Semiparametic Regression - Assignment 3

Szymon Armata | 341593 | Data Science

2 November 2022

Contents

1 Task 1.

We will begin this exercise by performing some linear algebra, including basis, determinants, etc. Then, we will implement OLS to determine a sample's regression coefficient.

Let's define following functions on $[1, 0]$

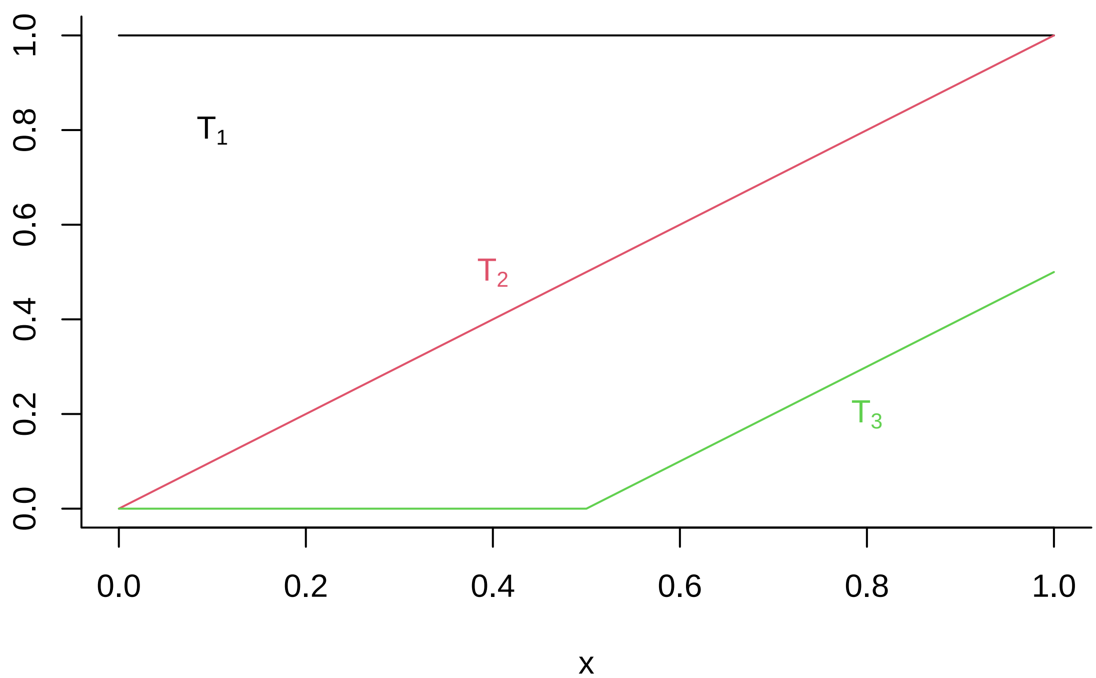
$$T_1(x) = 1, \qquad T_2(x) = x, \qquad T_3(x) = \left(x - \frac{1}{2}\right)_+$$

