JYX JYVÄSKYLÄN YLIOPISTO UNIVERSITY OF JYVÄSKYLÄ

This is a self-archived version of an original article. This version may differ from the original in pagination and typographic details.

Author(s): Nordhausen, Klaus; Taskinen, Sara

Title: Locally Weighted Scatterplot Smoother

Year: 2021

Version: Accepted version (Final draft)

Copyright: © 2021 Springer Nature Switzerland AG

Rights: In Copyright

Rights url: http://rightsstatements.org/page/InC/1.0/?language=en

Please cite the original version:

Nordhausen, K., & Taskinen, S. (2021). Locally Weighted Scatterplot Smoother. In B. S. Daya Sagar, Q. Cheng, J. McKinley, & F. Agterberg (Eds.), Encyclopedia of Mathematical Geosciences. Springer. Encyclopedia of Earth Sciences Series. https://doi.org/10.1007/978-3-030-26050-7_184-1

Chapter 1

Locally Weighted Scatterplot Smoother

Klaus Nordhausen and Sara Taskinen
Department of Mathematics and Statistics, University of Jyväskylä
klaus.k.nordhausen@jyu.fi, sara.l.taskinen@jyu.fi,

Definition

The locally weight scatterplot smoother (LOWESS) uses local regression models to obtain a smooth estimator for the conditional expected value of response variable. The locality is defined based on a nearest neighbour type approach assigning weights to the observations based on their distances from the point of interest as well as on their outlyingness which are then used in an iterated weighted least squares approach.

1 Introduction

A common problem in statistics is to identify the relationship between two variables, *i.e.*, between an explaining variable x and a response variable y based on a sample (x_i, y_i), $i, \dots, n$. Traditionally the conditional expectation of y given x is modelled as a function of x, and we assume that the data are generated by

$$y_i = g(x_i) + \epsilon_i,$$

where $g(\cdot)$ is a smooth function and ϵ_i is some random fluctuation with $E(\epsilon_i) = 0$ and $Var(\epsilon_i) = \sigma^2$. When visualizing a scatterplot of x_i against y_i one thus wants to see a smooth relationship in the conditional mean of y as a function of x.

K. Nordhausen & S. Taskinen. Locally Weighted Scatterplot Smoother. Submitted to: Encyclopedia of Mathematical Geosciences

Assuming $g(x) = \beta_0 + \beta_1 x$, where β_0 and β_1 are constants, a straight line is used to model the conditional mean of y, which is known as linear regression. However, in many practical problems it seems an oversimplification to approximate $g(\cdot)$ with a linear function. The use of a polynomial of order d, *i.e.*, $g(x) = \beta_0 + \beta_1 x + \ldots + \beta_d x^d$ may also not be reasonable and may overfit the data and cause problems at the boundaries of x.

In many cases it may be reasonable to assume that locally, in a neighbourhood of a point x, a linear regression provides a good approximation for the conditional mean of y. For such a local regression observations close to x are naturally more relevant than observations further away, thus leading to the following weighted regression problem

$$\underset{\beta_0,\beta_1}{\operatorname{argmin}} \sum_{i=1}^{n} K\left(\frac{x_i - x}{h}\right)(y_i - \beta_0 - \beta_1 x_i)^2 , \tag{1.1}$$

where K is a non-negative weight function called kernel, and the parameter h which defines locality is called the bandwidth. The bandwidth can be chosen for example using cross validation. This modelling approach is also known as local linear regression and is easily extended to the case of several predictors and to local polynomial regression.

Local linear regression has a long tradition and is discussed in detail in Loader (1999), for example. The main advantages of local linear regression are, according to Cleveland & Loader (1996), that (i) it is easy to understand and interpret, (ii) it adapts well to problems at the boundaries and regions of high curvature, and (iii) it does not require smoothness or regularity conditions. Many computational tools for local linear regression exist. In R, the method is available for example in the package `locfit` (Loader 2020).

