〈Research Paper〉

Journal of the Korean
Statistical Society
Vol. 19, No. 2, 1990

Optimal Rates of Convergence for Tensor Spline Regression Estimators [+]

Ja Yong Koo[*]

ABSTRACT

Let (X, Y) be a pair random variables and let f denote the regression function of the response Y on the measurement variable X. Let K(f) denote a derivative of f. The least squares method is used to obtain a tensor spline estimator $\hat{f}$ of f based on a random sample of size n from the distribution of (X, Y). Under some mild conditions, it is shown that $K(\hat{f})$ achieves the optimal rate of convergence for the estimation of K(f) in L_2 and L_∞ norms.

1. Introduction

Let (X, Y) be a pair of random variables and let f denote the regression function of the response Y on the measurement variable X. Given a random sample $(X_1, Y_1), \cdots, (X_n, Y_n)$ of size n from the distribution of (X, Y), a statistical problem is to obtain an estimator of K(f), where K(f) is a derivative of f of order m. The regression function f is assumed to belong to a space $\mathcal{S}$ is finite-dimensional, the problem of estimating K(f) is parametric. We eschew such an assumption and thus the given problem is nonparametric.

Several families of nonparametric estimators for the regression function including nearest neighbor, kernel, partition, and piecewise polynomial estimators have been proposed. For consistency of each method, see Stone (1977), Devroye and Wagner(1980), Spiegelman and Sacks(1980), Gordon and Olshen(1980), Stone(1982) respectively. An important question regarding optimal rates of convergence under smoothness assumptions on f has been studied in great detail by Stone(1982).

Stone(1986) proposed the method of parametric spline where we confine our attention to spline estimates with fixed degree and a modest number of knots N depending on the sample size n. Let $\mathcal{S}$ be a space of tensor-product splines, which is a linear vector space of dimension J. The tensor spline regression estimator $\hat{f}$ of the regression function f is defined as the minimizer over $s \in \mathcal{S}$ of the residual sum of squares $RSS(s) = \Sigma^n_{i=1}(Y_i - s(X_i))^2$. Theoretically J should tend to infinity as the sample size n tends to infinity. The proposed estimators have the smoothness advantage of splines and they achieve the optimal rates of convergence as shown in Stone(1982). Under mild conditions, the derivatives of the tensor spline estimators can achieve the optimal rate of convergence in L_2 and L_∞ norms. Moreover, they are easy to compute since they are computed by the usual least squares formula.

In parallel with this analytic approach, computer simulation will be used to determine the finite-sample performance of inference based on tensor spline models elsewhere. So far, we have used cubic spline for each coordinate. For the r-th coordinate, $1 \leq r \leq d$, the first knot has been placed at the minimum value of the r-th ccordinates $X_{r1}, \cdots, X_{rn}$, the last knot has been placed at the maximum value, and a number of intermediate knots have been placed at selected order statistics of the r-th coordinate. Also, the tentative rule of selecting the number of intermediate knots and the order statistics used in placing them has been made in a data-dependent manner. The number J depends on the smoothness of unknown f and thus a study on the data-dependent selection rule for J is currently under way.

Throughout the paper M, M_1, M_2, $\cdots$ denote unspecified positive constants independent of n whose values may change at their occurrences.

The organization of the paper is as follows. In Section 2, we introduce the tensor spline regression estimators. The main result is stated in Section 2. Some technical lemmas are stated and proved in the appendix.

2. Statement of Results

Let (X, Y) be a pair of random variables such that X ranges over a known compact set C having finite volume ; without loss of generality, let $C=[0, 1]^d$. Here Y is called a response variable and X is referred to as a vector of covariates. Let f be the regression function of Y on X ; so that $E(Y \mid X) = f(X)$.

Condition 1. f is bounded on C.

Condition 2. The marginal distribution G of X is absolutely continuous and its density g is bounded away from zero and infinity on C.

