Name:...

Functional Data Analysis

Assignment 2

Splines and regularization

Assignments constitute part of the examination and must be handed in time. You are asked to hand in the solutions during a week following the week on which the assignment has been discussed in classes. You can submit either an electronic copy or a hard copy of your work. In the latter case, staple your solutions together.

Problem 1 – B-splines In the lecture we have introduced B-splines as a basis for cubic splines. They were defined as follows

- Assume $\xi_1, \dots, \xi_K$ internal knots and two endpoints ξ_0 and ξ_{K+1}.
- Add three more knots that are equal to ξ_0 and additional three knots that are equal to ξ_{K+1} for the total of $K+8$ knots that from now are denoted by τ_i, $i = 1, \dots, K+8$.
- Define recursively functions $B_{i,m}$, that are splines of the $m-1$th order of smoothness (0 smoothness is discontinuity), $i = 1, \dots, K+8$, $m = 1, \dots, 4$
- For the knots τ_i, $i = 1, \dots, K+8$ we define $B_{i,m}$, $i = 1, \dots, K+8$, $m = 1, \dots, 4$
- The piecewise constant (0-smooth), $i = 1, \dots, K+7$,

$$B_{i,1}(x) = \begin{cases} 1 & \text{if } \tau_i \le x < \tau_{i+1} \\ 0 & \text{otherwise} \end{cases}$$

- Higher order of smoothness , $i = 1, \dots, K+8-m$,

$$B_{i,m}(x) = \frac{x - \tau_i}{\tau_{i+m-1} - \tau_i} B_{i,m-1}(x) + \frac{\tau_{i+m} - x}{\tau_{i+m} - \tau_{i+1}} B_{i+1,m-1}(x).$$

- $B_{i,4}$ are cubic order splines that constitutes basis for all cubic splines.

Consider an interval $[0,1]$ and only one internal point $\xi_1 = 1/2$.

1. How many total knots for the cubic B-splines are considered according to the above definition?
2. Write down explicitly all knots τ_i ='s.
3. Write down explicitly all functions for each recursion step.
4. Sketch the obtained functions.
5. Which of them constitutes the basis for all cubic splines with the given initial internal knot $\xi_1 = 1/2$.
6. How the computations change if $\xi_1 = \xi$ is another internal point, i.e. it is not equal to 1/2?
7.[1] Argue that cubic B-splines are still piecewise cubic polynomials on $[0,1]$ with continuous second derivative, i.e. that they are indeed splines and that they are linearly independent, i.e. one cannot be expressed by a linear combination of others. Consequently, any piecewise cubic spline can be expressed by a linear combination of the B-splines.

[1] This is a more challenging problem for PhD and more mathematically inclined students, for the rest it is an extra (not required) problem.

Solution

1. There will be $K+8=9$ knots, since $K=1$.
2. The knots are: $0,0,0,0,1/2,1,1,1,1$.
3. We have the following four recursion steps:
   (a) 0-smoothness, $m=1$. Piecewise constant functions $B_{i,0}(x)$, $i=1,\dots,8$. Only two are non-zero $B_{4,1}(x)=\mathbf{1}_{[0,1/2)}(x)$ and $B_{5,1}(x)=\mathbf{1}_{[1/2,1)}(x)$.
   (b) 1-smoothness, $m=2$. Piecewise linear continuous functions $B_{i,1}(x)$, $i=1,\dots,7$. Only three are non-zero.

$$B_{3,2}(x)=\begin{cases}1-2x & : x\in[0,1/2)\\ 0 & : x\in[1/2,1]\end{cases}$$

$$B_{4,2}(x)=\begin{cases}2x & : x\in[0,1/2)\\ 2-2x & : x\in[1/2,1]\end{cases}$$

$$B_{5,2}(x)=\begin{cases}0 & : x\in[0,1/2)\\ 2x-1 & : x\in[1/2,1]\end{cases}$$

   (c) 2-smoothness, $m=3$. Piecewise quadratic, continuously differentiable functions $B_{i,2}$, $i=1,\dots,6$. These are non-zero for $i=2,\dots,5$.

$$B_{2,3}(x)=\begin{cases}(1-2x)^2 & : x\in[0,1/2)\\ 0 & : x\in[1/2,1]\end{cases}$$

$$B_{3,3}(x)=\begin{cases}2x(2-3x) & : x\in[0,1/2)\\ 2(x-1)^2 & : x\in[1/2,1]\end{cases}$$

$$B_{4,3}(x)=\begin{cases}2x^2 & : x\in[0,1/2)\\ 2(1-x)(3x-1) & : x\in[1/2,1]\end{cases}$$

$$B_{5,3}(x)=\begin{cases}0 & : x\in[0,1/2)\\ (2x-1)^2 & : x\in[1/2,1]\end{cases}$$

