Bias and Variance Trade off

ITBodhi Jul 1, 2020 · 8 min read ★

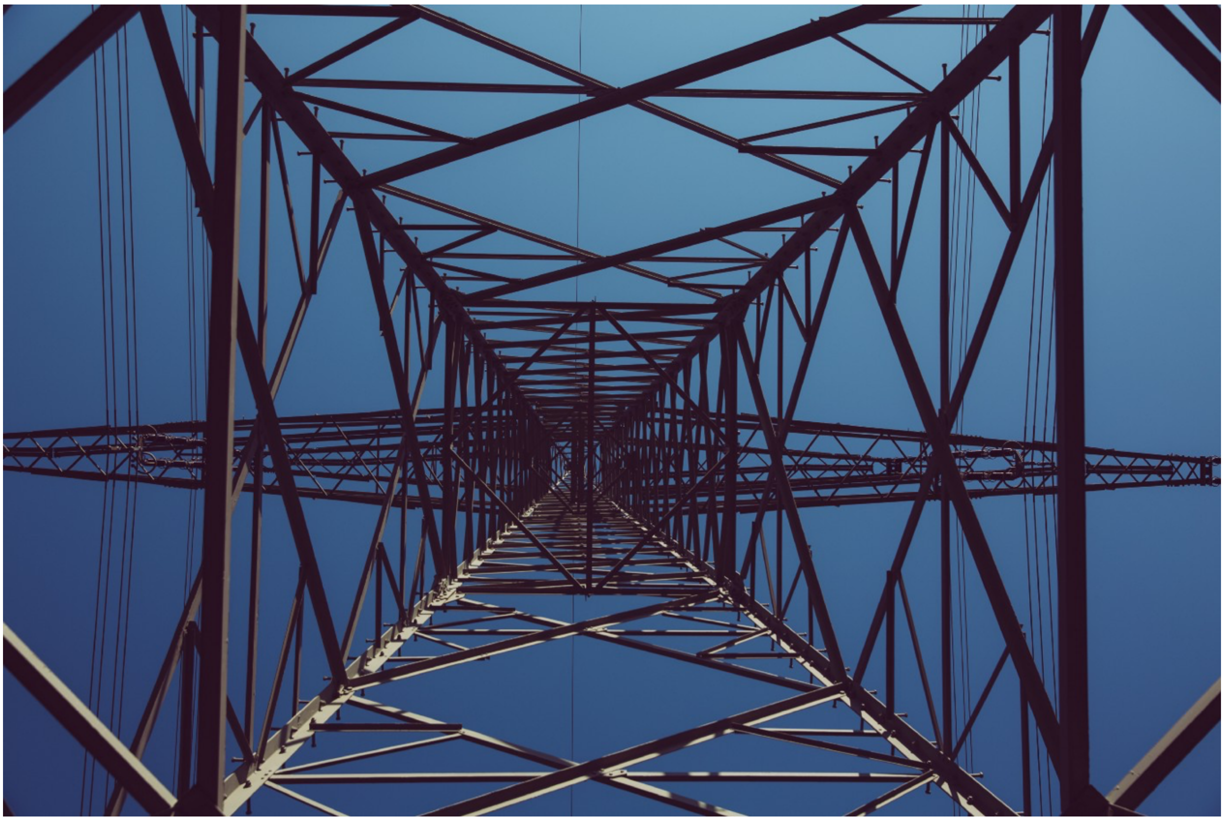


Photo by Markus Spiske on Unsplash

In Machine Learning, when we want to optimise model prediction, it is very important to understand the parameters which describe prediction errors and accuracy . In Machine Learning, any model's performance is based on its correct predictions and how well it is generalised i.e how accurately model gives correct predictions on seen(Training data) and unseen data (Test data, Validation data and Real time data). Best way to measure and optimise a model's accuracy is by keeping account of the ***Bias and Variance*** in the model.

So let's start with the basic understanding of the concepts and see how they help to make model which higher accuracy and minimum error.

Error in Machine Learning Model

In supervised machine learning, an algorithm is trained on the training data to build model that is well designed to make correct prediction on the unseen data that is not available for training.

A machine learning model is nothing but a mathematical function which describes relationship between Predictors (Features and Machine Learning terminology) and Traget variable.

