Want help with algorithms? Take the FREE Mini-Course.

Search...

Gentle Introduction to the Bias-Variance Trade-Off in Machine Learning

by **Jason Brownlee** on March 18, 2016 in **Understand Machine Learning Algorithms**

Supervised machine learning algorithms can best be understood through the lens of the bias-variance trade-off.

In this post, you will discover the Bias-Variance Trade-Off and how to use it to better understand machine learning algorithms and get better performance on your data.

Let's get started.

Gentle Introduction to the Bias-Variance Trade-Off in Machine Learning
Photo by Matt Biddulph, some rights reserved.

Overview of Bias and Variance

In supervised machine learning an algorithm learns a model from training data.

The goal of any supervised machine learning algorithm is to best estimate the mapping function (f) for the output variable (Y) given the input data (X). The mapping function is often called the target function because it is the function that a given supervised machine learning algorithm aims to approximate.

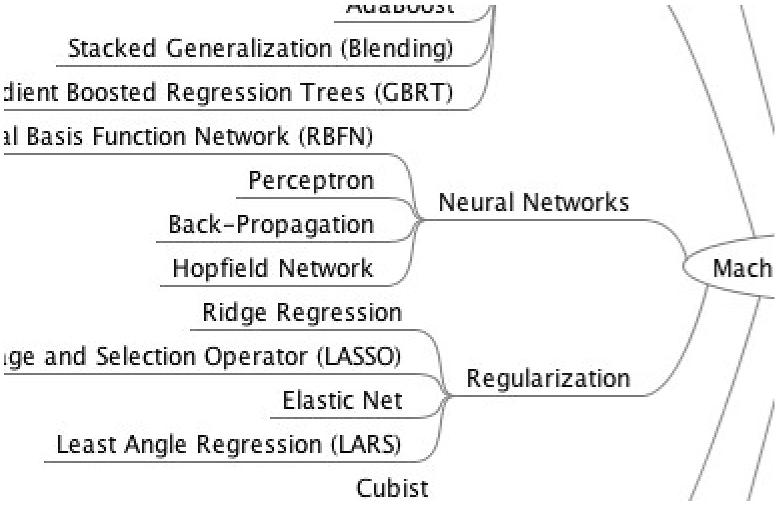
The prediction error for any machine learning algorithm can be broken down into three parts:

- Bias Error
- Variance Error
- Irreducible Error

The irreducible error cannot be reduced regardless of what algorithm is used. It is the error introduced from the chosen framing of the problem and may be caused by factors like unknown variables that influence the mapping of the input variables to the output variable.

In this post, we will focus on the two parts we can influence with our machine learning algorithms. The bias error and the variance error.

Get your FREE Algorithms Mind Map



Sample of the handy machine learning algorithms mind map.

I've created a handy mind map of 60+ algorithms organized by type.

Download it, print it and use it.

Download For Free

Also get exclusive access to the machine learning algorithms email mini-course.

Bias Error

Bias are the simplifying assumptions made by a model to make the target function easier to learn.

Generally, parametric algorithms have a high bias making them fast to learn and easier to understand but generally less flexible. In turn, 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 form of the target function.
- **High-Bias**: Suggests more assumptions about the form of the target function.

Examples of low-bias machine learning algorithms include: Decision Trees, k-Nearest Neighbors and Support Vector Machines.

Examples of high-bias machine learning algorithms include: Linear Regression, Linear Discriminant Analysis and Logistic Regression.

Variance Error

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

The target function is estimated 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.

Machine learning algorithms that have a high variance are strongly influenced by the specifics of the training data. This means that the specifics of the training have influences the number and types of parameters used to characterize the mapping function.

- **Low Variance**: Suggests small changes to the estimate of the target function with changes to the training dataset.
- **High Variance**: Suggests large changes to the estimate of the target function with changes to the training dataset.

Generally, nonparametric machine learning algorithms that have a lot of flexibility have a high variance. For example, decision trees have a high variance, that is even higher if the trees are not pruned before use.

