P}\{Y = 0\}$$

Assume that the density f of X exists. Let f_+ and f_- the conditional densities defined by

$$\mathbb{P}\{X \in A \mid Y = 1\} = \int_A f_+(x)\ dx$$

and

$$\mathbb{P}\{X \in A \mid Y = 0\} = \int_A f_-(x)\ dx.$$

Then

$$f(x) = p_+ \cdot f_+(x) + p_- \cdot f_-(x)$$

and

$$D(x) = \frac{\tilde{f}(x)}{f(x)},$$

where

$$\tilde{f}(x) := p_+ \cdot f_+(x) - p_- \cdot f_-(x).$$

Moreover,

$$f_+(x) = \frac{f(x) \cdot \big(1 + D(x)\big)}{2p_+},\ f_-(x) = \frac{f(x) \cdot \big(1 - D(x)\big)}{2p_-}.$$

Lemma 3. *Assume that $\tilde{f}$ is two-times differentiable with bounded second order partial derivatives. If D satisfies the weak margin condition and the density f is bounded below by f_{min}, then*

$$\bar{I}_h \leq c_{12} h^{2(1+\alpha)}.$$

Proof. Let $H_{\tilde{f}}$ be the Hessian-matrix of $\tilde{f}$. Then the conditions of the lemma imply that

$$\sup_{0 \leq t \leq 1} \left| (x-z)^T H_{\tilde{f}}\big(x + t(z-x)\big)(x-z) \right| \leq c_{13} \|x - z\|^2 \tag{11}$$

with $0 < c_{13} < \infty$. We have the decomposition

$$\begin{aligned}
\bar{D}_h(x) &= \frac{\int_{S_{x,h}} D(z)\mu(dz)}{\mu(S_{x,h})} \\
&= \frac{\int_{S_{x,h}} \frac{\tilde{f}(z)}{f(z)} f(z) dz}{\mu(S_{x,h})} \\
&= \frac{\int_{S_{x,h}} \Big(\tilde{f}(z) - \tilde{f}(x)\Big)\, dz}{\mu(S_{x,h})} + \frac{\tilde{f}(x)\lambda(S_{x,h})}{\mu(S_{x,h})} \\
&= \frac{\int_{S_{x,h}} \Big(\tilde{f}(z) - \tilde{f}(x)\Big)\, dz}{\mu(S_{x,h})} + D(x)\frac{f(x)\lambda(S_{x,h})}{\mu(S_{x,h})}.
\end{aligned}$$

The second order Taylor expansion

$$\tilde{f}(z) - \tilde{f}(x) = (z-x)^T \nabla \tilde{f}(x) + (z-x)^T H_{\tilde{f}}(\tilde{x}_z)(z-x)/2$$

with $\tilde{x}_z \in S_{x,h}$ implies that

$$\begin{aligned}
&\bar{D}_h(x) \\
&= \frac{\int_{S_{x,h}} \left((z-x)^T \nabla \tilde{f}(x) + (z-x)^T H_{\tilde{f}}(\tilde{x}_z)(z-x)/2 \right) \, dz}{\mu(S_{x,h})} + D(x) \frac{f(x)\lambda(S_{x,h})}{\mu(S_{x,h})} \\
&= \frac{\int_{S_{x,h}} (z-x)^T H_{\tilde{f}}(\tilde{x}_z)(z-x)/2 \, dz}{\mu(S_{x,h})} + D(x) \frac{f(x)\lambda(S_{x,h})}{\mu(S_{x,h})}.
\end{aligned}$$

Therefore, from (11) we get that

$$\bar{D}_h(x) \geq -\frac{c_{13} h^2 \lambda(S_{x,h})/2}{\mu(S_{x,h})} + D(x) \frac{f(x)\lambda(S_{x,h})}{\mu(S_{x,h})}.$$

Thus, for $D(x) \geq 0 > \bar{D}_h(x)$, we have

$$|D(x)| \leq \frac{c_{13} h^2}{2 f(x)} \leq \frac{c_{13} h^2}{2 f_{min}}.$$

The same inequality holds for $D(x) < 0 \leq \bar{D}_h(x)$. From the weak margin condition we get

$$\begin{aligned}
\bar{I}_h &= \int_{\left\{ sign\left(\bar{D}_h(x)\right) \neq sign\left(D(x)\right), x \in B_{0,h} \right\}} |D(x)| \mu(dx) \\
&\leq \int_{\left\{ |D(x)| \leq \frac{c_{13} h^2}{2 f_{min}} \right\}} |D(x)| \mu(dx) \\
&\leq c_{12} h^{2(1+\alpha)}.
\end{aligned}$$

■

Under the assumption of Lemma 2 we have that the weak margin condition holds with $\alpha = 1$. Hence by Lemma 3 and Proposition 1 we get that the approximation error $I_{n,h}$ is upper bounded by a multiple of h^4.

