ISSN 1512-1232

Bias-Variance Trade-Off for Machine Learning Algorithms

Magda Tsintsadze[1], Archil Elizbarashvili[1],

[1]Ivane Javakhishvili Tbilisi State University (TSU)

Abstract

Supervised machine learning algorithms use past knowledge to predict future events based on new data. In this process, the model acquires knowledge of the past thanks to the selected labeled examples. Having labeled data is essential in supervised machine learning, and as more varied the data is, the more effective the models are [1] [2] [3].

Different metrics are used to evaluate the complex model and it depends on the task itself. In binary classification tasks, metrics can be accuracy, precision, recall and F1-score. The following loss functions are used to evaluate the performance of the regression model: Mean Error (ME), Mean Absolute Error (MAE), Mean Squared Error (MSE) or Root Mean Square Error (RMSE),

In order to better train the model on existing data and acquire flexibility, we must artificially introduce some noise into the model, the so-called bias and elaborate on these data. A model trained by introducing noise too quickly or too slowly can underfit or overfit the training data, which is why a good prediction cannot be made based on the test data. On one hand, the model must learn well on the training data and on the another - it must generalize this knowledge well to new data as well. Achieving it perfectly is a difficult task. It is important to balance both the bias and the variance and choose a compromised option.

The bias-variance tradeoff is a fundamental challenge in supervised machine learning models because it involves finding the right balance between model complexity and generalization performance. In general, a model with high bias (i.e., underfitting) is too simplistic and may not capture the underlying patterns in the data, while a model with high variance (i.e., overfitting) is too complex and may fit the noise in the data rather than the true patterns.

A model with high bias typically has a high training error because of underfitting and high-test error, indicating that the model fails to learn important patterns and relationships in the data leading to inadequate predictive power. On the other hand, a model with high variance has a low training error because of overfitting by closely fitting the training data but a high test error because the model fails to generalize well beyond the training data.

The Bias-Variance Trade-Off is a widely recognized challenge in machine learning algorithms. In the work we present one of the possible solutions how the bias and variance can be optimally changed to obtain an acceptable predicting model.

In our experiment, we identified the point where the algorithm makes accurate predictions without having excessive bias or variance, thus we can optimize the model's performance and enhance its generalization ability. This is essential for developing machine learning models that can perform well on both the training set and unseen data.

ISSN 1512-1232

Additionally, our experiment can provide insights into the model selection process and aid in determining the appropriate complexity of the model, which is a critical aspect of developing effective machine learning models.

Keywords: bias, variance, machine learning, trade-off.

1. Introduction

Machine learning is a field of computer science that uses artificial intelligence to enable machines to learn and improve their methods through experience, without requiring direct programming. The term "machine learning" was coined in 1959 by Arthur Samuel in his article "Some Studies in Machine Learning Using the Game of Checkers" [4].

Unlike traditional computer programming, where the programmer determines the solution or decision, machine learning uses data as input to build decision-making models. These models make decisions by analyzing the interrelationships between the data, using statistical methods. This means that the output of decision models is based on the input data, rather than on predefined rules set by the programmer.

The learning process for a machine starts with observations or data, from which it looks for patterns and makes decisions based on the examples provided. This allows computers to learn automatically without human intervention and adjust their behavior accordingly.

Although the machine is responsible for making the final decision, the human programmer is responsible for providing the model with input data, selecting the appropriate algorithm, and specifying its parameters.

Machine learning models can improve their predictions based on experience, just as humans do, by considering the success or failure of previous attempts before making the next decision [5]. Today, machine learning is used in various applications such as search engines to find relevant information, fraud detection in information systems, recommendation engines [6], and sentiment analysis in texts [7].

For instance, YouTube uses machine learning models to offer users personalized recommendations for videos. The algorithm determines the user's preferences by analyzing their engagement with videos - which videos they liked, rewatched, or subscribed to. The model can also evaluate the content of the watched videos, such as sports footage, fantasy genre, or historical videos, to match these characteristics to the user's taste. With each login to the channel, the user is offered new videos of their taste, based on the observed behavior of the user.

Several techniques can be used for decision-making in machine learning, such as statistics-based algorithms, automated analytical models, and neural networks - a machine learning architecture based on a biological neuron. However, choosing the appropriate algorithm for a specific task remains a challenge in this field [8].

