

Semiparametric Regression - Assignment 6

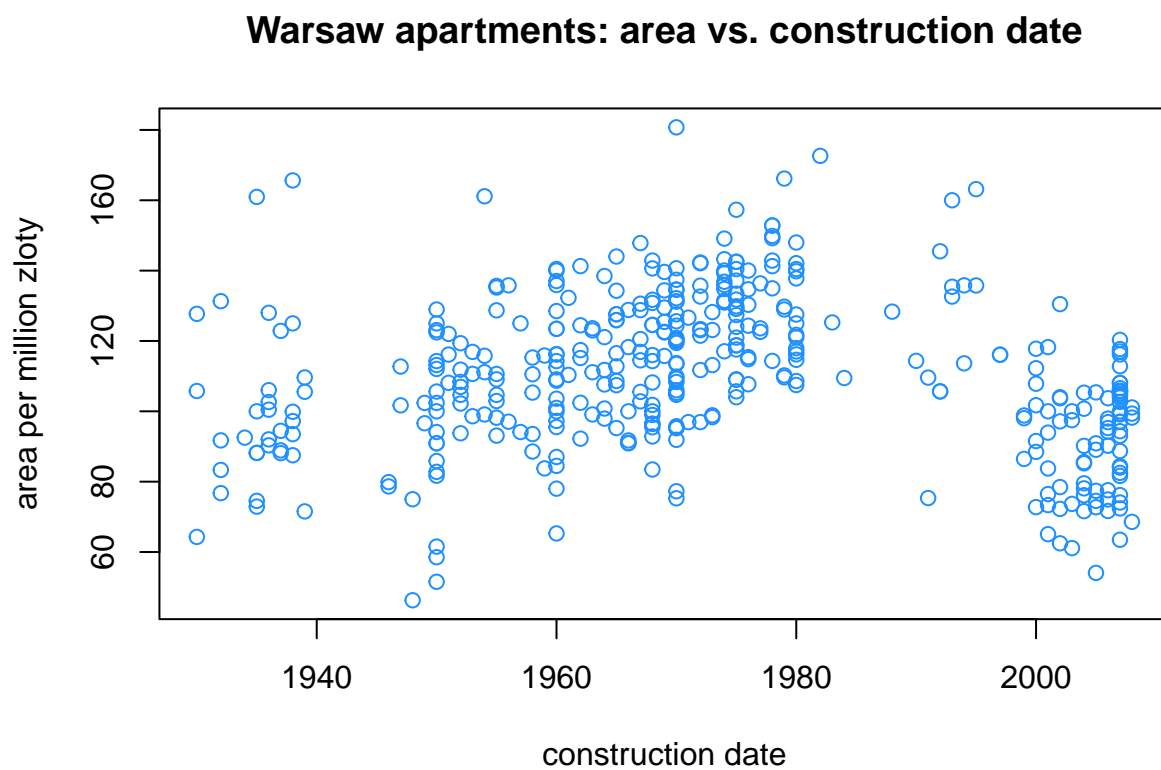
Szymon Armata | 341593 | Data Science

17 November 2022

Contents

1	Plot of the data	2
2	Task 1.	3
2.1	a	3
2.2	b	4
2.3	c	5

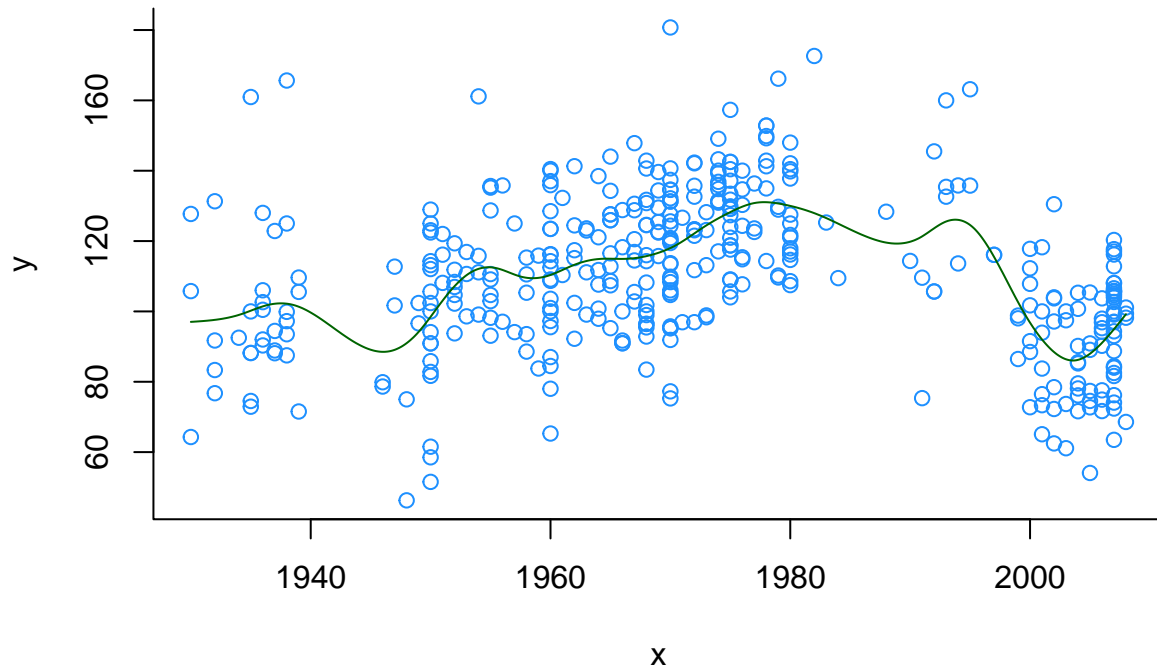
1 Plot of the data



