Vorlesung *Nichtparametrische Statistik*
Wintersemester 2010/11
Humboldt-Universität zu Berlin

Prof. Dr. Markus Reiß
Dipl. Math. Johanna Kappus

1. Übungsblatt

1. Es seien $X_1, \cdots, X_n$ unabhängig identisch verteilte Zufallsvariablen mit Werten in einem messbaren Raum $(\mathcal{X}, \mathcal{A})$. $\mathcal{F}$ sei eine Klasse von reellwertigen, messbaren Funktionen auf $(\mathcal{X}, \mathcal{A})$, so dass für alle $f \in \mathcal{F}$ gilt: $\mathbb{E}\,|f(X_1)| < \infty$. Für Funktionen $l, u \in \mathrm{L}^1(\mathcal{X})$ bezeichnet man die Menge von Funktionen

$$[l, u] := \{f : \mathcal{X} \to \mathbb{R} : l(x) \le f(x) \le u(x) \text{ für alle } x \in \mathcal{X}\}$$

als zugehörige *Klammer*.
Insbesondere heißt $[l, u]$ eine ε-*Klammer*, falls $\mathbb{E}\,|u(X_1) - l(X_1)| \le \varepsilon$ gilt.

Zeige: Falls für jedes $\varepsilon > 0$ endlich viele ε-Klammern existieren, die $\mathcal{F}$ überdecken, so gilt die *Glivenko-Cantelli-Eigenschaft*

$$\sup_{f \in F} \left| \frac{1}{n} \sum_{k=1}^{n} f(X_k) - \mathbb{E}[f(X_1)] \right| \xrightarrow{\text{f.s.}} 0 \quad .$$

2. Benutze Aufgabe 1, um den *Satz von Glivenko-Cantelli* zu beweisen:
Die fast sichere Konvergenz der empirischen gegen die wahre Verteilungsfunktion gilt gleichmäßig in $\mathbb{R}$, d.h. für $n \to \infty$ gilt

$$\sup_{x \in \mathbb{R}} |\hat{F}_n(x) - F(x)| \overset{\text{f.s.}}{\to} 0.$$

3.* Es sei Θ eine kompakte Teilmenge des $\mathbb{R}^d$. $Q : \mathbb{R}^k \times \Theta \to \mathbb{R}$ sei für jedes x stetig in ϑ und für jedes ϑ messbar in x. $X_1, \cdots, X_n$ seien u.i.v. im $\mathbb{R}^k$.

Zeige: Falls

$$\mathbb{E}\left[\sup_{\vartheta \in \Theta} |Q(X, \vartheta)|\right] < \infty$$

gilt, so folgt

$$\sup_{\vartheta \in \Theta} \left| \frac{1}{n} \sum_{j=1}^{n} \left(Q(X_j, \vartheta) - \mathbb{E}\left[Q(X_1, \vartheta)\right]\right) \right| \xrightarrow{\text{f.s.}} 0.$$

4. Für einen Kern $K \in L^1(\mathbb{R})$ und eine beschränkte Funktion f, die in x stetig ist, gilt Konvergenz $(K_h * f - f)(x) \to 0$ für $h \to 0$.
Man zeige anhand eines Beispiels, dass auf die Stetigkeitsforderung nicht verzichtet werden kann.

5. Zeige, dass für jedes $m \in \mathbb{N}$ genau ein Polynom P_m vom Grad $\leq m$ existiert, so dass
$$K(x) := \begin{cases} P_m(x), & \text{für } x \in [-1,1] \\ 0, & \text{sonst} \end{cases}$$
ein quadratintegrierbarer Kern m-ter Ordnung ist.

6. Es sei K der sinc-Kern. Zeige für $f \in H^s(\mathbb{R})$ die Abschätzung für den MISE
$$R_{\mathbb{R}}(\hat{f}_{n,h}, f) \leq C\|f\|_{H^s}^2 h^{2s} + n^{-1}h^{-1}\|K\|_{L^2}^2.$$
mit einer Konstanten C.

7.** Recherchiere folgende Ansätze zur nichtparametrischen Dichteschätzung: *Histogrammschätzer* und *kNN-Schätzer* (engl.: k-th nearest neighbour).
Implementiere diese Verfahren, sowie den Kerndichteschätzer für $n = 100, 1000, 10000$ Beobachtungen mit $\mathcal{N}(0,1)$-Verteilung und $\mathcal{U}([0,1])$ Verteilung für den Rechteckskern und den Dreieckskern, verschiedene Wahl der Bandweite, verschieden feine Partitionen und unterschiedliche Wahl von k.
Diskutiere die Ergebnisse.

Abgabe der Aufgaben in der Vorlesung am Freitag, dem 03.12.2010

Die mit * gekennzeichneten Aufgaben sind freiwillige Zusatzaufgaben.
Die mit ** gekennzeichneten praktischen Aufgaben sind innerhalb von 14 Tagen zu bearbeiten. Für die Scheinvergabe müssen insgesamt mindestens 4 Punkte aus den praktischen Aufgaben erzielt werden.