Despite its flexibility, local linear regression as defined in (1.1) has some shortcomings. One disadvantage is that the method uses the quadratic loss function, which is well-known to be sensitive to outliers. It is therefore natural to replace the quadratic loss function with a robust loss function and fit locally some robust regression model instead of a least squares regression. Local robust polynomial regression of order d can then defined as

$$\underset{\beta_0,\beta_1,\ldots,\beta_d}{\operatorname{argmin}} \sum_{i=1}^{n} K\left(\frac{x_i - x}{h}\right) \rho\left(y_i - \beta_0 - \beta_1 x_i - \cdots - \beta_d x_i^d\right), \tag{1.2}$$

where ρ is a loss function. The data analyst must now choose the kernel function

K, the loss function ρ, the order of the polynominal d and the bandwidth h.

2 Locally weight scatterplot smoother

In this context, Cleveland (1979) suggested the locally weighted scatterplot smoother (LOWESS) which uses as a kernel the tricube weight function

$$K(x) = (1 - |d|^3)^3,$$

when $|d| < 1$, and 0 otherwise, where d is the distance from the neighbourhood sample points to the value x. Cleveland (1979) argues that such kernel is beneficial as it assigns points far away a zero weight and is therefore better than kernels based on symmetric densities which give usually all points some positive weight. Depending on the scale of x, it may then however be advisable to scale the observations. For the neighbourhood structure LOWESS suggests an adaptive nearest neighbour method. Let $0 < f \leq 1$ be a tuning parameter and denote then $r = [fn]$ as the number of local data points to be used. At point x, those r points for which $|x - x_i|$ is the smallest, then form the neighbourhood. As the next step initial fits $\hat{y}_i$ at each data point are obtained using the local polynomial least square regression in order to obtain the residuals $r_i = y_i - \hat{y}_i$. These residuals are then robustly scaled using six times the median absolute deviation (mad), *i.e.*, the median of $\{|r_1|, \dots, |r_n|\}$. To downweight large residuals biweight kernel $B(t) = (1 - t^2)^2 I_{[-1,1]}(t)$ is then used to assign robust weights $\delta_i = B(r_i/(6\,\text{mad}))$. Based on this initial step the procedure is then iterated M times so that the kernel $K(x)$ is multiplied with robust weights δ_i. Thus, this corresponds to local iterated weighted least squares (IWLS) where the weights depend on the distance from the point of interest and whether the point is considered as outlying.
The parameter values for LOWESS that need to be chosen are f, d and M. Cleveland (1979) suggested to choose d from 0, 1 and 2 with a tendency to $d = 1$ as higher order polynomials tend to overfit the data and the method becomes numerically unstable. It is also argued that the number of iterations can be small, say $M = 2$ or $M = 3$, and no convergence criterion is needed. Parameter f has most impact on the results, as, when the value is increased, the smoothness of the fit increases. Thus it is a trade off between variability and recognizing the pattern. Cleveland (1979) suggests to start with $f = 0.5$ and limit the value to the range of $[0.2, 0.8]$. For automatic procedures Cleveland (1979) suggests to first choose the initial f which minimizes $\sum_{i=1}^{n}(y_i - \hat{y}_{i,f})^2$ and then to choose the final f as the minimizer of $\sum_{i=1}^{n} \delta_i(y_i - \hat{y}_{i,f})^2$. Based on the final robust weights

one can then compute the fitted values $\hat{y}_1^*, \ldots, \hat{y}_m^*$ for a grid of, not necessarily observed values, $x_1^*, \ldots, x_m^*$ covering the range of the observed x's. Using linear interpolation, these points can then be used to visualize the smooth function. For the theoretical properties of LOWESS under the assumption of iid errors following a Gaussian distribution, we refer to Cleveland (1979). The procedure in an algorithmic form is summarized in Algorithm 1.

Algorithm 1: LOWESS algorithm

Input: n data points (y_i, x_i), $i = 1, \ldots, n$;