The question left is that whether the upper bounds in Lemmas 1 and 3 are tight. Consider some examples, where $\alpha = 1$ and $\bar{I}_h$ can be calculated showing that the order of the lower bounds have the order of the upper bounds.

Example 1. Assume that f is the uniform density on $[-1, 1]^d$. Let $h < 1$, $\beta \geq 1$, $a > 0$, $b > 0$, $a + b < 1$. Choose

$$p_+ = \frac{1}{2} + \frac{b}{4(\beta + 1)}$$

and

$$D(x) = ax_1 + bx_1^\beta \cdot \mathbb{I}_{(0,1]}(x_1),$$

where $x = (x_1, \dots, x_d)$. Then

$$\begin{aligned}
\tilde f(x) &= \frac{ax_1 + bx_1^\beta \cdot \mathbb{I}_{(0,1]}(x_1)}{2^d} \cdot \mathbb{I}_{[-1,1]^d}(x)\\
f_+(x) &= \frac{1 + ax_1 + bx_1^\beta \cdot \mathbb{I}_{(0,1]}(x_1)}{2^{d+1}p_+} \cdot \mathbb{I}_{[-1,1]^d}(x)\\
f_-(x) &= \frac{1 - ax_1 - bx_1^\beta \cdot \mathbb{I}_{(0,1]}(x_1)}{2^{d+1}p_-} \cdot \mathbb{I}_{[-1,1]^d}(x).
\end{aligned}$$

One can check that D satisfies the weak margin condition with $\alpha = 1$. If $x_1 > 0$ then $sign\, \bar D_h(x) = sign\, D(x)$. Let V_d be the volume of the d-dimensional unit sphere, i.e. $V_d = \pi^{d/2}/\Gamma(d/2+1)$. For $-h < x_1 \le 0$

$$\begin{aligned}
&\int_{S_{x,h}} D(x)\ \mu(dx) = \int_{S_{x,h}} \frac{ax_1 + a(z_1 - x_1) + bz_1^\beta \cdot \mathbb{I}_{(0,1]}(z_1)}{2^d} \cdot \mathbb{I}_{[-1,1]^d}(z)\ dz\\
&= \frac{aV_d}{2^d} h^d x_1 + \int_{x_1-h}^{x_1+h} \frac{a(z_1 - x_1) + bz_1^\beta \cdot \mathbb{I}_{(0,1]}(z_1)}{2^d} V_{d-1}\big(h^2 - (z_1 - x_1)^2\big)^{(d-1)/2} dz_1\\
&= \frac{aV_d}{2^d} h^d x_1 + \frac{bV_{d-1}}{2^d} \int_0^{x_1+h} z_1^\beta \big(h^2 - (z_1 - x_1)^2\big)^{(d-1)/2}\ dz_1\\
&\le \frac{aV_d}{2^d} h^d x_1 + \frac{bV_{d-1}}{2^d} h^{d-1} \int_0^{x_1+h} z_1^\beta\ dz_1\\
&\le \frac{h^{d-1}}{2^d}\Big(aV_d h x_1 + \frac{bV_{d-1}}{\beta+1} h^{\beta+1}\Big).
\end{aligned}$$

