SDLU simulation on different number of nodes with constraint

Hyewon Lee

SDM LAB

March 12, 2021

Contents

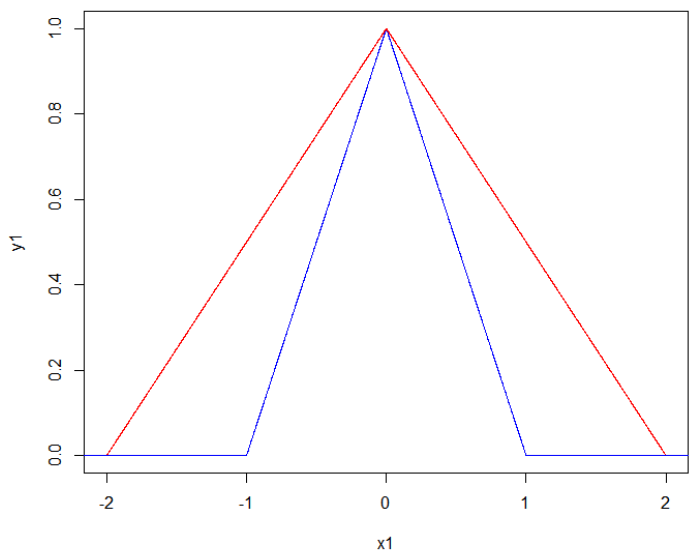
- Background information about SDLU
- Settings for simulation
- Test with different number of nodes
- To-do

Background information about SDLU

- SDLU model for estimating true function f

$$\mathrm{f}(x;\theta) = \beta_0 + \sum_{m=1}^{M} \beta_m \sigma(\alpha_{0m} + \alpha_{1m} x)$$

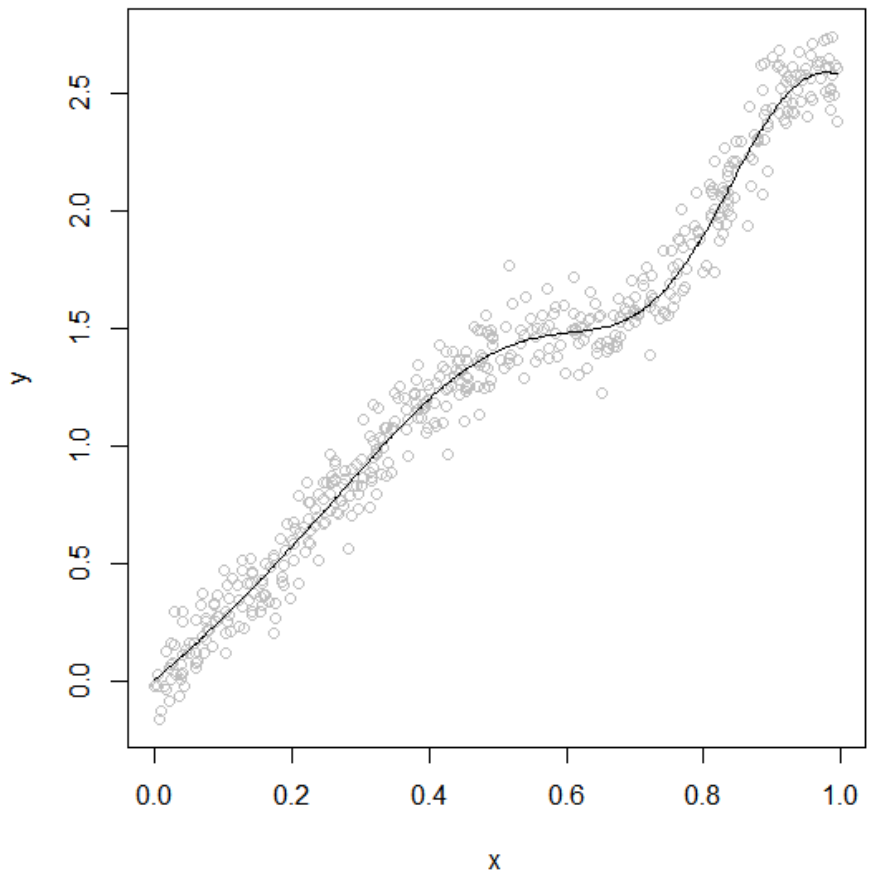
  where $x \in \mathbb{R}$
- M: the number of nodes
- $\frac{2}{|\alpha_{1m}|}$: width of B-spline activation function