Condition 3. There are positive constants t_0 and M such that

$$E[exp(tY) \mid X=x] < M \ for \ \mid t \mid < t_0 \ and \ x \in C.$$

Condition 3 is satisfied in the following examples [for several other examples, see page 1350 of Stone(1980)].

Example 1(Normal). Let φ denote Lebesgue measure on **R**, and let the conditional density $h(y \mid x, t) = h(y \mid x, f(x))\varphi(dy)$ of Y given X=x be

$$h(y \mid x, \ t) = (2\pi\sigma^2(x))^{1/2} \ exp\{-(y-t)^2/2\sigma^2(x)\}$$

where $\sigma^2(\cdot)$ is bounded away from zero and infinity on C.

Example 2(Bernoulli). Let φ be a counting measure on {0, 1} and let

$$h(y \mid t) = t^y(1-t)^{1-y}$$

where t ranges over a relatively compact open subinterval of (0, 1).

For simplicity in notation, we suppress the dependence on n of various quantities after these quantities are defined.

Let q denote a positive integer. Given $n \geq 1$, let $N=N_n$ denote a positive integer. Let $\mathcal{I}=[0, 1]$ be partitioned into subintervals

$$\mathcal{I}_k = [(k-1)/N, \ k/N), \ 1 \leq k < N, \ and \ \mathcal{I}_N = [(N-1)/N, \ 1].$$

$\mathcal{S}=\mathcal{S}_n$ denote the collection of functions s on $\mathcal{I}$ satisfying the following two properties : s is a polynomial of order q (degree less than q) on each of the subintervals $\mathcal{I}_1, \cdots, \mathcal{I}_N$; if $q>2$, s is $(q-2)$-times continuously differentiable on $\mathcal{I}$. Then $\mathcal{S}$ is a vector space of dimension $M=M_n=q+N-1$, which is a space of polynomial splines of order q with simple knots at j/N for $1 \leq j < N$. The functions in $\mathcal{S}$ are piecewise constant, usual B-splines [see de Boor(1978)].

Let $\mathcal{T}=\mathcal{T}_n$ denote the space of tensor-product splines of order q which is defined by the tensor product of $\mathcal{S}_1, \cdots, \mathcal{S}_d$. That is

$$\mathcal{T}=span\{B_{j_1}(x_1) \cdots B_{j_d}(x_d) : 1\leq j_r \leq M=q+N-1,\ 1\leq r\leq d\}.$$

The functions $\{B_{j_1 \cdots j_d}(x)=B_{j_1}(x_1) \cdots B_{j_d}(x_d) : 1\leq j_r \leq M,\ 1\leq r\leq d\}$ are called the tensor-product B-splines [see Schumaker(1981)]. The number $J=J_n$ of the tensor-product B-splines is $M^d=(q+N-1)^d$. Let $\{B_j : 1\leq j\leq J\}$ denote the tensor-product B-splines in some order from now on. These tensor-product B-splines are nonnegative and sum to 1 on C. Each B_r is zero outside a rectangle R_r of volume at most $O(J^{-1})$.

Let $\Theta=\Theta_n$ denote the collection of vectors $\theta\in\mathbf{R}^J$. For $\theta\in\Theta$ set

$$|\theta| =(\Sigma_j\ \theta_j^2)^{1/2}$$

$$|\theta|_\infty=max_{1\leq j\leq J}\ |\theta_j|$$

and

$$s(\cdot\ ;\theta)=s_n(\cdot\ ;\theta)=\Sigma_{j=1}^J\ \theta_j B_j.$$

Given $\theta\in\Theta$, set

$$\Lambda(\theta)=E[Y-s(X\ ;\theta)]^2.$$

Let $\mathbf{H}=\mathbf{H}_n$ denote the Hessian matrix of $-\Lambda(\theta)$ at $\theta\in\Theta$, which is the $J\times J$ matrix having entry $EB_j(X)B_k(X)$ in row j and column k for $1\leq j,\ k\leq J$. If τ is a nonzero element of Θ then $\tau^T\mathbf{H}\tau=E[s(X\ ;\tau)]^2>0$ from Condition 1 and the fact that $s(\cdot\ ;\tau)$ is not almost everywhere equal to zero on C. It follows that for each $n\geq1$, there is a unique $\theta^*=\theta^*_n\in\Theta$ that minimizes $\Lambda(\cdot)$ on Θ. Set $f^*=f_n^*=s(\cdot\ ;\theta^*)$.