Set $f \in (0, 1]$ to specify the neighbourhood;
Set integer $d \geq 0$ as the degree of the local polynomial;
Set integer $M \geq 1$ for the number of iterations;
Compute the size of the nearest neighbourhood $r = [fn]$;
for $i \leftarrow 1$ **to** n **do**
  For x_i identify the r points forming the neighbourhood and denote them $x_{r_l}^i$, where $r_l \in \{1, \ldots, i-1, i+1, \ldots, n\}$ and $l = 1, \ldots, r$;
  Use these r points to fit a local polynomial regression model of order d using the kernel $K(x_{r_l}^i) = (1 - |x_{r_l}^i - x_i|^3)^3$;
  Obtain the initial fit $\hat{y}_i$;
for *iter* $\leftarrow 1$ **to** M **do**
  Calculate the residuals $r_i = y_i - \hat{y}_i$ and the median of absolute deviations (mad) of the residuals;
  Compute the robust weights $\delta_i = B(r_i/(6\,\text{mad}))$;
  for $j \leftarrow 1$ **to** n **do**
    Use the neighbourhood points of x_j, $x_{r_l}^j$ and their corresponding robust weights $\delta_{r_l}^j$ to fit a local polynomial regression model of order d using the kernel $K(x_{r_l}^j) = \delta_{r_l}^j(1 - |x_{r_l}^j - x_j|^3)^3$;
    Obtain the robust fit $\hat{y}_j$;
Predict y for any point of interest x;

In R, LOWESS with $d = 1$ can be applied using for example the function `lowess`.

3 Extensions of LOWESS

Cleveland (1979) pointed out that LOWESS can be used with any kernel function and also variants exist to speed up the computations for large data sets and for

accommodating prior weights for the observations.
The main extension of LOWESS is however suggested in Cleveland & Devlin (1988) as locally estimated scatterplot smoothing (LOESS). LOESS generalizes LOWESS to the case of several predictors by replacing the simple regression with a multiple regression. Implementions such as `loess` in R let the user choose the degree of the polynominal d as well as the need for robust updates.

4 Example

To demonstrate LOWESS we consider average yearly temperatures (in centigrades) of the German state Baden-Wuerttemberg from 1880 to 2020. The data can be obtained from the CDC-portal of the Deutscher Wetterdienst. Figure 1 shows the data together with three global polynomial least squares fits. All fits indicate an increase in average temperature but seem not to fit the data very well for early temperature records. We then apply LOWESS via the R function `loess` and visualize the results in Figure 2 for different degrees of the polynomial (d) and different parameters defining the neighbourhood (f). Both, the robust variant with the IWLS steps as well when pure local LS fits are illustrated. The figure clearly shows that the differences between robust fits and LS fits are quite small indicating the lack of outlying observations. A quadratic polynomial seems however preferable, especially with a large value of f, and the LOWESS fit shows that after a long period of slow rise in average temperature, the rise has accelerated in the last 50 years. A more thorough analysis of temperature data using LOWESS is available in Wanishsakpong & Notodiputro (2018), for example.

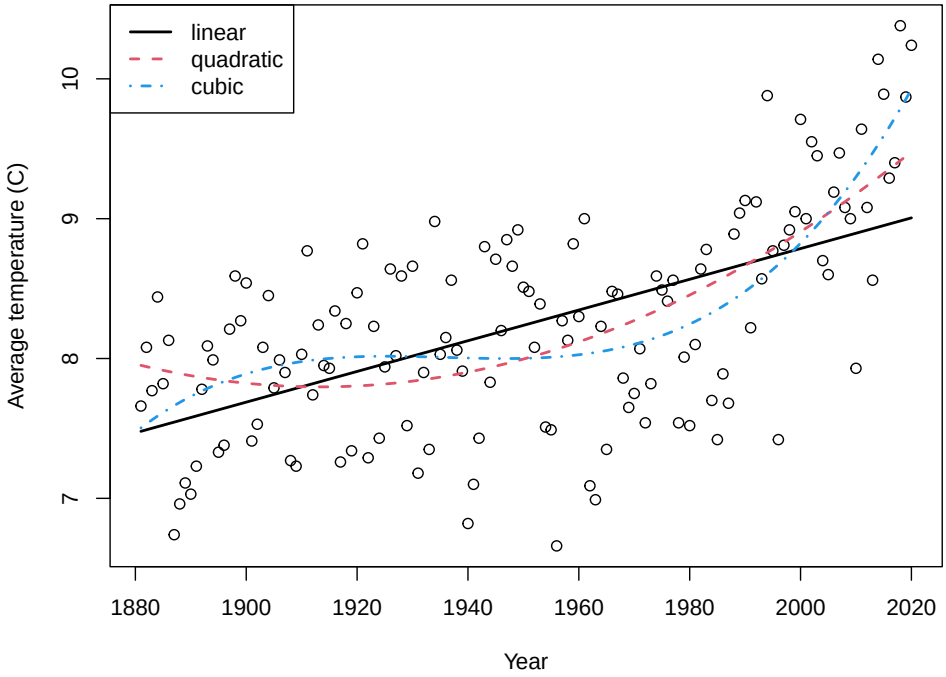

Figure 1: Yearly average temperatures (°C) for Baden-Wuerttemberg in Germany together with three global fits.

