Semiparametric regression - Homework 3

Klaudia Weigel

1 Exercise 1

We define the following functions on $[0, 1]$:

$$T_1(x) = 1, \qquad T_2(x) = x, \qquad T_3(x) = (x - \frac{1}{2})_+ = \max(x - \frac{1}{2}, 0),$$

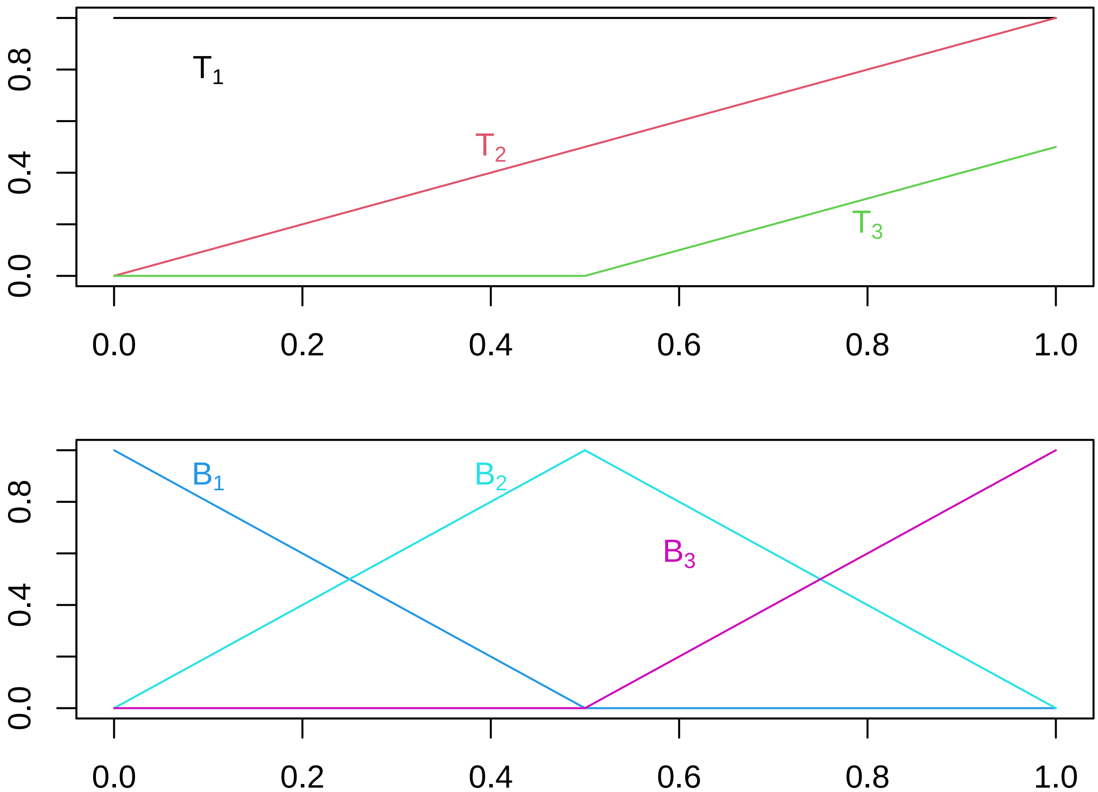
$$B_1(x) = (1 - 2x)_+, \qquad B_2(x) = 1 - |2x - 1|, \qquad B_3(x) = (2x - 1)_+,$$

where $x \in \mathbb{R}$.

1.1 (a)

We will plot the previously defined functions.

```
ng <- 101
xg <- seq(0, 1, length=ng)
T1g <- rep(1, ng)
T2g <- xg
T3g <- (xg - .5)*(xg - .5>0)
B1g <- (1 - 2*xg)*(1 - 2*xg>0)
B2g <- 1 - abs(2*xg - 1)
B3g <- 2*T3g
par(mfrow=c(2,1), mar=c(2, 4, 2, 2))
plot(0, type = "n", xlim=c(0,1), ylim=c(0,1), xlab="x", ylab="") #, bty="1")
lines(xg, T1g, col=1)
lines(xg, T2g, col=2)
lines(xg, T3g, col=3)
text(0.1, 0.8, expression(T[1]), col=1)
text(0.4, 0.5, expression(T[2]), col=2)
text(0.8, 0.2, expression(T[3]), col=3)
plot (0, type = "n", xlim=c(0,1), ylim=c(0,1), xlab="x", ylab="") #, bty="1")
lines(xg, B1g, col=4)
lines(xg, B2g, col=5)
lines(xg, B3g, col=6)
text(0.1, 0.9, expression(B[1]), col=4)
text(0.4, 0.9, expression(B[2]), col=5)
text(0.6, 0.6, expression(B[3]), col=6)
```



1.2 (b)

We will now find expressions for B_1, B_2, B_3 in terms of T_1, T_2 and T_3.