2. Machine learning methods and bias-variance trade-off

The bias-variance trade-off is an important concept in machine learning, and it applies to a wide range of machine learning methods, including both traditional statistical methods and modern deep learning models. For example, in linear regression, increasing the number of features or the complexity of the model can increase variance, while decreasing bias. In contrast, reducing the number of features or using a simpler model can reduce variance, but increase bias. Similarly, in decision trees, increasing the depth of the tree or the number of splits can increase variance, while

ISSN 1512-1232

decreasing bias. Reducing the depth or complexity of the tree can reduce variance but increase bias. In deep learning, the bias-variance trade-off is often addressed using regularization techniques, such as L1 or L2 regularization, dropout, or early stopping. These techniques help to control model complexity and prevent overfitting, which can increase variance. Bias-variance trade-off is a fundamental concept in machine learning, and it is important to carefully balance bias and variance when designing and training machine learning models.

Balancing deviation and variance is crucial to achieving good predictions on new data. The Bias-Variance Tradeoff is one of the challenges in machine learning, as models must learn well on training data while also generalizing that knowledge to new data. The model's errors may decrease if we slowly introduce noise during training, but at a certain point, the errors in predicting test data will increase, indicating overfitting. The point at which predictions deteriorate is called the bias-variance trade-off point. If the model does not fit the training data well, indicating underfitting, we cannot achieve good predictions on the test data for complex tasks. We must find a balance between overfitting and underfitting the existing data in order to achieve good predictions on new data. It is important for a machine learning model to learn from the training data and not from the test data to avoid overfitting and ensure good generalization performance on new, unseen data. Figure 1 illustrates this concept. Introducing some noise (deviation or bias) during training can improve the model's flexibility on different data. Ideally, the trained model should have low bias and low variance.

There are three main approaches in machine learning:

- supervised learning
- unsupervised learning
- reinforcement learning

ISSN 1512-1232

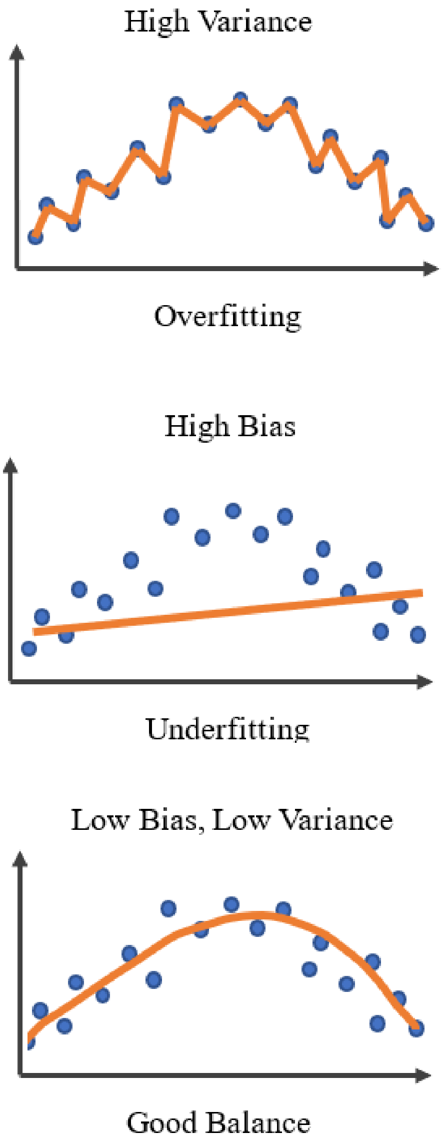

Fig 1: Variance and Bias different cases

In supervised learning, the algorithm learns from labeled training data to create a mapping function that can make good predictions on test data. This mapping function is also called the target function, as it approximates the correct results. The prediction error of a machine learning model can be decomposed into three parts: bias error, variance error, and irreducible error.

2.1. Irreducible error

Irreducible error, as the name suggests, cannot be reduced no matter how good the algorithm is. This error stems from the intrinsic structure of the task and may be caused by various factors, such as unknown variables affecting the mapping function.

2.2. Bias Error

Bias error refers to the difference between the mean prediction of our model and the actual value we are trying to predict. A model with high bias underfits the training data by oversimplifying the model, resulting in large errors on both training and test data. In general, parametric algorithms

ISSN 1512-1232

have high bias, which makes the model computationally efficient, but not flexible enough to fit the test data. On the other hand, a low bias indicates a more complex model that can make accurate predictions on a complex task. A low bias suggests fewer assumptions about the shape of the target function, whereas a high bias suggests more assumptions about it.

