

Smoothing spline

Smoothing splines are function estimates, $\hat{f}(x)$, obtained from a set of noisy observations y_i of the target $f(x_i)$, in order to balance a measure of goodness of fit of $\hat{f}(x_i)$ to y_i with a derivative based measure of the smoothness of $\hat{f}(x)$. They provide a means for smoothing noisy x_i, y_i data. The most familiar example is the cubic smoothing spline, but there are many other possibilities, including for the case where x is a vector quantity.

Contents

Cubic spline definition

Derivation of the cubic smoothing spline

De Boor's approach

Multidimensional splines

Related methods

Source code

References

Further reading

Cubic spline definition

Let $\{x_i, Y_i : i = 1, \dots, n\}$ be a set of observations, modeled by the relation $Y_i = f(x_i) + \epsilon_i$ where the ϵ_i are independent, zero mean random variables (usually assumed to have constant variance). The cubic smoothing spline estimate \hat{f} of the function f is defined to be the minimizer (over the class of twice differentiable functions) of^{[1][2]}

$$\sum_{i=1}^n \{Y_i - \hat{f}(x_i)\}^2 + \lambda \int \hat{f}''(x)^2 dx.$$

Remarks:

- $\lambda \geq 0$ is a smoothing parameter, controlling the trade-off between fidelity to the data and roughness of the function estimate. This is often estimated by generalized cross-validation,^[3] or by restricted marginal likelihood (REML) which exploits the link between spline smoothing and Bayesian estimation (the smoothing penalty can be viewed as being induced by a prior on the f).^[4]
- The integral is often evaluated over the whole real line although it is also possible to restrict the range to that of x_i .
- As $\lambda \rightarrow 0$ (no smoothing), the smoothing spline converges to the interpolating spline.
- As $\lambda \rightarrow \infty$ (infinite smoothing), the roughness penalty becomes paramount and the estimate converges to a linear least squares estimate.
- The roughness penalty based on the second derivative is the most common in modern statistics literature, although the method can easily be adapted to penalties based on other derivatives.
- In early literature, with equally-spaced ordered x_i , second or third-order differences were used in the penalty, rather than derivatives.^[5]
- The penalized sum of squares smoothing objective can be replaced by a *penalized likelihood* objective in which the sum of squares terms is replaced by another log-likelihood based measure of fidelity to the data.^[1] The sum of squares term corresponds to penalized likelihood with a Gaussian assumption on the ϵ_i .

Derivation of the cubic smoothing spline

It is useful to think of fitting a smoothing spline in two steps:

1. First, derive the values $\hat{f}(x_i); i = 1, \dots, n$.
2. From these values, derive $\hat{f}(x)$ for all x .

Now, treat the second step first.

Given the vector $\hat{\mathbf{m}} = (\hat{f}(x_1), \dots, \hat{f}(x_n))^T$ of fitted values, the sum-of-squares part of the spline criterion is fixed. It remains only to minimize $\int \hat{f}''(x)^2 dx$, and the minimizer is a natural cubic spline that interpolates the points $(x_i, \hat{f}(x_i))$. This interpolating spline is a linear operator, and can be written in the form

$$\hat{f}(x) = \sum_{i=1}^n \hat{f}(x_i) f_i(x)$$

where $f_i(x)$ are a set of spline basis functions. As a result, the roughness penalty has the form

$$\int \hat{f}''(x)^2 dx = \hat{\mathbf{m}}^T \mathbf{A} \hat{\mathbf{m}}.$$

where the elements of \mathbf{A} are $\int f_i''(x) f_j''(x) dx$. The basis functions, and hence the matrix \mathbf{A} , depend on the configuration of the predictor variables x_i , but not on the responses Y_i or $\hat{\mathbf{m}}$.

\mathbf{A} is an $n \times n$ matrix given by $\mathbf{A} = \Delta^T \mathbf{W}^{-1} \Delta$.

Δ is an $(n-2) \times n$ matrix of second differences with elements:

$$\Delta_{ii} = 1/h_i, \Delta_{i,i+1} = -1/h_i - 1/h_{i+1}, \Delta_{i,i+2} = 1/h_{i+1}$$

\mathbf{W} is an $(n-2) \times (n-2)$ symmetric tri-diagonal matrix with elements:

$W_{i-1,i} = W_{i,i-1} = h_i/6$, $W_{ii} = (h_i + h_{i+1})/3$ and $h_i = x_{i+1} - x_i$, the distances between successive knots (or x values).