Let the tensor spline $Pf=P_nf$ on C be defined by $Pf=f^*$. The regression function f is a tensor spline model if and only if $f=Pf$. When f is not a tensor spline model, the function $f-Pf$ roughly plays the role of a bias term.

Let $\|\cdot\|_2$ and $\|\cdot\|_\infty$ denote the usual L_2 and L_∞ norms of functions on C. Set

$$\delta=\delta_n=inf_{s\in\mathcal{T}}\ \|f-s\|_\infty.$$

By Theorem 12.7 of Schumaker(1981) δ is bounded under Condition 1.

Theorem 1. Assume that Conditions 1 and 2 hold. Then there exists a positive constant M which depends only on the order q of $\mathcal{T}$ such that

$$\|f-f^*\|_\infty\leq M\delta.$$

Proof. For $h\in\mathcal{G}$, let $h^*=s(\cdot\ ;\tau^*)$ be a function in $\mathcal{T}$ satisfying

$$E(h(X)-h^*(X))s(X)=0 \text{ for all } s\in\mathcal{T}.$$

Since $B_j\geq0$, all j, and $\Sigma_jB_j\leq1$, $\|h^*\|_\infty\leq|\tau^*|_\infty$, while by Condition 2, there exists a positive constant M_1 such that

$$\Sigma_jEB_i(X)B_j(X)\tau_j^* =EB_i(X)h(X)\leq\|h\|_\infty\ EB_i(X)\leq M_1J^{-1}\|h\|_\infty.$$

By Condition 2, Lemma 2 in the appendix, Theorem 1 and 3 of Descloux(1972) [see Koo(1991) for more elementary proof], there exists a positive constant M_2 such that

$$\|\mathbf{H}^{-1}\|_\infty=sup_i\Sigma_j\ |(\mathbf{H}^{-1})_{ij}|\leq M_2J.$$

Therefore $\|h^*\|_\infty\leq M_1M_2\|h\|_\infty$. Set $h=f-s$, where $s\in\mathcal{T}$ and $\|f-s\|_\infty\leq2\delta$. Since $h^*=f^*-s$, we get

$$\|f^* - f\|_\infty \leq \|f^* - s\|_\infty + \|s - f\|_\infty \leq M\delta$$

This completes the proof of Theorem 1.

Let $(X_1, Y_1), \cdots, (X_n, Y_n)$ denote a random sample of size n from the distribution of (X, Y). If $\hat{f}=\hat{f}_n=s(\cdot\ ;\hat{\theta})$ minimizes the residual sum of squares over the space of tensor product splines $\Sigma^n_{i=1}(Y_i - s(X_i\ ;\theta))^2$, then $\hat{f}$ is called the tensor spline regression estimate of f. The estimate $\hat{f}$ can be implemented using tensor-product B-splines [see Schumaker(1981)] and standard regression programs such as *S* or GLIM [see Becker and Chambers(1984), Baker and Nelder(1978)]. Let $\hat{\mathbf{H}}=\hat{\mathbf{H}}_n$ denote the J×J dimensional matrix with elements $\Sigma_j B_r(X_j)B_s(X_j)$ and let $\mathbf{S}(\theta)=\mathbf{S}_n(\theta)$ denote the J-dimensional vector of elements $\Sigma^n_{i=1}B_j(X_i)[Y_i - s(X_i\ ;\theta)]$.

