Journal of Advances in Mathematical & Computational Sciences
An International Pan-African Multidisciplinary Journal of the SMART Research Group
International Centre for IT & Development (ICITD) USA
© Creative Research Publishers
Available online at https://www.isteams.net/ mathematics-computationaljournal.info
CrossREF Member Listing - https://www.crossref.org/06members/50go-live.html

Controlling Bias-Variance Trade-offs for Predictive Models

[1]Agu Sunday Clement & [2]Elugwu Felix
[1]Department of Computer Science, Benson Idahosa University, Benin City, Nigeria
[2]Department of Computer Science, Delta State Polytechnic, Otefe-Oghara, Nigeria
E-mail: sagu@biu.edu.ng; https://orcid.org/0000-0003-0635-1939
E-mail: felixelugwu@gmail.com; https://orcid.org/0009-0007-3153-5955

ABSTRACT

This study addressed underexplored areas in predictive model performance by reviewing methods for controlling bias and variability, using a state-of-the-art smoothing spline algorithm (SSA) to preserve data quality. To confirm the tenacity of the SSA to enhancing high accuracy of a model, we studied three factors that impact the performance of predictive models whose outcome revealed that (a) the highest bias in a predictive function g, produced an inflexible line that under-fits the model, (b) a very high variance yielded a wiggly curve, touching every data-point that over-fits the model, and (c) simulating the predictive function g with different values of $\boldsymbol{\lambda}$ aided in determining the most appropriate balance between the bias and variance that yielded optimal performance. The study also revealed that $\boldsymbol{\lambda} = \mathbf{0}$ produced high variability model that interpolates every data point. And $\boldsymbol{\lambda}$ tending towards infinity, $\boldsymbol{\lambda} \rightarrow \infty$, completely biased the model producing a linear fit, highlighting that not considering bias-variance trade-offs with smoothing spline would either over-fit or under-fit the model, resulting in low performance. We also pointed out that the leave-one-out cross-validation (LOOCV) algorithm performed best in selecting the value of $\boldsymbol{\lambda}$ that optimizes the model's fit performance by minimizing the LOOCV error.

Keywords: Smoothing spline, Bias-variance trade-offs, Turning parameter, Predictive model, Loss function, Overfitting, Underfitting

Agu Sunday Clement & Elugwu Felix (2026): Controlling Bias-Variance Trade-offs for Predictive Models. Journal of Advances in Mathematical & Computational Science. Vol. 14, No. 1. Pp 85-93. Available online at www.isteams.net/mathematics-computationaljournal. dx.doi.org/10.22624/AIMS/MATHS/V14N1P6

1. INTRODUCTION

Predictive models in healthcare are making a revolutionary change with regards to evidence-based decision-making, exploiting sophisticated data analysis to enhance healthcare delivery and patient results [1]. And the applications of the predictive models are gaining traction in recent times, that could be applied in other areas outside medicine [2], ranging from protraction of disease prediction to Risk stratification to identify patients needing early intervention [3], and in all the phases of production, implementing proactive risk stratification with integrated quality control and process improvement, leveraging failure mode prediction/prevention and continuous safety monitoring [4]. Comprehending prediction errors (bias and variance) is crucial in model predictions as there is a balance between a model's capability to reduce bias and variance, respectively, in building accurate models and preventing the risk of over-fitting and under-fitting the model [5]. Model's intricacies and universality have to be considered after minimizing the testing error of a predictive model [6]. As a result, bias-variance decomposition was established as the first stage for curbing the challenges faced in selecting a model based on the highest testing performance. Moreover, applying bias-variance trade-offs to learning theories and pedagogy yielded insights for productive educational debates [7]. Classic bias-variance tradeoff that underpins predictive performance addressed models selection that optimally balanced complexity with generalization capability [8], which deployment in real educational settings could lead to unintended harm if the predictions are biased [9].