Now back to the first step. The penalized sum-of-squares can be written as

$$\{Y - \hat{\mathbf{m}}\}^T \{Y - \hat{\mathbf{m}}\} + \lambda \hat{\mathbf{m}}^T \mathbf{A} \hat{\mathbf{m}},$$

where $Y = (Y_1, \dots, Y_n)^T$.

Minimizing over $\hat{\mathbf{m}}$ by differentiating against $\hat{\mathbf{m}}$. This results in: $-2\{Y - \hat{\mathbf{m}}\} + 2\lambda \mathbf{A} \hat{\mathbf{m}} = \mathbf{0}$ [6] and $\hat{\mathbf{m}} = (\mathbf{I} + \lambda \mathbf{A})^{-1} Y$.

De Boor's approach

De Boor's approach exploits the same idea, of finding a balance between having a smooth curve and being close to the given data. [7]

$$p \sum_{i=1}^n \left(\frac{Y_i - \hat{f}(x_i)}{\delta_i} \right)^2 + (1-p) \int \left(\hat{f}^{(m)}(x) \right)^2 dx$$

where p is a parameter called smooth factor and belongs to the interval $[0, 1]$, and $\delta_i; i = 1, \dots, n$ are the quantities controlling the extent of smoothing (they represent the weight δ_i^{-2} of each point Y_i). In practice, since cubic splines are mostly used, m is usually 2. The solution for $m = 2$ was proposed by Reinsch in 1967.^[8] For $m = 2$, when p approaches 1, \hat{f} converges to the "natural" spline interpolant to the given data.^[7] As p approaches 0, \hat{f} converges to a straight line (the smoothest curve). Since finding a suitable value of p is a task of trial and error, a redundant constant S was introduced for convenience.^[8] S is used to numerically determine the value of p so that the function \hat{f} meets the following condition:

$$\sum_{i=1}^n \left(\frac{Y_i - \hat{f}(x_i)}{\delta_i} \right)^2 \leq S$$

The algorithm described by de Boor starts with $p = 0$ and increases p until the condition is met.^[7] If δ_i is an estimation of the standard deviation for Y_i , the constant S is recommended to be chosen in the interval $[n - \sqrt{2n}, n + \sqrt{2n}]$. Having $S = 0$ means the solution is the "natural" spline interpolant.^[8] Increasing S means we obtain a smoother curve by getting farther from the given data.

Multidimensional splines

There are two main classes of method for generalizing from smoothing with respect to a scalar x to smoothing with respect to a vector \mathbf{x} . The first approach simply generalizes the spline smoothing penalty to the multidimensional setting. For example, if trying to estimate $f(\mathbf{x}, z)$ we might use the Thin plate spline penalty and find the $\hat{f}(\mathbf{x}, z)$ minimizing

$$\sum_{i=1}^n \{y_i - f(x_i, z_i)\}^2 + \lambda \int \left[\left(\frac{\partial^2 f}{\partial x^2} \right)^2 + 2 \left(\frac{\partial^2 f}{\partial x \partial z} \right)^2 + \left(\frac{\partial^2 f}{\partial z^2} \right)^2 \right] d\mathbf{x} dz.$$

The thin plate spline approach can be generalized to smoothing with respect to more than two dimensions and to other orders of differentiation in the penalty.^[1] As the dimension increases there are some restrictions on the smallest order of differential that can be used,^[1] but actually Duchon's original paper,^[9] gives slightly more complicated penalties that can avoid this restriction.

The thin plate splines are isotropic, meaning that if we rotate the \mathbf{x}, z co-ordinate system the estimate will not change, but also that we are assuming that the same level of smoothing is appropriate in all directions. This is often considered reasonable when smoothing with respect to spatial location, but in many other cases isotropy is not an appropriate assumption and can lead to sensitivity to apparently arbitrary choices of measurement units. For example, if smoothing with respect to distance and time an isotropic smoother will give different results if distance is measure in metres and time in seconds, to what will occur if we change the units to centimetres and hours.

The second class of generalizations to multi-dimensional smoothing deals directly with this scale invariance issue using tensor product spline constructions.^{[10][11][12]} Such splines have smoothing penalties with multiple smoothing parameters, which is the price that must be paid for not assuming that the same degree of smoothness is appropriate in all directions.