From now on it is assumed that

$$J\to\infty \text{ and } J^2/n\to 0 \text{ as } n\to\infty.$$

Theorem 2. Suppose that Conditions 1-3 hold. Then

$$\|f - f^*\|_2 = O_{pr}((J/n)^{1/2}).$$

Proof. Let $\mathbf{S}^*=\mathbf{S}(\theta^*)$. Observe that $E\mathbf{S}^*_j=0$. Consequently,

$$E\,|\,\mathbf{S}^*\,|^2 = n\Sigma_j EB_j^2(X)[Y-f^*(X)]^2 \leq M_1 n\Sigma_j EB_j^2(X) = O(n) \tag{1}$$

by Conditions 1 and 2, Theorem 1 and Lemma 2. By Lemma 3, there exists a positive constant M_2 such that

$$(\hat{\theta}-\theta^*)^T\,\hat{\mathbf{H}}(\hat{\theta}-\theta^*) \geq M_2(n/J)\,|\,\hat{\theta}-\theta^*\,|^2, \tag{2}$$

except on an event whose probability tends to zero with n. Observe that

$$\hat{\mathbf{H}}(\hat{\theta}-\theta^*) = \mathbf{S}^*. \tag{3}$$

It follows from (1), (2) and (3) and Lemma 1 that $\|\hat{f}-f^*\|_2 = O_{pr}((J/n)^{1/2})$. This completes the proof of Theorem 2.

Remark 1. For the proof of Theorem 2, Condition 3 can be replaced with the weaker condition that $\mathrm{Var}[Y\,|\,X=x]$ is bounded on *C*.

Theorem 3. Suppose that Conditions 1−3 hold. Then

$$\|\hat{f}-f^*\|_\infty = O_{pr}([n^{-1}J\log(J)]^{1/2}).$$

Proof. Let $S_j=\Sigma^n_{i=1} B_j(X_i)(Y_i - f(X_i))$ be the j-th element of $\overline{\mathbf{S}}$. By Conditon 3 [see the proof of Lemma 12.27 in Breiman et al. (1984)], there are positive constant M_1 and M_2 such that

$$E[exp\{tn^{-1}\overline{S}_j\}\,|\,X_1, \cdots, X_n] \leq exp\{M_1t^2n^{-2}\Sigma^n_{i=1} B_j^2(X_i)\}$$

provided that

$$|\,tn^{-1}B_j(X_i)\,| \leq M_2 \text{ for } 1\leq i\leq n.$$

Therefore

$$P[n^{-1}\,|\,\overline{\mathbf{S}}\,|_\infty > s\,|\,X_1, \cdots, X_n] \leq 2\ \Sigma_j exp\{M_1t^2n^{-2}A_j - ts\},$$

where $A_j=\Sigma_i B_j^2(X_i)$. Observe that if $A_j=0$, then $\hat{\theta}_j-\theta^*_j=0$. It follows by choosing $t=n^2s/(2M_1A_j)^{-1}$

$$P[n^{-1}\,|\,\overline{\mathbf{S}}\,|_\infty > s] \leq 2\ \Sigma_j E\ exp\{-s^2n^2/(4M_1A_j)\}.$$

By Condition 2 and Bernstein's inequality [(2.13) of Hoeffding (1963)], Since $\exp\{-s^2n^2/(4M_1A_j)\}\leq 1$, we get that

$$P[n^{-1}|\bar{S}|_\infty > M_6(n^{-1}J^{-1}log(J))^{1/2}] \leq 2\ J\ exp\{-(M_6)^2/M_4 log(J)\} + 2\ J\ exp\{-M_5 n/J\}$$

and thus

$$|\bar{S}|_\infty = O_{pr}([nJ^{-1}log(J)]^{1/2}).$$

Let

$$G_j(X_i) = B_j(X_i)(f(X_i) - f^*(X_i))\ for\ 1\leq i\leq n\ and\ 1\leq j\leq J.$$

By Condition 1, Theorem 1 and Lemma 2,

$$|G_j(X_i)| \leq M_7\ and\ Var(G_j(X_i)) \leq M_8 J^{-1}.$$

By Bernstein's inequality,

$$P[|\Sigma_i G_j(X_i)| \geq M_9(nJ^{-1}log(J))^{1/2}] \leq 2exp\{-M_{10}log(J)\})\ for\ M_{10}>1.$$