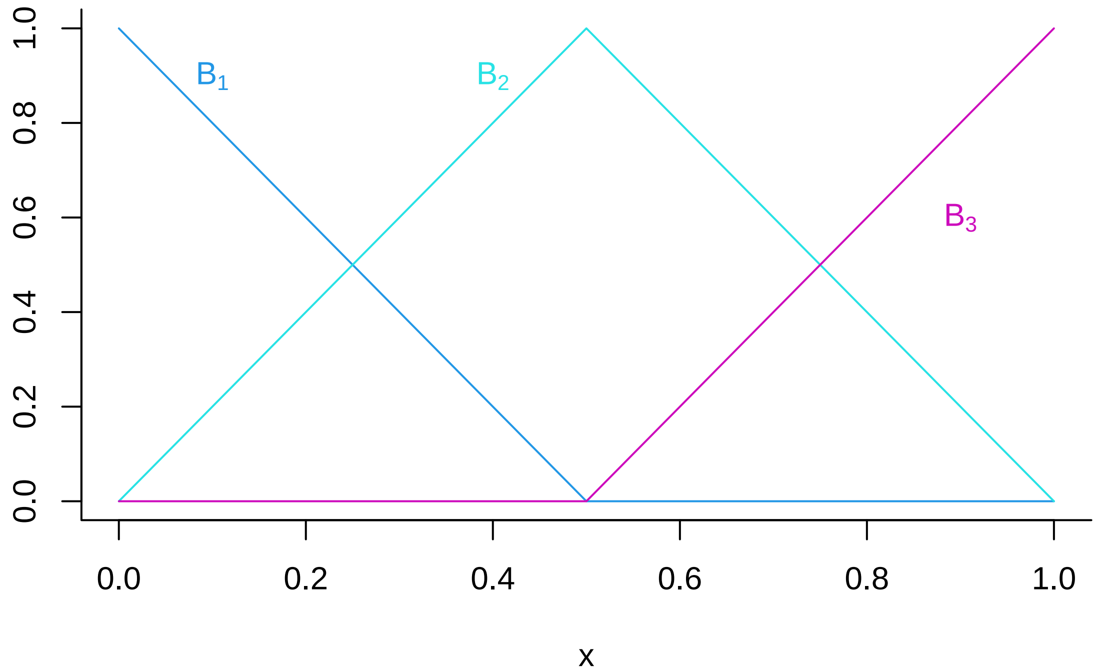
and

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

1.1 a

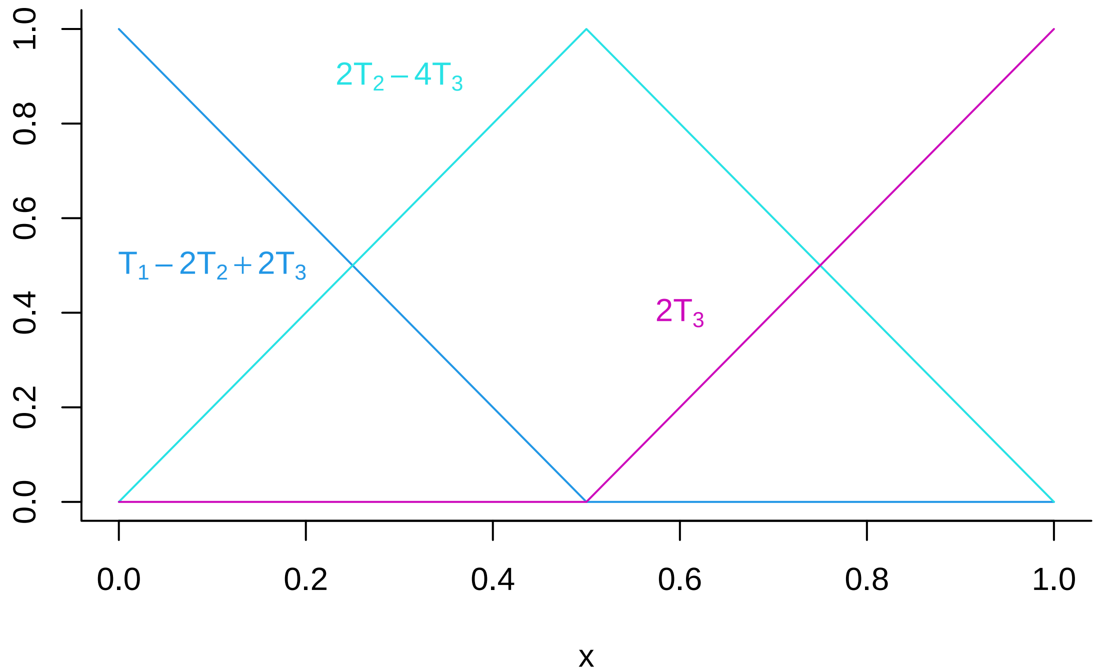
At first we have to obtain plots of the T_i and B_i, $i = 1, 2, 3$.




B1
B2
B3
0.0
0.2
0.4
0.6
0.8
1.0
x

1.2 b

Next, we have to find expressions for B_1, B_2, B_3 in terms of T_1, T_2, T_3.