Related methods

Smoothing splines are related to, but distinct from:

- Regression splines. In this method, the data is fitted to a set of spline basis functions with a reduced set of knots, typically by least squares. No roughness penalty is used. (See also multivariate adaptive regression splines.)
- Penalized Splines. This combines the reduced knots of regression splines, with the roughness penalty of smoothing splines.^{[13] [14]}

- Elastic maps method for manifold learning. This method combines the least squares penalty for approximation error with the bending and stretching penalty of the approximating manifold and uses the coarse discretization of the optimization problem; see thin plate splines.

Source code

Source code for spline smoothing can be found in the examples from Carl de Boor's book *A Practical Guide to Splines*. The examples are in the Fortran programming language. The updated sources are available also on Carl de Boor's official site [1] (<http://pages.cs.wisc.edu/~deboor/>).

References

1. Green, P. J.; Silverman, B.W. (1994). *Nonparametric Regression and Generalized Linear Models: A roughness penalty approach*. Chapman and Hall.
2. Hastie, T. J.; Tibshirani, R. J. (1990). *Generalized Additive Models*. Chapman and Hall. ISBN 978-0-412-34390-2.
3. Craven, P.; Wahba, G. (1979). "Smoothing noisy data with spline functions". *Numerische Mathematik*. **31** (4): 377–403. doi:10.1007/bf01404567 (<https://doi.org/10.1007%2Fbf01404567>).
4. Kimeldorf, G.S.; Wahba, G. (1970). "A Correspondence between Bayesian Estimation on Stochastic Processes and Smoothing by Splines". *The Annals of Mathematical Statistics*. **41** (2): 495–502. doi:10.1214/aoms/1177697089 (<https://doi.org/10.1214%2Faoms%2F1177697089>).
5. Whittaker, E.T. (1922). "On a new method of graduation". *Proceedings of the Edinburgh Mathematical Society*. **41**: 63–75.
6. Rodriguez, German (Spring 2001). "Smoothing and Non-Parametric Regression" (<http://data.princeton.edu/eco572/smoothing.pdf>) (PDF). 2.3.1 Computation. p. 12. Retrieved 28 August 2017.
7. De Boor, C. (2001). *A Practical Guide to Splines (Revised Edition)*. Springer. pp. 207–214. ISBN 978-0-387-90356-9.
8. Reinsch, Christian H (1967). "Smoothing by Spline Functions". *Numerische Mathematik*. **10** (3): 177–183. doi:10.1007/BF02162161 (<https://doi.org/10.1007%2Fbf02162161>).
9. J. Duchon, 1976, Splines minimizing rotation invariant semi-norms in Sobolev spaces. pp 85–100, In: Constructive Theory of Functions of Several Variables, Oberwolfach 1976, W. Schempp and K. Zeller, eds., Lecture Notes in Math., Vol. 571, Springer, Berlin, 1977
10. Wahba, Grace. *Spline Models for Observational Data*. SIAM.
11. Gu, Chong (2013). *Smoothing Spline ANOVA Models (2nd ed.)*. Springer.
12. Wood, S. N. (2017). *Generalized Additive Models: An Introduction with R (2nd ed.)*. Chapman & Hall/CRC. ISBN 978-1-58488-474-3.
13. Eilers, P.H.C. and Marx B. (1996). "Flexible smoothing with B-splines and penalties". *Statistical Science*. **11** (2): 89–121.
14. Ruppert, David; Wand, M. P.; Carroll, R. J. (2003). *Semiparametric Regression*. Cambridge University Press. ISBN 978-0-521-78050-6.

Further reading

- Wahba, G. (1990). *Spline Models for Observational Data*. SIAM, Philadelphia.
- Green, P. J. and Silverman, B. W. (1994). *Nonparametric Regression and Generalized Linear Models*. CRC Press.
- De Boor, C. (2001). *A Practical Guide to Splines (Revised Edition)*. Springer.

Retrieved from "https://en.wikipedia.org/w/index.php?title=Smoothing_spline&oldid=925559684"

This page was last edited on 10 November 2019, at 22:10 (UTC).

Text is available under the Creative Commons Attribution-ShareAlike License; additional terms may apply. By using this site, you agree to the Terms of Use and Privacy Policy. Wikipedia® is a registered trademark of the Wikimedia

Foundation, Inc., a non-profit organization.