Machine learning algorithms such as decision trees, K-nearest neighbors, and support vector machines are known for having low bias. On the other hand, linear regression and logistic regression are characterized by high bias, which may result in oversimplification of the model and poor predictions.

2.3. Variance Error

Variance is a measure of how much the predictions of a model vary when different training data are used. A model with high variance pays close attention to the training data and may not generalize well to new, unseen data. This results in good performance on the training data but high error rates on the test data. On the other hand, a low variance model is less influenced by the training data and is more likely to generalize well to new data. It produces consistent predictions with small changes in the training data. In general, non-parametric machine learning algorithms that are highly flexible have high variance, such as decision trees. In contrast, linear regression, linear discriminant analysis, and logistic regression have low variance, and they produce consistent predictions with small changes in the training data. The ultimate goal of any machine learning algorithm is to achieve low bias and low variance, i.e., to be able to make accurate predictions on both the training and test data. Figure 2 illustrates the ideal prediction, which is the center of the target. Moving away from the center indicates that the predictability of the model is decreasing.

ISSN 1512-1232

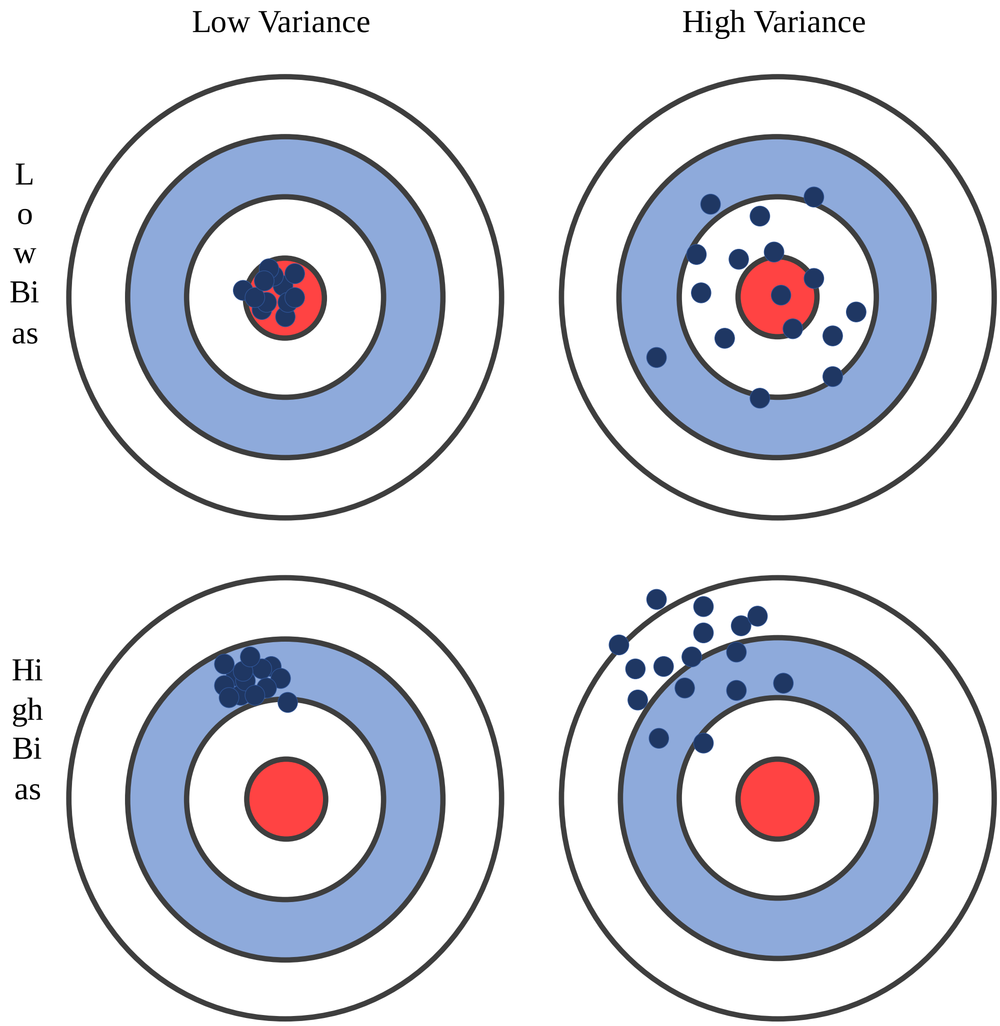

Fig 2: low/high bias/variance

In machine learning, there is often a trade-off between bias and variance, with parametric or linear algorithms typically having high bias and low variance, while non-parametric and non-linear algorithms have low bias and high variance.