Suppose that we have a training set consisting of a set of points x1, x2,x3….xn and labelled output as y1,y2,y3…..yn ted with each record. We assume that there is a function which describes this relationship as follows:

> *y=f(x) + e*
>
> *where e: — is noise in the data due to various reasons and that has 0 mean and some varaince.*

Now purpose of model creation is to learn pattern from the given data and approximate this function f(x) using various machine learning algorithms. This is kind of reverse engineering where data is given and by merely looking at the data we need to find some function f^(x) that approximate the behaviour of true function f(x). That is always not possible to due to limitation of algorithms and limited size of given data to learn pattern.

On given training data set D=(x1,y1), (x2,y2) ,(x3,y3)……(xn,yn) we need to create a Machine Learning model that predicts correct (approx.) value y for given data (x_i,y_i) from ***seen data*** (Training data) and ***unseen data***(Test/Real Time).

There would always besome error between actual value of y and predicted y^ since real data contain noise e so we has to ready to accept some irreducible predcition error. Let us measure it using mean squared error (For Regression problem when using Linear Regression or SVM Algorithm) that is difference between actual and prediction value.

> *MSE = (y — y^)²*

If we further decompose this equation, we reach at

> *MSE = $Bias^2$ + Variance + Irreducible error*

Irreducible error is the error that can't be ZERO by creating any good models. It is a measure of the amount of noise in our training data. No matter how good we make our prediction model using any robust algorithm, irreducible error that can not be removed.

> Here we can see prediction error is made of two components which can be controlled: Bias and Variance

What is Bias?

Bias is the difference between the average prediction of our model and the correct target value which model is trying to predict.

Bias is inherent to the algorithm we choose to make the Model. A biased model is one that makes incorrect assumptions about the dataset to make the target function easier to learn.For example, suppose that we use a linear regression model on a data that has cubic relationhip. This model will be biased because we have taken wrong assumption about the data and model would be trained using this wrong assumptions about data.

> Low Bias is desired : High Bias means UnderFitting of the model on training data.

> *Given points x, a true value f(x), and a model $\hat{f}(x)$, bias can be expressed mathematically as:*
>
> $Bias[\hat{f}(x)]=E[\hat{f}(x)-f(x)]$
>
> *Where E[·] is the expected value function (e.g. the mean value).*

Model with high bias is very simple because of the assumptions made about the data(which may or may not be true) ***pays very little attention to the training data and oversimplifies the model. High Bias model always leads to high error on training as well as on test data.***

In supervised learning, **Underfitting/High Bias** happens when algorithms are simple(Linear Algorithms) making them fast to learn and easier to understand but generally less flexible to learn complex patterns in data. They have lower predictive performance on complex problems that fail to meet the simplifying assumptions of the algorithms bias.

- **Low Bias** : Suggests less assumptions about the relation in data to predict the target function. *Example: SVM, Decision Tree, KNN*
- **High-Bias** : Suggests more assumptions about the relation in data to predict the target function. *Example: Linear Regression, Logistic Regression*

What is Variance?

Variance is the amount that the estimate of the target function will change if different training data was used.

The *variance* of the learning method, how much the learning method will move around its mean.

Variance is error from sensitivity to small fluctuations in the training set. High variance can cause an algorithm to model the random noise in the training data, rather than the intended result.

> Low Variance is desired : High Variance means OverFitting of the model on training data.

> *It can expressed mathematically as:*

> $\mathrm{Var}(\hat{f}(x)) = E[\hat{f}(x)^2] - E[\hat{f}(x)]^2$
>
> *Where E[·] is the expected value function (e.g. the mean value).*

Model with high variance pays a lot of attention to training data and fail to generalize on the unseen data. As a result, such models perform very well on training data but has high prediction error on unseen/test data.

We are doing the reverse engineering and trying to learn true relationship function from the training data by a machine learning algorithm, so we should expect the algorithm to have some variance. Ideally, it should not change too much from one training dataset to the next, meaning that the algorithm is good at picking out the hidden underlying mapping between the inputs and the output variables.

In supervised learning, **overfitting** happens when algorithms(Non Linear Algorithms) have a high variance are strongly influenced by the specifics of the training data and try to learn patterns which are noisy and not generalized and only limited to training data set.