The majority of the machine learning methods either over-fit or under-fit predictive models in the presence of autocorrelated errors due to the presence of noise, which the smoothing spline tends to filter this when predicting nonparametric regression, forecasting, spatial, and survival models whose performance is based on the selected smoothing parameter [10]. Computing a non-parametric path function to determine the causal relationship between extrinsic and intrinsic factors using a smoothing spline yielded more flexible function estimations for data patterns with close accuracy to the real data pattern [11]. An adaptive smoothing spline (AdaSS) estimator algorithm with optimization tuning parameters comparatively adapted better to the true coefficient function over regions of large curvature without being under-smoothed over the remaining part of the domain [12]. According to [13], the design of standard growth charts to assess the nutritional status of toddlers using a multiresponse semiparametric regression (MSR) model was estimated using a smoothing spline.

[14] and [10] established that most of the regression splines such as piecewise-polynomial, cubic spline, and natural splines produce smooth curves for predictive models but they do not provide mechanisms for controlling the bias-variance trade-offs giving the propensity of yielding overly flexibly model Hence, this study reviews balancing reduction in the bias and variance of a predictive model using the smoothing spline method that reduces the residual sum of squares (RSS), as well as produces a smooth curve by adding some constraints that controls the bias-variance trade-offs to maximize predictive accuracy.

2. MATERIALS AND METHODS

Predictive modeling is a part of Machine learning that allows accurate prediction of outcomes, and its techniques use existing data as input to forecast new output values [15]. It centers on fitting models that enhance accuracy based on the accepted data [16]. Generally, an optimal model is influenced by balancing the bias and the variability of a model which has been featured in a few machine learning approaches, such as ridge and lasso regression, and also in statistical thresholds, which revealed that adjusting the probability values plays the role of harmonizing the bias and the variance of the model [14]. The high-low algorithm uses the highest and lowest task levels over the period to approximate models, compute features, and estimate fixed costs.

Doubtlessly, this could be deceptive when the high and low task levels are aberrant from the normal activity at the majority of the data points [17]. To overcome under-fitting/over-fitting and produce an essential equation, the model's feature complexity needs to be reduced, and powers have to be added to transform the original features into polynomial features [18].

2.1 Bias-Variance Trade-Offs

The interrelation of the target variable Y and the other covariates X in Eq. (1)

$$\mathrm{Y} = \mathrm{f}(\mathrm{X}) + \mathrm{e} \qquad (1)$$

where the error term e is normally distributed with mean zero. Fitting a model $\hat{\mathrm{f}}(\mathrm{X})$ of $\mathrm{f}(\mathrm{X})$, the expected squared error at a point x_0 is Eq. (2) [7].

$$\mathrm{Err}(\mathrm{x}_0) = \mathrm{E}\left[\left(\mathrm{Y} - \hat{\mathrm{f}}(\mathrm{x}_0)\right)^2\right] \qquad (2)$$

According to [12], Eq. (2) is the mean square error (MSE), which is the result of two competing properties that can be broken into the sum of three basic quantities: the variance of $\hat{\mathrm{f}}(\mathrm{x}_0)$, the squared bias of $\hat{\mathrm{f}}(\mathrm{x}_0)$ and the variance of the irreducible error ϵ, Eq. (3)

$$\mathrm{E}(\mathrm{y}_0 - \hat{\mathrm{f}}(\mathrm{x}_0))^2 = \mathrm{Var}\left(\hat{\mathrm{f}}(\mathrm{x}_0)\right) + [\mathrm{Bias}(\hat{\mathrm{f}}(\mathrm{x}_0))]^2 + \mathrm{Var}(\epsilon) \qquad (3)$$

Remark I:

According to [7], bias occurs due to the inability of a model to comprehend the hidden pattern of the data especially when little amount of data is used to fit the model, leading to increase in bias and reduced variance, producing linear fit, and under-fitting the model Figure 1a, whereas high variability results when the model captures the babble as well as the core principles in the data, leading to reduced bias and increased variance, yielding wiggly fit in an attempt to interpolate all the data-points, and over-fits the model Figure 1b. Balancing the bias and variance generates a good balance, giving an optimal model with a smooth curve Figure 1c