2 Task 1.

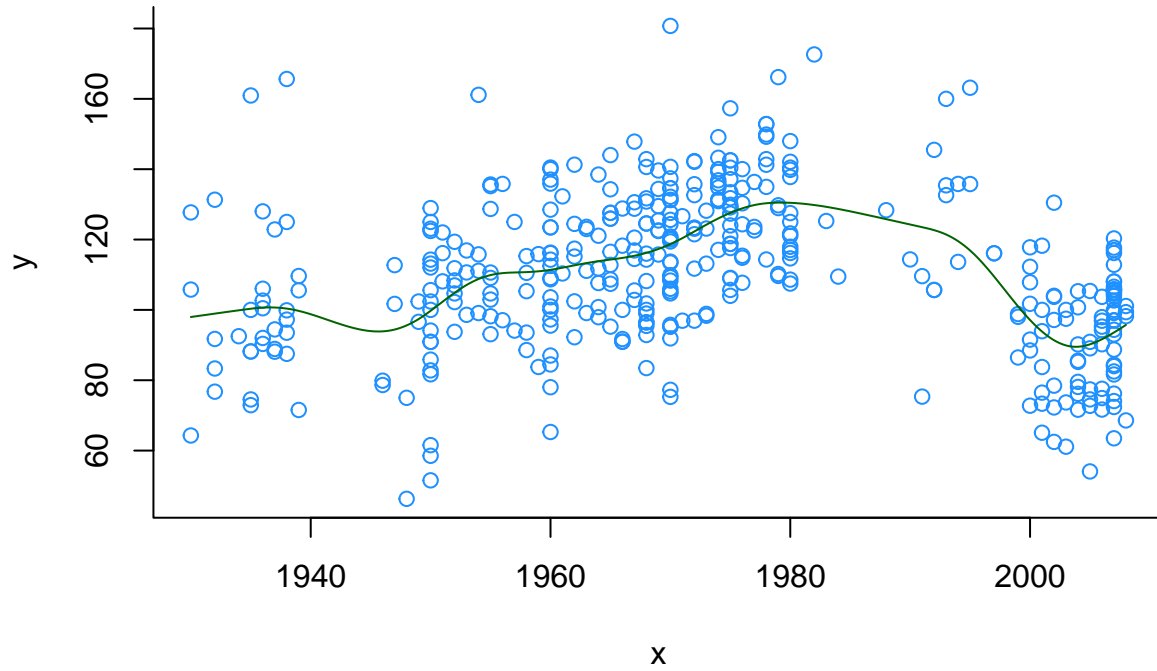
2.1 a

In this point we will fit and plot a penalized spline fit to the running non-parametric regression example data with 30 cubic regression spline basis functions and GCV-based smoothing parameter selection.



