9. Find a set of basis functions for the natural (linear outside the interval \((\xi_1, \xi_K)\)) quadratic regression splines with knots at \(\xi_1 < \xi_2 < \cdots < \xi_K\).

(10 pts) Based on HTF’s construction, the general form for the truncated-power basis set, let

\[
\begin{align*}
    h_j(x) &= x^{j-1}, \quad j = 1, \ldots, 2, \\
    h_{2+l}(x) &= (x - \xi_l)^2, \quad l = 1, \ldots, K.
\end{align*}
\]

Then, \(H \equiv \{h_i(x)\}_{i=1}^{2+K}\) is a basis set of piecewise linear function. For a natural quadratic spline, the functions are asked to be linear outside the interval \((\xi_1, \xi_K)\). Therefore, two more constraints at each side are required. As Ex.5.4 of HTF, I have

\[
f(x) = \sum_{j=1}^{2} \alpha_j x^{j-1} + \sum_{k=1}^{K} \beta_k (x - \xi_k)^2_+,
\]

and let

\[
\frac{d^2}{dx^2} f(x) = 2 \sum_{k=1}^{K} \beta_k \equiv 0
\]

yields \(\beta_K = -\sum_{k=1}^{K-1} \beta_k\). Then,

\[
f(x) = \sum_{j=1}^{2} \alpha_j x^{j-1} + \sum_{k=1}^{K-1} \beta_k [(x - \xi_k)^2_+ - (x - \xi_K)^2_+]
\]

indicates \(N \equiv \{1, x, \{(x - \xi_k)^2_+ - (x - \xi_K)^2_+\}_{k=1,\ldots,K-1}\}\) is a basis for the natural quadratic spline. Note that \(N\) is dependent on the given starting set \(H\).

10. (B-Splines) For \(a < \xi_1 < \xi_2 < \cdots < \xi_K < b\) consider the B-spline bases of order \(m\), \(\{B_{i,m}(x)\}\) defined recursively as follows. For \(j < 1\) define \(\xi_j = a\), and for \(j > K\) let \(\xi_j = b\). Define

\[
B_{i,1}(x) = I[\xi_i \leq x < \xi_{i+1}]
\]

(in case \(\xi_i = \xi_{i+1}\) take \(B_{i,1}(x) \equiv 0\)) and then

\[
B_{i,m}(x) = \frac{x - \xi_i}{\xi_{i+m-1} - \xi_i} B_{i,1} \left( B_{i,(m-1)}(x) + \frac{\xi_{i+m} - x}{\xi_{i+m} - \xi_{i+1}} B_{i+1,(m-1)}(x) \right)
\]

(where we understand that if \(B_{i,1}(x) \equiv 0\) its term drops out of the expression above). For \(a = -0.1\) and \(b = 1.1\) and \(\xi_i = (i - 1)/10\) for \(i = 1, 2, \ldots, 11\), plot the non-zero \(B_{i,3}(x)\). Consider all linear
combinations of these functions. Argue that any such linear combination is piecewise quadratic with first derivatives at every $\xi_i$. If it is possible to do so, identify one or more linear constraints on the coefficients (call them $c_i$) that will make $q_c(x) = \sum_i c_i B_{i,3}(x)$ linear to the left of $\xi_1$ (but otherwise minimally constrain the form of $q_c(x)$).

(10 pts) The attached code gives a recursive function defining the formula $B_{i,m}(x)$ and generates a plot for $i = -2, -1, \ldots, 11$, $m = 3$, and $x \in [-0.1, 1.1]$ where the vertical dot lines indicate the $\xi_i$'s.
```r
### Compute B_{i, 3}(x)
N.x <- 200
x.all <- seq(a, b, length = N.x)
ret.B <- matrix(0, nrow = N.x, ncol = K + 3)  # Extra column for i = -2, -1, and 0.
m <- 3
for(i in 1:ncol(ret.B)){
  for(j in 1:N.x){
    ret.B[j, i] <- B(x.all[j], i - 3, m) # Input i - 3 such that from -2 to 11.
  }
}

### Generate plot
colors <- c("orange", "green3", "blue")
par(mar = c(3, 3, 2, 1), mgp = c(2, 1, 0))
plot(NULL, NULL, type = "n", xlab = "x", ylab = expression(B["i, 3"](x)),
     main = paste("B-Spline bases of order ", m, sep = ""),
     xlim = c(a - 0.3, b), ylim = range(ret.B, na.rm = TRUE))
ret <- lapply(1:ncol(ret.B),
  function(i) lines(x.all, ret.B[,i], lwd = 2,
     col = colors[i %% 3 + 1], lty = i %% 5 + 1))
abline(v = xi, lty = 3)
legend(-0.4, 0.9, paste("i = ", 1:ncol(ret.B) - 3, sep = ""),
       col = colors[1:ncol(ret.B) %% 3 + 1], cex = 0.7,
       lty = 1:ncol(ret.B) %% 5 + 1, lwd = 2)
```