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 $\varepsilon \sim N(0,1)$
- List of number of nodes: $4, 8, 16, 32, 64, 128$
- Widths: Generate widths from uniform distribution $U(1,2), U(0.25, 0.5), U(0.0625, 0.125)$
- No Lasso penalty given

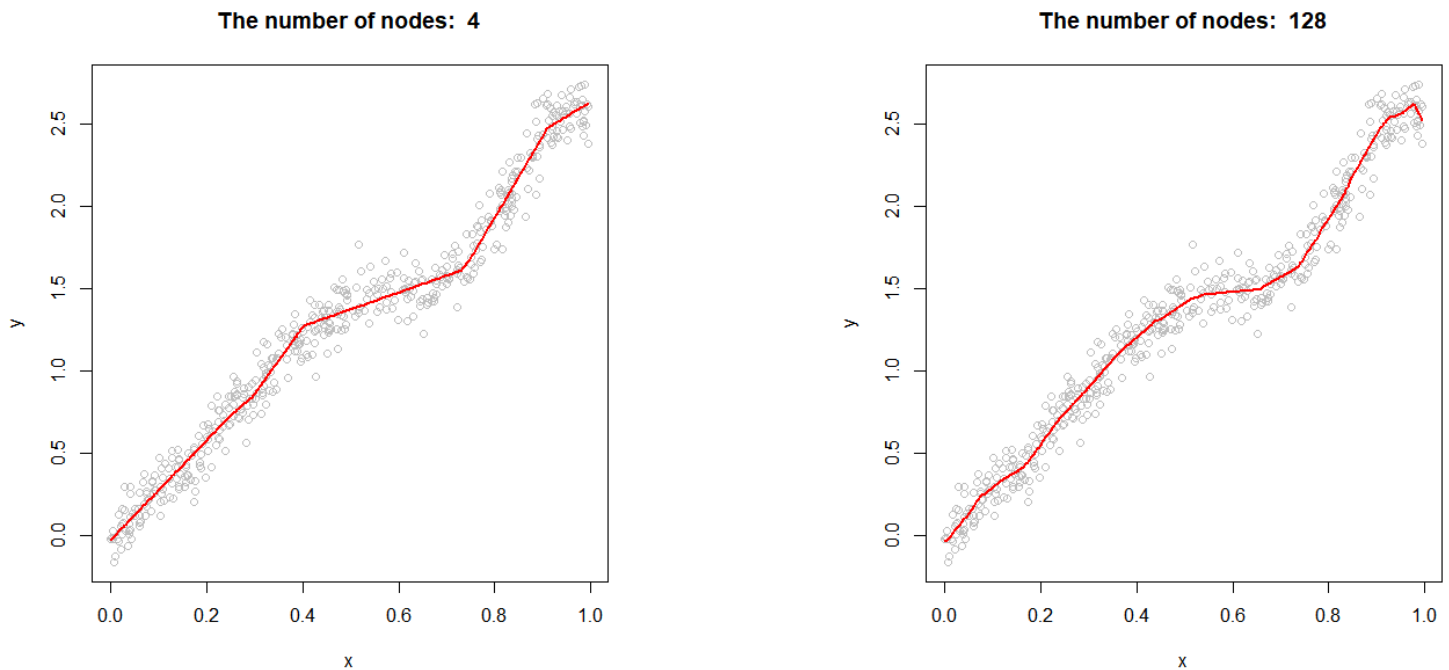
$$f(x) = 2.5x + 0.2\sin((3x)^2)$$



Test with width [1,2]

- For all number of nodes(4, 8, 16, 32, 64, 128), estimated function captures global trend of data.
- There is no such number of nodes that captures locality of data.

