Lecture 8

Nonparametric regression: splines

Denote
$$\mathcal{P}_m = \{p : \ p(x) = \sum_{i=1}^{m} c_i x^{i-1}, \ c_1, \dots, c_m, x \in \mathbb{R}\}.$$
Next, consider an interval $[a, b)$ $(a, b \in \mathbb{R}, -\infty < a < b < \infty)$. Let $a = \tau_0 < \tau_1 < \dots < \tau_k < \tau_{k+1} = b$, $\tau_i \in \mathbb{R}$ and denote $\Delta_k = \{\tau_i\}_{i=0}^{k+1}$ a **partition of** $[a, b)$ **into** $k + 1$ **subintervals** $[\tau_i, \tau_{i+1})$**,** $i = 0, \dots, k$**.**

Definition Let Δ_k be a partition of $[a, b)$, then
$$\mathcal{PP}_m(\Delta_k) = \{p : \ \exists\, p_0, \dots, p_k \in \mathcal{P}_m \text{ such that } p(x) = p_i(x) \text{ for } x \in [\tau_i, \tau_{i+1}), \ i = 0, \dots, k\}$$
is the **space of piecewise polynomials of order** m **based on** Δ_k.

Definition

Let Δ_k be a partition of $[a, b)$. Let $m \in \mathbb{N}$ and $M = (m_1, \dots, m_k)$ be a vector of integers with $1 \leq m_i \leq m$, $i = 1, \dots, k$. Then, the space
$$\begin{aligned}\mathcal{S}_m(M, \Delta_k) \ = \ \Big\{ & s : \ \exists\, s_0, \dots, s_k \in \mathcal{P}_m \text{ such that } s(x) = s_i(x) \text{ for } x \in [\tau_i, \tau_{i+1}), \ i = 0, \dots, k \\ & \text{and } s_{i-1}^{(j)}(\tau_i) = s_i^{(j)}(\tau_i), \ j = 0, 1, \dots, m-1-m_i, \ i = 1, \dots, k \Big\}\end{aligned}$$
is the **space of polynomial splines of order** m **and multiplicities** M **based on** Δ_k.

Thus, a polynomial spline is a piecewise polynomial with the additional conditions on connection of pieces: $s_{i-1}^{(j)}(\tau_i) = s_i^{(j)}(\tau_i)$, $j = 0, 1, \dots, m - 1 - m_i$.
If all $m_i = m$, then $s_{i-1}(\tau_i)$ and $s_i(\tau_i)$ are unrelated and $\mathcal{S}_m(M = (m, \dots, m), \Delta_k) = \mathcal{PP}_m(\Delta_k)$.
If all $m_i = 1$, then $\mathcal{S}_m(M = (1, \dots, 1), \Delta_k) \subset \mathcal{C}^{m-2}[a, b]$ is the *smoothest space of polynomial splines*. Indeed, if the pieces were joined together any smoother, then the knots would disappear.
Further we will deal with $\mathcal{S}_m(M = (1, \dots, 1), \Delta_k)$ only and will denote it just $\mathcal{S}_m(\Delta_k)$. Note that one can define $\mathcal{S}_m(\Delta_k)$ also by $\mathcal{S}_m(\Delta_k) = \mathcal{PP}_m \cap \mathcal{C}^{m-2}[a, b]$.

Theorem $\mathcal{S}_m(\Delta_k)$ *is a linear vector space of dimension* $k + m$. Proof: See Schumaker (2007) *Spline functions: basic theory*, p.110.
Idea: to see that the dimension is $k + m$ note that each piece $s_i(x) = \sum_{j=1}^{m} c_{ij} x^{j-1}$, $i = 0, \dots, k$, so that there are $m \cdot (k + 1)$ parameters c_{ij}. Thereby, there are $(m - 1) \cdot k$

continuity conditions on c_{ij}: $s_{i-1}^{(j)}(\tau_i) = s_i^{(j)}(\tau_i)$, $j = 0, 1, \ldots, m-1-1$, $i = 1, \ldots, k$. Hence, the space dimension is $m(k+1) - (m-1)k = mk + m - mk + k = m + k$ (=number of parameters minus number of constraints).

Let now consider a basis on this space. Denote for some $j \in \{0, 1, \ldots\}$

$$(x-\tau)_+^j = \begin{cases} (x-\tau)^j, & x \geq \tau \\ 0, & x < \tau, \end{cases}$$

with the convention $0^0 = 1$.

