COMSTA: 6545 Model 3G pp. 1–21 (col. fig: NIL)

ARTICLE IN PRESS

Computational Statistics and Data Analysis xx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

Goodness-of-fit test for nonparametric regression models: Smoothing spline ANOVA models as example

Sebastian J. Teran Hidalgo [a,*], Michael C. Wu [b], Stephanie M. Engel [c], Michael R. Kosorok [d]

[a] Department of Biostatistics, Yale University, New Haven, CT, USA
[b] Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA
[c] Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA
[d] Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

ARTICLE INFO

Article history:
Received 3 October 2016
Received in revised form 10 January 2018
Accepted 12 January 2018
Available online xxxx

Keywords:
Goodness-of-fit
Interaction testing
Smoothing spline models
Bootstrap
Test of independence

ABSTRACT

Nonparametric regression models do not require the specification of the functional form between the outcome and the covariates. Despite their popularity, the amount of diagnostic statistics, in comparison to their parametric counterparts, is small. We propose a goodness-of-fit test for nonparametric regression models with linear smoother form. In particular, we apply this testing framework to smoothing spline ANOVA models. The test can consider two sources of lack-of-fit: whether covariates that are not currently in the model need to be included, and whether the current model fits the data well. The proposed method derives estimated residuals from the model. Then, statistical dependence is assessed between the estimated residuals and the covariates using the HSIC. If dependence exists, the model does not capture all the variability in the outcome associated with the covariates, otherwise the model fits the data well. The bootstrap is used to obtain p-values. Application of the method is demonstrated with a neonatal mental development data analysis. We demonstrate correct type I error as well as power performance through simulations.

1. Introduction

Nonparametric regression models can provide a better fit when parametric assumptions are too restrictive (e.g., linearity of the mean). A popular nonparametric model from the machine learning literature is kernel ridge regression (KR) (Liu et al., 2007; Shawe-Taylor and Cristianini, 2004). In KR regression, the input covariates are mapped to a high (possibly infinite) dimensional space through a kernel function, but which is only represent through a kernel matrix of the sample points. Another popular method is k-nearest neighbor (KNN) regression (Hastie et al., 2005), in which for each observation, an average of the outcome is taken across all k-closest values with respect to the covariates. The fitted values KNN regression look jagged and are hard to interpret. Smoothing spline ANOVA models (SS-ANOVA) are also a popular nonparametric regression methodology this time arising from the statistical literature (Craven and Wahba, 1978; Golub et al., 1979; Gu, 2013; Kimeldorf and Wahba, 1971; Wahba, 1990). SS-ANOVA models estimate the mean of an outcome as a smooth function with an ANOVA decomposition which partitions the variation of the outcome attributed to the covariates into main effects, two-way interactions, and all other higher-level interactions, but as a summation of functions, not constants, as with classical

* Corresponding author.
E-mail address: sebastian.teranhidalgo@yale.edu (S.J. Teran Hidalgo).

https://doi.org/10.1016/j.csda.2018.01.004
0167-9473/

Please cite this article in press as: Teran Hidalgo S.J., et al., Goodness-of-fit test for nonparametric regression models: Smoothing spline ANOVA models as example. Computational Statistics and Data Analysis (2018), https://doi.org/10.1016/j.csda.2018.01.004.

ANOVA. The KR, KNN, SS-ANOVA regressions as well as kernel smooth regression, local polynomial regression and others are *linear smoothers*, meaning that the fitted values are a linear function of the outcome.

Several goodness-of-fit tests exist for nonparametric regression models. The literature contains results on testing a parametric form under the null hypothesis against a nonparametric (semiparametric) one under the alternative hypothesis. Examples include tests of deviation from the parametric linear regression (Fan and Huang, 2001; Kuchibhatla and Hart, 1996). Another method tests the goodness-of-fit of a linear model, which can potentially detect a nonparametric form under the alternative (Sen and Sen, 2014). A test that allows a nonparametric form under the null exists (Einmahl and Van Keilegom, 2008b), but it is only defined for a model with one covariate. A test that allows for multiple covariates in the Nadaraya–Watson (NW) regression model (Neumeyer, 2009) exists in the literature. Although similar to our method, the usage of the NW regression model is different from our current research for two fundamental reasons. First, the NW regression jointly models the mean of the outcome conditional on the covariates. Therefore, it is difficult to build examples where the conditional mean suffers from lack-of-fit and the test is essentially for heteroscedasticity of the residuals. The simulation scenarios are examples where the lack-of-fit comes only from the variance of the residuals being dependent on the covariates, and lack-of-fit coming from modeling the mean is never explored. Second, nonparametric methods like GAMs and SS-ANOVA models are not considered. Examples where the lack-of-fit comes from modeling the mean incorrectly could be explored by using these models. Given the popularity of these regression models, we believe it is important to develop a goodness-of-fit methodology for them.