Therefore

$$sup_j |\Sigma_i G_j(X_i)| = O_{pr}([nJ^{-1}log(J)]^{1/2}).$$

By Lemma 3 and Theorem 3 of Descloux(1972), $\|H^{-1}\|_\infty = O_{pr}(J/n)$. Therefore

$$\|\hat{f} - f^*\|_\infty \leq |\hat{\theta} - \theta^*|_\infty \leq \|H^{-1}\|_\infty\ |S^*|_\infty = O_{pr}(J/n[nJ^{-1}log(J)]^{1/2}).$$

This completes the proof of Theorem 3.

The rate of convergence of $\hat{f}$ to f will now be determined. To this end, given positive numbers a_n and b_n for $n>1$, let $a_n \approx b_n$ mean that a_n/b_n is bounded away from zero and infinity. Let $\alpha = (\alpha_1, \cdots, \alpha_d)$ denote a d-tuple of nonnegative integers and set $[\alpha] = \alpha_1 + \cdots + \alpha_d$. Let D^α denote the differential operator defined by

$$D^\alpha = \partial^{[\alpha]}/\partial x_1^{\alpha_1} \cdots\cdots \partial x_d^{\alpha_d}.$$

Let k be a nonnegative integer, let $\beta \in [0, 1)$ be such that $p = k + \beta > d/2$, and let $L \in (0, \infty)$. Let $\mathcal{G}$ denote the collection of k-times continuously differentiable functions f on C such that

$$|D^\alpha f(x) - D^\alpha f(x_0)| < L\,|x - x_0|^\beta\ for\ x_0,\ x\in C\ and\ [\alpha] = k.$$

Let K(f) be a linear combination with constant coefficients of $D^\alpha f$, $[\alpha]\leq k$; so that $K(f) = Uf$, where

$$U = \Sigma_{[\alpha]\leq k}\ q_\alpha D^\alpha,$$

q_α being real constant for $[\alpha]\leq k$. Let m denote the order of K, defined by

$$m = max([\alpha] : [\alpha]\leq k\ and\ q_\alpha \neq 0).$$

Think of $p = k + \beta$ as a measure of the smoothness of the functions in $\mathcal{G}$ and set $\gamma = 1/(2p+d)$, $r = p\gamma$, and $s = (p-m)\gamma$.

Theorem 4. Suppose that the regression function f belongs to $\mathcal{G}$ and $q \geq k$. Let K be a m-th order differential operator.

(a) If $N \approx n^\gamma$, then

$$\|\hat{f} - f\|_2 = O_{pr}(n^{-r})$$

and

(b) If $N \approx (n/\log n)^{\gamma}$, then

$$\|\hat{f}-f\|_{\infty}=O_{pr}((n^{-1}log(n))^{r})$$

and

$$\|K(\hat{f})-K(f)\|_{\infty}=O_{pr}((n^{-1}log(n))^{s}).$$

Proof. By Theorem 2 and 3 we get that

$$\|\hat{f}-f^{*}\|_{2}=O_{pr}([J/n]^{1/2}) \text{ and } \|\hat{f}-f^{*}\|_{\infty}=O_{pr}([Jlog(J)/n]^{1/2}).$$

By Lemma 1, we get that for $f \in \mathcal{F}$,

$$\|f-f^{*}\|_{\infty}=O(N^{-p}).$$

By choosing $N \approx n^{\gamma}$, we get that

$$\|\hat{f}-f\|_{2}=O_{pr}(n^{-r}).$$

By choosing $N \approx (n/\log(n))^{\gamma}$, we get that

$$\|\hat{f}-f\|_{\infty}=O_{pr}([log(n)/n]^{r}).$$

The desired result of Theorem 4 now follows from these and Lemma 4.

Remark 2. The rates given in Theorem 4 are the well-known optimal rates of convergence for nonparametric regression function estimation; see Stone(1982).