(5 pts) Argue for piecewise quadratic with first derivatives at every $\xi_i$. Let $I_i(x) \equiv B_{i,1}(x)$ and $I_i(x) = 0$ for any $i$ is out of expression. I have
\[
B_{i,3}(x) = \frac{x - \xi_i}{\xi_{i+2} - \xi_i} B_{i,2}(x) + \frac{\xi_{i+3} - x}{\xi_{i+3} - \xi_{i+1}} B_{i+1,2}(x) \\
= \frac{x - \xi_i}{\xi_{i+2} - \xi_i} \left[ \frac{x - \xi_i}{\xi_{i+1} - \xi_i} I_i(x) + \frac{\xi_{i+2} - x}{\xi_{i+2} - \xi_i} I_{i+1}(x) \right] \\
+ \frac{\xi_{i+3} - x}{\xi_{i+3} - \xi_{i+1}} \left[ \frac{x - \xi_{i+1}}{\xi_{i+2} - \xi_{i+1}} I_{i+1}(x) + \frac{\xi_{i+3} - x}{\xi_{i+3} - \xi_{i+2}} I_{i+2}(x) \right]
\]
which is in a quadratic form of $x$. Also, for $j \neq i + 1, i + 2$,
\[
\lim_{x \to \xi_j} \frac{d}{dx} B_{i,3}(x) = \lim_{x \to \xi_j^+} \frac{d}{dx} B_{i,3}(x) = 0.
\]
For $\xi_{i+1}$,

$$\lim_{x \to \xi_{i+1}} \frac{d}{dx} B_{i,3}(x) = \lim_{x \to \xi_{i+1}} \left[ \frac{1}{\xi_{i+2} - \xi_{i+1}} \frac{x - \xi_i}{\xi_{i+2} - \xi_i} + \frac{x - \xi_i}{\xi_{i+2} - \xi_{i+1}} \right] = 10,$$

$$\lim_{x \to \xi_{i+1}^+} \frac{d}{dx} B_{i,3}(x) = \lim_{x \to \xi_{i+1}^+} \left[ \frac{(\xi_{i+2} - x) - (x - \xi_i)}{(\xi_{i+2} - \xi_i)(\xi_{i+2} - \xi_{i+1})} + \frac{-(x - \xi_{i+1}) + (\xi_{i+3} - x)}{(\xi_{i+3} - \xi_{i+1})(\xi_{i+2} - \xi_{i+1})} \right] = 10.$$ 

For $\xi_{i+2}$, similarly,

$$\lim_{x \to \xi_{i+2}^-} \frac{d}{dx} B_{i,3}(x) = \lim_{x \to \xi_{i+2}^-} \frac{d}{dx} B_{i,3}(x) = -10.$$

So, the first derivative exists for all $x \in [a, b]$. Therefore, the same conclusion holds for any linear combination of $B_{i,3}(x)$. (The derivatives work for $i = 0, \ldots, 9$, but the same argument can be made for all $i$)

(5 pts) Find $c_i$ such that $q_c(x) = \sum_i c_i B_{i,3}(x)$ is linear to the left of $\xi_1 = 0$. Note that $B_{i,3}(\xi_1^+) = 0$ for $i \neq -2, -1, 0$, so let $c_i = 0$ for the corresponding $i$. For linearity, I only need a constraint on $c_{-2}$, $c_{-1}$, and $c_0$ which require

$$\lim_{x \to \xi_1^-} \frac{d^2}{dx^2} q_c(x) = \lim_{x \to \xi_1^-} \left[ c_{-2} \frac{d^2}{dx^2} B_{-2,3}(x) + c_{-1} \frac{d^2}{dx^2} B_{-1,3}(x) + c_0 \frac{d^2}{dx^2} B_{0,3}(x) \right] \equiv 0.$$ 