So here are the representations we developed:

$$B_1 = T_1 - 2(T_2 - T_3)$$

$$B_2 = 2(T_2 - 2T_3)$$

$$B_3 = 2T_3$$

Now we have to find the result of this equation: $B_1 + B_2 + B_3$.

$$B_1 + B_2 + B_3 = T_1 - 2T_2 + 2T_3 + 2(T_2 - 2T_3) + 2T_3 = T_1 = 1 \tag{1}$$

1.3 c

To find matrix L_{TB}

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

we have to solve equation like this:

$$[T_1 - 2(T_2 - T_3) \quad 2(T_2 - 2T_3) \quad 2T_3] = [T_1 \quad T_2 \quad T_3]\, L_{TB}$$

One of the solutions of this equation is following matrix:

$$\begin{bmatrix} 1 & 0 & 0 \\ -2 & 2 & 0 \\ 2 & -4 & 2 \end{bmatrix}$$

1.4 d

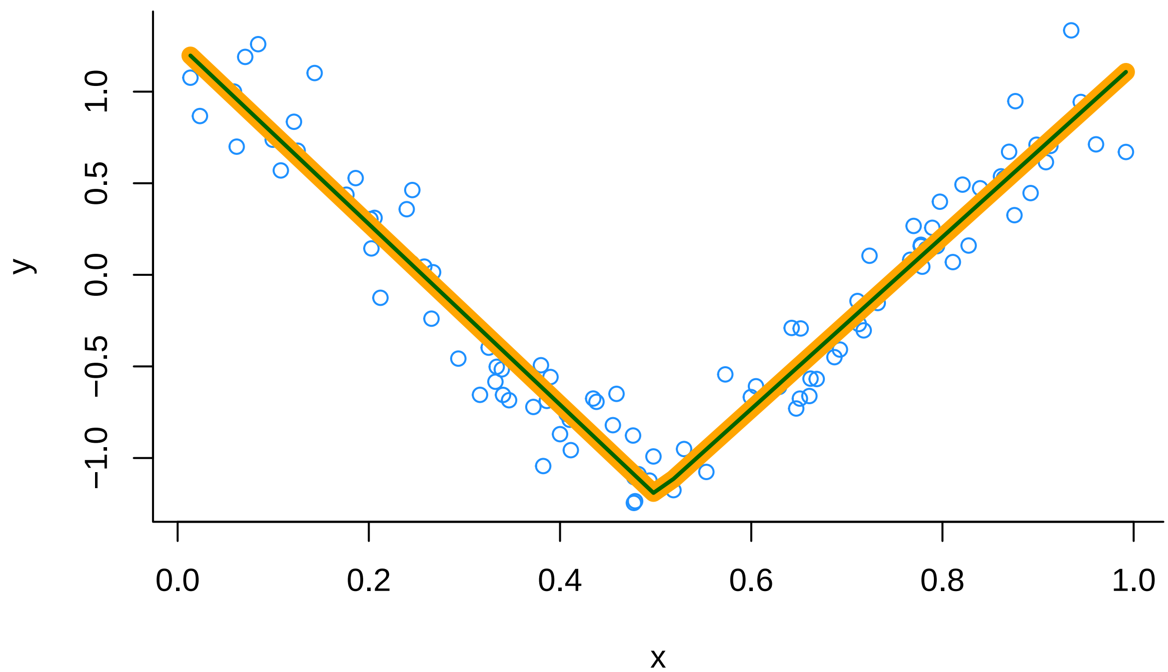
Now we have to find a determinant of L_TB and establish that L_TB is invertible. In linear algebra language, this implies that $\{B_1, B_2, B_3\}$ is an alternative basis for the vector space of functions spanned by $\{T_1, T_2, T_3\}$

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

Determinant of L_TB is not equal to 0, which means that matrix L_TB is invertible.

1.5 e

Below we compare two models, first build using basis $\{T_1, T_2, T_3\}$, another using $\{B_1, B_2, B_3\}$. We expect that they are visually undifferentiated.



As seen above, fitted lines overlap each other. It confirms that ordinary least squares regression with design matrix X_T leads to same fit as that with design matrix X_B.