For example, the K nearest neighbor algorithm tends to have low bias and high variance, but the balance between the two can be adjusted by changing the value of K. Increasing K, increases the number of neighbors used to make a prediction, which can improve the accuracy of the model, but may also increase the bias.

Similarly, the support vector machine algorithm typically has low bias and high variance, but adjusting the C parameter can change this balance. Increasing C can reduce the variance but increase the bias by allowing the model to tolerate more boundary violations in the training data.

This relationship between bias and variance is unavoidable, as increasing one tends to decrease the other. See Figure 3 for a visual illustration.

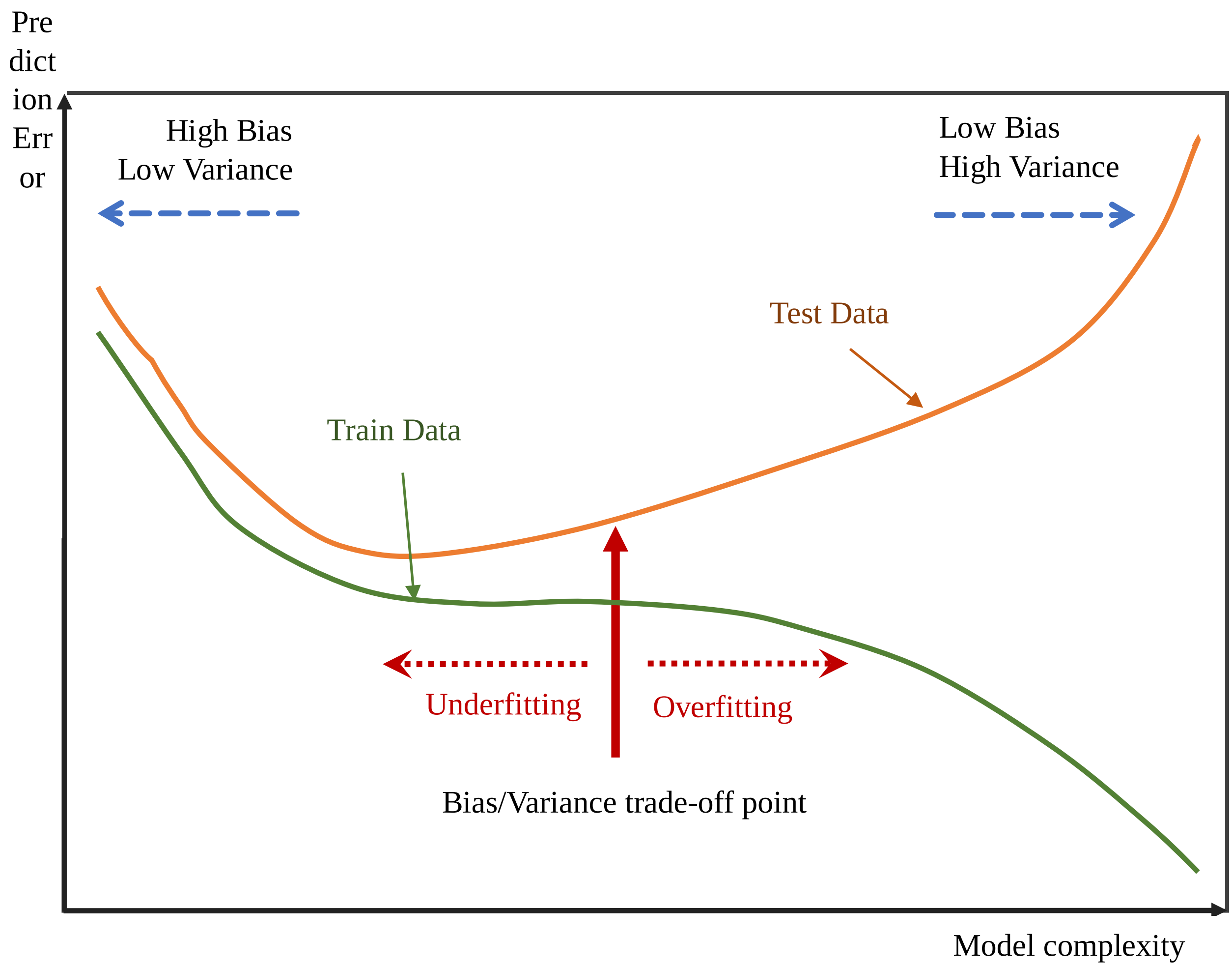

Fig 3: Bias-Variance trade-off

In reality, we cannot calculate the exact values of bias and variance errors because we do not have access to the true target function. Nonetheless, understanding the concepts of bias and variance can help us adjust the behavior of a machine learning algorithm in order to improve its predictive performance. By modifying the balance between bias and variance, we can aim to create a model that performs well not only on the training data but also on new, unseen data.