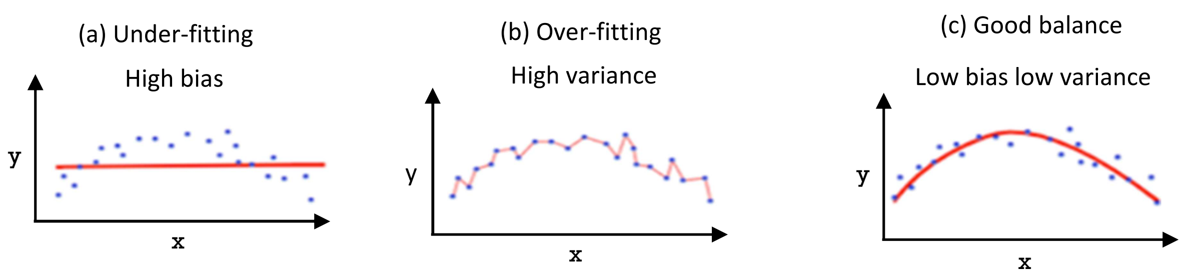

Figure 1 (A) Under-Fitted Model (B) Over-Fitted Model (C) Optimal Model [5]

Thus, the bias-variance trade-off theory seeks a method that reduces the bias and reduces the variance introduced by a predictive model's fit for optimal performance.

2.2 Smoothing Spline

Based on the concepts of bias-variance trade-offs, the smoothing spline improves on other regression spline approaches by finding a function g(x) that aligns the historic data such that the RSS is reduced. However, without constraining $g(x_i)$, the value of RSS can be reduced to zero by selecting g that touches all of the y_i. A function such as this would disappointingly over-fit the data. The limitation is resolved by adding some constraints on $g(x_i)$. Hence, the function g (also known as the smoothing spline) is sought to minimize Eq. (4) [14].

$$\sum_{\mathbf{i}}^{\mathbf{n}}\left(\mathbf{y_i}-\mathbf{g}(\mathbf{x_i})^{\mathbf{2}}\right) + \boldsymbol{\lambda}\int \mathbf{g''t^2dt} \qquad (4)$$

Where
λ is a positive tuning argument.

Applying the "Loss + Penalty" formulation, the term $\sum_{i=1}^{n}(y_i - g(x_i))^2$ becomes the RSS, popularly known as the loss function, that empowers g to align the data appropriately while $\lambda\int g''(t)^2 dt$ is the added constraint, popularly known as the penalty term, that penalizes variability in g and encourages g to be smooth.

Remark II:

The higher the λ , the smoother the g. However, if (lambda tends towards infinity), then g becomes a smooth straight line that approximates the training data points as closely as possible, Figure 2. This implies that we have the highest bias in g, which makes g rigid (inflexible) and under-fits the model. As the loss function in Eq. (4) only results in reducing the RSS, the scenario models g as a linear least squares fit

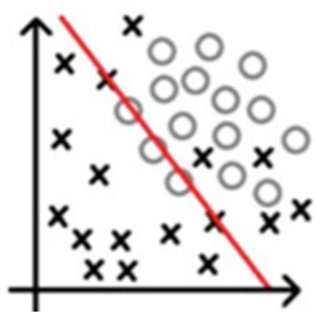

Figure 2: $\lambda \to \infty$ Under-Fits The Model [1]

Remark III:

The penalty in Eq. 4 has no impact when λ is 0, thereby making the function g become too jumpy, so that it interpolates every observation in the training dataset. Figure 3. This implies that g has very high variance, which makes g highly flexible. This case builds g as a wiggle curve, since the penalty term in Eq. (1) only amounts to maximizing the variability in g.

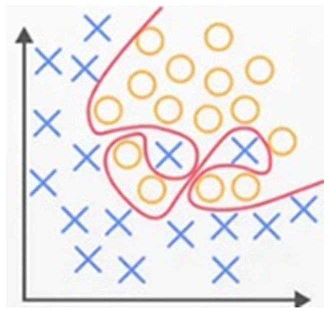

Figure 3: λ = 0 Over-Fits The Model [19]

Remark IV:
When λ is tuned to an intermediate value, g estimates the training observations and ensures that the curve is somewhat smooth Figure 4. This implies that the λ can be turned by simulating/selecting different values and observing which of the values regulates the bias-variance trade-offs of the smoothing spline g, to yield optimal performance.