And we have a lower bound by

$$\begin{aligned}
&\int_{S_{x,h}} D(x)\mu(dx) = \frac{aV_d}{2^d} h^d x_1 + \frac{bV_{d-1}}{2^d} \int_0^{x_1+h} z_1^\beta \big(h^2 - (z_1 - x_1)^2\big)^{(d-1)/2}\ dz_1\\
&= \frac{aV_d}{2^d} h^d x_1 + \frac{bV_{d-1}}{2^d} \int_0^h \mathbb{I}_{(-x_1,\infty)}(\tilde z_1)(\tilde z_1 + x_1)^\beta \big(h^2 - \tilde z_1^2\big)^{(d-1)/2}\ d\tilde z_1\\
&= \frac{aV_d}{2^d} h^d x_1 + \frac{bV_{d-1}}{2^d} \int_0^h \Big(\mathbb{I}_{(0,\infty)}(\tilde z_1)\tilde z_1^\beta \big(h^2 - \tilde z_1^2\big)^{(d-1)/2}\\
&\qquad + x_1 \mathbb{I}_{(-\tilde x_1,\infty)}(\tilde z_1)\beta(\tilde z_1 + \tilde x_1)^{\beta-1}\big(h^2 - \tilde z_1^2\big)^{(d-1)/2}\Big)\ d\tilde z_1\\
&\ge \frac{aV_d}{2^d} h^d x_1 + \frac{bV_{d-1}}{2^d} \int_0^h \Big(\tilde z_1^\beta \big(h^2 - \tilde z_1^2\big)^{(d-1)/2} + x_1 \beta \tilde z_1^{\beta-1} h^{d-1}\Big)\ d\tilde z_1\\
&\ge \frac{aV_d}{2^d} h^d x_1 + \frac{bV_{d-1}}{2^d} \Big(\int_0^{h/2} \tilde z_1^\beta \big(h^2 - (h/2)^2\big)^{(d-1)/2}\ d\tilde z_1\\
&\qquad + \int_{h/2}^h (h/2)^{\beta-1} \tilde z_1 \big(h^2 - \tilde z_1^2\big)^{(d-1)/2}\ d\tilde z_1 + x_1 h^d\Big)
\end{aligned}$$

$$= \frac{aV_d}{2^d} h^d x_1 + \frac{bV_{d-1}}{2^d} \Big(\frac{1}{\beta+1} (h/2)^{\beta+1} \big(h^2 - (h/2)^2 \big)^{(d-1)/2}$$
$$+ (h/2)^{\beta-1} \big(h^2 - (h/2)^2 \big)^{(d+1)/2} \frac{1}{d+1} + x_1 h^d \Big)$$
$$= \frac{h^{d-1}}{2^d} \Big((aV_d + bV_{d-1}) h x_1 + bV_{d-1} \big(\frac{(3/4)^{(d-1)/2}}{(\beta+1)2^{\beta+1}} + \frac{(3/4)^{(d+1)/2}}{(d+1)2^{\beta-1}} \big) h^{\beta+1} \Big).$$

For the notations

$$c_{14} = \frac{bV_{d-1}}{aV_d + bV_{d-1}} \big(\frac{(3/4)^{(d-1)/2}}{(\beta+1)2^{\beta+1}} + \frac{(3/4)^{(d+1)/2}}{(d+1)2^{\beta-1}} \big)$$
$$c_{15} = \frac{bV_{d-1}}{aV_d \cdot (\beta+1)}$$

we get that

$$-c_{14} h^\beta < x_1 \quad \Longrightarrow \quad \bar{D}_h(x) > 0 \quad \Longrightarrow \quad -c_{15} \cdot h^\beta < x_1$$

Therefore

$$\bar{I}_h = \int_{\{sign\ \bar{D}_h(x) \neq sign\ D(x)\}} |D(x)|\ \mu(dx)$$
$$\geq \int_{-c_{14} \cdot h^\beta}^{0} \int_{[-1,1]^{d-1}} -\frac{a}{2^d} x_1\ d(x_2, \ldots x_d)\ dx_1 = \frac{ac_{14}^2}{4} \cdot h^{2\beta}$$

Analogously

$$\frac{ac_{14}^2}{4} \cdot h^{2\beta} \leq \bar{I}_h \leq \frac{ac_{15}^2}{4} \cdot h^{2\beta}$$

- If $\beta = 1$, then D, $\tilde{f}$, f_+ and f_- are Lipschitz continuous and
$$\bar{I}_h \geq c_{16} h^2.$$
- If $\beta = 1 + \epsilon/2$ for $\epsilon > 0$, then D, $\tilde{f}$, f_+ and f_- are continuously differentiable and
$$\bar{I}_h \geq c_{17} h^{2+\epsilon}.$$
- If $\beta = 2 + \epsilon/2$ for $\epsilon > 0$, then D, $\tilde{f}$, f_+ and f_- are two times continuously differentiable and
$$\bar{I}_h \geq c_{18} h^{4+\epsilon}.$$

5 Estimation error

Next we consider the estimation error. Introduce the notations

$$N_{x,h} = \frac{\mu(S_{x,h}) \bar{D}_h(x)^2}{1 - \mu(S_{x,h}) \bar{D}_h(x)^2}$$

and

$$R_{x,h} = \frac{c_{19}}{\sqrt{\mu(S_{x,h})(1 - \mu(S_{x,h})\bar{D}_h(x)^2)^3}}$$

with a universal constant $c_{19} > 0$. Put

$$\bar{J}_{n,h} = \int_{\{sign\, \bar{D}_h(x) = sign\, D(x)\}} |D(x)| \Phi\left(-\sqrt{n \cdot N_{x,h}}\right) \mu(dx),$$

where Φ stands for the standard Gaussian distribution function.