Note that only $I_0(x) \neq 0$ for $x \to \xi_1^-$. Then, $\xi_{-1} = \xi_0 = a = -0.1$

$$\lim_{x \to \xi_1^-} \frac{d^2}{dx^2} B_{0,3}(x) = \lim_{x \to \xi_1^-} \frac{2}{(\xi_2 - \xi_0)(\xi_1 - \xi_0)} = 100,$$

$$\lim_{x \to \xi_1^-} \frac{d^2}{dx^2} B_{-1,3}(x) = \frac{-2}{(\xi_1 - \xi_{-1})(\xi_1 - \xi_0)} + \frac{-2}{(\xi_2 - \xi_0)(\xi_1 - \xi_0)} = -300,$$

and

$$\lim_{x \to \xi_1^-} \frac{d^2}{dx^2} B_{-2,3}(x) = \frac{-2}{(\xi_1 - \xi_{-1})(\xi_1 - \xi_0)} = -200.$$ 

So, the constraint is the set $\{c_i\}$ such that $-2c_{-2} - 3c_{-1} + c_0 = 0$ and $c_j = 0$ for $j \neq 0, -1, -2$.

11. (3.23 of HTF) Suppose that columns of $X$ with rank $p$ have been standardized, as has $Y$. Suppose also that

$$\frac{1}{N} |\langle x_j, Y \rangle| = \lambda \quad \forall j = 1, \ldots, p$$

Let $\hat{\beta}^{\text{ols}}$ be the usual least squares coefficient vector and $\hat{Y}^{\text{ols}}$ be the usual projection of $Y$ onto the
column space of $X$. Define $\hat{Y}(\alpha) = \alpha \hat{X} \hat{\beta}^{\text{ols}}$ for $\alpha \in [0, 1]$. Find

$$\frac{1}{N} \left| \langle x_j, Y - \hat{Y}(\alpha) \rangle \right| \quad \forall j = 1, \ldots, p$$

in terms of $\alpha$, $\lambda$, and $(Y - \hat{Y}^{\text{ols}})'(Y - \hat{Y}^{\text{ols}})$. Show this is decreasing in $\alpha$. What is the implication of this as regards the LAR algorithm?

(5 pts) Note that $\hat{Y}^{\text{ols}} = X \beta^{\text{ols}}$ and $\hat{Y}(\alpha) = \alpha \hat{Y}^{\text{ols}}$. For all $j = 1, \ldots, p$,

$$\frac{1}{N} \left| \langle x_j, Y - \hat{Y}(\alpha) \rangle \right| = \frac{1}{N} \left| \langle x_j, Y - \alpha \hat{Y}^{\text{ols}} \rangle \right| = \frac{1}{N} \left| \langle x_j, Y - \hat{Y}^{\text{ols}} + \hat{Y}^{\text{ols}} - \alpha \hat{Y}^{\text{ols}} \rangle \right|$$

($\because x_j \perp Y - \hat{Y}^{\text{ols}}$)

$$= \frac{1}{N} \left| \langle x_j, (1 - \alpha) \hat{Y}^{\text{ols}} \rangle \right| = (1 - \alpha)\lambda.$$

(No graded) See 3.23 of HTF for the complete question. Need to show: the correlation is

$$\lambda(\alpha) = \frac{(1 - \alpha)\lambda}{\sqrt{(1 - \alpha)^2 + \frac{\alpha(2 - \alpha)}{N} RSS}}$$

where $RSS = (Y - \hat{Y}^{\text{ols}})'(Y - \hat{Y}^{\text{ols}})$. Note that $\text{Var}(Y) = \text{Var}(x_j) = 1$ for all $j$ since standardized, and $\text{Cov}(\hat{Y}, Y - \hat{Y}) = 0$ since perpendicular. Then, we have

$$\text{Var}(Y - \hat{Y}(\alpha)) = \text{Var}(Y - \alpha Y + \alpha Y - \alpha \hat{Y})$$