Examples of low-variance machine learning algorithms include: Linear Regression, Linear Discriminant Analysis and Logistic Regression.

Examples of high-variance machine learning algorithms include: Decision Trees, k-Nearest Neighbors and Support Vector Machines.

Bias-Variance Trade-Off

The goal of any supervised machine learning algorithm is to achieve low bias and low variance. In turn the algorithm should achieve good prediction performance.

You can see a general trend in the examples above:

- Parametric or linear machine learning algorithms often have a high bias but a low variance.
- Non-parametric or non-linear machine learning algorithms often have a low bias but a high variance.

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

Below are two examples of configuring the bias-variance trade-off for specific algorithms:

- The k-nearest neighbors algorithm has low bias and high variance, but the trade-off can be changed by increasing the value of k which increases the number of neighbors that contribute t the prediction and in turn increases the bias of the model.
- The support vector machine algorithm has low bias and high variance, but the trade-off can be changed by increasing the C parameter that influences the number of violations of the margin allowed in the training data which increases the bias but decreases the variance.

There is no escaping the relationship between bias and variance in machine learning.

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

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

In reality, we cannot calculate the real bias and variance error terms because we do not know the actual underlying target function. Nevertheless, as a framework, bias and variance provide the tools to understand the behavior of machine learning algorithms in the pursuit of predictive performance.

Further Reading

This section lists some recommend resources if you are looking to learn more about bias, variance and the bias-variance trade-off.

- Bias-variance tradeoff on Wikipedia
- Understanding the Bias-Variance Tradeoff
- Inductive Bias on Wikipedia

Summary

In this post, you discovered bias, variance and the bias-variance trade-off for machine learning algorithms.

You now know that:

- Bias is the simplifying assumptions made by the model to make the target function easier to approximate.
- Variance is the amount that the estimate of the target function will change given different training data.

- Trade-off is tension between the error introduced by the bias and the variance.

Do you have any questions about bias, variance or the bias-variance trade-off. Leave a comment and ask your question and I will do my best to answer.

About Jason Brownlee

Jason Brownlee, PhD is a machine learning specialist who teaches developers how to get results with modern machine learning methods via hands-on tutorials.
View all posts by Jason Brownlee →

‹ Supervised and Unsupervised Machine Learning Algorithms
Overfitting and Underfitting With Machine Learning Algorithms ›

Your Start in Machine Learning

55 Responses to *Gentle Introduction to the Bias-Variance Trade-Off in Machine Learning*

Akash Ansari March 28, 2016 at 9:21 pm #

REPLY

I have just read a nice and delicate blogpost about bias-variance tradeoff. Looking forward to learn more Machine Learning Algorithms in a simpler fashion.

R M Jain September 15, 2016 at 7:44 pm #

REPLY

Hi, I want to check bias-variance tradeoff for iris dataset.. Is anyone knows how to find it>?????????
please tell the solution…

R M Jain September 15, 2016 at 7:45 pm #

REPLY

Hi, I required to find using r programming functions.. please reply…

R Karthik February 2, 2017 at 5:12 pm #

REPLY

There is a typo i guess…. High bias means more assumptions for the target function. Eg: Linear regression.

But in the article it is specified as opposite,

” High-Bias: Suggests less assumptions about the form of the target function. “

Jason Brownlee February 3, 2017 at 9:57 am #

REPLY

Thanks, fixed.

Fei Du February 25, 2017 at 12:31 pm #

REPLY

Hi, less assumptions probably mean less complex model, so I guess high-bias may suggest less complex model and less assumptions.

Jason Brownlee February 26, 2017 at 5:28 am #

REPLY

A high bias assumes a strong assumption or strong restrictions on the model.