$$
\begin{aligned}
B_1(x) &= \max(1-2x, 0) = 1-2x-(1-2x)\mathbb{I}_{\{x>1/2\}} = 1-2x+2(x-\frac{1}{2})\mathbb{I}_{\{x>1/2\}} \\
&= 1-2x+2\max(x-\frac{1}{2}, 0) = T_1(x) - 2T_2(x) + 2T_3(x).
\end{aligned}
$$

$$
\begin{aligned}
B_2(x) &= 1-|2x-1| = 1+2x-1-2(2x-1)\mathbb{I}_{\{x>1/2\}} = 2x-4(x-\frac{1}{2})\mathbb{I}_{\{x>1/2\}} \\
&= 2x-4\max(x-\frac{1}{2}, 0) = 2T_2(x) - 4T_3(x),
\end{aligned}
$$

$$
B_3(x) = \max(2x-1, 0) = 2\max(x-\frac{1}{2}, 0) = 2T_3(x).
$$

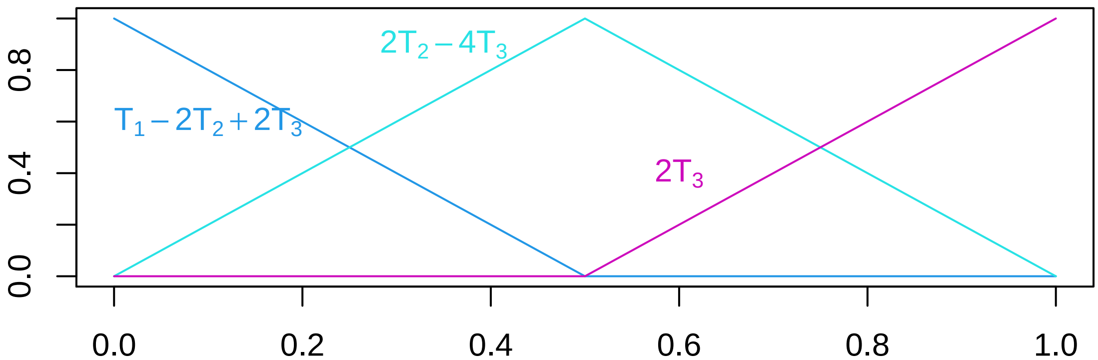
Next we'll check what $B_1 + B_2 + B_3$ is:

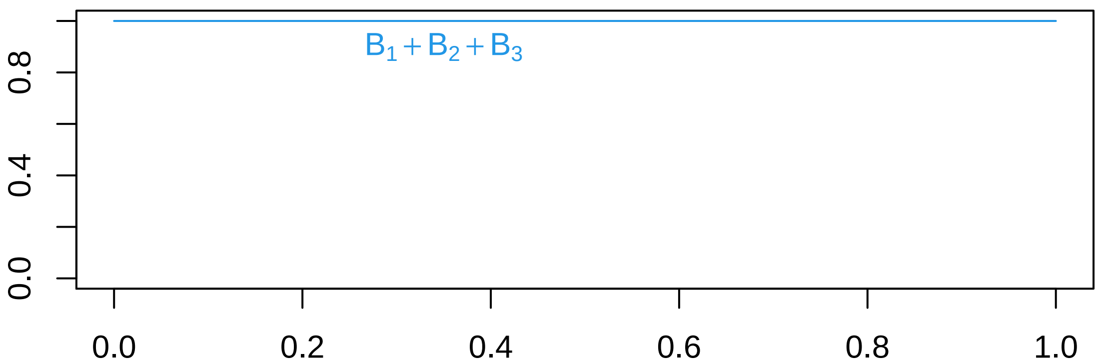
$$
B_1 + B_2 + B_3 = T_1 - 2T_2 + 2T_3 + 2T_2 - 4T_3 + 2T_3 = T_1 = 1.
$$

We can check the results with a plot.

```
par(mfrow=c(2,1), mar=c(2, 4, 2, 2))
plot (0, type = "n", xlim=c(0,1), ylim=c(0,1), xlab="x", ylab="") #, bty="1")
lines(xg, T1g - 2*T2g + 2*T3g, col=4)
lines(xg, 2*T2g - 4*T3g, col=5)
lines(xg, 2*T3g, col=6)
text(0.1, 0.6, expression(T[1] - 2*T[2] + 2*T[3]), col=4)
```

```r
text(0.35, 0.9, expression(2*T[2] - 4*T[3]), col=5)
text(0.6, 0.4, expression(2*T[3]), col=6)
plot (0, type = "n", xlim=c(0,1), ylim=c(0,1), xlab="x", ylab="") #, bty="1")
lines(xg, B1g + B2g + B3g, col=4)
text(0.35, 0.9, expression(B[1] + B[2] + B[3]), col=4)
```





1.3 (c)

We will obtain 3×3 matrix L_{TB} such that

$$[B_1(x) \quad B_2(x) \quad B_3(x)] = [T_1(x) \quad T_2(x) \quad T_3(x)]L_{TB}$$

for any $x \in [0,1]$.