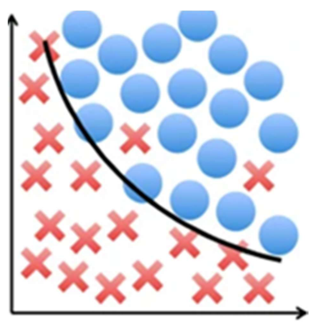

Figure 4: An Ideal Model, λIs Adjusted To Best Fit Curve [20]

2.3. Justifications

Smoothing spline presumably experiences excessive degrees of freedom, like other regression splines, where for each data point, a knot is granted much pliability. However, λ applied the effective degrees of freedom [14]. Ordinarily, the degree of freedom refers to the number of adjustable variables n, such as the number of coefficients in a polynomial or cubic spline. In contrast, effective degrees of freedom constrain or shrink the n parameter. In fact, increasing λ increases from 0 to ∞, the model's effective degrees of freedom steadily decline - from n down toward 2. Thus, the effective degrees of freedom quantify the flexibility of a smoothing spline - the more the effective degrees of freedom, the higher the variance (more flexibility), and the lower the bias of the smoothing spline. Just as other regression splines, the smoothing spline g(x) that gives Eq. (4) the minimum has specific unique attributes: (a) It is a piecewise cubic polynomial with knots positioned at the distinct values of the $x_1, . . . ,\ x_n$. (b) It marks continuity of both the first and second derivatives at each knot, and (c) It behaves as a linear function beyond the outermost knots.

This implies that the smoothing spline g(x) that reduces Eq. (1) is a natural cubic spline with knots located at each observed $x_1, . . . , x_n$. but it differs from the natural cubic spline found in the regression spline.

2.4. Choosing Lambda λ

Unlike other regression splines, which place knots at each training observation $x_1, . . . , x_n$. and require selecting both the number and locations of these knots during model fitting, the smoothing spline, in contrast, chooses a value of lambda that optimizes the model's fit performance by minimizing the leave-one-out cross-validated (LOOCV) error, RSS. Using LOOCV, Figure 5 depicts the smoothing spline fits for the Wage dataset [14]. The red curve shows the smoothing spline fitted with 16 effective degrees of freedom. The blue curve presents the smoothing spline when λ was automatically found by LOOCV, resulting to 6.8 effective degrees of freedom. The red curve appeared slightly wigglier. Thus, the smoothing spline fit with 6.8 degrees of freedom is recommended because it offers simpler models and is generally superior, except when the dataset supports a more complex model.

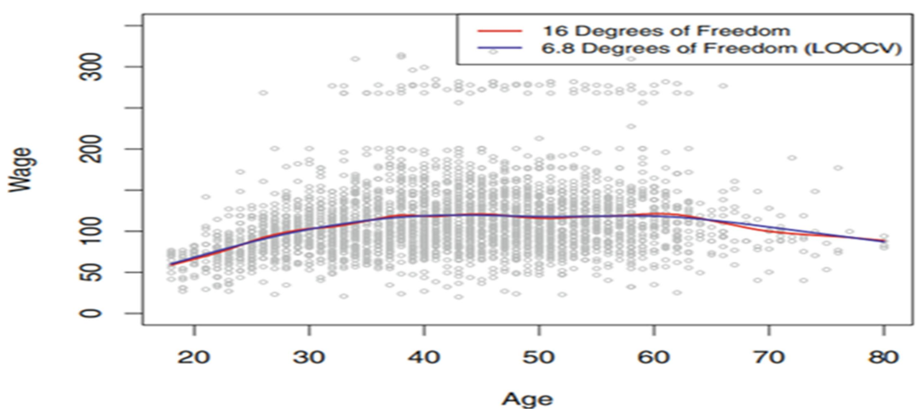

Figure 5: Smoothing spline fits to the Wage dataset [12]