Remark 3. These rates are optimal when the conditional densiy $h(y \mid x, f(x))\varphi(dy)$ of Y given X=x satisfies

$$\int h(y \mid x,f_1(x))log[h(y \mid x,f_1(x))/h(y \mid x,f_2(x))]\varphi(dy) \leq M \mid f_1(x)-f_2(x) \mid^{\lambda},$$

with $\lambda=2$ and $\mathcal{F}$ satisfies the smoothness conditions given above. The optimal rates of convergence will have the form n^{-r} with $r=p/(\lambda p+d)$ for general λ. If $\mathcal{F}$ is a tensor Sobolev space, the optimal rates will be different [Koo(1991)].

Appendix

Recall the definition of $\mathcal{F}$ defined in Section 2.

Lemma 1. There exists a positive constant M depending only on q such that for $f \in \mathcal{F}$

$$\delta=inf_{s\in\mathcal{F}} \|f-s\|_{\infty} \leq MN^{-p}.$$

Proof. Let $i=(i_1, \cdots, i_d)$, where $i_j \in \{0, \cdots, (N-1)\}$ for $j=1, \cdots, d$ and let

$$\mathcal{C}_i=\{x=(x_1, \cdots, x_d): i_j/N \leq x_j \leq (i_j+1)/N, j=1, \cdots, d\}.$$

Let $\|\cdot\|_i$ denote the L_{∞} norm on $\mathcal{C}_i$ and define $\tilde{f}$ to be the quasi-interpolant of f denfined by (12.29) of Schumaker(1981). Fix $q \leq i_j \leq q+N-1$, $j=1, \cdots, d$. An error bound for the difference $\tilde{f}-f$ on the rectangle $\mathcal{C}_i$ will be obtained at first. By Theorem 13.18 of Schumaker(1981), there exists a polynomial $\tilde{u}_i$ of total order k such that

$$\|f-u_i\|_i \leq M_1 N^{-p},$$

The constant M_1 depends only on k, d. Now since the quasi-interpolant operator reproduces polynomials, i.e. $\tilde{u}_i=u_i$ and the quasi-interpolant operator is a bounded linear operator [see (12.31) of Schumaker (1981)]

$$\|f-\tilde{f}\|_i \le \|f-u_i\|_i + \|\tilde{f}-\tilde{u}_i\|_i \le M_2\|f-u_i\|_i \le M_1M_2N^{-p}.$$

Summing over all $q \le i_j \le q+N-1$, $j=1, \cdots, d$, we obtain the desired result.

Lemma 2. Let $s=s(\cdot\,;\theta)\in\mathcal{J}$. There exist positive constants M_1, M_2 and M_3 such that

$$M_1J^{-1}|\theta|^2 \le \|s(\cdot\,;\theta)\|_2 \le M_2J^{-1}|\theta|^2.$$

Proof. The second inequality is straightforward. As to the first, observe that there exists a positive constant M independent of n such that

$$|\theta_j|^2 \le MJ\int_{R_j} s^2,$$

where R_j is the support of B_j [see page 489 of Schumaker(1981)]. From this Lemma 2 follows.

Lemma 3. If n/J^2 increases as n increases, then there exist positive constants M_1 and M_2 such that uniformly for all $\theta\in\mathbf{R}^J$,

$$M_1nJ^{-1}|\theta|^2 \le \Sigma^n_{i=1} s^2(X_i\,;\theta) \le M_2nJ^{-1}|\theta|^2,$$

except on an event whose probability tends to zero as n increases.

Proof. Let $Z_i=s^2(X_i\,;\theta)$ and let $\mu=EZ_i$ for $1\le i\le n$. By Condition 2 and Lemma 1, we get that $\mu\approx\int s^2(\cdot\,;\theta)\approx|\theta|^2/J$. Observe that $|Z_i|=O(|\theta|^2)$. From the Hoeffding's inequality [see Theorem 2 of Hoeffding(1963)], the desired result now follows.