   (d) 3-smoothness, $m=4$. Piecewise cubic, twice continuously differentiable function $B_{i,3}$, $i = 1, \ldots, 5$. These are non-zero for all $i = 1, \ldots, 5$.

$$B_{1,4}(x) = \begin{cases} (1-2x)^3 & : x \in [0, 1/2) \\ 0 & : x \in [1/2, 1] \end{cases}$$

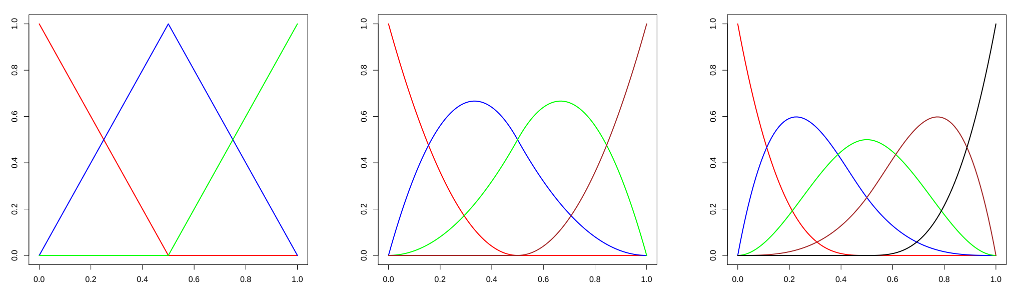
$$B_{2,4}(x) = \begin{cases} 2x\left((1-2x)^2 + (1-x)(2-3x)\right) & : x \in [0, 1/2) \\ 2(1-x)^3 & : x \in [1/2, 1] \end{cases}$$

$$B_{3,4}(x) = \begin{cases} 2x^2(3-4x) & : x \in [0, 1/2) \\ 2(x-1)^2(4x-1) & : x \in [1/2, 1] \end{cases}$$

$$B_{4,4}(x) = \begin{cases} 2x^3 & : x \in [0, 1/2) \\ (1-x)\left((2x-1)^2 + 2x(3x-1)\right) & : x \in [1/2, 1] \end{cases}$$

$$B_{5,4}(x) = \begin{cases} 0 & : x \in [0, 1/2) \\ (2x-1)^3 & : x \in [1/2, 1] \end{cases}$$

4. The following are the graphs of the above functions



They were obtained using the package instead directly evaluating the above functions. Here is *R*-code.

```
install.packages("splines")
 library(splines)
 help(splines)
 library(help = "splines")
 help(bs)
 x=seq(0,1,by=0.01)

 BS=bs(x,knots=c(0.5),intercept=TRUE)
 plot(x,BS[,1])
 plot(x,pmax(0,(1-2*x)^3)) #pointwise maximum
 dev.set(2)
 plot(x,pmax(0,(1-2*x)^3),col="red")

 pdf("BSplines.pdf")

 Colors=c("red","blue","green","brown","black")
 plot(x,BS[,1], col=Colors[1], type=’l’,ylab="",lwd=2)
 for(i in 2:5)
 {
  lines(x,BS[,i], col=Colors[i],lwd=2)
}

dev.off()
```

5. This is the case of $m = 4$, presented on the right hand side graph.

6. The main change is in the middle value of knots τ_5. The rest of the computations will be analogous, although more tedious. Utilizing some software is

highly recommended.

7. The fact that each B-spline is indeed a spline is quite obvious since the first order B-splines are indeed splines and the subsequent orders are obtained by multiplication by a linear function.

   The linear independence can be shown by recurrence. First, we note that the first order splines $B_{4,1}, \dots, B_{K+1,1}$ are clearly linearly independent (why?). We can argue that $B_{3,2}, \dots, B_{K+2,2}$ are linearly independent as follows. Any combination linear of them is expressed as a product of a linear function and a linear combination of the B-splines of the first order. However, for the first $B_{3,2}, \dots, B_{K+1,2}$ to be dependent, a linear combination of $B_{4,1}, \dots, B_{K,1}$ multiplied by a linear function, would have to be zero, which is impossible. The same applies to $B_{4,2}, \dots, B_{K+2,2}$. Eventually, by the identical argument and recurrence, one gets linear independence of $B_{1,4}, \dots, B_{K+4,4}$. For details, if interested, ask your instructors.

Problem 2 – Smoothing splines The following is a simplified account of using smoothing splines to provide generalized additive fit to the regression problem with one predictor

$$y = \alpha + f(x) + \epsilon.$$

- A spline basis method that avoids any knot selection
- It is using the maximal set of knots (knot is located at each location that is given in the data)
- It is not overfitting because irregularity is penalized
- It is estimated by a linear function outside the range of predictors (smoothing on the boundaries)
- It minimizes the penalized residual sum of squares

$$PRSS(f, \lambda) = \sum_{i=1}^{N} (y_i - f(x_i))^2 + \lambda \int f''(t)^2 dt$$

- $\lambda = 0$: any fit that interpolates data exactly.
- $\lambda = \infty$: the least square fit (second derivative is zero)
- We fit by the cubic splines with the maximal number of knots equal to the values of x's and

$$f(x) = \sum_{j=1}^{N+4} \gamma_j B_j(x) \tag{1}$$

  $B_j(x)$ are natural splines: linear outside the data range and the cubic polynomial inside of it.