Proposition 2. *We have that*

$$|J_{n,h} - \bar{J}_{n,h}| \le \int_{\{sign\, \bar{D}_h(x) = sign\, D(x)\}} \frac{R_{x,h} \cdot |D(x)|}{\sqrt{n} + n^2 N_{x,h}^{3/2}} \mu(dx).$$

Put $\varepsilon > 0$. If the density of X exists then, for h small enough,

$$\begin{aligned}
&\int_{\{sign\, \bar{D}_h(x) = sign\, D(x)\}} |D(x)| \Phi\left(-\sqrt{(1+\varepsilon) n \cdot \mu(S_{x,h})} |\bar{D}_h(x)|\right) \mu(dx) \\
&\le \bar{J}_{n,h} \\
&= \int_{\{sign\, \bar{D}_h(x) = sign\, D(x)\}} |D(x)| \Phi\left(-\sqrt{n \cdot \mu(S_{x,h})} |\bar{D}_h(x)|\right) \mu(dx),
\end{aligned}$$

and

$$\begin{aligned}
&\int_{\{sign\, \bar{D}_h(x) = sign\, D(x)\}} \frac{R_{x,h} \cdot |D(x)|}{\sqrt{n} + n^2 N_{x,h}^{3/2}} \mu(dx) \\
&\le \int_{\{sign\, \bar{D}_h(x) = sign\, D(x)\}} \frac{2c_{19} \cdot |D(x)|}{\sqrt{n\mu(S_{x,h})}(1 + (\sqrt{n \cdot \mu(S_{x,h})} |\bar{D}_h(x)|)^3)} \mu(dx)
\end{aligned}$$

with a universal constant $c_{19} > 0$.

Proof. First we show the following: For fixed x and h, under $0 < \bar{D}_h(x)$ we have that

$$|\mathbb{P}\{D_{n,h}(x) \le 0\} - \Phi\left(-\sqrt{n \cdot N_{x,h}}\right)| \le \frac{R_{x,h}}{\sqrt{n} + n^2 N_{x,h}^{3/2}},$$

which implies the first half of the proposition. (The case $\bar{D}_h(x) \le 0$ and $D_{n,h}(x) > 0$ is completely analogous.) Introduce the notation

$$Z_i = -(2Y_i - 1)\mathbb{I}_{\{X_i \in S_{x,h}\}}.$$

Then

$$\mathbb{P}\{D_{n,h}(x) \le 0\} = \mathbb{P}\left\{\sum_{i=1}^n Z_i \ge 0\right\} = \mathbb{P}\left\{\frac{\sum_{i=1}^n (Z_i - \mathbb{E}\{Z_i\})}{\sqrt{n \mathbb{V}ar(Z_1)}} \ge -\frac{\sqrt{n}\mathbb{E}\{Z_1\}}{\sqrt{\mathbb{V}ar(Z_1)}}\right\}.$$

Because of

$$\mathbb{V}ar(Z_1) = \mathbb{E}\{|Z_1|^2\} - (\mathbb{E}\{Z_1\})^2 = \mu(S_{x,h}) - \mu(S_{x,h})^2 \bar{D}_h(x)^2$$

and by $0 < \bar{D}_h(x)$ we have that

$$\frac{\mathbb{E}\{Z_1\}}{\sqrt{\mathbb{V}ar(Z_1)}} = -\frac{\sqrt{\mu(S_{x,h})}\bar{D}_h(x)}{\sqrt{1-\mu(S_{x,h})\bar{D}_h(x)^2}} = -\sqrt{N_{x,h}}$$

Therefore the central limit theorem for the probability $\mathbb{P}\{D_{n,h}(x) \leq 0\}$ implies that

$$\mathbb{P}\{D_{n,h}(x) \leq 0\} = \mathbb{P}\left\{-\frac{\sum_{i=1}^n (Z_i - \mathbb{E}\{Z_i\})}{\sqrt{n\mathbb{V}ar(Z_1)}} \leq -\sqrt{nN_{x,h}}\right\} \approx \Phi\left(-\sqrt{nN_{x,h}}\right).$$

