Universidad Politécnica de Yucatán

Computational Robotics Engineering

Machine Learning

1LU - Task: Research on Dimensionality Reduction

Teacher Victor Ortiz Santiago

Avila Chan Gabriela Elizabeth
Castillo Fernandez Brenda Estefania
Baas Cabañas Raymundo Dariel
Cach Rosas Jhair Alejandro
Chi Centeno Mariana Guadalupe
Chi Gongora Enrique Arturo Emmanuel

May 21, 2023

9B

What is the Bias-Variance Trade-Off?

- **<u>Bias:</u>** The bias of our model has to do with the assumptions that it makes about the data, and how well it fits to it when it is trained. A model with high bias doesn't fit well the training data, has limited flexibility, or is extremely simple for the data that we have, resulting generally on a high training error.

The bias tells us how well our model approximates reality.

The error due to the Bias of a model is simply the difference between the expected value of the estimator (i.e. the average prediction of the model) and the actual value. When it is said that a model has a very high bias it means that the model is very simple and has not been adjusted to the training data (it is usually underfitting), so it produces a high error in all samples: training, validation, and test. Linear models often suffer from errors by Bias.

- **<u>Variance</u>**: The variance of our model has to do with how it varies its results depending on the sample of data that it uses for its training. A model with high variance can fit specific to data well, so it has problems generalizing to unseen data, resulting on a high-test error.

The variance tells us how sensible our model is to the training data.

The variance of an estimator is how much the prediction varies depending on the data we use for training.

As we well know, most machine learning algorithms learn as training data comes in. So it is normal for all models to have some variance. Although if we create a robust model, it should learn the relationships between the variables and the target.

- **<u>Bias-variance trade-off:</u>** The bias-variance trade-off in machine learning (ML) is a foundational concept that affects a supervised model's predictive performance and accuracy.

The training dataset and the algorithm(s) will work together to produce results, but ML models aren't 'black box', and humans must understand the ensemble of interactions and tensions that affect their predictive capabilities.

The bias-variance trade-off helps describe prediction errors in supervised models.

The trade-off is also linked to the concepts of overfitting and underfitting. Together, these concepts help explain common issues in machine learning projects and illuminate methods to optimize the model.

The bias-variance trade-off is inescapable – there is no avoiding the tension between the two:

- Increasing the bias will always decrease the variance.
- Increasing the variance will always decrease the bias.

In an ideal world, a model would accurately capture the regularities and trends of its training data but generalize well to training sets or unseen data. Here, the model would have low bias and low variance.

However, this is usually not the case, and different algorithms tend to fall in one of the two camps, i.e., linear models risk high bias and low variance. In contrast, nonlinear models risk high variance and low bias. There are positives to glean from both bias and variance, so long as they're balanced for the problem space.

The bias-variance tradeoff refers to the balance that is needed between bias and variance to build a model that can generalize well to new data. A model that is too simple will have high bias but low variance, while a model that is too complex will have low bias but high variance. The goal is to find the right level of complexity that minimizes both bias and variance, resulting in a model that can accurately generalize to new data.

The Effects of High Bias

Underfitting: Underfitting is a scenario in data science where a data model is unable to capture the relationship between the input and output variables accurately, generating a high error rate on both the training set and unseen data. It occurs when a model is too simple, which can be a result of a model needing more training time, more input features, or less regularization. When a model is underfitted, it cannot establish the dominant trend within the data, resulting in training errors and poor performance of the model. If a model cannot generalize well to new data, then it cannot be leveraged for classification or prediction tasks. Generalization of a model to new data is ultimately what allows us to use machine learning algorithms every day to make predictions and classify data.

High bias and low variance are good indicators of underfitting. Since this behavior can be seen while using the training dataset, underfitted models are usually easier to identify than overfitted ones.

Low complexity: Model complexity refers to how intricate a machine learning model is in terms of its structure and the number of parameters it possesses. In simpler terms, it relates to how well a model can fit the training data and potentially generalize to new, unseen data. A model that is too simple might underfit the data and fail to capture important patterns, while a model that is overly complex might overfit the data and capture noise instead of true relationships. Due to the model being too simple, the biased model is unable to learn complex features of a training data, thus, making it inefficient when solving complex problems.

Low training accuracy: Due to the inability to correctly process training data, the biased model shows high-training loss resulting in low-training accuracy. It is also important to know that a model with high bias can get stuck in local minimums, leading to poor optimization and an inability to capture the most important features in the data.