3. The mathematical formalization of the trade-off problem between variance and bias

Let's assume that we have a variable Y we want to predict based on some data X. Their relationship can be expressed as $Y = f(X) + e$, where e is a normally distributed error term with a mean of 0. Our goal is to approximate $f(X)$ using a model such as linear regression, denoted as $\hat{f}(X)$. The expected squared error at a point x is given by:

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

Expanding this expression, we obtain:

$$Err(x) = (E[\hat{f}(x)] - f(x))^2 + E\left[(\hat{f}(x) - E[\hat{f}(x)])^2\right] + \sigma_e^2$$

$$Err(x) = Bias^2 + Variance + Irreducible\ Error$$

This equation shows that the error at a point x is the sum of three terms: the squared bias, the variance, and the irreducible error (σ_e^2). The bias represents the error due to the difference between the expected value of the model's predictions and the true value of $f(x)$). The variance represents the error due to the variability of the model's predictions when trained on different datasets. The irreducible error is the error that can't be reduced by any model, regardless of its complexity.

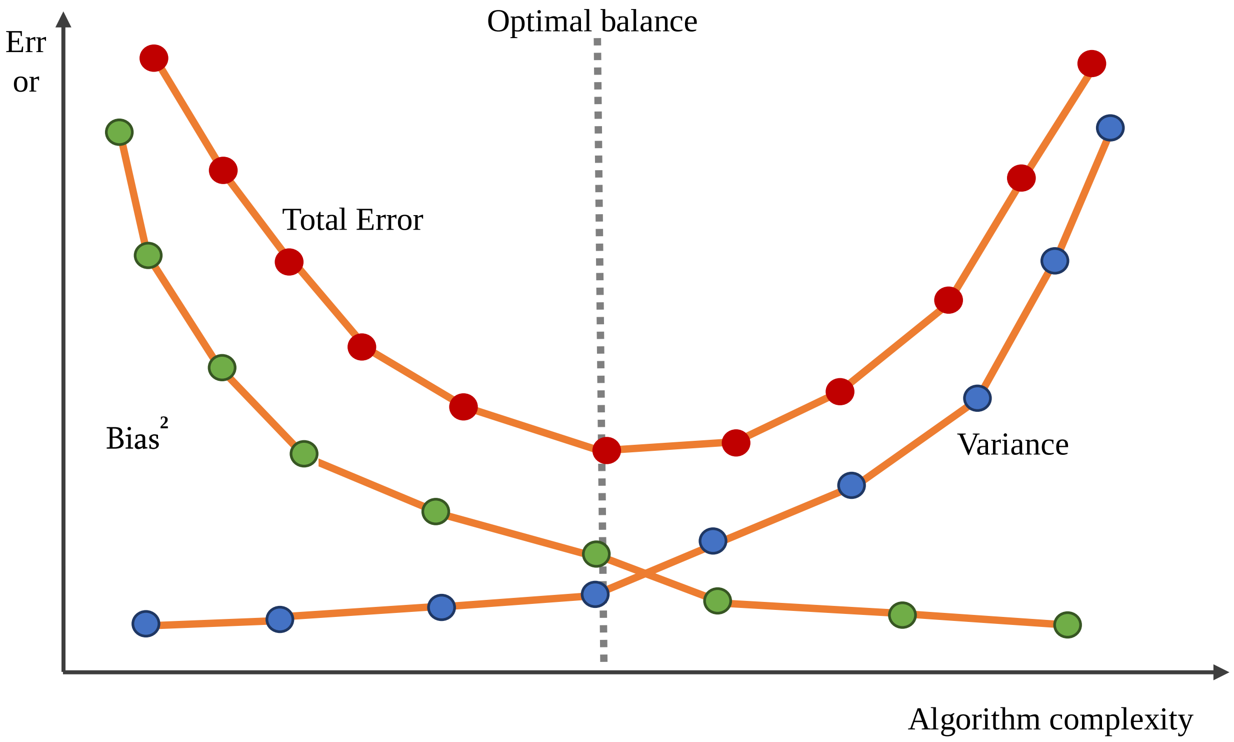

Fig 4: Bias-Variance Trade-off

Therefore, the total error of the model is the sum of the squared bias, variance, and irreducible error.