Notice that it is only an approximation. In order to make bounds out of the normal approximation, we refer to Berry-Esseen type central limit theorem (see Theorem 14 in Petrov [12]). Thus,

$$\left|\mathbb{P}\{D_{n,h}(x) \leq 0\} - \Phi\left(-\sqrt{nN_{x,h}}\right)\right| \leq \frac{c_{19}\frac{\mathbb{E}\{|Z_1|^3\}}{\mathbb{V}ar(Z_1)^{3/2}}}{\sqrt{n}\left(1+\left(\sqrt{nN_{x,h}}\right)^3\right)},$$

with the universal constant $30.84 \geq c_{19} > 0$ (cf. Michel [11]). Because of $\mathbb{E}\{|Z_1|^3\} = \mu(S_{x,h})$ we get that

$$c_{19}\frac{\mathbb{E}\{|Z_1|^3\}}{\mathbb{V}ar(Z_1)^{3/2}} = \frac{c_{19}}{\mu(S_{x,h})^{1/2}\left(1-\mu(S_{x,h})\bar{D}_h(x)^2\right)^{3/2}} = R_{x,h},$$

hence

$$\left|\mathbb{P}\{D_{n,h}(x) \leq 0\} - \Phi\left(-\sqrt{nN_{x,h}}\right)\right| \leq \frac{R_{x,h}}{\sqrt{n}(1+(n \cdot N_{x,h})^{3/2})}.$$

Concerning the second half of the proposition notice that if the density of X exists then

$$R_{x,h} \leq \frac{2c_{19}}{\sqrt{\mu(S_{x,h})}},$$

and, for any $\varepsilon > 0$,

$$\mu(S_{x,h})\bar{D}_h(x)^2 \leq N_{x,h} \leq (1+\varepsilon)\mu(S_{x,h})\bar{D}_h(x)^2$$

if h is small enough. ∎

Next we show that the upper bound of the error term in the second half in Proposition 2 is of order $o\left(\frac{1}{nh_n^d}\right)$.

Lemma 4. *Assume that*

$$\lim_{n\to\infty} h_n = 0 \quad \textit{and} \quad \lim_{n\to\infty} nh_n^d = \infty, \tag{12}$$

and that there is a $\tilde{c} > 0$ *such that* $sign\,\bar{D}_h(x) = sign\,D(x)$ *implies* $|\bar{D}_h(x)| \geq \tilde{c}|D(x)|$. *If* X *has a density* f *with bounded support then*

$$\begin{aligned} A_n &:= \int_{\{sign\,\bar{D}_{h_n}(x)=sign\,D(x)\}} \frac{|D(x)|}{\sqrt{n\mu(S_{x,h_n})}(1+(\sqrt{n\mu(S_{x,h_n})}|\bar{D}_{h_n}(x)|)^3)}\mu(dx) \\ &= o\left(\frac{1}{nh_n^d}\right). \end{aligned} \tag{13}$$

Proof. Let B be the bounded support of f. Then $f(x) > 0$ on B and $f(x) = 0$ on B^c. Introduce the notation

$$f_n(x) = \frac{\mu(S_{x,h_n})}{\lambda(S_{x,h_n})},$$

where λ stands for the Lebesgue measure. Under the conditions of the lemma we have that

$$\begin{aligned} A_n &\leq \int \frac{|D(x)|}{\sqrt{n\mu(S_{x,h_n})}(1+(\sqrt{n\mu(S_{x,h_n})}\tilde{c}|D(x)|)^3)}\mu(dx) \\ &= \int \frac{|D(x)|}{\sqrt{n\lambda(S_{x,h_n})f_n(x)}(1+(\sqrt{n\lambda(S_{x,h_n})f_n(x)}\tilde{c}|D(x)|)^3)}\mu(dx) \\ &= \int \frac{1}{f_n(x)} \frac{1}{\tilde{c}n\lambda(S_{x,h_n})} \frac{\sqrt{n\lambda(S_{x,h_n})f_n(x)}\tilde{c}|D(x)|}{1+(\sqrt{n\lambda(S_{x,h_n})f_n(x)}\tilde{c}|D(x)|)^3}\mu(dx) \\ &= \frac{1}{\tilde{c}nh_n^d V_d} \int_B \frac{f(x)}{f_n(x)} r_n(x)dx \end{aligned}$$

with

$$r_n(x) = \frac{\sqrt{n\lambda(S_{x,h_n})f_n(x)}\tilde{c}|D(x)|}{1+(\sqrt{n\lambda(S_{x,h_n})f_n(x)}\tilde{c}|D(x)|)^3}.$$