Massi December 10, 2017 at 8:28 am #

REPLY

I agree with Fei Du

Arindam Paul April 25, 2019 at 12:28 pm #

REPLY

Fei Du and Massi are wrong. A simpler model is one with more assumptions. This is how a generalized linear model becomes linear when we remove non-linear terms. Less complex and fewer assumptions are opposite in this case. A linear model is a special case of a polynomial and thus puts more restrictions on the model.

Jason Brownlee April 25, 2019 at 2:42 pm #

Model complexity is a sticky concept, there are many ways to measure it (e.g. freedom, capacity, parameters, etc).

A high bias model can be complex or simple, complexity is a different axis of consideration.

Soghra. July 6, 2017 at 4:27 am #

REPLY

This is perfect explanation. Thanks for your efforts.

Jason Brownlee July 6, 2017 at 10:26 am #

REPLY

Thank you.

sam August 1, 2017 at 10:14 pm # REPLY

Very simple explanation. thanks.

Jason Brownlee August 2, 2017 at 7:51 am # REPLY

I'm glad it helped Sam.

Raj September 6, 2017 at 8:46 am # REPLY

Good explanation, Thank you!

Jason Brownlee September 7, 2017 at 12:48 pm # REPLY

I'm glad it helped.

Ashwin Agrawal October 13, 2017 at 10:12 pm # REPLY

Nice work!!! I have one query if we decrease the variance then we observe the bias increases and vice versa, but is the rate of fall and rise of these parameters is same or constant or it is dependent of specific algorithms used. Can we tune a model based on bias-variance trade off?

Jason Brownlee October 14, 2017 at 5:45 am # REPLY

They are tied together.

Yes, one must tune the trade-off on a specific problem to get the right level of generalization required.

soumya December 7, 2017 at 1:30 am # REPLY while designing any model which must be considered to minimize first bias or variance so as to get a better model ?

Jason Brownlee December 7, 2017 at 8:05 am #

REPLY

Perhaps start with something really high bias and slowly move toward higher variance?

soumya December 7, 2017 at 1:35 am #

REPLY

Is there any limit or a scale to know the errors are minimum or maximum in bias and variance?

Jason Brownlee December 7, 2017 at 8:05 am #

REPLY

No, it is problem/metrics/algorithm specific.

soumya December 7, 2017 at 5:10 pm #

REPLY

Thank you

Keval December 8, 2017 at 8:01 am #

REPLY

What are some the measures for understanding bias and variance in our model? How can we quantify bias-variance trade-off? Thanks.

Jason Brownlee December 8, 2017 at 2:28 pm #

REPLY

Good question, it may be possible for specific algorithms, such as in knn increasing k from 1 to n (number of patterns) and plotting model skill in a test set.

Your Start in Machine Learning

Bilawal Wajid Ali January 25, 2018 at 11:21 am # REPLY

hello

I guess this post is too old , but still I am giving a try may be if you can answer my question it would be helpful for me

What role exactly Bias and variance could play on the result for training data

what is actually means by the trade-off between variance and bias
how it could effect the approach overall

Thanks

Jason Brownlee January 26, 2018 at 5:36 am # REPLY

It is an abstract idea to help understand how machine learning algorithms work in general. The tension that exists between prior knowledge/bias and learning from data/variance.

Manish Sihag February 14, 2018 at 11:08 am # REPLY

Hello, I am still a little confused about this. Please help me out by reading the following situation.

Suppose there are 5 parties standing in election and we want to make prediction beforehand about who will win. We choose 5 people from a community and ask them separately, the name of the party for which they are going to vote. Suppose all 5 of them chooses all 5 different parties. If we treat each person as a machine learning model, their answer as training data and make predictions accordingly then for all 5 models, we will end up predicting 5 different outputs. This makes it a high variance case because results are varying and high bias because we chose only a particular type of people (all are from a particular community). Next if we choose 1000 people for poll and there are suppose 200 people for each party. Now if we treat all 1000 people as model, we will have 200 predictions for each party. We can call it a low bias because we chose a larger group now and hence they are more representative of their population but this is still high variance, right? Because there are still equally varied results. Lastly, if 700 of those people chose one party (and that party actually wins) and rest 300 are distributed in other parties, is this what we call low variance? What will we call it if that party loses?

