Statistical Learning

Group 19 - Le superchicche

Cammarota Sara, Iadisernia Giulia, Petrucci Ilaria

May 15th, 2023

Part I

1.1 "Be linear in transformed feature space"

bla bla bla

1.2 Truncated power basis functions

Truncated power basis functions $G_{d,q}$ are defined as $\{g_1(x) = 1, g_2(x) = x, ..., g_{d+1}(x) = x^d\}$ and $\{g_{(d+1)+j}(x) = (x - \xi_j)_+^d\}_{j=1}^q$ where $(x)_+ = max\{0, x\}$
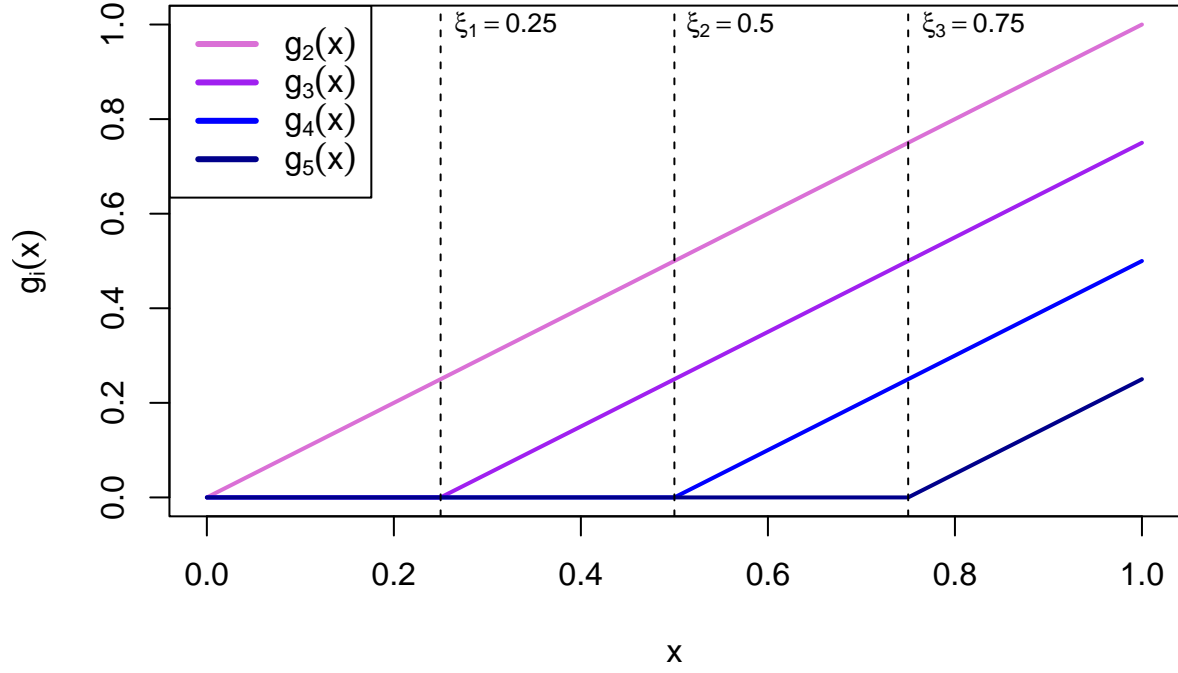
We choose *degree* $d \in \{1, 3\}$ and the number of equi-spaced *knots* $q \in \{3, 10\}$ where $\{\xi_j\}_{j=1}^q \in (0, 1)$.