Theorem
A basis for $\mathcal{S}_m(\Delta_k)$ is given by

$$\left\{1, (x-\tau_0), \ldots, (x-\tau_0)^{m-1}, (x-\tau_1)_+^{m-1}, \ldots (x-\tau_k)_+^{m-1}\right\}.$$

Proof: See Schumaker (2007), pp. 111 – 112; see also Lemma 14.1 in Györfi et al. (2002). Idea: Each element of this basis vector is obviously from $\mathcal{S}_m(\Delta_k)$ and the dimension of this basis is $k + m$, it remains to check linear independence using its definition.

This basis is known as **truncated polynomial basis**.

This theorem says that any $s \in \mathcal{S}_m(\Delta_k)$ can be represented as

$$s(x) = \sum_{i=1}^{m} \alpha_i (x-\tau_0)^{i-1} + \sum_{i=1}^{k} \alpha_{m+i}(x-\tau_i)_+^{m-1}$$

for some suitable $\alpha_i \in \mathbb{R}$. However, the truncated polynomial basis is not well suited for numerical calculations: to evaluate $s(x)$ at $x \in [\tau_k, \tau_{k+1})$ one needs to calculate *all* basis elements. Additionally, this basis is far from being orthogonal and is often bad conditioned. Therefore, we consider another basis, which is more "local".

To Δ_k let define $\tau_{-m+1} = \tau_{-m+2} = \ldots = \tau_0 = a$ and $b = \tau_{k+1} = \tau_{k+2} = \ldots = \tau_{k+m}$ and call $\widetilde{\Delta}_k = \{\tau_i\}_{i=-m+1}^{k+m}$ **an extended partition** for $\mathcal{S}_m(\Delta_k)$.

Definition Let $\widetilde{\Delta}_k$ be an extended partition associated with $\mathcal{S}_m(\Delta_k)$, then the functions

$$N_i(x) = (-1)^m(\tau_{i+m} - \tau_i)[\tau_i, \ldots, \tau_{i+m}](x - \cdot)_+^{m-1}, \quad i = -m+1, \ldots, k, \quad x \in [a, b)$$

are called **B-splines**. Here

$$[\tau_1, \ldots, \tau_k]f = \frac{[\tau_2, \ldots, \tau_k]f - [\tau_1, \ldots, \tau_{k-1}]f}{\tau_k - \tau_1},$$

with $[\tau_i]f := f(\tau_i)$ defines the k-th order divided difference of f.

Schumaker (2007) proves on p. 116, that $N_i(x)$, $i = -m+1, \ldots, k$ form a basis for $\mathcal{S}_m(\Delta_k)$.

Example: $m = 2$, $i = 1$

$$\begin{aligned} N_1(x) &= (-1)^2(\tau_3 - \tau_1)[\tau_1, \tau_2, \tau_3](x - \cdot)_+ \\ &= (\tau_3 - \tau_1)\frac{[\tau_2, \tau_3](x - \cdot)_+ - [\tau_1, \tau_2](x - \cdot)_+}{\tau_3 - \tau_1} \\ &= \frac{(x - \tau_3)_+ - (x - \tau_2)_+}{\tau_3 - \tau_2} - \frac{(x - \tau_2)_+ - (x - \tau_1)_+}{\tau_2 - \tau_1}, \end{aligned}$$

so that for $x \in [\tau_1, \tau_2)$ we find $N_1(x) = (x - \tau_1)/(\tau_2 - \tau_1)$, while for $x \in [\tau_2, \tau_3)$ it holds $N_1(x) = -(x - \tau_3)/(\tau_3 - \tau_2)$. For x outside $[\tau_1, \tau_3)$ $N_1(x) = 0$.

Another, recursive definition of B-splines

$$\begin{aligned} N_{i,1}(x) &= \begin{cases} 1, & x \in [\tau_i, \tau_{i+1}) \\ 0, & \text{else} \end{cases} \\ N_{i,m}(x) &= \begin{cases} \frac{x-\tau_i}{\tau_{i+m-1}-\tau_i} N_{i,m-1}(x) + \frac{\tau_{i+m}-x}{\tau_{i+m}-\tau_{i+1}} N_{i+1,m-1}(x), & x \in [\tau_i, \tau_{i+m}) \\ 0, & \text{else} \end{cases}, \end{aligned}$$

for $i = -m+1, \ldots, k$ and with convention $0/0 = 0$. Here $N_{i,j}(x)$ denotes the i-th basis function for $\mathcal{S}_j(\Delta_k)$.