Techniques to Reduce Bias in Models

High bias happens because of a *high training error*. There are multiple ways to reduce the bias of a model, such as:

1. By **adding more features** from the data to make the model more complex. Including more complex features can help the model capture more complexity and reduce bias.
2. By **increasing training iterations** so that more complex models and relevant data can be learned.
3. Replacing current model with **more complex model** can reduce the bias.
4. Using **non-linear algorithms.**
5. Using **non-parameterized algorithms.**

6. By **decreasing regularization** on inputs at different levels, the model can learn the training set more efficiently and prevent underfitting.
7. By using a **new model architecture**. However, this should only be used as a last resort if none of the methods above give satisfactory results.

The Bias-Variance Trade-Off in Action

1. **Low-Bias, Low-Variance:**
   The combination of low bias and low variance shows an ideal machine learning model. However, it is not possible practically.

A model that has low bias and low variance means that the model is able to capture the underlying patterns in the data (low bias) and is not too sensitive to change in the training data (low variance). This is the ideal scenario for a machine learning model, as it is able to generalize well to new, unseen data and produce consistent and accurate predictions. But in practice, it's not possible.

2. **Low-Bias, High-Variance:** With low bias and high variance, model predictions are inconsistent and accurate on average. This case occurs when the model learns with a large number of parameters and hence leads to an **overfitting.**

A model with high variance and low bias is said to be overfitting.

3. **High-Bias, Low-Variance:** With High bias and low variance, predictions are consistent but inaccurate on average. This case occurs when a model does not learn well with the training dataset or uses few numbers of the parameter. It leads to **underfitting** problems in the model.

A model with high bias and low variance is said to be underfitting.

4. **High-Bias, High-Variance:**
   With high bias and high variance, predictions are inconsistent and inaccurate on average.

A model has both high bias and high variance, which means that the model is not able to capture the underlying patterns in the data (high bias) and is also too sensitive to change in the training data (high variance). As a result, the model will produce inconsistent and inaccurate predictions on average.

Methods for Variance Reduction in Machine Learning Models

Introduction: Addressing variance in machine learning models is pivotal for achieving robust and reliable predictions. This report outlines three key techniques employed to reduce variance in models, namely, the incorporation of more data, feature selection, and regularization. Each method is discussed in depth, highlighting its significance and advantages.

1. **Incorporating More Data**

In machine learning, a model's performance is often contingent on the volume of data used for training. Augmenting the dataset, or "adding more data," entails expanding the dataset to encompass a larger number of data points. This approach offers several noteworthy advantages:

a. **Increased Generalization**

Expanding the dataset exposes the model to a broader spectrum of examples and variations within the data, thereby enhancing its generalization capabilities. A larger dataset enables the model to better discern underlying patterns and relationships that exhibit greater consistency across a larger sample. Consequently, the risk of the model overfitting or memorizing noise and outliers in the training data is diminished.

b. **Enhanced Representation**

A larger dataset provides a more faithful representation of the real-world phenomenon under scrutiny, reducing the likelihood of the model making unwarranted assumptions or generalizations rooted in a restricted sample of examples.

c. **Improved Robustness**

The inclusion of more data augments the model's resilience against variations and shifts in the data distribution. As a result, the model becomes less sensitive to individual data points or minor fluctuations, yielding more stable and dependable predictions.

2. **Feature Selection**

Feature selection is a fundamental technique in machine learning, encompassing the process of selecting a subset of the most pertinent and informative features from the original set. The objective is to retain only those features that contribute significantly to the model's predictive power, while disregarding less relevant or redundant ones.

a. **Importance of Feature Selection**

The significance of feature selection lies in its capacity to reduce the number of input variables considered by the model during predictions. This yields several advantages:

b. **Reduced Complexity**

A reduced set of features results in a simpler model with fewer parameters to estimate. Simpler models are less prone to overfitting as they possess limited capacity to memorize noise in the training data.

c. **Improved Generalization**

By focusing on the most relevant features, the model is more adept at capturing underlying patterns and relationships in the data, culminating in improved generalization performance on unseen data.

d. **Efficiency**

Training and deploying models with fewer features often confer computational efficiency and speed benefits, particularly when handling voluminous datasets.