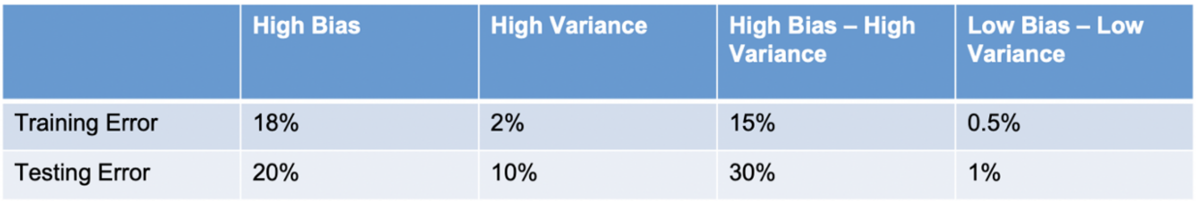
- **Low Variance**: Small changes to the estimate of the target function with changes to the training dataset. *Example: Linear Regression, Logistic Regression*
- **High Variance**: Large changes to the estimate of the target function with changes to the training dataset.*Example: SVM,*

Decision Tree, KNN

Bias-Variance Trade-Off

> Ultimate goel of any machine learning algorithm is to build a prediction model with Low Bias and Low Variance that is measure of reduced error and correct prediction power of any model.

1. High Variance-High Bias — The model is inconsistent and also inaccurate on prediction
2. Low Variance-High Bias — Models is consistent but low on prediction
3. High Variance-Low Bias — Somewhat accurate but inconsistent on prediction
4. Low Variance-Low Bias — ideal scenario, the model is consistent and highly accurate on prediction



	High Bias	High Variance	High Bias – High Variance	Low Bias – Low Variance
Training Error	18%	2%	15%	0.5%
Testing Error	20%	10%	30%	1%

Bias and Variance in Machine Learning Models

Generally, You can see a general trend in the examples above:

- **Linear** machine learning algorithms often have a high bias but a low variance. *Example:Linear Regression, Logistic Regression*
- **Nonlinear** machine learning algorithms often have a low bias but a high variance. *Example: Decision Tree, SVM, Neural Networks*

The parameterization of machine learning algorithms is often a battle to balance out bias and variance.

Finding the right balance between the bias and variance of the model is called the Bias-Variance trade-off.

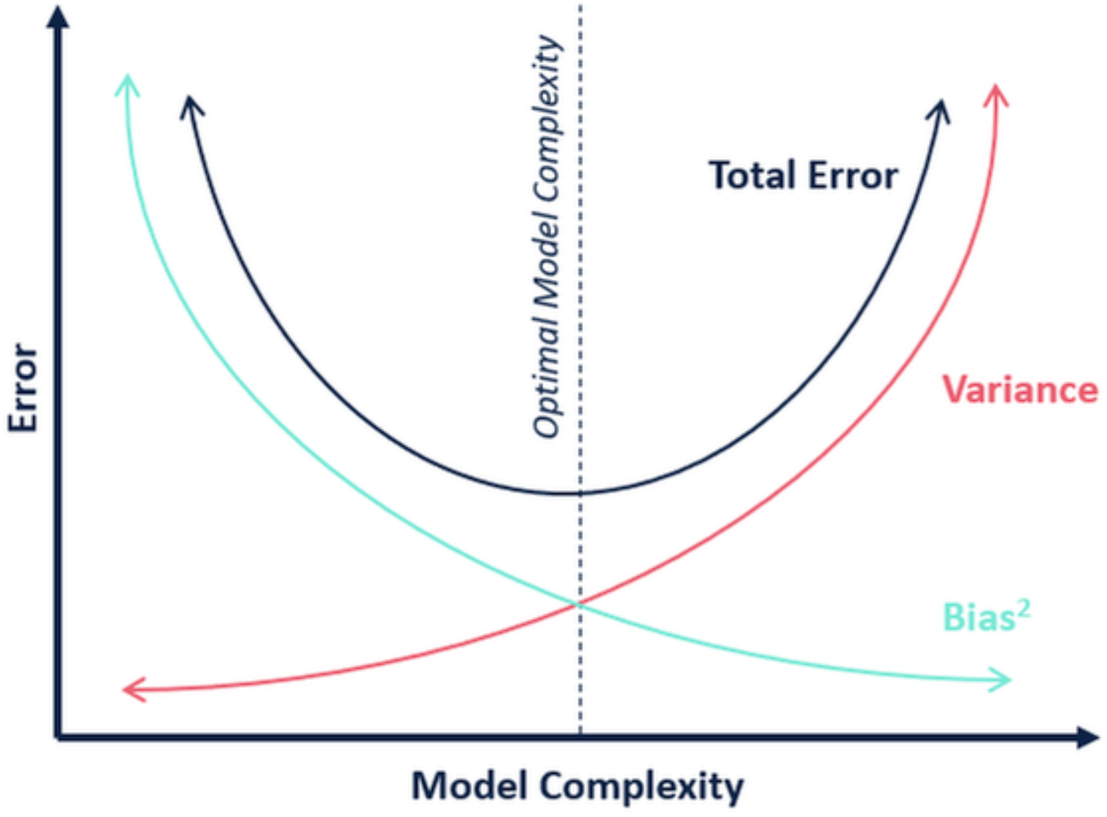
There is inverse relationship between bias and variance in machine learning.