Vorlesung *Nichtparametrische Statistik*
Wintersemester 2010/11
Humboldt-Universität zu Berlin

Prof. Dr. Markus Reiß
Dipl. Math. Johanna Kappus

2. Übungsblatt

1. Gib einen sinnvollen Kernschätzer $\hat{f'}$ für die erste Ableitung der Dichtefunktion $f : \mathbb{R} \to \mathbb{R}^+$ an und analysiere unter geeigneten Voraussetzungen den Bias- und den Varianzterm für $\hat{f}_n'(x_0)$.

2. In welchen der Sobolev-Räume $H^s, s \geq 0$, liegen die folgenden Funktionen?

   a) Die Gleichverteilungsdichte $f(x) = 1([0,1]^d)(x),\ x \in \mathbb{R}^d$

   b) Die $\Gamma(\alpha,\beta)$-Dichte $f(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta},\ \alpha, \beta > 0,\ x \in \mathbb{R}^+$.

   Gib gegebenenfalls unter den üblichen Voraussetzungen eine obere Schranke für den MISE des Kerndichteschätzers an.

3. Zeige für den Nadaraya-Watson-Schätzer die Darstellung

$$\hat{f}_{n,h}^{NW}(x) = \arg\min_{y \in \mathbb{R}} \left(\sum_{i=1}^n (Y_i - y)^2 K_h(X_i - x) \right).$$

4. Für Zufallsvariablen $X_1, \cdots, X_n$ mögen Konstanten $a > 0$ und $C < \infty$ existieren, so dass $\max_{1 \leq k \leq n} \mathbb{E}\left[\exp(aX_k^2)\right] \leq C$ gilt.
   Zeige:

$$\mathbb{E}\left[\max_{1 \leq k \leq n} X_k^2\right] \leq \frac{1}{a} \log(Cn).$$

   Hinweis: Benutze die Jensensche Ungleichung.

5. Betrachte das nichtparametrische Regressionsmodell auf dem Einheitsintervall mit äquidistantem Design, Regressionsfunktion $f \in \mathscr{H}_{[0,1]}(\alpha; R; L)$ und standardnormalverteilten Fehlern. Es sei $\hat{f}_{n,h}$ der lokal-polynomiale Schätzer vom Grad $\lfloor \alpha \rfloor$. Der Kern K sei Lipschitz-stetig mit Träger in $[0,1]$. Weiter seien die Eigenwerte von $B(x)$ nach unten beschränkt.
   Zeige für das gleichmäßige quadratische Risiko die Abschätzung

$$\mathbb{E}\,\|\hat{f}_{n,h} - f\|_\infty^2 \leq C_1 h^{2\alpha} + C_2 \frac{\log n}{nh}$$

   mit geeigneten Konstanten C_1 und C_2.
   Anleitung: Für die Analyse des Varianzterms betrachtet man die Abweichung an den Gitterpunkten und benutzt ein Stetigkeitsargument.

Abgabe der Aufgaben in der Vorlesung am Freitag, dem 10.12.2010

Vorlesung *Nichtparametrische Statistik*
Wintersemester 2010/11
Humboldt-Universität zu Berlin

Prof. Dr. Markus Reiß
Dipl. Math. Johanna Kappus

3. Übungsblatt

1. Betrachte das Dichteschätzproblem auf einer offenen Teilmenge $D \subseteq \mathbb{R}^d$ und setze $f|_D \in L^2(D)$ voraus. Es sei $S_M \subseteq L^2(D) \cap C(D)$ ein M-dimensionaler linearer Unterraum mit Orthonormalbasis $\varphi_1, \cdots, \varphi_M$ und $\hat{f}_{n,M}$ der zugehörige Projektionsschätzer.
   Zeige: Es gilt $\hat{f}_{n,M} \in S_M$ und für alle $\varphi \in S_M$ ist $\int \varphi(x)\hat{f}_{n,M}(x)\,\mathrm{d}x = \int \varphi(x)\hat{\mu}_n(\,\mathrm{d}x)$.
   Schließe daraus, dass der Projektionsschätzer eindeutig und unabhängig von der Wahl einer Orthonormalbasis ist.

2. Zeige (mit den Bezeichnungen aus Aufgabe 1) folgende Bias-Varianz-Zerlegung für den Projektionsschätzer: $\hat{f}_{n,M}$:

$$R_D(\hat{f}_{n,M}, f) = \|f - P_{S_M} f\|^2_{L^2(D)} + n^{-1}\left(\int_D \sum_{m=1}^{M} \varphi_m(x)^2 f(x)\,\mathrm{d}x - \|P_{S_M} f\|^2_{L^2(D)}\right),$$

   wobei $P_{S_M} : L^2(D) \to S_M$ die L^2-Orthogonalprojektion auf S_M bezeichnet und f kurz für die Einschränkung von f auf D steht.