Also let me know if I made any incorrect remarks.
Thanks.

Jason Brownlee February 14, 2018 at 2:42 pm #

REPLY

There would be some variance, but why high variance? If they are from the same geographical region, watch the same news, then the bias may be high but the variance low.

Also, if there are only 2 choices, e.g. binary, then things like bias and variance don't mean much. Perhaps you want an example such as guessing the number of coins in a jar?

Jeff Nyman June 14, 2018 at 5:44 am #

REPLY

A question on this statement. You are saying if there are only two choices, then bias and variance don't mean much. Do you mean two choices as in the outputs? The reason I'm asking is for various sentiment analysis ideas, wherein you have two choices (outputs): 0 or 1. Basically "does not predict" or "does predict."

But it would seem that in such binary situations, bias could creep in. Such as in data that uses movie reviews, wherein a bias may creep in if the word "bad" was previously only used in clearly negative reviews. But then a new review comes in that says "Wanted to see this movie really bad. Not disappointed." Here the use of "bad" is not in a negative context, but might be subject to bias.

Am I just totally missing the point of your comment?

Manish Sihag February 15, 2018 at 12:51 am #

REPLY

Yes, I suppose.The reason I am saying it a high variance is because high variance is the spread of predictions by different models from target output. And in this case out of 1000 only 200 will be on target and rest 800 varied. But probably I am getting it all wrong, I will think about it some more. Thanks for the quick reply.

Doug March 9, 2018 at 7:17 am #

REPLY

Is this bias concept different than the bias added into neural network model y=mx + b(bias of 1)?

Jason Brownlee March 10, 2018 at 6:11 am #

REPLY

Yes, different concepts. The bias in a model is a specific manipulation of the model, the bias in the tradeoff is an abstract concept regarding model behavior.

mars March 29, 2018 at 6:35 pm # REPLY

Does the variance here correspond to the RMSE ?

Jason Brownlee March 30, 2018 at 6:34 am # REPLY

No, but perhaps the variance in RMSE when the model is trained on different samples of training data (e.g. variance in skill given initial conditions).

Bibek Subedi May 8, 2018 at 4:00 am # REPLY

Hi Jason,

Your explanation is clear and concise. A nice thing about you is you make complex math heavy topics simple and easier to understand.

Jason Brownlee May 8, 2018 at 6:16 am # REPLY

Thanks.

Igor June 6, 2018 at 2:03 am # REPLY

Hi Jason,

Is my understanding correct regarding the bias?

Bias is when we assume certain things about the training data (it's shape for example) and we choose a model accordingly. But, then we get predictions far away from the exptected values and we realise that we did a mistake in certain assumptions of our training data?

Jason Brownlee June 6, 2018 at 6:41 am # REPLY

Sounds good.

Yongtao Ding July 20, 2018 at 5:55 am #

REPLY

Hi Jason,

This post is clear and easy to understand. I have one question regarding your statement about how SVM manages variance issue. As you said in this article, through increasing penalty parameter C, SVM could decrease its variance.

From my perspective, C is the penalty parameter, and it is different from the regularization lambda, through decreasing C, we could narrow the margin, and the learner could go a little underfitting, which would decrease the variance.

Look forward to hearing from you!

Best,
Yongtao

Jason Brownlee July 20, 2018 at 6:21 am #

REPLY

Not sure, I need to think about it.

Perhaps run some experiments to confirm your framing.

Yongtao Ding July 20, 2018 at 6:02 am #

REPLY

just wanna check if my comment is submitted

Jason Brownlee July 20, 2018 at 6:20 am #

REPLY

Comments are moderated, learn more here:
https://machinelearningmastery.com/faq/single-faq/where-is-my-blog-comment

Khalid Usman November 8, 2018 at 1:14 am #

REPLY

Dear @Jason Brownlee : Thanks alot for another nice and easy tutorial.