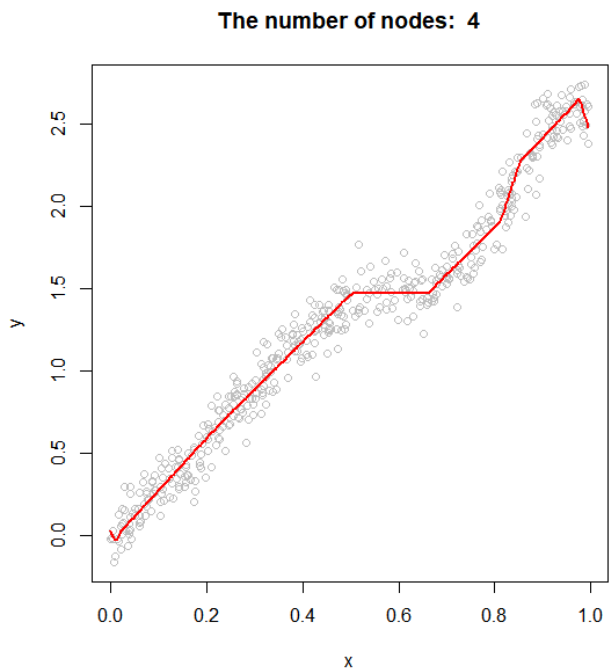
Figure: Comparison of fit depending on the number of nodes