3. Die Fourierbasis des komplexwertigen Funktionenraums $L^2([0,1])$ ist definiert durch $\varphi_m(x) := e^{2\pi i m x}, m \in \mathbb{Z}$. Man definiert den periodischen L^2-Sobolevraum der Ordnung $s > 0$ auf $[0,1]$ als

$$H^s_{per} := \left\{ f \in L^2([0,1]) : \sum_{m\in\mathbb{Z}} (1+m^2)^s |\langle f, \varphi_k\rangle_{L^2([0,1])}|^2 < \infty \right\}.$$

   Für ungerade $M \in \mathbb{N}$ ist

$$S_M := \operatorname{span}\{\varphi_m, m = -(M-1)/2, \cdots, (M-1)/2\}.$$

   Gib für Dichten $f \in H^s_{per}([0,1])$ eine Abschätzung für den Bias-Term $\|f - P_{S_M}\|^2_{L^2}$ an. Schließe daraus, dass die $(S_M)_{M\geq 1}$ Approximationsräume beliebiger Ordnung bilden.

4. Es seien $\mathbb{P}$ und $\mathbb{Q}$ Wahrscheinlichkeitsmaße auf $(\mathcal{X}, \mathcal{A})$ mit Dichten p und q bezüglich eines dominierenden Maßes μ. Zeige folgende Aussagen für den Totalvariationsabstand:

   a) $0 \leq \|P - Q\|_{TV} \leq 1$.
   b) $\|P - Q\|_{TV} = \frac{1}{2}\int_{\mathcal{X}} |p - q|(x)\mu(\,\mathrm{d}x) = \int_{\mathcal{X}} \max(p,q)(x)\mu(\,\mathrm{d}x) - 1.$

5. Zeige folgende Ungleichung für den Totalvariationsabstand und die Kullback-Leibler-Divergenz:
$$\|P-Q\|_{TV} \leq \sqrt{\mathrm{KL}(\mathbb{P} \,|\, \mathbb{Q})/2}$$
Anleitung: Betrachte die Funktion
$$h(z) := z\log(z) - z + 1,\ \ z \geq 0,$$
wobei man in der Null stetig ergänzt.
Zeige
$$\forall z \geq 0 : (\frac{4}{3} + \frac{2}{3}z)h(z) \geq (z-1)^2.$$
Benutze die Cauchy-Schwarz-Ungleichung.

6.** Die Regressionsfunktion f sei gegeben durch
$$f(x) := \begin{cases} 0, & \text{falls } x \in [0, \frac{1}{3}] \cup [\frac{2}{3}, 1] \\ 5 - 30|x - \frac{1}{2}|, & \text{falls } x \in (\frac{1}{3}, \frac{2}{3}). \end{cases}.$$
Erzeuge Beobachtungen
$$Y_i = f(X_i) + \varepsilon_i,\ \ i = 1, \cdots, n$$
mit zufälligen Designpunkten $X_i \sim U([0,1])$ und gaußverteilten Fehlern für $n = 10, n = 100, n = 1000$.
Implementiere den Nadaraya-Watson-Schätzer und den $LP(1)$-Schätzer mit Rechteckskern sowie den Projektionsschätzer mit Fourierbasis.
Veranschauliche den Zusammenhang zwischen der Bandweite h bzw. der Dimension M und dem $MISE$ graphisch und zeige für die Orakelwahl von h bzw. M ein Bild des Schätzers und der Regressionsfunktion.

Abgabe der Aufgaben in der Vorlesung am Freitag, dem 17.12.2010

Vorlesung *Nichtparametrische Statistik*
Wintersemester 2010/11
Humboldt-Universität zu Berlin

Prof. Dr. Markus Reiß
Dipl. Math. Johanna Kappus

4. Übungsblatt

1. In der Vorlesung wurde für Dichten f aus der Hölderklasse $\mathscr{H}_D(\alpha, L, R)$, $D \subseteq \mathbb{R}$, die optimale Minimaxrate $n^{-\alpha/(2\alpha+1)}$ für punktweisen Verlust hergeleitet.
   Zeige: Betrachten derselben Alternativen liefert für die untere Schranke des MISE lediglich die suboptimale Rate $n^{-1/2}$.