1.2.1 Our implementation vs. . .

Let's suppose $d = 1$ and $q = 3$

$\{g_i(x)\}_{i=1}^2 = \{g_1(x) = 1, g_2(x) = x\}$ and $\{g_{2+j}(x)\}_{j=1}^3 = \{g_3(x) = (x - \xi_1)_+, g_4(x) = (x - \xi_2)_+, g_5(x) = (x - \xi_3)_+\}$ where the knots $\{\xi_j\}_{j=1}^3$ are respectively equal to $0.25, 0.50, 0.75$.
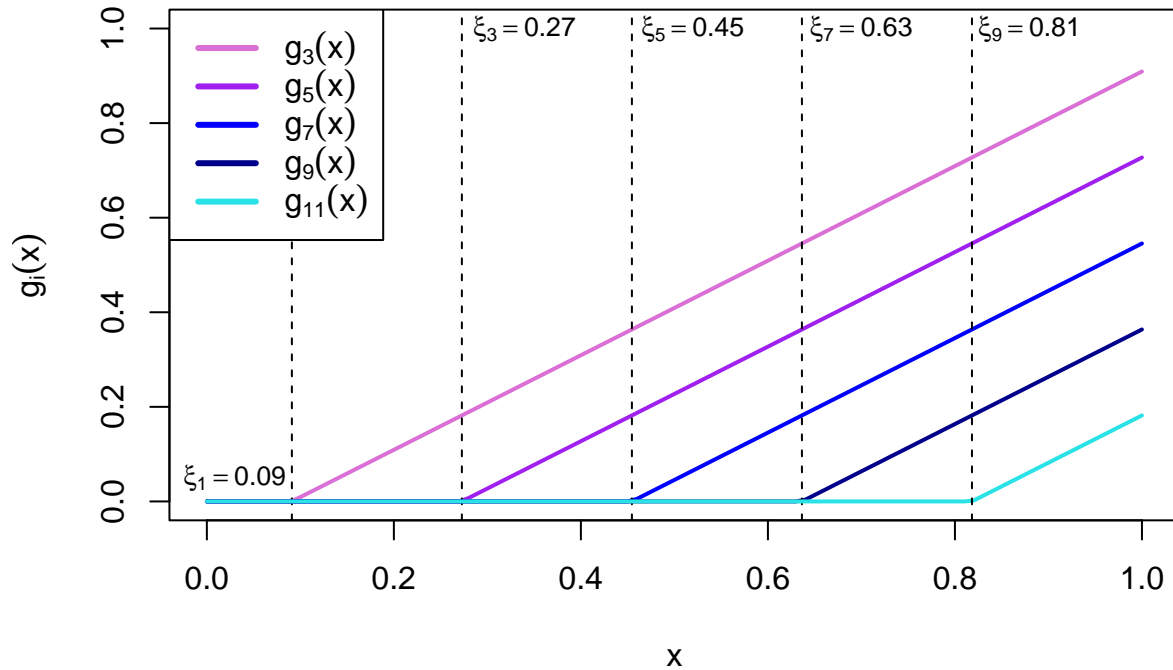
Truncated power basis with d=1 & q=3

When $d = 1$ and $q = 10$

$$\{g_i(x)\}_{i=1}^2 = \{g_1(x) = 1, g_2(x) = x\} \quad \text{and} \quad \{g_{2+j}(x)\}_{j=1}^{10}$$

where the knots $\{\xi_j\}_{j=1}^3$ are respectively equal to $0.09, 0.18, 0.27, 0.36, 0.45, 0.54, 0.63, 0.72, 0.81, 0.91$. In the figure below we plot some of these basis functions.

Truncated power basis with d=1 & q=10

$g_3(x)$
$g_5(x)$
$g_7(x)$
$g_9(x)$
$g_{11}(x)$

$\xi_1 = 0.09$
$\xi_3 = 0.27$
$\xi_5 = 0.45$
$\xi_7 = 0.63$
$\xi_9 = 0.81$

$g_i(x)$

x

Now, let's have a look at what changes when we increase the degree to $d = 3$. For simplicity, let's choose $q = 3$.

rifare plots con x, x**2, x**3 + i tre knots

1.2.2 . . . ChatGPT's reply

1.3 Our very first Kaggle competition

```
library(caret)
library(glmnet)
library(elasticnet)
library(doParallel)
library(dplyr)
```

1.3.1 Import modules

3

```
train <- read.csv('train.csv')
test <- read.csv('test.csv')
```

1.3.2 The `wmap` dataset

Part II