Our goal is to find a model that has low total error. See Figure 4. A model with an optimal balance of bias and variance can generalize well to new data and avoid overfitting or underfitting. Hence, the bias-variance trade-off is crucial for building accurate predictive models [11].

4. The Bias-Variance Trade-off project

The goal of supervised machine learning is to find a function (f) that accurately predicts the relationship between the input predictor (x) and the observed output (y):

$$y = f(x) + \varepsilon$$

where ε is the noise of the data.

We created our synthetic x and y by choosing a cosine wave function:

$$y = cos(2\pi x) + \varepsilon$$

We assume a normally distributed noise, with mean = 0 and variance = 1.

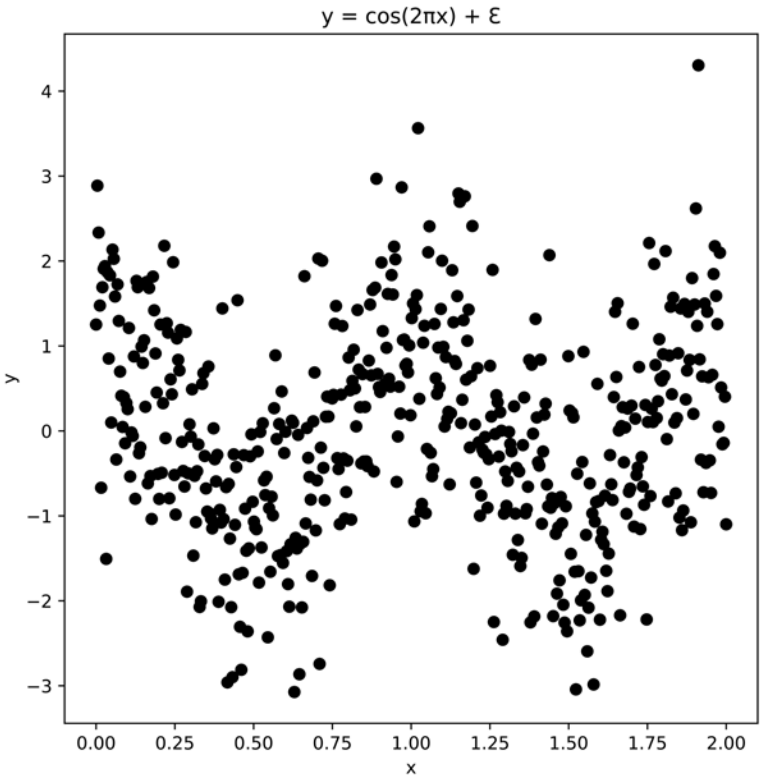

Fig 5: cosine wave function

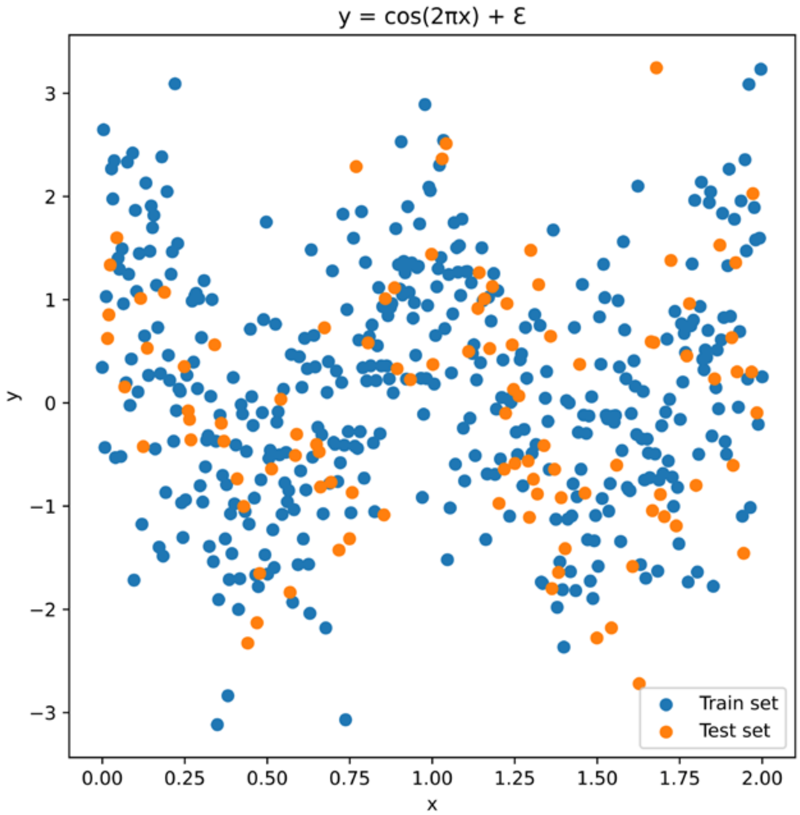

Fig 6: cosine function splatted into train and test data sets.

ISSN 1512-1232

Our objective is to minimize the deviation of the predictor output value from the observed value. Let's take the mean squared error (MSE) as a metric for this. And show that this error of the model consists of the square of the deviation, the variance, and the irreducible error. Since we cannot change the irreducible error, let's focus on the deviation and variance errors. To illustrate the bias-variance trade-off, let's fit several different polynomial functions (models) to the training data with increasing polynomial degree and observe the trend in MSE error as model complexity increases. See Figure 7.