2. Betrachte das Modell der nichtparametrischen Regression mit Regressionsfunktion $f : [0,1] \to \mathbb{R}$, deterministischem, äquidistantem Design und standardnormalverteilten Fehlern.
   Zeige für $s \in \mathbb{N}_+$ und $L > 0$ die untere Schranke

$$\liminf_{n\to\infty} \inf_{\tilde{f}_n} \sup_{f \in H^s_{per}(L)} n^{\frac{2s}{2s+1}} \, \mathbb{E}_f \left[\|\tilde{f}_n - f\|^2_{L^2([0,1])} \right] > 0.$$

3. Zeige unter denselben Voraussetzungen wie in Aufgabe 2 für $\alpha > 0$ und $L > 0$ die untere Schranke

$$\liminf_{n\to\infty} \inf_{\tilde{f}_n} \sup_{f \in \mathscr{H}_{[0,1]}(\alpha,L,R)} \left(\frac{n}{\log n} \right)^{\frac{2\alpha}{2\alpha+1}} \mathbb{E}_f \left[\|\tilde{f}_n - f\|^2_\infty \right] > 0.$$

   Hinweis: Betrachte die Alternativen

$$f_{0,n} \equiv 0 \text{ und } f_{j,n}(x) = \gamma_0 h_n^\alpha \varphi\left(\frac{x - x_j}{h_n} \right), \; j = 1, \cdots, M_n$$

   mit einer Konstanten γ_0, $h_n = \frac{1}{M_n}$ für ein ganzzahliges $M_n > 1$, $x_j = \frac{j-1/2}{M_n}$ und einer geeigneten regulären Funktion φ mit $\varphi(x) > 0 \Leftrightarrow x \in]-1/2, 1/2[$.

4. Im Modell der nichtparametrischen Regression auf dem Einheitsintervall mit äquidistantem Design und standardnormalverteilten Fehlern sei

$$f_\tau(x) := \sum_{i=1}^n W_{n,i}(x,\tau) Y_i$$

   ein linearer Schätzer von f mit Gewichten, die von x und einem Parameter τ abhängen. Betrachte den empirischen L^2-Verlust.
   Zeige, dass

$$C(\tau) := \frac{1}{n} \sum_{i=1}^n (Y_i - f_\tau(X_i))^2 + \frac{2}{n} \sum_{i=1}^n W_{n,i}(x_i, \tau)$$

bis auf einen Summanden, der nicht von τ abhängt, ein unverzerrter Schätzer für das Risiko $\mathbb{E}\left[\|f_\tau - f\|_n^2\right]$ ist.
Es sei $\{\varphi_j\}_{j\in\mathbb{Z}}$ die Fourierbasis. $\hat{f}_{n,M}$ möge den Projektionsschätzer bezeichnen. Es sei $\tau = M$. Gib in dieser Situation $C(\tau)$ konkret an.

5. Es sei $(\alpha_m)_{m\in\mathbb{N}}$ eine Folge von Gewichten in $[0,1]$. Betrachte im Dichteschätzmodell für eine ONB (φ_m) in $L^2(D)$ den *verallgemeinerten Projektionsschätzer*

$$\hat{f}_{n,\alpha}(x) := \sum_{m=1}^{\infty} \alpha_m \langle \varphi_m, \hat{\mu}_n \rangle \varphi_m(x).$$

   a) Zeige, dass $\hat{f}_{n,\alpha} \in L^2(D)$ fast sicher wohldefiniert ist, falls $\sup\limits_{x\in D} \sum_{m=1}^{\infty} \alpha_m^2 \varphi_m(x)^2 < \infty$ gilt, und gib die Bias-Varianz-Zerlegung an.

   b) Gib unter der der Restriktion $\alpha_m \in \{0,1\}\ \forall m \geq 1$ eine Orakelwahl von (α_m) an, d.h. denjenigen Filter $(\alpha_m(f))$, der für eine bekannte Dichte $f \in L^2(D)$ den MISE minimiert.

Abgabe der Aufgaben in der Vorlesung am Freitag, dem 14.01.2011

Vorlesung *Nichtparametrische Statistik*
Wintersemester 2010/11
Humboldt-Universität zu Berlin

Prof. Dr. Markus Reiß
Dipl. Math. Johanna Kappus

5. Übungsblatt

1. Es sei X binomialverteilt mit Parametern n und p. Leite mit der Hoeffding-Ungleichung eine Abschätzung für $\mathbb{P}\left(|X-np| \geq \kappa\sqrt{np(1-p)}\right)$ her.
   Anleitung: Stelle X als Summe unabhängiger Zufallsvariablen dar und erhalte daraus durch geeignete Normierung ein Martingal.

2. Leite in der Situation von Aufgabe 1 eine Abschätzung mittels Bernstein-Ungleichung her. In welchen Fällen ist dies eine Verbesserung der Abschätzung mit Hoeffding-Ungleichung?

3. In der Vorlesung wurde das CV-Kriterium

$$CV_n(h) := \int \hat{f}_{n,h}(x)^2 \,\mathrm{d}x - 2\hat{I}_n(h)$$

   mit

$$\hat{I}_n(h) := \frac{1}{n}\sum_{i=1}^{n} \hat{f}_{n,h}^{(-i)}(X_i) \text{ und } \hat{f}_{n,h}^{(-i)}(x) := \frac{1}{n-1}\sum_{j\neq i} K_h(x-X_j)$$

   eingeführt. Begründe, warum der Ansatz, statt $\hat{I}_n(h)$ den naiven Schätzer

$$\hat{I}_n(h)^{naiv} = \frac{1}{n}\sum_{i=1}^{n} \hat{f}_{n,h}(X_i)$$

   zu betrachten, nicht zielführend ist und systematisch zum Unterglätten führt.