- I*ncreasing the bias will decrease the variance.*
- *Increasing the variance will decrease the bias.*

There is a trade-off at play between these two parameters and the algorithms you choose and the way you choose to configure them are finding different balances in this trade-off for your problem

> ***High Model Complexity : High Variance , Low Bias***

> *Low Model Complexity : High Bias , Low Variance*



https://ai-pool.com/a/s/bias-variance-tradeoff-in-machine-learning.

In a hypothetical example when model is perfect with ZERO error, will have perfect combination of Zero Bias and Zero Variance. In reality, it is very difficult to calculate the real bias and variance because we do not know the actual underlying target function.

If the model is too simple and has very few parameters, it will suffer from high bias and low variance. On the other hand, if the model has a large number of parameters, it will have high variance and low bias. This trade-off should result in a perfectly balanced relationship between the two.

Bulls Eye Diagram

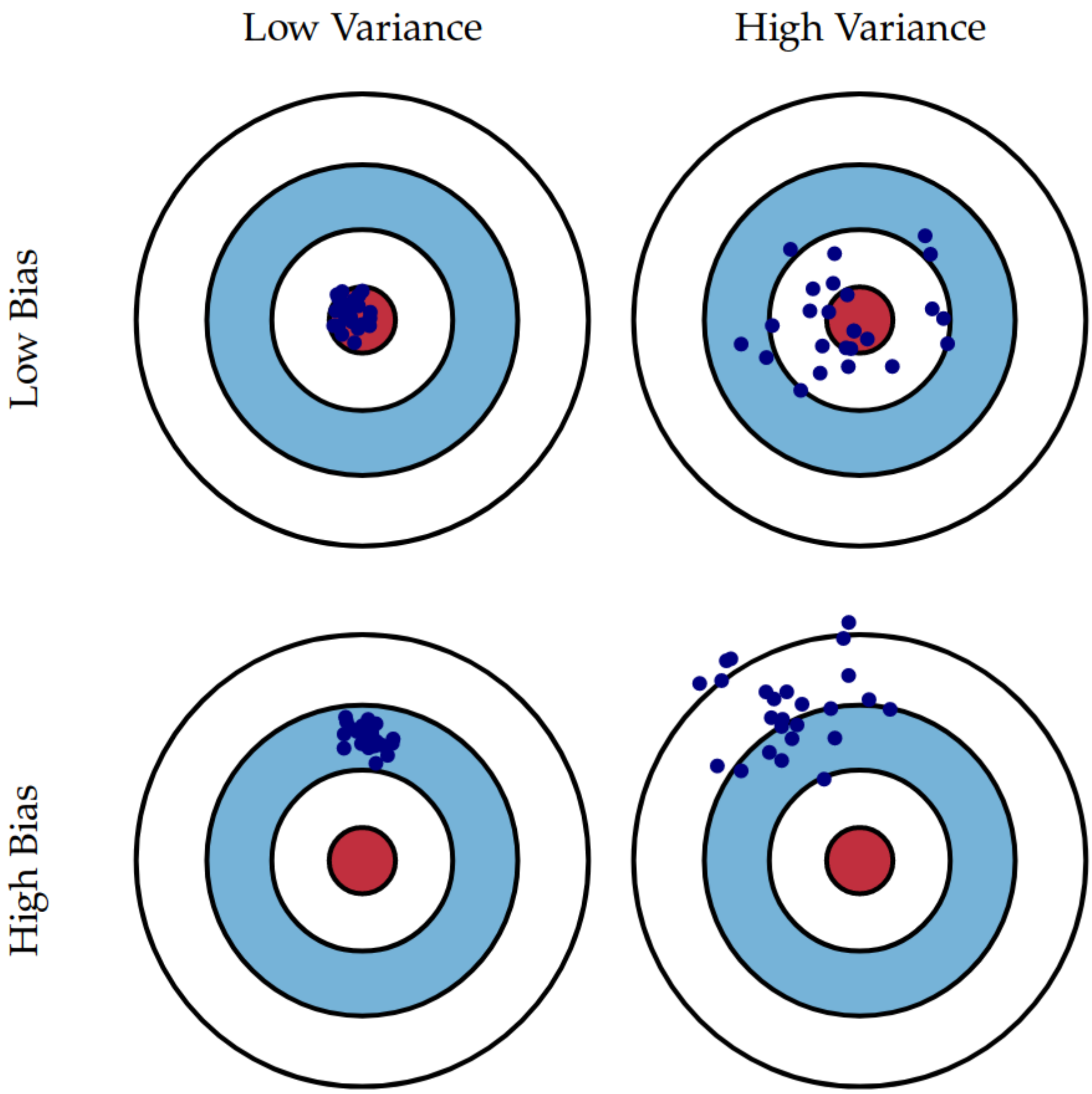
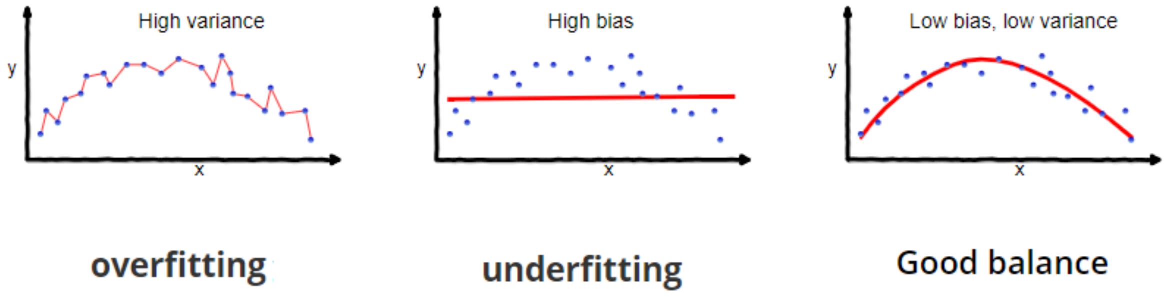