3. **Regularization**

Regularization is a vital strategy for averting overfitting by introducing a penalty term into the model's loss function. This penalty encourages the model to maintain smaller, more manageable parameter values. Two prevalent types of regularization, L1 and L2 regularization, are explored:

a. **L1 Regularization (Lasso)**

L1 regularization appends a penalty term to the loss function proportional to the absolute values of the model's parameters. It promotes the sparsity of the model by driving some parameters to precisely zero, thereby performing feature selection and eliminating less critical features.

b. **L2 Regularization (Ridge)**

L2 regularization inserts a penalty term proportional to the square of the model's parameters into the loss function. While it curbs parameter values, it rarely forces them to absolute zero. Instead, it prevents the model from becoming overly complex, reducing the impact of individual features.

Both L1 and L2 regularization harmonize the model's aspiration to fit the training data with the objective of maintaining simpler models characterized by smaller parameter values. These techniques, in turn, mitigate overfitting and enhance the model's capacity to generalize effectively to new, unseen data.

In summary, the methods discussed in this report, namely, the inclusion of more data, feature selection, and regularization, constitute indispensable tools for reducing variance in machine learning models. Each approach serves a unique purpose, collectively contributing to the production of models that are robust, accurate, and capable of delivering reliable predictions.

Ensemble methods, bias, and variance

- **Bagging and variance:** Bagging is meant to reduce the variance without increasing the bias. This technique is especially effective where minute changes in a learner's training set lead to huge changes in the predicted output. Bagging reduces the variance by aggregating individual models. These models have dissimilar statistical properties like the means and standard deviations, among others.

  If we measure a random variable (x) with a normal distribution, which is denoted as $N(\mu, \sigma^2) \cdot \mu$ is the mean of the distribution. It could also represent its median or mode. The parameter σ is the standard deviation. If we carry out only one measurement of the mean and variance of variable x. The mean we expect for variable x_1 is μ. On the other hand, the variance of the distribution will be the square of σ.

  Suppose we measure our random variable (x), P times. That is, measurement in the form of $(x_1, x_2 \ldots x_P)/P$. The mean will still be μ . However, as per the equation below, the variance will be smaller.

$$\frac{Var(x_1) + \cdots Var(x_P)}{P^2} = \frac{P\sigma^2}{P^2} = \frac{\sigma^2}{P}$$

It is evident that the mean stays the same, while the variance is averaged. Hence the variance is reduced. Bagging performs well on high variance models like decision trees. On lower variance models such as linear regression, it is not expected to affect the learning process. However, the accuracy reduces when bagging is carried out on models with high bias.

Carrying out bagging on models with high bias leads to a drop in accuracy. This is clear when comparing the performance of the model with and without bagging. Without bagging, the accuracy will be higher than when we implement bagging on such a model. I encourage checking out the article mentioned above to understand the experiment and findings in detail.

- **Boosting and bias:** Boosting is especially useful in models that exhibit underfitting. These models are highly biased and have low variance. They show how poorly a function fits the given data points. To deal with this error, we train a learner and identify where it exhibits bias errors. The observations that are wrongly classified are assigned higher weights.

  The weighting allows each new model to concentrate its efforts on the observations that have proven difficult to fit correctly. A new classifier is then introduced and intuitively used to make predictions on the same data. With each iteration, these misclassified/difficult-to-fit data points can fit better, and the error will be reduced. This is how bias is reduced through boosting. The steps involved in the boosting process are outlined in the article linked in the previous paragraph. It is worth noting that boosting can also affect lowering variance but has a focus on reducing bias.

What is High Variance?

High variance, also known as overfitting, means the model focuses too much on specific patterns in the training dataset and does not generalize well on unseen data. Overfitting can happen when models are too complex.

A model that shows high variance learns a lot and perform well with the training dataset, and does not generalize well with the unseen dataset. As a result, such a model gives good results with the training dataset but shows high error rates on the test dataset.
Since, with high variance, the model learns too much from the dataset, it leads to overfitting of the model. A model with high variance has the below problems:

- **A high variance model leads to overfitting:** Overfitting is a modeling error in statistics that occurs when a function is too closely aligned to a limited set of data points. As a result, the model is useful in reference only to its initial data set, and not to any other data sets.

Overfitting the model generally takes the form of making an overly complex model to explain idiosyncrasies in the data under study. In reality, the data often studied has some degree of error or random noise within it. Thus, attempting to make the model conform too closely to slightly inaccurate data can infect the model with substantial errors and reduce its predictive power.