4. In der Vorlesung ergab sich bei der Analyse der Kreuzvalidierung der Term

$$S_3 := 2\frac{(2\pi)^{-d}}{n}\sum_{k=1}^{n}\int Re\left((\mathcal{F}K(hu)^2 - \mathcal{F}K(hu))\varphi(u)\psi_k(u)\right)\,\mathrm{d}u.$$

   Zeige die obere Schranke

$$\mathbb{P}\left(|S_3| \geq \kappa_3\right) \leq 2\exp\left(-c_0\kappa_3/(h^{d/4}MISE_n(h) + n^{-1/2}MISE_n(h))\right)$$

   mit einer Konstanten $c_0 > 0$.
   Anleitung: Schätze zunächst das Integral mit Hilfe der Cauchy-Schwarz-Ungleichung durch ein Vielfaches des $BIAS$ ab. Zeige die Abschätzung

$$\begin{aligned}&\mathrm{Var}\left(2(2\pi)^{-d}\int Re\left((\mathcal{F}K(hu)^2 - \mathcal{F}K(hu))\varphi(u)\psi_k(u)\right)\,\mathrm{d}u\right)\\ \leq\ & 8BIAS_n(h)^2h^{-d/2}\|K\|_{L^2}\|K\|_{L^1}\|f\|_{L^2}\end{aligned}$$

   und wende dann die Bernstein-Ungleichung an.

5. Betrachte das Modell der nichtparametrischen Regression mit zufälligem Design und

$$\mathbb{E}_f[\varepsilon_i | X_1, \cdots, X_n] = 0 \quad \text{und} \quad \mathbb{E}_f[\varepsilon_i \varepsilon_j | X_1, \cdots, X_n] = \sigma^2 \delta_{i,j}, \; i, j = 1, \cdots, n.$$

Es sei $\hat{f}_{n,\tau}(x) = \sum_{i=1}^{n} Y_i w_i(x, \tau)$ ein linearer Schätzer der Regressionsfunktion f. Definiere das Kreuzvalidierungs-Kriterium

$$CV(\tau) := \frac{1}{n} \sum_{i=1}^{n} (Y_i - \hat{f}_{n,\tau}^{(-i)}(X_i))^2$$

mit dem Leave-one-out-Schätzer

$$\hat{f}_{n,\tau}^{(-i)}(x) := \sum_{j \neq i} Y_j w_j(x, \tau).$$

Es möge

$$\|\hat{f}_{n,\tau}^{(-i)}\|_X^2 := \int \hat{f}_{n,\tau}^{(-i)2}(x) \, \mathbb{P}^X(\,\mathrm{d}x) < \infty, \quad i = 1, \cdots, n$$

gelten.
Zeige: Es gilt

$$\mathbb{E}_f[CV(\tau)] = \mathbb{E}_f\left[\|\hat{f}_{n,\tau}^{(-1)} - f\|_X^2\right] + \sigma^2$$

unter der Voraussetzung, dass die $(X_i, Y_i), i = 1, \cdots, n$ u.i.v. sind und dass $\hat{f}_{n,\tau}^{(-i)}(x)$ für alle i, x dieselbe Verteilung wie $\hat{f}_{n,\tau}^{(-1)}(x)$ besitzt.
Diskutiere die Voraussetzung.

6.** Wende die Kreuzvalidierungs-Methode auf die Kerndichteschätzer aus Aufgabe 1.7 an. Stelle für $n = 100, 1000, 10000$ Beobachtungen den Schätzer mit Orakelwahl der Bandweite sowie mit Bandweitenwahl mittels Kreuzvalidierung graphisch dar.

Abgabe der Aufgaben in der Vorlesung am Freitag, dem 21.01.2010

Vorlesung *Nichtparametrische Statistik*
Wintersemester 2010/11
Humboldt-Universität zu Berlin

Prof. Dr. Markus Reiß
Dipl. Math. Johanna Kappus

6. Übungsblatt

Man betrachtet das Gaußsche Folgenraummodell. Für eine Folge $\vartheta = (\vartheta_j)_{j\geq 1} \in \ell^2(\mathbb{N})$ verfügt man über die Beobachtungen

$$y_j = \vartheta_j + \sigma\varepsilon_j, \quad j = 1, 2, \cdots$$

mit unabhängigen standardnormalverteilten Fehlern ε_i und $0 < \sigma < 1$.
Für eine Folge von Koeffizienten $a_j \geq 0$ definiert man das *Ellipsoid*

$$\Theta = \Theta(a, Q) := \{\vartheta = (\vartheta_j)_{j\geq 1} : \sum_{j=1}^{\infty} a_j^2 \vartheta_j^2 \leq Q\}.$$