Lemma 4. Let Q be a m-th order differential operator. Then for $h\in\mathcal{G}$ there are $M_1, \cdots, M_4\in(0, \infty)$ such that for $s\in\mathcal{J}$

$$\|Qs-Qh\|_2^2 \le M_1N^{2(m-p)} + M_2N^{2m}\|s-h\|_2^2 \quad \text{(a)}$$

$$\|Qs-Qh\|_\infty \le M_3N^{(m-p)} + M_4N^m\|s-h\|_\infty. \quad \text{(b)}$$

Proof. Let $x_0\in R=R_{nj}\subset[0, 1]^d$, where R_j has volume N^{-d}. Let $\tau=N^{-1}$. There exists a polynomial u of total order k such that for a function h in $\mathcal{G}$, $|h-u|\le M_5\tau^p$ and $\int_R|h-u|^2\le M^2_5\tau^{2p+d}$. By an extension of Lemma 11 of Stone(1985) using compactness arguments,

$$(Qs(x_0)-Qh(x_0))^2=(Qs(x_0)-Qu(x_0))^2\le M_6\tau^{-(2m+d)}(M_3^2\tau^{2p+d}+\int_R(s-h)^2)$$

and hence

$$\int_R(Qs-Qh)^2\le M_6\tau^{-2m}(M_3^2\tau^{2p+d}+\int_R(s-h)^2).$$

Consequently,

$$\|Qs-Qh\|_2^2\le M_6(M_3^2\tau^{2(p-m)}+\tau^{-2m}\|s-h\|_2^2),$$

which completes the proof of (a). The proof of (b) is obvious

References

1. Baker, R.J. and Nelder, J.A.(1978). *The GLIM Systems, Release 3, Generalized Linear Interactive Modelling*. Numerical Analysis Group, Oxford.
2. Becker, R.A. and Chambers, J.M.(1984). *S: An Interactive Environment for Data Analysis and Graphics*, Wadsworth, California.
3. De Boor, C.(1978). A Bound on the L_∞-norm of L_2-approximation by Splines in Terms of a Global Mesh Ratio. Math. Comp. Vol. 30, 765-771.
4. De Boor, C.(1978). *A Practical Guide to Splines*. Springer-Berlag, New York.
5. Breiman, L., Friedman, H.H., Olshen, R.A. and Stone, C.J.(1984). *Classification and Regression Trees*.

6. Craven, P. and Wahba, G. (1979). Smoothing Noisy Data with Spline Functions, Estimating the Correct Degree of Smoothing by the Method of Generalized Cross Validation. *Numer. Math.*, Vol. 31, 377-403.
7. Descloux, J. (1972). On Finite Element Matrices. *SIAM J. Numer. Anal.*, Vol. 9, 260-265.
8. Devroye, L.P. and Wagner, T.J. (1980). Distribution-free Consistancy Results in Nonparametric Discrimination and Regression Function Estimation. *Ann. Statist.*, Vol. 8, 231-239.
9. Koo, J.Y. (1991). Optimal Rates of Convergence in Tensor Sobolev Spaces. Manuscript.
10. Gordon, L. and Olshen, R.A. (1980). Consistent Nonparametric Regression from Recursive Partitioning Schemes. *J. Mult. Anal.*, Vol. 10, 611-627.
11. Schumaker, L.L. (1981). *Spline Functions: Basic Theory*. Wiley, New York.
12. Spiegelman, C. and Sacks, J. (1980). Consistent Window Estimation in Nonparametric Regression. *Ann. Statist.*, Vol. 8, 240-814.
13. Stone, C.J. (1977). Consistent Nonparametric Regression. *Ann. Statist.*, Vol. 8, 1348-1360.
14. Stone, C.J. (1982). Optimal Global Rates of Convergence for Nonparametric Regression. *Ann. Statist.*, Vol. 10, 1040-1053.
15. Stone, C.J. (1986). Contribution to the Discussion of the Paper by T. Hastie and R. Tibshirani. *J. Statist. Science*.