3. RESULT AND DISCUSSION

Turning parameter λ was applied in RR and LR in predicting Nigeria's GDP [21]. This study used 5-fold Cross-validation (5-fold CV) metrics to evaluate models developed using three regression approaches for the analysis of Nigeria's GDP data, showing the 5-fold CV Mean Squared Errors (MSEs) for ordinary least squares, ridge regression, and lasso regression. The findings revealed that the Ridge regression has the minimal MSE with a turning parameter = 100. Predictive models play a central role in timely, effective decision-making across science, medicine, and other domains. Evaluation metrics must meet significance criteria and, in the presence of target variables, also demonstrate predictive power. Primarily, seven bias-variance trade-off optimization techniques, namely, kernel parameters C, hypertuning parameters, λ tuning parameter, probability threshold, Unbiased Risk (UBR) smoothing parameter, Generalized Maximum Likelihood (GML), and Generalized Cross-Validation (GCV) have been reviewed in relation to the smoothing spline technique under study. In agreement with this work, the support vector machine (SVM) applied kernel parameters C and σ\sigmaσ to regulate the bargain between increasing the margin/decreasing classification error and the complexity of the decision boundary to obtain the highest accuracy of 98.8% over other methods in early detection of heart disease [22]. According to [23], the hypertuning parameters were employed to control early diagnosis of breast cancer, achieving accuracies of 100% and 99.30% in the random forest model, respectively, on a holdout set. According to [10], in a time series observations with auto-correlated error levels of (ρ=0.2, 0.5, and 0.8) using spline technique out-weighted the contemporary approaches with the MSE of 0.04.

CONCLUSION

With progress on both theoretical and computational fronts, smoothing spline modeling has become an established tool in machine learning regression and classification. Higher predictive potency might be achieved in some cases, . Smoothing splines offer a captivating equilibrium of predictability in line with the accuracy of predictions by controlling the two vital prediction errors (bias and variance) to strike a balance between under-fitting and over-fitting the predictive model in the fields of science and medicine to ensure timely and informed decisions.

ACKNOWLEDGMENT

The authors wish to acknowledge the educational resources provided by our various institutions and friends who have contributed to the actualization of this journal.

Ethical Statement
The study is proper with ethical standards.

Financial Support for the Study (Funding)
This study did not receive any financial support.

Presentation Information
The data from this study has not been presented anywhere

Conflicts of Interest
The authors declare no conflicts of interest regarding this study.

Author Contributions
Agu Sunday Clement contributed to the conceptualization of the study, methodology/design, data collection, and writing – original drafting. Felix Elugwu contributed to the writing – reviewing and editing, Investigation, and Validation.

REFERENCES

[1] Montesinos, L. O. A., Montesinos, L. A., & Crossa, J. (2022). Overfitting, model tuning, and evaluation of prediction performance. In Multivariate statistical machine learning methods for genomic prediction (pp. 59-83). Springer. https://doi.org/10.1007/978-3-030-89010-0_4

[2] Lingxiao, C. (2020). Overview of clinical prediction models. Annals of Translational Medicine 8(4) 71. doi: 10.21037/atm.2019.11.121. PMID: 32175364; PMCID: PMC7049012.

[3] Toma, M., & Ong, C. W. (2023). Predictive modeling in medicine. Encyclopedia, 3(2), 590-601. https://doi.org/10.3390/encyclopedia3020042

[4] Dorota, O. (2021). Advanced Predictive Analytics in Chemical Manufacturing: Improving Your Chemical Operations. https://nexocode.com/blog/posts/predictive-analytics-in-chemical-manufacturing/ Accessed 31/10/2024

[5] Vishnuprasanth, V. K. (2024, January 10). Understanding the bias-variance tradeoff in machine learning. AI Accelerator Institute. https://www.aiacceleratorinstitute.com/understanding-the-bias-variance-tradeoff-in-machine-learning/

[6] Xingquan, G. and Henry, B. (2022). Bias-variance tradeoff in machine learning: Theoretical formulation and implications to structural engineering applications. Structures. 46(2022).17-30. https://doi.org/10.1016/j.istruc.2022.10.004Chen, L. (2020). Overview of clinical prediction models. Annals of Translational Medicine, 8(4), 71. https://doi.org/10.21037/atm.2019.11.121