- The solution has the form

$$\hat{\boldsymbol{\gamma}} = \left(\mathbf{B}^T\mathbf{B} + \lambda\boldsymbol{\Omega}_B\right)^{-1}\mathbf{B}^T\mathbf{y},$$

  where

$$\boldsymbol{\Omega}_B = \left[\int B_i''(t)B_j''(t)\ dt\right]$$

- To see this substitute (**??**) to the PRSS – it becomes a regular least squares problem

Discuss the following properties.

1. Explain why if $\lambda = 0$ the optimal fit will interpolate data exactly.
2. Explain why if $\lambda = \infty$ the optimal fit will be the regular least squares fit of the regression line.

3. Substitute (**??**) to the PRSS and express the latter using vector $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_{N+4})$ and the matrices

$$\mathbf{B} = [B_{ij}] = [B_j(x_i)],$$
$$\boldsymbol{\Omega}_B = \left[\int B_i''(t) B_j''(t)\ dt\right].$$

4.[2] Compute the derivative of the PRSS with respect to $\boldsymbol{\gamma}$ and check that it is equal to zero if

$$\boldsymbol{\gamma} = \left(\mathbf{B}^T\mathbf{B} + \lambda \boldsymbol{\Omega}_B\right)^{-1} \mathbf{B}^T\mathbf{y}.$$

Solution

1. Taking $\lambda = 0$ and assuming that there is as many knot points as the size of data, leads to a curve that goes through all the data points, i.e. data are interpolated exactly.
2. Taking $\lambda = \infty$ makes any non-linear but smooth fit penalized by infinity. Thus the only fit that is acceptable and smooth is the linear one.
3. Let us denote $\mathbf{B}''(t) = [B_1''(t) \dots B_{N+4}''(t)]$, $\mathbf{B} = [B_{ij}]$, where $B_{ij} = B_j(x_i)$. With this notation we have

$$\begin{aligned} PRSS &= \|\mathbf{y} - \mathbf{B}\boldsymbol{\gamma}\|^2 + \lambda \int |\mathbf{B}''(t)\boldsymbol{\gamma}|^2 dt \\ &= (\mathbf{y} - \mathbf{B}\boldsymbol{\gamma})^T (\mathbf{y} - \mathbf{B}\boldsymbol{\gamma}) + \lambda \boldsymbol{\gamma}^T \boldsymbol{\Omega}_B \boldsymbol{\gamma}. \end{aligned}$$

4. We start with basic facts about the derivative of vector valued function depending on vector valued argument. For such a function $\mathbf{g}(\boldsymbol{\gamma})$ we denote by $\mathbf{g}'$ the matrix of partial derivatives $\partial g_i / \partial \gamma_j$, where index i runs through rows and j through columns. We have the following product rule

$$\left(\mathbf{f}^T\mathbf{g}\right)' = \left(\mathbf{f}'\right)^T \mathbf{g} + \left(\mathbf{g}'\right)^T \mathbf{f}.$$

Let us now consider $PRSS$ and apply this rule to it to get

$$\begin{aligned} \left((\mathbf{y} - \mathbf{B}\boldsymbol{\gamma})^T (\mathbf{y} - \mathbf{B}\boldsymbol{\gamma}) + \lambda \boldsymbol{\gamma}^T \boldsymbol{\Omega}_B \boldsymbol{\gamma}\right)' &= \left((\mathbf{y} - \mathbf{B}\boldsymbol{\gamma})^T (\mathbf{y} - \mathbf{B}\boldsymbol{\gamma})\right)' + \lambda \left(\boldsymbol{\gamma}^T \boldsymbol{\Omega}_B \boldsymbol{\gamma}\right)' \\ &= -2\mathbf{B}^T(\mathbf{y} - \mathbf{B}\boldsymbol{\gamma}) + 2\lambda \boldsymbol{\Omega}_B \boldsymbol{\gamma}. \end{aligned}$$

Here we use that $\boldsymbol{\Omega}_B = \boldsymbol{\Omega}_B^T$. Thus the vector of derivatives is equal to zero if $\left(\mathbf{B}^{\mathbf{B}} + \lambda \boldsymbol{\Omega}_B\right) \boldsymbol{\gamma} = \mathbf{B}^T\mathbf{y}$, which concludes the argument.

[2] This is a more challenging problem for PhD and more mathematically inclined students, for the rest it is an extra (not required) problem.