SDLU simulation on different number of nodes with residual fitting

Hyewon Lee

SDM LAB

March 19, 2021

Contents

- Review and Goal
- Settings for simulation
- Residual fitting process
- Test with different number of nodes
- To-do

Review and Goal

- Last meeting
  - Conducted SDLU simulation on different number of nodes with constraint on width of activation function.
- Goal for today's meeting
  - SDLU simulation with residual fitting on different number of nodes and compare results.

Settings for simulation

- Size of data: 500
- x: 500 data randomly chosen from uniform distribution $U(0,1)$
- f(true function): $2.5x + 0.2\sin((3x)^2)$
- y: $f + 0.1\varepsilon$ where $\varepsilon \sim N(0,1)$
- Widths: Gradually decrease the width of activation function. $([1,2] \rightarrow [0.5,1] \rightarrow [0.25,1] \rightarrow \cdots)$
- No Lasso penalty given

$$f(x) = 2.5x + 0.2\mathsf{sin}((3x)^2)$$
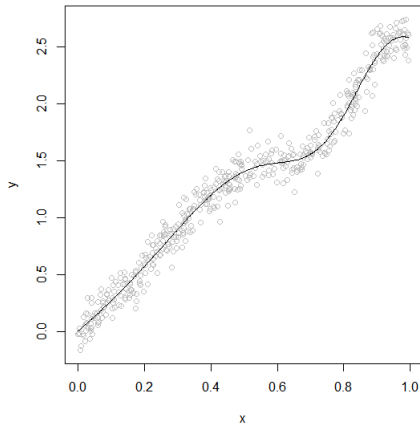
Residual fitting process

- Sequentially fit model to residuals of previous model in decreasing order with respect to the width of nodes

$$y \sim x \rightarrow f_1$$
$$y - f_1 \sim x \rightarrow f_2$$
$$(y - f_1) - f_2 \sim x \rightarrow f_3$$
$$\vdots$$
$$(y - f_1 - f_2 - f_3 - f_4) - f_5 \sim x \rightarrow f_6$$

$$f = f_1 + \cdots + f_6$$

First Simulation

The number of nodes assigned to each step:

- Baseline: 4   constraint : $[1, 2]$
- 1st residual fitting: 8   constraint : $[1, 2]/2$
- 2nd residual fitting: 16   constraint : $[1, 2]/4$
- 3rd residual fitting: 32   constraint : $[1, 2]/8$
- 4th residual fitting: 64   constraint : $[1, 2]/16$
- 5th residual fitting: 128   constraint : $[1, 2]/32$
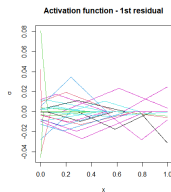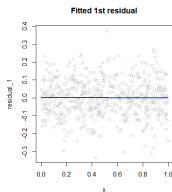
Total: 252 nodes

Simulation animation

Second Simulation

The number of nodes assigned to each step:

- Baseline: 42   constraint : $[1, 2]$
- 1st residual fitting: 42   constraint : $[1, 2]/2$
- 2nd residual fitting: 42   constraint : $[1, 2]/4$
- 3rd residual fitting: 42   constraint : $[1, 2]/8$
- 4th residual fitting: 42   constraint : $[1, 2]/16$
- 5th residual fitting: 42   constraint : $[1, 2]/32$

Total: 252 nodes

Simulation animation

Figure: (a) Total 12 nodes(8 nodes) (b) Total 84 nodes(42 nodes)

(a)   (b)

Figure: (a) Total 12 nodes(8 nodes) (b) Total 84 nodes(42 nodes)

(a)

(b)

Comparison of residual activation function - width $[1,2]/32$

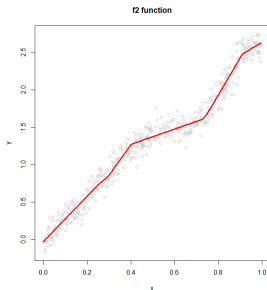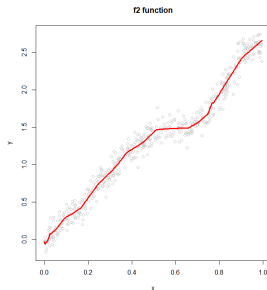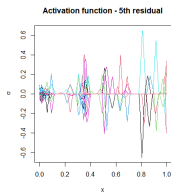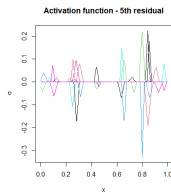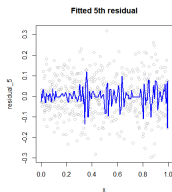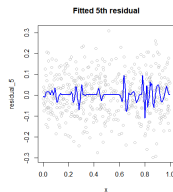Figure: (a) Total 256 nodes(128 nodes) (b) Total 256 nodes(42 nodes)
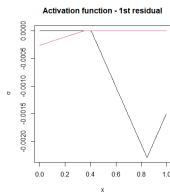
(a)   (b)
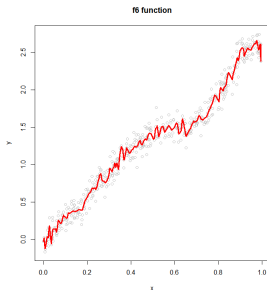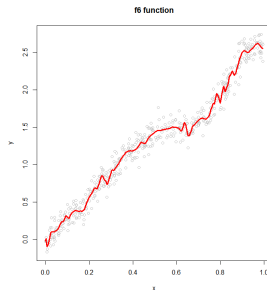
Comparison of fit - width $[1,2]/32$

Figure: (a) Total 256 nodes(128 nodes) (b) Total 256 nodes(42 nodes)

(a)   (b)

Results summary

- Large width
  - No need to use large number of nodes to fit residuals when the width of activation function is big.
- Narrow width
  - Well captured overall locality of data when many nodes are used to fit residuals.