[7] Doroudi, S. (2020). The bias-variance tradeoff: How data science can inform educational debates. AERA Open, 6(4). https://doi.org/10.1177/2332858420977208

[8] Mistry H. (2025) Mastering Model Selection for AI/ML Models, European Journal of Computer Science and Information Technology,13(14),55-67

[9] Baker, R. S. & Hawn, A. (2022). Algorithmic bias in education. International Journal of Artificial Intelligence in Education, 32(4), 1052–1092. https://doi.org/10.1007/s40593-021-00285-9

[10] Adams, S. O., & Asemota, O. J. (2023). The efficiency of the proposed smoothing method over the classical cubic smoothing spline regression model with autocorrelated residual. Journal of Mathematics and Statistics Studies, 4(1), 26-40. https://doi.org/10.32996/jmss.2023.4.1.3

[11] Iriany, A., & Fernandes, A. A. R. (2023). Hybrid Fourier series and smoothing spline path non-parametrics estimation model. Frontiers in Applied Mathematics and Statistics, 8, 1045098. https://doi.org/10.3389/fams.2022.1045098

[12] Centofanti, F., Lepore, A., & Menafoglio, A. (2023). Adaptive smoothing spline estimator for the function-on-function linear regression model. Computational Statistics, 38, 191-216. https://doi.org/10.1007/s00180-022-01223-6

[13] Budi, L., Chamidah, I. N., Nyoman, B., & Dursun, A. (2023). Determining confidence interval and asymptotic distribution for parameters of multiresponse semiparametric regression model using smoothing spline estimator. *Journal of King Saud University - Science, 35*(5), 102664. https://doi.org/10.1016/j.jksus.2023.102664

[14] Gareth, J., Daniela, W., Trevor, H., and Robert T. (2017). An Introduction to Statistical Learning with Application in R. Springer, New York

[15] TechTarget. (2020). In-depth guide to machine learning in the enterprise [White paper]. https://media.techtarget.com/digitalguide/images/Misc/EA-Marketing/Eguides/In-depth_Guide_to_Machine_Learning_in_the_Enterprise.pdf
[16] OCI. (2020). What is Data Sciene? https://www.oracle.com/ng/what-is-data-science/#challenges Accessed 17 December 2024.
[17] Isaac, O. A., Adedeji, A. A., & Ismail, I. R. (2012). Polynomial regression model of making cost prediction in mixed cost analysis. Mathematical Theory and Modeling, 2(2), 14-24. https://www.iiste.org/Journals/index.php/MTM/article/view/1240
[18] Nikhil, Arushi, S., Santu, P., & Neha, G. (2021). Polynomial based linear regression model to predict COVID-19 cases. In Proceedings of the 6th International Conference on Recent Trends on Electronics, Information, Communication & Technology (pp. 28-29). https://scite.ai/reports/polynomial-based-linear-regression-model-LegERdnm
[19] Ambarish, S. (2022). Overfitting and Underfitting in practice: common scenarios and solutions.https://medium.com/@Ambarish_224/overfitting-and-underfitting-in-practice-common-scenarios-and-solutions-407d5a3de05a. Accessed 20 October 2024.
[20] Radiya-Dixit, E., Zhu, D., & Beck, A. H. (2017). Automated classification of benign and malignant proliferative breast lesions. Scientific Reports, 7, 11990. https://doi.org/10.1038/s41598-017-10324-y
[21] Agu, S. C., Onu, F. U., Ezemagu, U. K., & Oden, D. (2022). Predicting gross domestic product to macroeconomic indicators. Intelligent Systems with Applications, 15, 200082. https://doi.org/10.1016/j.iswa.2022.200082
[22] Ahmad, A. A., & Polat, H. (2023). Prediction of heart disease based on machine learning using jellyfish optimization algorithm. Diagnostics, 13(14), 2392. https://doi.org/10.3390/diagnostics13142392
[23] Manas, M., & Batra, V. (2023). Diagnosis of breast cancer using random forests. Procedia Computer Science, 218, 429-437. https://doi.org/10.1016/j.procs.2023.01.025