Summary

LOWESS and its extension LOESS are powerful and simple regression methods ideal for situations where an initial model assumption is not available, as the methods do not use any special model specification. LOWESS exploits locality which means that it requires dense observations to fully achieve its strengths. While LOWESS was originally designed for models with a signal predictors LOESS allows more predictors meaning that the method creates smooth surfaces which then need even denser data and a careful choice for the notion of distance.

Cross References

Smoothing Filter, Splines

*

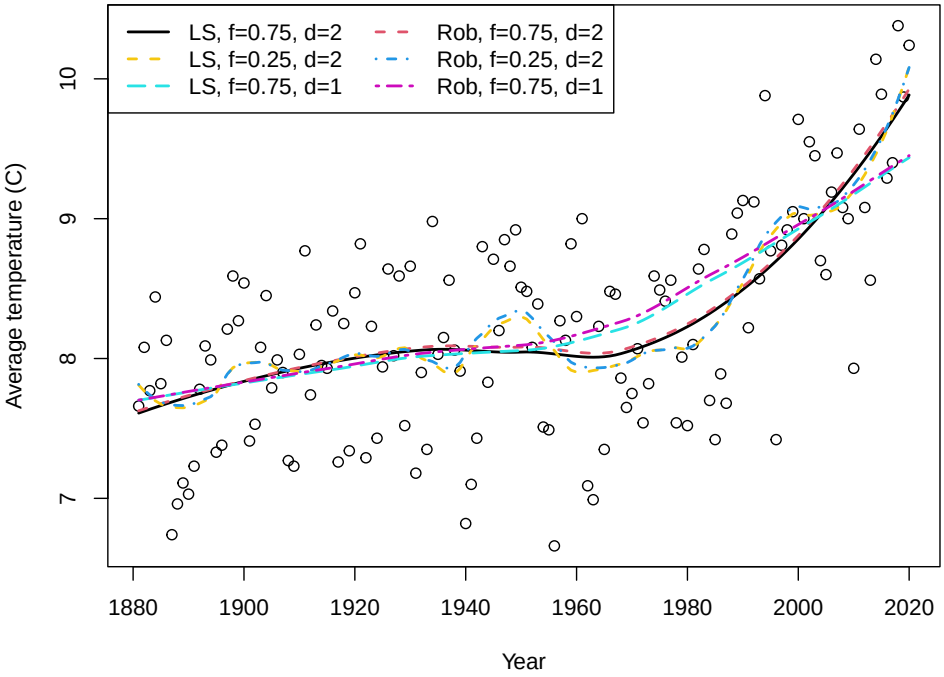

Figure 2: Yearly average temperatures (°C) for Baden-Wuerttemberg in Germany together with locally smoothed fits.

Acknowledgements

References

Cleveland, William S. 1979. Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association* 74(368). 829–836. DOI: 10.1080/01621459.1979.10481038.

Cleveland, William S. & Susan J. Devlin. 1988. Locally weighted regression: An approach to regression analysis by local fitting. *Journal of the American Statistical Association* 83. 596–610.

Cleveland, William S. & Clive Loader. 1996. Smoothing by local regression: principles and methods. In Wolfgang Härdle & Michael G. Schimek (eds.), *Statistical theory and computational aspects of smoothing*, 10–49. Heidelberg: Physica-Verlag HD. DOI: 10.1007/978-3-642-48425-4_2.

Loader, Catherine. 2020. *locfit: Local Regression, Likelihood and Density Estimation*. R package version 1.5-9.4. https://CRAN.R-project.org/package=locfit.

Loader, Clive. 1999. *Local regression and likelihood*. New York: Springer-Verlag.
Wanishsakpong, Wandee & Khairil Anwar Notodiputro. 2018. Locally weighted scatter-plot smoothing for analysing temperature changes and patterns in australia. *Meteorological Applications* 25(3). 357–364. DOI: 10.1002/met.1702.