Thus, we have to show that

$$\int_B \frac{f(x)}{f_n(x)} r_n(x)dx \to 0.$$

By the Lebesgue density theorem $f_n(x) \to f(x)$ and therefore $f_n(x)/f(x) \to 1$ λ - a.e. on B, and so $r_n(x) \to 0$ - a.e. on B. Moreover, this convergence is dominated:

$$r_n(x) \leq \max_{0\leq u} \frac{u}{1+u^3} =: r_{max}.$$

Thus,

$$\int_B r_n(x)dx \to 0.$$

Apply the decomposition

$$\int_B \frac{f(x)}{f_n(x)} r_n(x)dx \leq \int_B \left| \frac{f(x)}{f_n(x)} - 1 \right| r_n(x)dx + \int_B r_n(x)dx$$
$$\leq r_{max} \int_B \left| \frac{f(x)}{f_n(x)} - 1 \right| dx + o(1).$$

In order to prove

$$\int_B \left| \frac{f(x)}{f_n(x)} - 1 \right| dx \to 0 \tag{14}$$

we refer to the Riesz-Vitali-Scheffé theorem, according to which (14) is satisfied if

$$\frac{f}{f_n} \geq 0,$$

$$\frac{f}{f_n} \to 1 \quad \lambda \text{ - a.e. on } B,$$

and

$$\int_B \frac{f(x)}{f_n(x)} dx \to \int_B 1\, dx = \lambda(B). \tag{15}$$

Thus, it remains to show (15). By the generalized Lebesgue density theorem (cf. Lemma 24.5 in Györfi et al. [5]), for each μ-integrable function m

$$\int_B \left| \frac{\int_{S_{x,h_n}} m(z)\mu(dz)}{\mu(S_{x,h_n})} - m(x) \right| \mu(dx) \to 0.$$

Therefore

$$\int_B \frac{\int_{S_{x,h_n}} m(z)\mu(dz)}{\mu(S_{x,h_n})} \mu(dx) \to \int_B m(x)\mu(dx).$$

Choose

$$m(x) = \frac{1}{f(x)},\ x \in B.$$

Then

$$\int_B \frac{f(x)}{f_n(x)} dx = \int_B \frac{\int_{S_{x,h_n}} m(z)\mu(dz)}{\mu(S_{x,h_n})} \mu(dx) \to \int_B m(x)\mu(dx) = \int_B 1\, dx = \lambda(B),$$

and the lemma is proved. ∎

As we already mentioned, using Hoeffding and Bernstein inequalities Kohler and Krzyżak [7] proved that under the condition (8) we have

$$J_{n,h_n} \leq \frac{c_6}{nh_n^d} \tag{16}$$

with $c_6 < \infty$.

We believe that applying Proposition 2 the condition (8) can be weakened such that the following conjecture holds: If X is bounded and it has a density then we have (16).

Because of Lemma 4, this conjecture means that

$$\int_{\{sign\, \bar{D}_h(x)=sign\, D(x)\}} |D(x)| \Phi\left(-\sqrt{n\mu(S_{x,h_n})}|\bar{D}_{h_n}(x)|\right)\mu(dx) \le \frac{c_6}{nh_n^d}. \quad (17)$$

Concerning a possible way to prove (17) we may apply the covering argument of (5.1) in Györfi et al. [5], which says that for bounded X,

$$\int \frac{1}{\mu(S_{x,h_n})}\mu(dx) \le \frac{c_{20}}{h_n^d}.$$

The bounded support of X can be covered by spheres $S_{x_j,h_n/2}, j = 1,\dots,M_n$ such that $M_n \le c_{21}/h_n^d$. Let $S'_{x,h_n} = S_{x,h_n} \cap \{y :\ sign\, \bar{D}_{h_n}(y) = sign\, D(y)\}$. If $x \in S'_{x_j,h_n/2}$ then $S'_{x_j,h_n/2} \subseteq S'_{x,h_n}$. For (17),

$$\begin{aligned}
&\int_{\{sign\, \bar{D}_h(x)=sign\, D(x)\}} |D(x)| \Phi\left(-\sqrt{n\mu(S_{x,h_n})}|\bar{D}_{h_n}(x)|\right)\mu(dx)\\
&\le \sum_{j=1}^{M_n} \int_{S'_{x_j,h_n/2}} |D(x)| \Phi\left(-\sqrt{n\mu(S_{x,h_n})}|\bar{D}_{h_n}(x)|\right)\mu(dx)\\
&\le \sum_{j=1}^{M_n} \int_{S'_{x_j,h_n/2}} |D(x)| \Phi\left(-\sqrt{n\mu(S_{x_j,h_n/2})}|\bar{D}_{h_n}(x)|\right)\mu(dx)\\
&\le \frac{1}{n}\sum_{j=1}^{M_n} \int_{S'_{x_j,h_n/2}} n|D(x)| e^{-n\mu(S_{x_j,h_n/2})|\bar{D}_{h_n}(x)|^2/2}\mu(dx)\\
&\le \frac{1}{n}\sum_{j=1}^{M_n} \int_{S'_{x_j,h_n/2}} n|D(x)| e^{-n\mu(S_{x_j,h_n/2})\tilde{c}^2|D(x)|^2/2}\mu(dx),
\end{aligned}$$