$$= (1 - \alpha)^2 \text{Var}(Y) + \alpha^2 \text{Var}(Y - \hat{Y}) + 2 \alpha (1 - \alpha) \text{Cov}(Y, Y - \hat{Y})$$

$$= (1 - \alpha)^2 + \alpha^2 \frac{RSS}{N} + 2 \alpha (1 - \alpha) \text{Cov}(Y, Y - \hat{Y})$$

$$= (1 - \alpha)^2 + \alpha^2 \frac{RSS}{N} + 2 \alpha (1 - \alpha) [\text{Cov}(Y - \hat{Y}, Y - \hat{Y}) - \text{Cov}(\hat{Y}, Y - \hat{Y})]$$

$$= (1 - \alpha)^2 + \alpha^2 \frac{RSS}{N} + 2 \alpha (1 - \alpha) \text{Var}(Y - \hat{Y})$$

$$= (1 - \alpha)^2 + \alpha (2 - \alpha) \frac{RSS}{N}.$$
Therefore, we have

\[ \lambda(\alpha) = \frac{\text{Cov}(x_j, Y - \hat{Y}(\alpha))}{\sqrt{\text{Var}(x_j)\text{Var}(Y - \hat{Y}(\alpha))}} = \frac{(1 - \alpha)\lambda}{\sqrt{(1 - \alpha)^2 + \frac{\alpha(2 - \alpha)}{N}\text{RSS}}} \]

(No graded) Need to show: \( \lambda(\alpha) \) is decreased in \( \alpha \).

\[ \frac{d}{d\alpha} \lambda(\alpha) = -\lambda \frac{1}{N} \text{RSS} < 0. \]

(No graded) The LAR algorithm along the selection process keeps the correlation of residuals and variables in the active set decreasing, and keep the correlation tied/equal for all variables in the active set.

12. a) Suppose that \( a < x_1 < x_2 < \cdots < x_N < b \) and \( s(x) \) is a natural cubic spline with knots at the \( x_i \) interpolating the points \( (x_i, y_i) \) (i.e. \( s(x_i) = y_i \)). Let \( z(x) \) be any twice continuously differentiable function on \( [a, b] \) also interpolating the points \( (x_i, y_i) \). Then

\[ \int_a^b (s''(x))^2dx \leq \int_a^b (z''(x))^2dx \]

(Hint: Consider \( d(x) = z(x) - s(x) \), write

\[ \int_a^b (d''(x))^2dx = \int_a^b (z''(x))^2dx - \int_a^b (s''(x))^2dx - 2 \int_a^b s''(x)d'(x)dx \]

and use integration by parts and the fact that \( s'''(x) \) is piecewise constant.)

(5 pts) By the hint, I need to show \( \int_a^b s''(x)d'(x)dx = 0 \). Let \( x_0 = a, x_{N+1} = b \), and \( s_i = s''(x) \)
for \( x \in [x_i, x_{i+1}] \). Also, we have \( d(x_i) = z(x_i) - s(x_i) = y_i - y_i = 0 \).
\[
\int_a^b s''(x)d''(x)dx = s''(x)d'(x)|_a^b - \int_a^b s'''(x)d'(x)dx
\]
(\( \because s''(a) = s''(b) = 0 \))
\[
= \sum_{i=0}^N \int_{x_i}^{x_{i+1}} s'''(x)d'(x)dx
\]
\[
= \sum_{i=0}^N s_i \int_{x_i}^{x_{i+1}} d'(x)dx
\]
\[
= \sum_{i=0}^N s_i[d(x_{i+1}) - d(x_i)]
\]
\[
= 0
\]

Therefore, \( \int_a^b (d''(x))^2dx = \int_a^b (z''(x))^2dx - \int_a^b (s''(x))^2dx \) and \( \int_a^b (d'(x))^2dx \geq 0 \), so \( \int_a^b (s''(x))^2dx \leq \int_a^b (z''(x))^2dx \).

b) Use a) and prove that the minimizer of \( \sum_{i=1}^N (y_i - h(x_i))^2 + \lambda \int_a^b (h''(x))^2dx \) over the set of twice continuously differentiable functions on \([a, b]\) is a natural cubic spline with knots at the \( x_i \).