Insbesondere sind für das *Sobolev-Ellipsoid* $\Theta(a(\beta), Q)$, $\beta > 0$, die Koeffizienten gegeben durch

$$a_j = \begin{cases} j^\beta, & j \text{ gerade} \\ (j-1)^\beta, & j \text{ ungerade.} \end{cases}$$

Das *lineare Minimax-Risiko* über Θ ist definiert als

$$R^L := \inf_\lambda R(\lambda, \vartheta) = \inf_\lambda \sup_{\vartheta\in\Theta} \mathbb{E}_\vartheta \, \|\hat{\vartheta}(\lambda) - \vartheta\|^2,$$

wobei das Infimum über allen linearen Schätzern $\hat{\vartheta}(\lambda) = (\lambda_1 y_1, \lambda_2 y_2, \cdots)$, $\lambda \in \ell^2(\mathbb{N})$ betrachtet wird.
Ein linearer Schätzer $\hat{\vartheta}(\lambda^*)$ heißt *linear minimax*, falls

$$\sup_{\vartheta\in\Theta} R(\lambda^*, \vartheta) = R^L$$

gilt. Ein linearer Schätzer $\hat{\vartheta}(\lambda^*)$ heißt *linear asymptotisch minimax*, falls

$$\lim_{\sigma\to 0} \frac{\sup_{\vartheta\in\Theta} R(\lambda^*, \vartheta)}{R^L} = 1$$

gilt. (Beachte, dass die Definition des Schätzers von σ abhängt.)

1. Es sei $(a_j)_{j\geq 0}$ eine wachsende Folge nichtnegativer Zahlen mit $a_j \to \infty$. Zeige, dass für $\sigma > 0$ und $Q > 0$ eine eindeutige Lösung $\kappa = \kappa(\sigma)$ der Gleichung

$$\frac{\sigma^2}{\kappa} \sum_{j=1}^{\infty} a_j (1 - \kappa a_j)_+ = Q \tag{0.1}$$

existiert, und dass

$$\kappa = \frac{\sigma^2 \sum_{m=1}^{N} a_m}{Q + \sigma^2 \sum_{m=1}^{N} a_m^2}$$

gilt mit

$$N = \max\left\{j : \sigma^2 \sum_{m=1}^{j} a_m(a_j - a_m) < Q\right\}.$$

2. Es sei $\Theta(a, Q)$ ein Ellipsoid mit $\mathrm{Card}\{j : a_j = 0\} < \infty$. Es möge eine Lösung $\kappa = \kappa(\sigma)$ von (0.1) existieren. Definiere die Gewichte

$$\ell_j := (1 - \kappa a_j)_+,\ j = 1, 2, \cdots, \quad \ell := (\ell_1, \ell_2, \cdots)$$

und setze

$$\mathcal{D}^* := \sigma^2 \sum_{j=1}^{\infty} \ell_j,$$

wobei $\mathcal{D}^* < \infty$ gelten möge. $\hat{\vartheta}(\ell)$ sei der zu ℓ gehörende lineare Schätzer im Gaußschen Folgenraummodell.
Zeige, dass für das lineare Minimax-Risiko die Gleichung

$$\inf_{\lambda} \sup_{\vartheta\in\Theta} R(\lambda, \vartheta) = \sup_{\vartheta\in\Theta} \inf_{\lambda} R(\lambda, \vartheta) = \sup_{\vartheta\in\Theta} R(\ell, \vartheta) = \mathcal{D}^*.$$

gilt.

Anleitung: Um

$$\sup_{\vartheta\in\Theta} R(\ell, \vartheta) \le \mathcal{D}^*$$

zu zeigen, betrachtet man für jedes ϑ die Bias-Varianz-Zerlegung des Risikos und verwendet die Tatsache, dass κ eine Lösung von (0.1) ist.
Um

$$\sup_{\vartheta\in\Theta} \inf_{\lambda} R(\lambda, \vartheta) \ge \mathcal{D}^*$$

nachzuweisen, betrachtet man die Menge V der Folgen $v = (v_1, v_2, \cdots)$ reeller Zahlen, für die v_j beliebig ist für $a_j = 0$ und

$$v_j := \frac{\sigma^2(1 - \kappa a_j)_+}{\kappa a_j} \text{ für } a_j > 0.$$

Man zeigt, dass $V \subseteq \Theta$ gilt und betrachtet das Supremum über der Menge V. Dabei benutzt man wieder die Bias-Varianz-Zerlegung sowie die Definition der v_j.