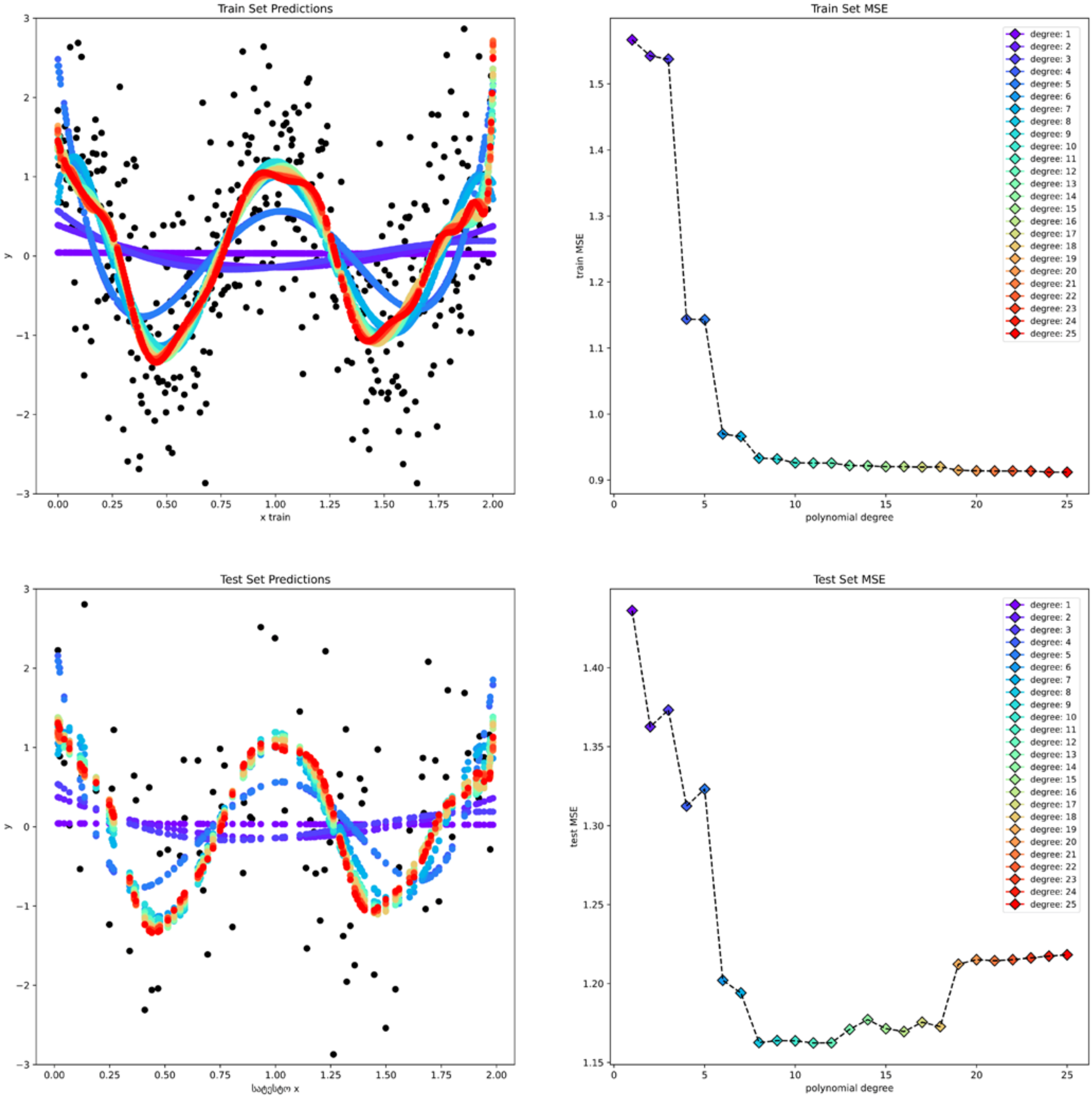

Fig 7: variation of MSE trough polynomial degree

By observing the MSE plots while varying the polynomial degree (model complexity) from left to right, we can see that as the models become more complex, the train error decreases (reducing bias), but the test error increases significantly (increasing variance). Conversely, when the model complexity decreases, the performance on the train set is poor (increasing bias) but the generalization on the test set improves (reducing variance).

The best fit is represented by the polynomial degree that minimizes the test error and its value is 10.

Conclusion

The paper provides an overview of machine learning approaches, their relevance, and the scope of their application. It discusses current challenges in the field, particularly the need to balance bias-variance in order to achieve better predictive accuracy.

While the actual deviation and variance error cannot be calculated in most cases due to the unknown target function, they can be used to modify the behavior of machine learning algorithms.

The paper describes a concrete example of how to balance bias-variance by dividing data into training and testing sets and reducing the bias of the output value of the predictor from the observed value.

The paper emphasizes the bias-variance trade-off because irreducible error cannot be changed. To illustrate this trade-off, the paper fits several polynomial functions with increasing degree to the training data and shows the trend in mean squared error (MSE) as model complexity increases.

The paper concludes by identifying a specific point in the bias-variance trade-off where decreasing model complexity results in poorer performance on the training data.

Acknowledgement: This research PHDF-22-1840 is supported by Shota Rustaveli National Science Foundation of Georgia (SRNSFG).

References:

1. Aurélien Géron. “Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow”. O’Reilly Media, Second Edition, 2019.