So from the description can i say, that Linear algorithms (Linear/Logistic Regression & LDA) will only under-fit and never face over-fitting problem, because you said that these algorithms have high bias and low variance and vice-versa for non-linear problems (decision tree, KNN, SVM)

Jason Brownlee November 8, 2018 at 6:11 am #

REPLY

A high bias algorithm can still overfit.

Thahimum Hassan December 25, 2018 at 5:13 am #

REPLY

Hi Jason,

I am struggling to calculate the bias/discrimination of the 'Adult Dataset', downloaded from the UCI machine learning repository. Do you know how to calculate the discrimination using a matrix? Thanks in advance.

Jason Brownlee December 25, 2018 at 7:25 am #

REPLY

Sorry, I don't have a tutorial on that dataset, maybe this process will help:
https://machinelearningmastery.com/start-here/#process

Akash Dubey January 14, 2019 at 1:44 am #

REPLY

I have one question though : For a model total error is calculated as

Total error = (Bias)^2 + Variance + Irreducible error

Why do we take (Bias)^2 for the calculation, why not just Bias ?

Jason Brownlee January 14, 2019 at 5:30 am #

REPLY

If bias is the variance then the units are squared units. It is a notation for the units.

Sandeep January 18, 2019 at 1:31 am #

REPLY

Hi Jason,

Thank you very much for your post. It helped me understanding variance and Bias in model.

How do we identify the bias and variance when we apply Random Forest or Logistic regression? I mean what type of graph should I perform to check it? What is x-axis and what is Y-axis?

Thank you for your help on this.

Sandeep

Jason Brownlee January 18, 2019 at 5:43 am # REPLY

Generally, we don't. It is a concept to help understand the model behavior.

Yousuf Azad March 1, 2019 at 8:35 pm # REPLY

Feedback : Few images would have helped a lot

Jason Brownlee March 2, 2019 at 9:32 am # REPLY

Thanks for the suggestion.

What would you like to see images of exactly?

e.khalili May 5, 2019 at 6:03 am # REPLY

your page dont have any mathematical bias.
please improve …

Jason Brownlee May 5, 2019 at 6:36 am # REPLY

Thanks for the suggestion.

Leave a Reply

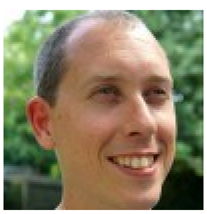
Your Start in Machine Learning

Name (required)

Email (will not be published) (required)

Website

Welcome to Machine Learning Mastery!



Hi, I'm **Jason Brownlee**, PhD.
I write tutorials to help developers (*like you*) get results with machine learning.

Read More

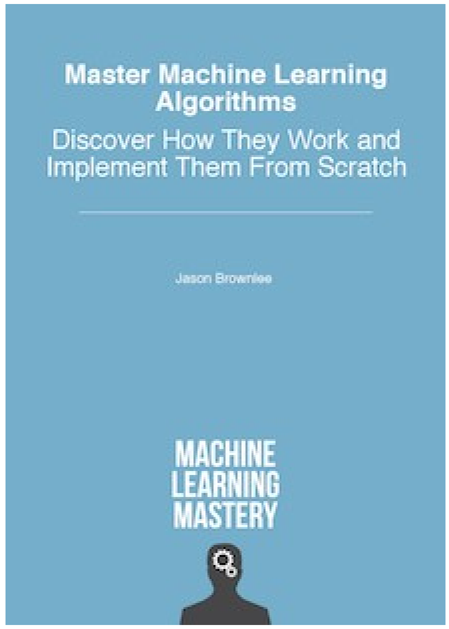
Need help?

Start Here
(over 15 quick-start guides!)

Understand How Algorithms Work
No theory, just clear descriptions.

Your Start in Machine Learning

Click to Get Started Now!



Address: PO Box 206, Vermont Victoria 3133, Australia. | ACN: 626 223 336.