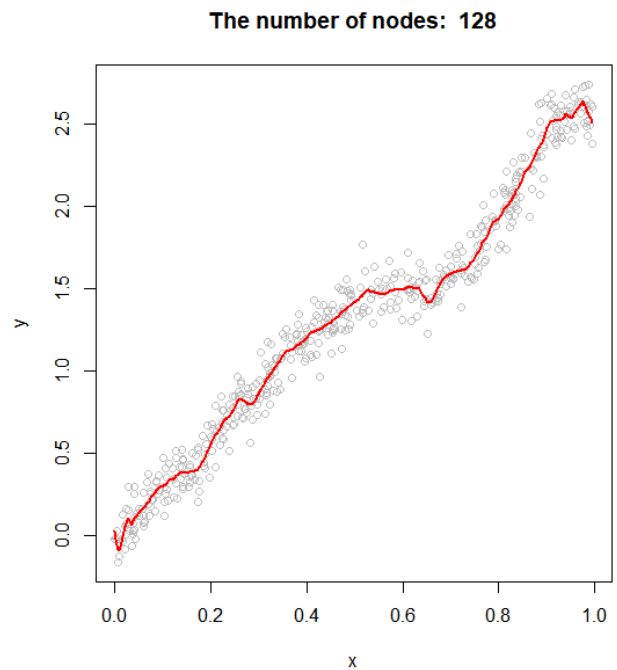


Test with width $[0.25, 0.5]$

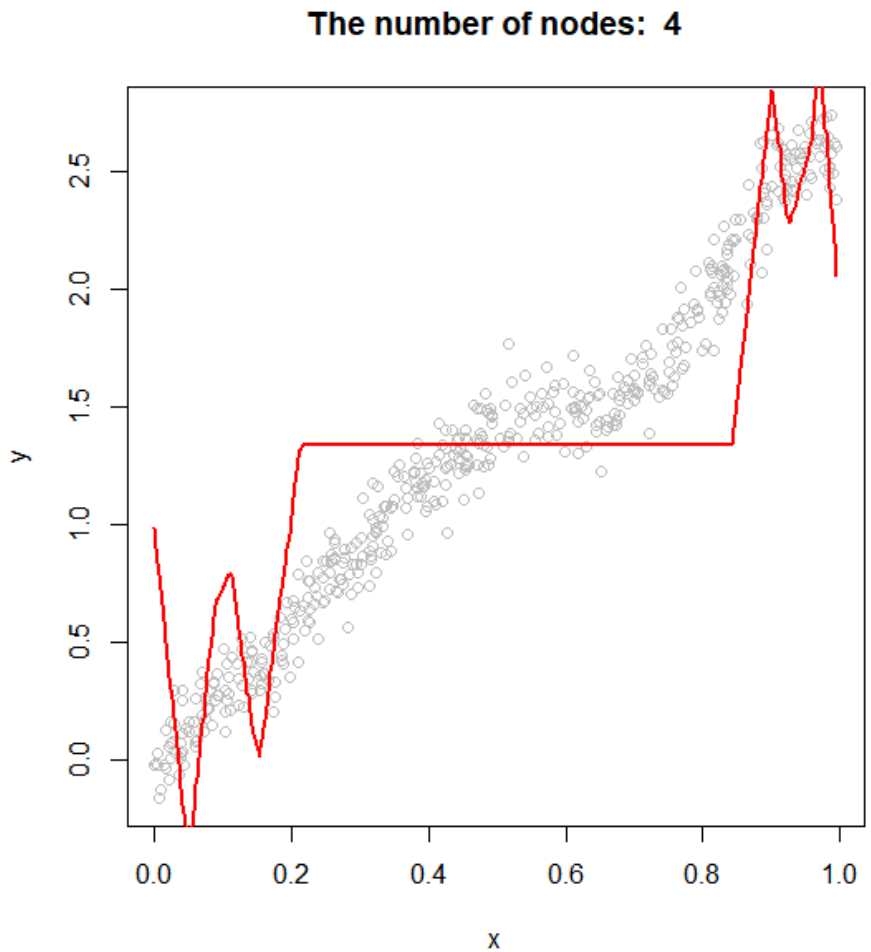
- Small number of nodes(4, 8, 16, 32) well captures global trend of data.
- Estimated function starts to overfit to data when the number of nodes exceeds 64.

