Journal of Statistical Planning and Inference 153 (2014) 56–74

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

# Performance criteria and discrimination of extreme undersmoothing in nonparametric regression

Mark A. Lukas *

*Mathematics and Statistics, Murdoch University, South Street, Murdoch, WA 6150, Australia*

ARTICLE INFO

*Article history:*
Received 16 February 2013
Received in revised form
14 May 2014
Accepted 15 May 2014
Available online 23 May 2014

*Keywords:*
Prediction error
Average squared error
Sobolev error
Optimality criterion
Smoothing spline
Kernel smoothing

ABSTRACT

The prediction error (average squared error) is the most commonly used performance criterion for the assessment of nonparametric regression estimators. However, there has been little investigation of the properties of the criterion itself. This paper shows that in certain situations the prediction error can be very misleading because it fails to discriminate an extreme undersmoothed estimate from a good estimate. For spline smoothing, we show, using asymptotic analysis and simulations, that there is poor discrimination of extreme undersmoothing in the following situations: small sample size or small error variance or a function with high curvature. To overcome this problem, we propose using the Sobolev error criterion. For spline smoothing, it is shown asymptotically and by simulations that the Sobolev error is significantly better than the prediction error in discriminating extreme undersmoothing. Similar results hold for other nonparametric regression estimators and for multivariate smoothing. For thin-plate smoothing splines, the prediction error's poor discrimination of extreme undersmoothing becomes significantly worse with increasing dimension.



## 1. Introduction

A fundamental problem of data analysis is the estimation of a smooth function $f:[a,b]\to\mathbb{R}$ from data $y_i$ that follow the model:

$$y_i = f(x_i) + \varepsilon_i, \quad i = 1, \dots, n,$$

where the design points $x_i$ satisfy $a \le x_1 < x_2 < \cdots < x_n \le b$ and the $\varepsilon_i$ are i.i.d. $N(0,\sigma^2)$ random errors. If a functional form for $f$ is not known, it is appropriate to use a nonparametric regression method. Several such methods have been proposed and studied, including kernel smoothing, local polynomial smoothing, series estimators, regression splines and smoothing splines; see Eubank (1988), Eubank (1999), Green and Silverman (1994), Hart (1997), Wahba (1990), and Wand and Jones (1995) for discussion of these methods.

We will mainly consider smoothing spline estimators. The smoothing spline of degree $2m-1$ is defined as the function $f_\lambda$ that minimizes

$$n^{-1}\sum_{i=1}^{n}(y_i - h(x_i))^2 + \lambda\int_a^b (h^{(m)}(x))^2\,dx$$

* Tel.: +61 893602423; fax: +61 893606332.
*E-mail address:* M.Lukas@murdoch.edu.au

http://dx.doi.org/10.1016/j.jspi.2014.05.006
0378-3758/

in the Sobolev space $\mathcal{W}^{m,2}[a,b]$. Here $\lambda > 0$ is the smoothing parameter; a larger value of $\lambda$ yields a smoother estimate. The most popular smoothing splines are cubic smoothing splines, for which $m = 2$.

To assess the quality of a spline estimator $f_\lambda$ and hence define the "best" choice of $\lambda$, one needs a suitable performance criterion (optimality criterion). The most commonly used performance measure is the prediction error (average squared error)

$$T(\lambda) = n^{-1} \sum_{i=1}^{n} (f(x_i) - f_\lambda(x_i))^2, \tag{1}$$

together with the prediction risk $ET(\lambda)$. A related measure is the (squared) $L_2$ error (integrated squared error)

$$I(\lambda) = \int_a^b (f(x) - f_\lambda(x))^2 \, dx \tag{2}$$

together with the associated risk $EI(\lambda)$. Note that, for equally spaced points $x_i$, $T(\lambda)$ in (1) is a standard quadrature approximation of $I(\lambda)$ in (2), and so they can be expected to be fairly close.

Most of the extensive literature on the performance of smoothing spline estimators, including various parameter selection methods, uses the prediction error or $L_2$ error (or risk) as the only performance measure. This includes both asymptotic analyses as well as simulation studies; see, e.g., Li (1986) and Lee (2003). However, there has not been much attention given to the question of whether the prediction error or $L_2$ error is actually the most suitable performance criterion. One of the aims of this paper is to shed light on this question.