where the last inequality follows by the assumptions of Lemma 4. If

$$\sup_j \int_{S'_{x_j,h_n/2}} n|D(x)| e^{-n\mu(S_{x_j,h_n/2})\tilde{c}^2|D(x)|^2/2}\mu(dx) < \infty$$

then

$$\int_{\{sign\, \bar{D}_h(x)=sign\, D(x)\}} |D(x)| \Phi\left(-\sqrt{n\mu(S_{x,h_n})}|\bar{D}_{h_n}(x)|\right)\mu(dx) \le c_{22}\frac{M_n}{n} \le \frac{c_6}{nh_n^d}.$$

Example 2. Notice that the upper bound in (16) is tight. As in the Example 1, if μ is the uniform distribution and

$$D(x) = x_1,$$

then $sign\,\bar{D}_{h_n}(x) = sign\,D(x)$ and $|\bar{D}_{h_n}(x)| = |D(x)|$ for $h_n < 1/2$ and $|x_1| < 1/2$. Thus

$$\begin{aligned}
&J_{n,h_n}\\
&\geq \int_{\{sign\,\bar{D}_{h_n}(x)=sign\,D(x)\}} |D(x)|\Phi\left(-\sqrt{(1+\varepsilon)n\mu(S_{x,h_n})}|\bar{D}_{h_n}(x)|\right)\mu(dx)\\
&\geq \int_0^{1/2} z\Phi\left(-\sqrt{(1+\varepsilon)V_d n h_n^d 2^{-d}}z\right)dz\\
&= \frac{1}{(1+\varepsilon)V_d n h_n^d 2^{-d}} \int_0^{\sqrt{(1+\varepsilon)V_d n h_n^d 2^{-d}}/2} u\Phi(-u)\,du\\
&= \frac{1}{(1+\varepsilon)V_d n h_n^d 2^{-d}} \left(\int_0^\infty u\Phi(-u)\,du + o(1)\right)\\
&\geq \frac{c_{23}}{nh_n^d},
\end{aligned}$$

with $0 < c_{23}$. Hall and Kang [6] proved that if X has a density, bounded from above and from below then

$$J_{n,h_n} = o\left(\frac{1}{nh_n^d}\right),$$

which contradicts the lower bound $\frac{c_{23}}{nh_n^d}$.

In general, we conjecture the following: If X has a density, which is bounded by f_{max}, then

$$\frac{c_{24}}{nh_n^d} \leq J_{n,h_n}$$

with $0 < c_{24}$. This conjecture is supported by the fact that

$$\mu(S_{x,h_n}) \leq f_{max} V_d h_n^d.$$

Therefore

$$\begin{aligned}
&\int_{\{sign\,\bar{D}_{h_n}(x)=sign\,D(x)\}} |D(x)|\Phi\left(-\sqrt{(1+\varepsilon)n\mu(S_{x,h_n})}|\bar{D}_{h_n}(x)|\right)\mu(dx)\\
&\geq \int_{\{sign\,\bar{D}_{h_n}(x)=sign\,D(x)\}} |D(x)|\Phi\left(-\sqrt{(1+\varepsilon)n f_{max} V_d h_n^d}|\bar{D}_{h_n}(x)|\right)\mu(dx).
\end{aligned}$$

6 Conclusion

We presented tight upper bounds for the rate of convergence of the error probability of kernel classification rule. Decomposing the excess error probability into the sum of approximation and estimation error, we derived approximate formulas both for approximation error and estimation error.

Under weak margin condition with α and Lipschitz condition on the regression function D, Kohler and Krzyżak [7] showed that the approximation error $I_{n,h}$ is upper bounded by $c_5 h^{\alpha+1}$. If, in addition, the conditional densities are twice continuously differentiable, then we proved that $I_{n,h} \leq c_{12} h^{2(\alpha+1)}$. Furthermore, we present an example, according to which these upper bounds are tight. Under the assumption that the Minkowski content of the decision boundary is finite and the lower Lipschitz inequality holds, the weak margin condition holds with $\alpha = 1$. Hence we get the upper bound $I_{n,h} \leq c_{11} h^4$ as in Hall and Kang [6] as a special case.