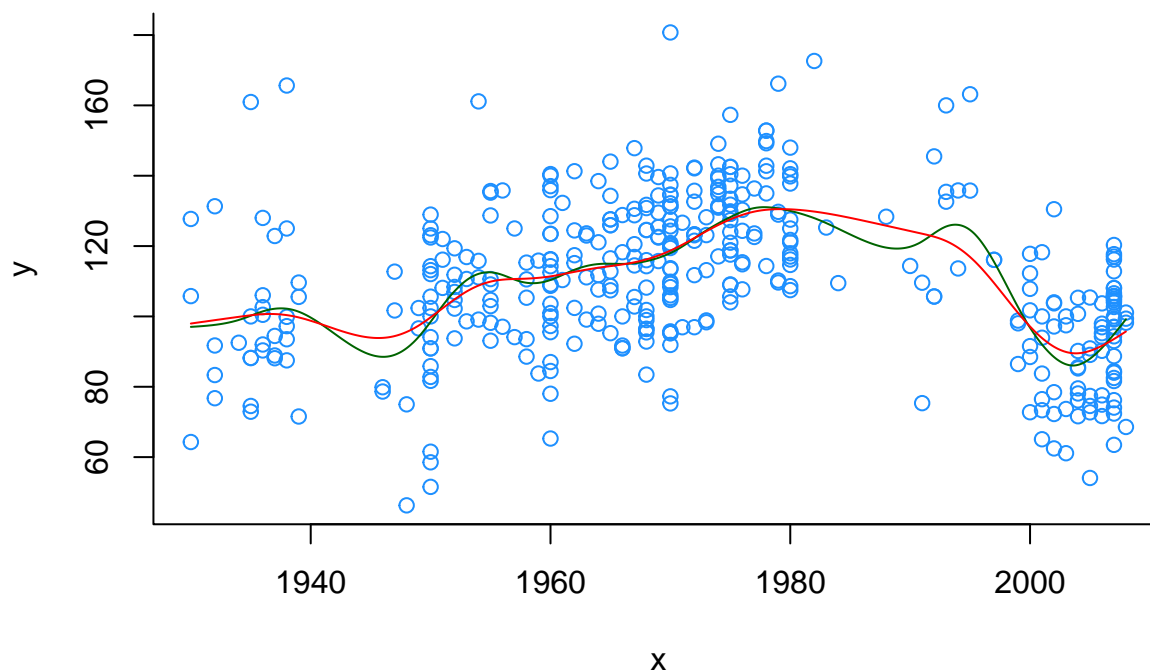
2.2 b

In this point we will fit and plot a penalized spline fit to the running non-parametric regression example data with 30 cubic regression spline basis functions and REML-based smoothing parameter selection.



2.3 c

Now we have to compare two smoothing parameter selection methods: GCV and REML. For a better comparison of the two methods, we will combine the two plots into one to see the similarities as well as the differences, which will best allow us to evaluate them.



On this plot the red line represents the curve plotted using the method of REML, while the green line represents the curve plotted with the method GCV.

Those two methods are based on different approaches, GCV is based on model selection, REML on likelihood. Comparing approaches based on model selection should result in nearly identical fits for all of them - they are equivalent.

As can be seen, the lines are very close to each other, almost overlapping in many places (such as between 1965 and 1975). Their greatest differences are in times when there is a shortage of data (for example, the years 1940-1945 or 1980-1995). With fewer knots, REML resembles GCV slightly but is less susceptible to outliers.

In summary, the choice of smoothing parameter selection method has a major impact on fit. The choice of a particular method will depend, among other things, on the initial dataset including the occurrence of outliers. In the case of our dataset, it is difficult to make a clear judgement as to which of the selected methods is more effective, but this does not affect the fact that the selection of the appropriate method has a significant impact on fit.