Fig. 1 Graphical illustration of bias and variance.

In the above diagram, center of the target is a model that perfectly predicts correct values. As we move away from the bulls-eye our predictions become get worse and worse. We can repeat our process of model building to get hits on the target.

How to find the Right Balance?

> Ideally, low bias and low variance is the target for any Machine Learning model.



Lowering high Bias or Underfitting:

1. Use non Parameterised Algorithms

2. Make model more complex with more features

3. Use Non Linear Algorithms *Example(Polynomial Regression, Kernel Function in SVM*

Lowering high Variance or Overfitting:

1. Use More Data for training to make model learn maximum hidden pattern from the training data and model becomes generalised.

2. Use Regularization Techniques *Example: L1 , L2, Drop Out, Early Stopping(in case of Neural Networks)etc.*

3. Hyper Parameter Tuning to avoid Overfitting *Example: Higher value of K in KNN, Tuning of C and Gama for SVM, Depth of Tree in Decision Tree*

4. Use less number of features — *Manual or Feature Selection Algorithms or automated using L1, L2 Regularization*

5. Reduce complexity of Model — *Reduce polynomial degree in case of Polynomial regression and Logistic regression*

6. Use Advance techniques like Cross Validation, Stratified Cross Validation etc.

> Finding the right balance with Low Bias and Low Variance is an iterative process where model is trained with different combination of features, Hyperparameters, different set of data set for training and test to find the right combination. We stop when we reach

> the point where Low Bias and Low variance is achieved and model is neither underfit and nor overfit.i.e Prediction accuracy is same on train and test data.

Understanding of the Bias and variance concept is fundamental to train Right prediction Model for your problem.

Thank you for reading this article and I hope it has helped you to understand the concept.

Happy Reading….

www.itbodhi.com

Machine Learning Bias Variance Tradeoff Bias And Variance

Data Science Training Machine Learning Training