- **Increase model complexities:** As model complexity increases, performance on the data used to build the model (training data) improves. However, performance on an independent set (validation data) improves up to a point, then starts to get worse. This is called overfitting.
- **Sensitivity to Training Data**: High variance models are highly sensitive to changes in the training data. Even a small change in the training set can lead to a significantly different model. This instability makes it challenging to have confidence in the model's predictions.

Example: Predicting Exam Scores

High Bias (Underfitting): Using a simple linear regression model to predict exam scores based on study hours, assuming a linear relationship. Results in systematic errors, underestimating or overestimating scores for all students, and oversimplification of student performance.

Balanced Bias and Variance (Good Fit): Choosing a more flexible model like polynomial regression, carefully tuning complexity, and regularization to balance bias and variance. Accurate predictions based on study hours, generalizing to different student populations.
Low Bias, High Variance (Overfitting): Implementing an overly complex deep neural network, capturing noise in the training data, and leading to excellent performance on training data but poor generalization to new data due to sensitivity to fluctuations.
In this teacher's example, the bias-variance tradeoff highlights the importance of selecting an appropriate level of model complexity to achieve accurate and reliable predictions for real-world scenarios, such as predicting student exam scores.

Conclusion

The bias-variance trade-off is a crucial concept in machine learning, and finding the optimal balance is necessary for developing accurate models. By understanding the implications and applications of the trade-off, we can create models that generalize well to new data and produce accurate predictions.

References:

(N.d.). Retrieved September 21, 2023, from Researchgate.net website: https://www.researchgate.net/figure/As-model-complexity-increases-performance-on-the-data-used-to-build-the-model-training_fig1_220741326#:~:text=by%20Jonathan%20Shapiro-,As%20model%20complexity%20increases%2C%20performance%20on%20the%20data%20used%20to,This%20is%20called%20overfitting.

Ayuya, C. (n.d.). Ensemble learning on bias and variance. Retrieved September 21, 2023, from Engineering Education (EngEd) Program | Section website: https://www.section.io/engineering-education/ensemble-bias-var/

Belkin, M., Hsu, D., Ma, S., & Mandal, S. (2019). Reconciling modern machine-learning practice and the classical bias–variance trade-off. Proceedings of the National Academy of Sciences, 116(32), 15849-15854.

Bias and Variance in Machine Learning. (n.d.). Retrieved September 21, 2023, from www.javatpoint.com website: https://www.javatpoint.com/bias-and-variance-in-machine-learning

Elliott, S. (2022, February 18). Overcoming bias & variance in machine learning. Retrieved September 21, 2023, from Towards Data Science website: https://towardsdatascience.com/overcoming-bias-variance-in-machine-learning-31169dc649ed

GeeksforGeeks. (2023). Bias and variance in machine learning. *GeeksforGeeks*. https://www.geeksforgeeks.org/bias-vs-variance-in-machine-learning/

Guan, X., & Burton, H. (2022). Bias-variance tradeoff in machine learning: Theoretical formulation and implications to structural engineering applications. In Structures (Vol. 46, pp. 17–30). Elsevier BV. https://doi.org/10.1016/j.istruc.2022.10.004

Dar, Y., Muthukumar, V., & Baraniuk, R. G. (2021). A farewell to the bias-variance tradeoff? an overview of the theory of overparameterized machine learning. arXiv preprint arXiv:2109.02355.

Doroudi, S. (2020). The bias-variance tradeoff: How data science can inform educational debates. AERA open, 6(4), 2332858420977208.

Just Understanding Data. (2023, 19 abril). *Introduction to the Bias-Variance Trade-Off in Machine Learning - Just understanding data*. Understanding Data. https://understandingdata.com/posts/introduction-to-the-bias-variance-trade-off-in-machine-learning/

Pham, K. (2023, 12 mayo). Overfitting, generalization, & the Bias-Variance tradeoff. *Medium*. https://medium.com/@khang.pham.exxact/overfitting-generalization-the-bias-variance-tradeoff-5800f8c2200

Sanz, F. (2021). Entiende de una vez qué es el tradeoff Bias-Variance. *The Machine Learners*. https://www.themachinelearners.com/tradeoff-bias-variance/

Thorn, J. (2021, 13 diciembre). Quick Bias/Variance Trade-Off - towards data science. *Medium*. https://towardsdatascience.com/quick-bias-variance-trade-off-d4895b126b08

Twin, A. (2010, August 14). Understanding overfitting and how to prevent it. Retrieved September 21, 2023, from Investopedia website: https://www.investopedia.com/terms/o/overfitting.asp