If X has a lower bounded density, then Kohler and Krzyżak proved the upper bound on the estimation error: $J_{n,h} \leq c_6/(nh^d)$. We show that this upper bound is tight, too.

References

1. J-Y. Audibert and A. B. Tsybakov Fast learning rates for plug-in classifiers, *Annals of Statistics*, 35:608–633, 2007.
2. L. Devroye On the almost everywhere convergence of nonparametric regression function estimates, *Annals of Statistics*, 9:1310–1319, 1981.
3. L. Devroye, L. Györfi and G. Lugosi, *A Probabilistic Theory of Pattern Recognition*, Springer–Verlag, New York, 1996.
4. L. Devroye and T. J. Wagner, Distribution-free consistency results in nonparametric discrimination and regression function estimation, *Annals of Statistics*, 8:231–239, 1980.
5. L. Györfi, M. Kohler, A. Krzyżak and H. Walk, *A Distribution-Free Theory of Nonparametric Regression*, Springer–Verlag, New York, 2002.
6. P. Hall and K. Kang, Bandwidth choice for nonparametric classification, *Annals of Statistics*, 33:284–306, 2005.
7. M. Kohler and A. Krzyżak, On the rate of convergence of local averaging plug-in classification rules under a margin condition. *IEEE Transactions on Information Theory* 53:1735–1742, 2007.
8. A. Krzyżak, The rates of convergence of kernel regression estimates and classification rules, *IEEE Transactions on Information Theory*, IT-32:668–679, 1986.
9. A. Krzyżak and M. Pawlak, Distribution-free consistency of a nonparametric kernel regression estimate and classification, *IEEE Transactions on Information Theory*, IT-30:78–81, 1984.
10. E. Mammen and A. B. Tsybakov, Smooth discrimination analysis, *Annals of Statistics*, 27:1808–1829, 1999.
11. R. Michel, On the constant in the non-uniform version of the Berry-Esseen theorem, *Z. Wahrsch. Verw. Gebiete*, 55:109–117, 1981.
12. V. V. Petrov, *Sums of Independent Random Variables*, Springer-Verlag, Berlin, 1975.
13. A. B. Tsybakov, Optimal aggregation of classifiers in statistical learning, *Annals of Statistics*, 32:135–166, 2004.

Maik Döring
Institut für Angewandte Mathematik und Statistik, Universität Hohenheim, Schloss Hohenheim, 70599 Stuttgart, Germany
E-Mail: maik.doering@uni-hohenheim.de

László Györfi
Department of Computer Science and Information Theory Budapest University of Technology and Economics Magyar Tudósok k"orútja 2., H-1117 Budapest, Hungary
E-Mail: gyorfi@cs.bme.hu

Harro Walk
Fachbereich Mathematik, Universität Stuttgart, Pfaffenwaldring 57, 70569 Stuttgart, Germany
E-Mail: walk@mathematik.uni-stuttgart.de
WWW: http://www.isa.uni-stuttgart.de/LstStoch/Walk/

Erschienene Preprints ab Nummer 2012-001

Komplette Liste: `http://www.mathematik.uni-stuttgart.de/preprints`

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

2012-013 *Diaz Ramos, J.C.; Dominguez Vázquez, M.; Kollross, A.:* Polar actions on complex hyperbolic spaces

2012-012 *Moroianu; A.; Semmelmann, U.:* Weakly complex homogeneous spaces

2012-011 *Moroianu; A.; Semmelmann, U.:* Invariant four-forms and symmetric pairs

2012-010 *Hamilton, M.J.D.:* The closure of the symplectic cone of elliptic surfaces

2012-009 *Hamilton, M.J.D.:* Iterated fibre sums of algebraic Lefschetz fibrations

2012-008 *Hamilton, M.J.D.:* The minimal genus problem for elliptic surfaces

2012-007 *Ferrario, P.:* Partitioning estimation of local variance based on nearest neighbors under censoring

2012-006 *Stroppel, M.:* Buttons, Holes and Loops of String: Lacing the Doily

2012-005 *Hantsch, F.:* Existence of Minimizers in Restricted Hartree-Fock Theory

2012-004 *Grundhöfer, T.; Stroppel, M.; Van Maldeghem, H.:* Unitals admitting all translations

2012-003 *Hamilton, M.J.D.:* Representing homology classes by symplectic surfaces

2012-002 *Hamilton, M.J.D.:* On certain exotic 4-manifolds of Akhmedov and Park

2012-001 *Jentsch, T.:* Parallel submanifolds of the real 2-Grassmannian