B-splines in a nutshell

Each $N_i(x)$ of order m

1. consists of m polynomial pieces of degree $m-1$, which join at $m-1$ inner knots
2. $N_i(x) > 0$ for $x \in [\tau_i, \tau_{i+m})$ and is zero for xs outside of this interval
3. $N_i(x)$ overlaps with $2(m-1)$ pieces of its neighbors
4. $\sum_{i=j-m+1}^{j} N_i(x) = 1$, $x \in [\tau_j, \tau_{j+1})$

Now consider a fixed design regression model with deterministic $\{x_i\}_{i=1}^n \in [0, 1]$

$$Y_i = f(x_i) + \epsilon_i,\ \mathrm{cov}(\epsilon_i \epsilon_j) = \sigma^2 \delta_{ij},\ \mathrm{E}(\epsilon_i) = 0,\ i = 1, \ldots, n.$$

Regression function f is estimated by regression splines, that is

$$\begin{aligned}\widehat{f}_n &= \arg\min_{s\in\mathcal{S}_m(\Delta_k)} \frac{1}{n}\sum_{i=1}^n \{Y_i - s(x_i)\}^2 = N(\cdot)\arg\min_{\beta\in\mathbb{R}^{k+m}} \frac{1}{n}\sum_{i=1}^n \left\{Y_i - \sum_{j=1}^{k+m} N_j(x_i)\beta_j\right\}^2 \\ &= N(\cdot)\arg\min_{\beta\in\mathbb{R}^{k+m}} (Y-N\beta)^t(Y-N\beta) = N(\cdot)(N^tN)^{-1}N^tY,\end{aligned}$$

where $N = \{N(x_1)^t, \ldots, N(x_n)^t\}^t$ is the basis matrix with $N(x) = \{N_1(x), \ldots, N_{k+m}(x)\}$ as some basis of $\mathcal{S}_m(\Delta_k)$.

For x_i it is assumed that there is a distribution function Q with a positive continuous density q, such that the empirical distribution of x_i, denoted by Q_n can be sufficiently good approximated by Q:

$$\sup_{x\in[0,1]} |Q_n(x) - Q(x)| = o(k^{-1}).$$

For the fixed knots τ_i it is assumed, that

$$\int_0^{\tau_i} p(x)dx = i/(k+1), \quad i = 0, \ldots, k+1$$

for some positive continuous density $p(x)$ on $[0,1]$. Such knot sequences are called **regular sequences generated by** p. Under this assumptions we get for $f \in \mathcal{C}^m$ that

$$\begin{aligned}\mathrm{E}\int_0^1 \left\{\widehat{f}_n(x) - f(x)\right\}^2 q(x)dx &= \left\{\frac{k\sigma^2}{n}\int_0^1 \frac{p(x)}{q(x)}dx + \frac{k^{-2m}\mathcal{B}_{2m}}{(2m)!}\int_0^1 \frac{\{f^{(m)}(x)\}^2}{\{p(x)\}^{2m}}dx\right\}\{1+o(1)\}, \\ &= AMISE\{\widehat{f}_n\}\{1+o(1)\}.\end{aligned}$$

where $\mathcal{B}_{2m}$ is the Bernoulli number. Moreover, this asymptotic mean integrated squared error is minimized by

(i) $q^a_{opt}(x) = \frac{\sqrt{p(x)}}{\int_0^1 \sqrt{p(x)}dx}$,

(ii) $p^a_{opt}(x) = \frac{\left\{f^{(m)}(x)\right\}^{\frac{4}{4m+1}}}{\int_0^1 \left\{f^{(m)}(x)\right\}^{\frac{4}{4m+1}} dx}$,

(iii) $k^a_{opt} = \left[\frac{\mathcal{B}_{2m}n}{(2m-1)!\sigma^2}\left(\int_0^1 \left\{f^{(m)}(x)\right\}^{\frac{4}{4m+1}} dx\right)^{-1}\right]^{\frac{1}{2m+1}} \int_0^1 \left\{f^{(m)}(x)\right\}^{\frac{4}{4m+1}} dx.$

For equidistant knots one can employ usual unbiased risk estimators to get the optimal number of knots. For example,

$$GCV(k) = \frac{\|\widehat{f}_n - f\|^2}{[1 - \mathrm{tr}\{N(N^tN)^{-1}N^t\}/n]^2} = \frac{\|\widehat{f}_n - f\|^2}{(1-k/n)^2}.$$