Another article (Eubank et al., 1995) generalizes a Tukey-type test of additivity proposed in Hastie and Tibshirani (1990). This test can only detect two-way interactions that are a product of the main effects, which limits the type of departures of goodness-of-fit it can detect. Also, using generalized additive models (GAMs) a test for specific interaction terms exists (Roca-Pardiñas et al., 2005). That means that under the null the GAM has some main effects and interactions, and under the alternative the same model holds but with the addition of one more interaction. This is not a goodness-of-fit test since the user of this test has to specify the model under the alternative. A methodology exists to test whether an extra set of variables should be included in a nonparametric regression model (Delgado and Manteiga, 2001).

Our literature review shows that a general test for nonparametric regression that detects lack-of-fit in modeling the mean does not exist. This paper resolves these issues by proposing a goodness-of-fit test for nonparametric regressions that are *linear smoothers* as defined in (2) (i.e., regression models that make use of a matrix to transform the outcome vector into a vector of fitted values). Our method is an innovation with respect to previous methods in that:

- Our goodness-of-fit methodology tests a nonparametric model under the null against a general alternative, which can include parametric as well as nonparametric forms. Other methods require the model under the null to be parametric (Fan, 1996; Fan and Huang, 2001; Sen and Sen, 2014; Kuchibhatla and Hart, 1996), the regression to be univariate (Einmahl and Van Keilegom, 2008b), require the form to be multiplicative, or have specific interactions (Eubank et al., 1995), none of which are limitations in the current research.
- Other methods exist for testing the independence between residuals and covariates, as in the current research, but only in the context where the lack-of-fit comes from departures from the homoscedasticity assumption (Einmahl and Van Keilegom, 2008a; Neumeyer, 2009). The case where the lack-of-fit comes from incorrectly modeling the mean has not yet been analyzed. This is of importance given the fact that models like GAMs and SS-ANOVA have become highly popular as nonparametric models and they can suffer from lack-of-fit of the mean.
- Our method can incorporate testing of external variables, as Delgado and Manteiga (2001). Thus, we present a unified framework to test both for goodness-of-fit with respect to variables used to build the model or a set of external variables, a unification that was not attempted in the literature cited above.
- Our methodology provides a degree of freedom adjustment for the bootstrap null distribution, which was missing from the reviewed literature (Sen and Sen, 2014; Neumeyer, 2009).

We will use SS-ANOVA models throughout in examples, theory and simulations, but we emphasize that this methodology can be applied to any nonparametric linear smoother. The assessment of goodness-of-fit will be accomplished by fitting the model of interest and obtaining estimated residuals. The residuals contain the leftover information that remains unexplained by the model. Statistical dependence is then assessed between the estimated residuals and the covariates in the model, with the Hilbert–Schmidt independence criterion (HSIC). If dependence exists, the model does not capture all the variability in the outcome associated with the covariates. If no dependence exists, the model fits the data well. This process can also be used with covariates that are not in the model, in order to assess whether their absence contributes to lack-of-fit. A test statistic is created from the HSIC between residuals and covariates to test for lack-of-fit. The bootstrap is used to derive p-values. The degrees of freedom of the model are calculated as the trace of the hat matrix and are used to adjusted the variance of the bootstrap distribution. The current article is an extension of the goodness-of-fit test for linear models proposed by Sen and Sen (2014). The major contributions we make to the literature include: identifying the need for assessing goodness-of-fit in a nonparametric regression, developing a test statistic, creating a variance adjustment to the bootstrap to improve the finite sample performance of the method, providing theoretical justification the use of the HSIC, and demonstrating correct type I error as well as power performance through numerical simulations.

This paper is organized as follows. In Section 2 the method for goodness-of-fit for linear smoothers is introduced. Section 2 includes an introduction to linear smoothers, a formal definition of SS-ANOVA, a description of the evaluation of goodness-of-fit using the HSIC, the bootstrap for deriving p-values for the test statistic, and illustrative cases of lack-of-fit. In Section 3,

Please cite this article in press as: Teran Hidalgo S.J., et al., Goodness-of-fit test for nonparametric regression models: Smoothing spline ANOVA models as example. Computational Statistics and Data Analysis (2018), https://doi.org/10.1016/j.csda.2018.01.004.