We have:

$$[B_1 \quad B_2 \quad B_3] = [T_1 - 2T_2 + 2T_3 \quad 2T_2 - 4T_3 \quad 2T_3] = [T_1 \quad T_2 \quad T_3] \begin{bmatrix} 1 & 0 & 0 \\ -2 & 2 & 0 \\ 2 & -4 & 2 \end{bmatrix}.$$

Therefore

$$L_{TB} = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 2 & 0 \\ 2 & -4 & 2 \end{bmatrix}.$$

1.4 (d)

```r
LTB <- matrix(c(1,0,0,  -2,2,0,  2,-4,2), byrow = TRUE, nrow = 3)
det(LTB)
```

```
## [1] 4
```

The determinant of L_{TB} is equal to 4, therefore the matrix is invertible.

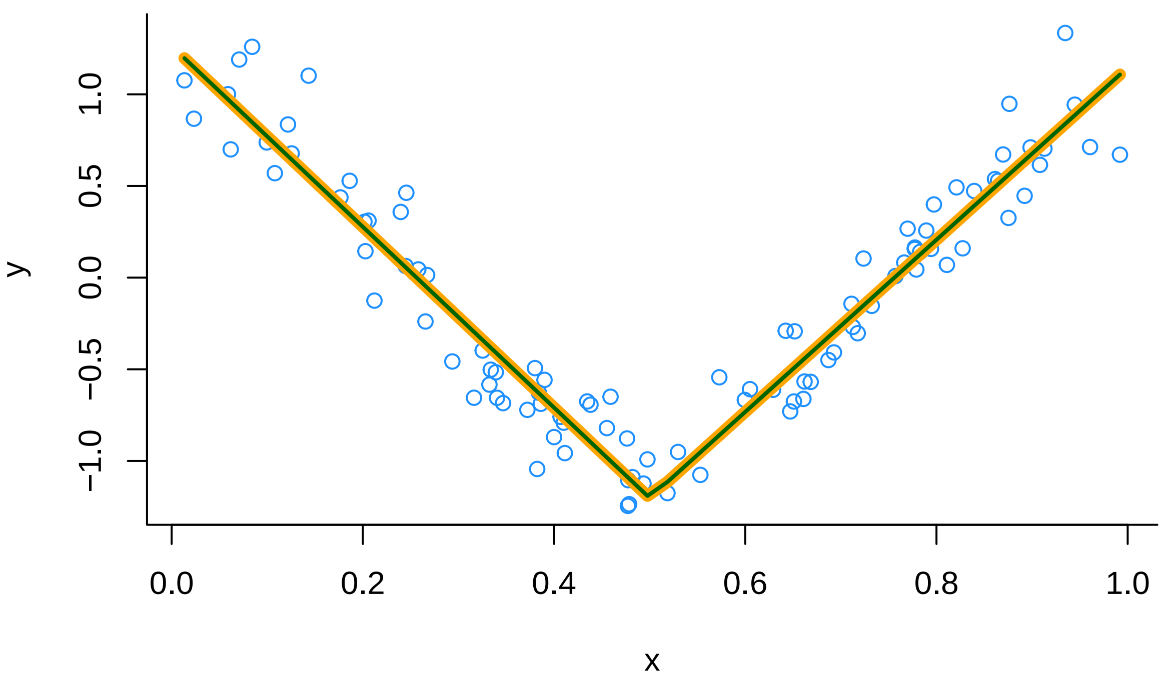
1.5 (e)

Let

$$X_T = \begin{bmatrix} T_1(x_1) & T_2(x_1) & T_3(x_1) \\ \vdots & \vdots & \vdots \\ T_1(x_n) & T_2(x_n) & T_3(x_n) \end{bmatrix}, \qquad X_T = \begin{bmatrix} B_1(x_1) & B_2(x_1) & B_3(x_1) \\ \vdots & \vdots & \vdots \\ B_1(x_n) & B_2(x_n) & B_3(x_n) \end{bmatrix}$$

be design matrices. We will check if the fit obtained using matrix X_T is the same as that with X_B.

```r
set.seed(1)
n <- 100
x <- sort(runif(n))
y <- cos(2*pi*x) + 0.2*rnorm(n)
plot(x,y,col='dodgerblue', bty='l')
XT <- cbind(rep(1,n), x, (x - .5)*(x - .5 > 0))
XB <- cbind((1 - 2*x)*(1 - 2*x > 0), 1 - abs(2*x - 1), (2*x - 1)*(2*x - 1 > 0))
fitT <- lm(y~XT-1)
fitB <- lm(y~XB-1)
lines(x, fitted(fitT), col = 'orange', lwd = 6)
lines(x, fitted(fitB), col = 'darkgreen', lwd = 2)
```



We see that both fitted lines are the same.