(5 pts) Let \( C^2[a, b] \) be the set of twice continuously differentiable functions on \([a, b]\). Assume \( z(x) \in C^2[a, b] \) is the minimizer rather than \( s(x) \). i.e.
\[
\sum_{i=1}^N (y_i - z(x_i))^2 + \lambda \int_a^b (z''(x))^2dx \leq \sum_{i=1}^N (y_i - h(x_i))^2 + \lambda \int_a^b (h''(x))^2dx
\]
for all \( h(x) \in C^2[a, b] \). But from a), I have \( s(x_i) = z(x_i) \) and \( \int_a^b (s''(x))^2dx \leq \int_a^b (z''(x))^2dx \) which yield
\[
\sum_{i=1}^N (y_i - s(x_i))^2 + \lambda \int_a^b (s''(x))^2dx \leq \sum_{i=1}^N (y_i - z(x_i))^2 + \lambda \int_a^b (z''(x))^2dx
\]
and contradict to the assumption. Therefore, \( s(x) \) has to be the minimizer.

13. Let \( \mathcal{H} \) be the set of absolutely continuous functions on \([0, 1]\) with square integrable first
derivatives (that exist except possibly at a set of measure 0). Equip \( \mathcal{H} \) with an inner product

\[
\langle f, g \rangle_{\mathcal{H}} = f(0)g(0) + \int_0^1 f'(x)g'(x)dx
\]

a) Show that

\[
R(x, z) = 1 + \min(x, z)
\]

is a reproducing kernel for this Hilbert space.

(5 pts)

\[
\langle R(x, z), f(z) \rangle_{\mathcal{H}} = R(x, 0)f(0) + \int_0^1 \frac{d}{dz}R(x, z)\frac{d}{dz}f(z)dz
\]

\[
= f(0) + \int_x^1 \frac{d}{dz}R(x, z)\frac{d}{dz}f(z)dz + \int_x^1 \frac{d}{dz}R(x, z)\frac{d}{dz}f(z)dz
\]

\[
= f(0) + \int_x^1 f'(z)dz + 0
\]

\[
= f(x)
\]

b) Using Heckman’s development, describe as completely as possible,

\[
\arg\min_{h \in \mathcal{H}} \left( \sum_{i=1}^N (y_i - h(x_i))^2 + \lambda \int_0^1 (h'(x))^2 dx \right)
\]

(10 pts) The followings are based on Heckman’s notion.

As the Equation (2), we can define \( \mu \equiv h(x) \), \( F_i(\mu) = \mu \) (i.e. \( F_i \) is an identity), and \( L\mu(x) \equiv h'(x) \), then this question fit in the general penalized least squares problem. This \( L\mu(x) \) is also of the form of the Equation (3). Therefore, by the Theorem 3, the minimizer exists and is of the form of the Equation (7). For this question, the solution is of the form \( \hat{\mu}(x) = \hat{\alpha}_1 u_1(x) + \sum_{i=1}^N \hat{\beta}_i \eta_i(x) \) for the given inner product. For finding \( \hat{\mu}(x) \), we can utilize the Equation (8) and it’s solutions, the Equation (11) and (12), but we have to change notations and dimensions.

For \( R_1(x, x_i) \), we already did in the part a).
For \( \eta_1(x) \), by the definition of the Theorem 3,
\[
\eta_1(x) = F_i(R_1(x, x_i)) = R_1(x, x_i) = \min(x, x_i).
\]
This can be verified from the Green’s function: \( G(x, z) = 1 \) for \( x \leq z \), 0 else. Also, \( R_1(x, x_i) = \int_0^1 G(x, u)G(x_i, u)du \).