Figure: Comparison of fit depending on the number of nodes





Test with width $[0.0625, 0.125]$ with animation

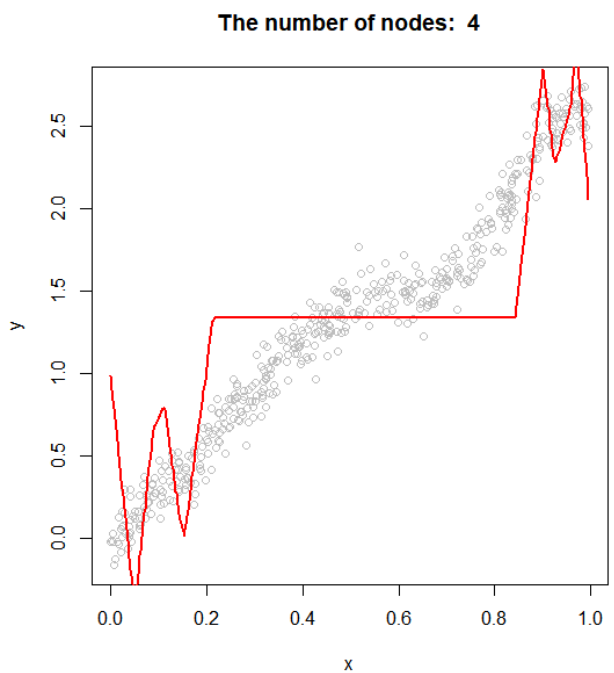


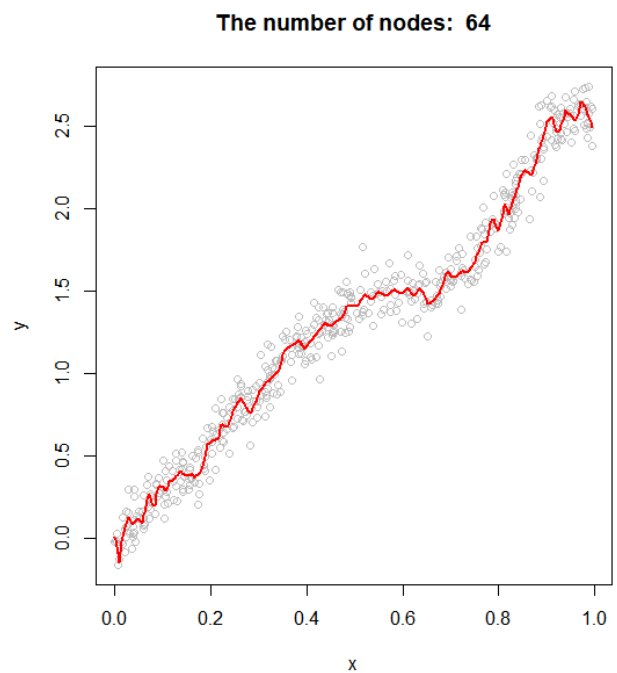
Play Pause

Test with width $[0.0625, 0.125]$

- When the number of nodes are small(4, 8), estimated function fails to capture global trend of data. Rather, it only captures locality at certain points.
- As the number of nodes increases, there is a tendency of overfitting.

Figure: Comparison of fit depending on the number of nodes



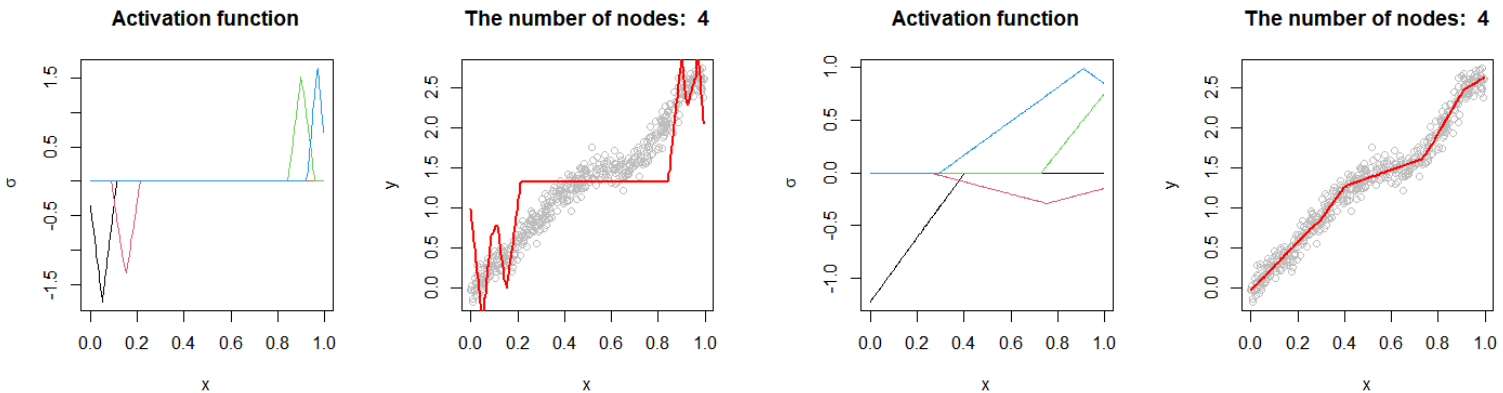


Conclusion

- Overall fit of data shows more interpretable patterns when constraint is given to width of B-spline activation function.
- If the width of activation function is small, we need sufficiently many nodes to capture global trend of data.
- If the width of activation function is relatively large, small number of nodes are enough to capture global trend.

Conclusion

Figure: (a) Narrow width (b) Large width



(a) (b)

To-do

- Fit SDLU model with residual fitting method and analyze results.
- Compare the effect of different number of nodes when residual fitting is added.