3. Betrachte nun das Sobolev-Ellipsoid $\Theta(a(\beta), Q)$. Zeige:

   (i) Es gibt eine eindeutige Lösung κ von (0.1), und κ hat die Gestalt

   $$\kappa = \kappa^*(1 + o(1)), \sigma \to 0$$

   für

   $$\kappa^* := \left(\frac{\beta}{(2\beta + 1)(\beta + 1)Q}\right)^{\frac{\beta}{2\beta+1}} \sigma^{\frac{2\beta}{2\beta+1}}.$$

(ii) Es gilt

$$\mathcal{D}^* = \mathcal{C}^* \sigma^{\frac{4\beta}{2\beta+1}}(1+o(1)),\ \sigma \to 0$$

mit

$$\mathcal{C}^* = (Q(2\beta+1))^{\frac{1}{2\beta+1}} \left(\frac{\beta}{\beta+1}\right)^{\frac{2\beta}{2\beta+1}}$$

(iii) Es gilt

$$\max_{j\geq 2} v_j^2 a_j^2 = O\left(\sigma^{\frac{2}{2\beta+1}}\right),\ \sigma \to 0$$

für

$$v_j^2 = \frac{\sigma^2(1-\kappa a_j)_+}{\kappa a_j}.$$

Hinweis: Verwende die Resultate aus Aufgaben 1 und 2.

4. Definiere für den Sobolev-Ellipsoid $\Theta(a(\beta), Q)$ die Gewichte

$$\ell_j^* = (1-\kappa^* a_j)_+$$

mit κ^* wie oben. Zeige für den zugehörigen linearen Schätzer die Asymptotik

$$\sup_{\vartheta \in \Theta(a(\beta),Q)} R(\ell^*, \vartheta) \leq \mathcal{D}^*(1+o(1)), \sigma \to 0.$$

Hinweis: Man verwendet das entsprechende Resultat aus Aufgabe 2 und die Tatsache, dass sich das maximale Risiko von $\hat{\vartheta}(\ell^*)$ asymptotisch wie das maximale Risiko von $\hat{\vartheta}(\ell)$ verhält.

5. (a) Betrachte das statistische Modell

$$x = a + \sigma\varepsilon$$

für eine reelle Zahl a, $\sigma > 0$ und standardnormalverteilten Fehler ε. Zeige, dass der Bayes-Schätzer zur a-priori-Verteilung $\mathcal{N}(0, s^2)$ gegeben ist durch

$$\hat{a}^B := \frac{s^2}{\sigma^2 + s^2} x.$$

(b) Man betrachtet das Modell des Gaußschen weißen Rauschens

$$dY(t) = f(t)\,\mathrm{d}t + \sigma dW(t),\ t \in [0,1],\ 0 < \sigma < 1.$$

Dem Ellipsoid $\Theta(a(\beta), Q)$ entspricht die Sobolev-Klasse

$$\tilde{H}(\beta, Q) := \{f \in L^2([0,1]) : \vartheta \in \Theta(a(\beta), Q)\},$$

wobei die ϑ_j die Koeffizienten bezüglich der trigonometrischen Basis

$$\varphi_1(x) \equiv 1,\ \varphi_{2j}(x) = \sqrt{2}\cos(2\pi jx),\ \varphi_{2j+1}(x) = \sqrt{2}\sin(2\pi jx),\ j \geq 1$$

bezeichnen mögen.
Weiter betrachtet man die parametrische Klasse

$$\Theta_N := \{\vartheta^N = (\vartheta_2, \cdots, \vartheta_N) : \sum_{j=2}^{N} a_j^2 \vartheta_j^2 \leq Q\}$$

und die zugehörige parametrische Teilmenge

$$\mathcal{F}_N = \left\{ f_{\vartheta^N}(x) = \sum_{j=2}^{N} \vartheta_j \varphi_j(x) : \vartheta^N \in \Theta_N \right\}$$

von $\tilde{H}(\beta, Q)$.
Zeige die Abschätzung

$$\inf_{\tilde{f}_\sigma} \sup_{f \in \tilde{H}(\beta.Q)} \mathbb{E}_f \, \|f - \tilde{f}_\sigma\|^2_{L^2[0,1]} \geq \inf_{\hat{\vartheta}^N} \sup_{\vartheta^N \in \Theta^N} \mathbb{E}_{\vartheta^N} \left[\sum_{j=2}^{N} (\hat{\vartheta}_j - \vartheta_j)^2 \right],$$

wobei man links das Infimum über allen (nicht notwendigerweise linearen) Schätzern im Modell des Gaußschen weißen Rauschens betrachtet und und rechts das Minimum über allen Zufallsgrößen $\hat{\vartheta}^N \in \Theta_N$.

(c) Betrachte auf $\mathbb{R}^N$ die Lebesgue-Dichte

$$\mu(\vartheta^N) = \prod_{k=2}^{N} \mu_{s_k}(\vartheta_k),$$

wobei μ_{s_k} die Dichte der $\mathcal{N}(0, s_k^2)$-Verteilung bezeichnet mit $s_k := (1 - \delta) v_k$ für ein $0 < \delta < 1$. Zeige die Abschätzung für das Minimax-Risiko

$$\inf_{\tilde{f}_\sigma} \sup_{f \in \tilde{H}(\beta,Q)} \mathbb{E}_f [\|f - \tilde{f}_\sigma\|^2_{L^2} \geq \inf_{\hat{\vartheta}^N} \sum_{j=2}^{N} \int_{\Theta_N} \mathbb{E}_\vartheta [(\hat{\vartheta}_k - \vartheta_k)^2] \mu(\vartheta^N) \, \mathrm{d}\vartheta^N$$

Hinweis: Benutze die Tatsache, dass man das Minimax-Risiko durch das Bayes-Risiko abschätzen kann.