For \( u_1(x) \), the Theorem 5 gives a way to obtain a basis for kernel of \( L \), and an example is given in Heckman’s page 10. We have \( m = 1 \) and \( L\mu(x) = h'(x) \), and the solution of \( x = 0 \) is \( r_1 = 0 \). Therefore,
\[
u_1(x) = \exp(r_1x) = 1.
\]

For \( \hat{\alpha} \) and \( \hat{\beta} \), as given in the Equation (4), we solve
\[
(Y - T\alpha - K\beta)'(Y - T\alpha - K\beta) + \lambda\beta'K\beta
\]
where \( \alpha = \alpha_0, \beta = (\beta_1, \ldots, \beta_n)' \), \( Y = (Y_1, \ldots, Y_n)' \), \( T_{i1} = 1 \), \( i = 1, \ldots, n \), \( K_{ij} = R_{1x_j}(x_i) \), \( i, j = 1, \ldots, n \), and \( R_{1x_j}(x_i) = \min(x_i, x_j) \) as given in part a). (i.e. \( K_{ij} = (\min(x_i, x_j))_{N \times N} \)).

\( D = I \) is an identity matrix in this setup.

Follow the same derivative in Heckman’s page 15, we have
\[
M = K + \lambda I
\]
and from the Equations (11) and (12)
\[
\hat{\alpha} = (T'M^{-1}T)^{-1}T'M^{-1}Y
\]
and
\[
\hat{\beta} = M^{-1}[I - T(T'M^{-1}T)^{-1}T'M^{-1}]Y.
\]

So, the solution is
\[
\hat{\mu}(x) = \hat{\alpha}_1u_1(x) + \sum_{i=1}^N \hat{\beta}_i \eta_1(x)
\]
\[
= \arg\min_{h \in \mathcal{H}} \left( \sum_{i=1}^N (y_i - h(x_i))^2 + \lambda \int_0^1 (h'(x))^2 dx \right).
\]
c) Using Heckman’s development, describe as completely as possible,

\[
\arg\min_{h \in H} \left( \sum_{i=1}^{N} (y_i - \int_0^{x_i} h(t)dt)^2 + \lambda \int_0^1 (h'(x))^2 dx \right)
\]

(10 pts) Similarly to the part b) and the example in Heckman’s page 16, we have to rewrite
the notation carefully. Let \( F_i(\mu(x)) \equiv \int_0^{x_i} h(x)dx \) (i.e. \( f_i(x) = 1 \)) and \( L\mu(x) \equiv h'(x) \).

For A, \( R_1(x, z) = \min(x, z) \) is the same in the part a).

For B,

\[
\eta_{i1}(x) = F_i(R_1(x, z)) = \int_0^{x_i} R_1(x, z)dz = \frac{1}{2} x_i^2 I(x \geq x_i) + \left[ xx_i - \frac{1}{2} x^2 \right] I(x < x_i).
\]

For C, \( u_1 = 1 \) as given in the part b),

\[ T_{i1} = F_i(u_1) = x_i \]

and

\[
K_{ij} = F_i(\eta_{j1}(x)) = \int_0^{x_i} \eta_{j1}(x)dx
\]

\[
= \int_0^{x_i} \frac{1}{2} x_j^2 I(x \geq x_j) + \left[ xx_j - \frac{1}{2} x^2 \right] I(x < x_j)dx
\]

\[
= \frac{1}{2} x_j^2 (x_i - x_j)I(x_i \geq x_j) + \left[ \left( \frac{1}{2} x_i^2 x_j - \frac{1}{6} x_i^3 \right) I[x_i < x_j] + \left( \frac{1}{2} x_j^3 - \frac{1}{6} x_j^3 \right) I[x_i \geq x_j] \right]
\]

\[
= \frac{1}{6} x_j^2 (3x_i - x_j)I(x_i \geq x_j) + \frac{1}{6} x_i^2 (3x_j - x_i)I(x_i < x_j).
\]

Note that \( K_{ij} = K_{ji} \).

For D, \( \hat{\alpha} \) and \( \hat{\beta} \) are the same formula as the part b) but substitute \( T, K \) and \( M \) by the
above settings A, B, and C.
For E, the solution is similarly

\[ \hat{\mu}(x) = \hat{\alpha}_1 u_1(x) + \sum_{i=1}^{N} \hat{\beta}_i \eta_i(x) \]

\[ = \arg\min_{h \in \mathcal{H}} \left( \sum_{i=1}^{N} (y_i - h(x_i))^2 + \lambda \int_{0}^{1} (h'(x))^2 dx \right). \]

with substitution by the above settings from A to E.

Total: 75 pts (Q9: 10, Q10: 20, Q11: 10, Q12:10, Q13:25). Any resonable solutions are acceptable.