It was shown in Marron and Tsybakov (1995) that the use of the $L_2$ error (as well as the $L_1$ and $L_\infty$ errors) in nonparametric regression does not properly match what the eye can see. The authors argue that a better qualitative assessment of an estimated curve, when compared to a true curve, is obtained by considering the distance between the graphs of the curves (as sets of points in $\mathbb{R}^2$), rather than the vertical distance associated with a function space norm, and they propose several visual error criteria in this way. However, as noted in Marron and Tsybakov (1995), this approach is not suitable if the regression is to be used for prediction purposes.

In this paper, we take the more general view that the performance criterion used in nonparametric regression should make the regression suitable for estimation of $f(x)$ on the whole interval $[a,b]$ and for associated prediction, while, at the same time, the criterion should be consistent with our visual perception of the quality of the estimator.

We will focus on a specific weakness of the prediction error as a performance measure. Clearly, the prediction error is a discrete measure in that it depends on $f - f_\lambda$ only through its values at the discrete set of points $x_i$, but this is not its main weakness. The main problem with the prediction error (as well as the $L_2$, $L_1$ and $L_\infty$ errors) is that it is insensitive to deviations in the derivative and (linearized) curvature (and higher derivatives) of the spline $f_\lambda$. This situation is inconsistent with our visual perception of the quality of a fitted curve. Because $f$ is smooth, large deviations in the derivative or curvature of $f_\lambda$ from those of $f$ can be easily identified visually from the graphs of $f$ and $f_\lambda$.

To make this more concrete, suppose that the error is approximately given by $f(x) - f_\lambda(x) \simeq c \cos k\pi x$ for $x \in [0,1]$, where $c > 0$ is small and $k$ is a large integer satisfying $k \geq c^{-1}$. Assume that the $x_i$ are equally spaced in $[0,1]$ and $n$ is sufficiently large so that $T(\lambda) \simeq I(\lambda)$. Then $T(\lambda) \simeq I(\lambda) \simeq c^2/2$, which is small even though $f - f_\lambda$ is very wiggly, with large integrated squared (linearized) curvature satisfying

$$\int_0^1 (f''(x) - f_\lambda''(x))^2 \, dx \geq K \simeq k^2 \pi^4 / 2.$$

Therefore, the prediction error (and $L_2$ error) fails to detect that $f - f_\lambda$ is very wiggly, and yet this would be easily seen from its graph.

Clearly, the same conclusion holds if $f - f_\lambda \simeq h$, where $h(x)$ is a short finite sum of the form

$$h(x) = \sum (c_j \cos k_j \pi x + d_j \sin l_j \pi x)$$

and $c_j > 0$ and $d_j > 0$ are small, with $k_j \geq c_j^{-1}$ and $l_j \geq d_j^{-1}$. Note that this assumption for $f - f_\lambda$ is quite plausible if $\lambda$ is very small, because, in this situation, $f - f_\lambda$ would be close to the spline that interpolates the errors $\varepsilon_i$ at the design points $x_i$. It is likely that this error vector is of high frequency, measured, say, by the number of sign changes. Then, using the Demmler–Reinsch basis for the space of smoothing splines, $f - f_\lambda$ would be approximately equal to a finite sum involving the high frequency basis functions, and it is known that, for equally spaced $x_i$, the basis functions are approximately equal to trigonometric functions; see Culpin (1986), Eubank (1988, Sect. 5.3).

The above reasoning indicates that the prediction error may fail to detect when a spline estimate $f_\lambda$ is very wiggly. It will be shown that this does actually occur in practice. Moreover, the prediction error can be a misleading performance measure because it can fail to discriminate an extreme undersmoothed spline estimate from a good estimate. Section 2.1 presents simulation results that illustrate this property and we identify the situations where it can occur; these are: small sample size $n$ or small error variance $\sigma^2$ or a function $f$ with high curvature.

In Section 2.2, we define a measure of the prediction error's capacity to discriminate extreme undersmoothing. This is defined as the probability that the value of the prediction error for the most extreme undersmoothed spline estimate, i.e., the interpolating spline, is relatively close to the minimum prediction risk (within a factor $D$). The larger the value of this probability, the more likely it is that the prediction error will fail to discriminate extreme undersmoothing. We investigate