2. Gregory Druck, Gideon Mann, Andrew McCallum. “Learning from Labeled Features using Generalized Expectation Criteria”. SIGIR 08: Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval. July 2008, Pages 595–602.
3. Malti Bansal and Apoorva Goyal and Apoorva Choudhary. „A comparative analysis of K-Nearest Neighbor, Genetic, Support Vector Machine, Decision Tree, and Long Short Term Memory algorithms in machine learning“, Decision Analytics Journal, 2022, Voluem 3, 100071, 2772-6622.

ISSN 1512-1232

4. Samuel, Arthur L. "Some Studies in Machine Learning Using the Game of Checkers". IBM Journal of Research and Development (1959). 44: 206-226. CiteSeerX 10.1.1.368.2254
5. Jeremy Watt, Reza Borhani, Aggelos Katsaggelos. Machine Learning Refined: Foundations, Algorithms, and Applications. Cambridge University Press; 2nd edition, 2020
6. Thomas, Binu, John, Amruth K. Machine Learning Techniques for Recommender Systems – A Comparative Case Analysis- IOP Conference Series: Materials Science and Engineering. IOP Publishing, 2021.
7. Nineli Lashkarashvili, Magda Tsintsadze, Toxicity detection in online Georgian discussions, International Journal of Information Management Data Insights,Volume 2, Issue 1,2022
8. Sarker, I.H. Machine Learning: Algorithms, Real-World Applications and Research Directions. SN COMPUT. SCI. 2, 160 (2021). https://doi.org/10.1007/s42979-021-00592-x
9. Erica Briscoe & Jacob Feldman. Conceptual complexity and the bias/variance tradeoff, Elsevier, Cognition, Volume 118, Issue 1, January 2011, Pages 2-16
10. Brady Neal , Sarthak Mittal, Aristide Baratin, Vinayak Tantia, Matthew Scicluna, Simon Lacoste-Julien, Ioannis Mitliagkas, A Modern Take on the Bias-Variance Tradeoff in Neural Networks, ICLR 2019 Conference, 2018
11. Mikhail Belkin, Daniel Hsu, Siyuan Ma, and Soumik Mandal, Reconciling modern machine-learning practice and the classical bias–variance trade-off, July 24, 2019, PNAS, Vol. 116 | No. 32

7 figures are used in the article.

Article received: 2023-02-26