6. Recherchiere Varianten der Kreuzvalidierung und erkläre die Anwendung der „leave-p-out CV“ und „k-fold CV “ auf die Dichteschätzung. Simuliere diese Verfahren und vergleiche sie mit den Resultaten für die herkömmliche Kreuzvalidierung.

Abgabe der Aufgaben in der Vorlesung am Freitag, dem 28.01.2011

In Aufgabe 6 können bis zu acht Punkte erreicht werden. Das gesamte Blatt geht mit insgesamt acht Punkten in die Bewertung ein, der Rest sind Zusatzpunkte.

Vorlesung *Nichtparametrische Statistik*
Wintersemester 2010/11
Humboldt-Universität zu Berlin

Prof. Dr. Markus Reiß
Dipl. Math. Johanna Kappus

7. Übungsblatt

1. In der Vorlesung wurde der Hard-Thresholding-Schätzer $\hat{f}_M^{(hard)}$ eingeführt. Zeige die Darstellung

$$\hat{f}_M^{(hard)} = \underset{g \in \operatorname{span}(\varphi_1, \cdots, \varphi_M)}{\operatorname{argmin}} \left(\sum_{j=1}^{M} (y_j - g_j)^2 + \lambda \left| \{1 \leq j \leq M : g_j \neq 0\} \right| \right)$$

für ein geeignetes λ.

2. a) Man definiert den Soft-Thresholding Schätzer

$$\hat{f}_M^{(soft)} := \sum_{j=1}^{M} \hat{f}_j^{(soft)} \varphi_j.$$

mit

$$\hat{f}_j^{(soft)} := \left((y_j - \kappa \frac{\sigma}{\sqrt{n}})_+ - (y_j + \kappa \frac{\sigma}{\sqrt{n}})_- \right).$$

Zeige die Darstellung

$$f_M^{(soft)} = \underset{g \in \operatorname{span}(\varphi_1, \cdots, \varphi_M)}{\operatorname{argmin}} \left(\sum_{j=1}^{M} (y_j - g_j)^2 + \lambda \sum_{j=1}^{M} |g_j| \right)$$

für ein geeignetes $\lambda > 0$.

b) Welcher Schätzer ergibt sich aus der ℓ^p-Penalisierung

$$\hat{f} = \underset{g \in \operatorname{span}(\varphi_1, \cdots, \varphi_M)}{\operatorname{argmin}} \left(\sum_{j=1}^{M} (y_j - g_j)^2 + \lambda \sum_{j=1}^{M} |g_j|^p \right)$$

für $p \in \mathbb{R}^+$?

3. Übertrage die Idee des Hard-Thresholding-Schätzers auf das Modell der nichtparametrischen Dichteschätzung mit Dichte $f : [0,1] \to \mathbb{R}^+$. Bestimme den entsprechenden Orakel-Schätzer.

4. Leite im Dichteschätz-Modell aus Aufgabe 3 für Threshold-Schätzer der Form $\hat{f}_j = y_j 1(|y_j| \geq \kappa \frac{\sigma}{\sqrt{n}})$ eine Orakel-Ungleichung für den MISE her. Diskutiere die Wahl von κ.

5. Betrachte $f(x) = |x - s|^\alpha$ für $s \in (0,1)$ und $\alpha \in (0,1]$

a) Zeichne f für $s = 1/3, \alpha = 1/2$ und die beste Approximation $\Pi_J f, J = 1, 3, 5$, durch Haar-Wavelets in ein gemeinsames Koordinatensystem.

b) Welche Ordnung besitzen die Haar-Wavelet-Koeffizienten $|\langle f, \psi_{j,k} \rangle|$ in j?

c) Vergleiche den Approximationsfehler $\|f - \Pi_J f\|_{L^2}$ mit dem Fehler bei bester N-Term Approximation mit $N = 2^J$.

6.** Betrachte das Modell der nichtparametrischen Regression mit deterministischem Design und standardnormalverteilten Fehlern mit der Regressionsfunktion $f(x) = |x - 1/3|^{1/2}$. Implementiere für $n = 1024$ Beobachtungen den Hard-/Soft-Thresholding-Schätzer für die Haar-Wavelet-Basis mit geeigneten Schwellenwerten κ. Diskutiere die Ergebnisse und vergleiche mit den Resultaten für den Projektionsschätzer auf die Haar-Wavelet-Basis.

Abgabe der Aufgaben in der Vorlesung am Freitag, dem 4.02.2011