
Welcome to Data Science Online Bootcamp

Week#4_Day#4

dφ

Democratizing Data Science Learning

Learning Objectives

Bias and Variance



Bias and Variance in real world

In dictionary terms :

Bias : Prejudice in favor of or against one thing, person, or group compared with another, usually in a way considered to be unfair.

Variance: The state or fact of disagreeing or quarreling.

In short, Bias represents how unfair is something towards others, and Variance represents how likely something changes with respect to others.

Confusing ? Worry not. The next example will clarify all your doubts.

Example

Let's assume you have called two weather examiners, Mr. Bishop and Mr. Varian to test if it will rain or not.

Mr. Bishop loves rain a lot. And Mr. Varian is a bookworm.

Let us talk about the conditions for rain.

- **It rains only if it's little humid.**
- **It does not rain if it's windy, hot or freezing.**



Mr. Bishop representing Bias

You ask Mr. Bishop (Despite of his training, he is too biased towards rain) :

Me :Sir, its extremely hot out here, will it rain ?

Mr. Bishop : Yup.

Me :Sir, its little windy, will it rain ?

Mr. Bishop : May be not.

Me :Sir, its freezing will it rain ?

Mr. Bishop : Yes of course.

Me :Sir, its little humid, will it rain ?

Mr. Bishop : Damn sure.

Did you notice, Mr. Bishop is highly Biased towards chances of having rain.
During the test, he is unable to predict most of them correctly.

This condition is called **under fitting**.



Mr. Varian representing Variance

Now let us see your conversation with Mr. Varian (a bookworm who completely remembers the training he had):

Me :Sir, its extremely hot out here, will it rain ?

Mr. Varian: Nope.

Me :Sir, its little windy, will it rain ?

Mr. Varian: No way.

Me :Sir, its freezing, will it rain ?

Mr. Varian: No way.

Me :Sir, its little humid, will it rain ?

Mr. Varian: Yes it will.



Mr. Varian representing Variance

Mr. Varian successfully predicted whether it will rain or not. But being a bookworm, Mr. Varian is unknown to the conditions not described in the book during training.

Now, we ask Mr. Varian :

Me :Sir, there is a giant sitting on the cloud who lost his candy. Will it rain ?
Mr. Varian: Not sure, since the answer is “No” to most of the conditions, there is a high possibility that it will not rain .

Now, although the decision of Mr. Varian varies perfectly with the input conditions, he is not able to predict for the new and unseen condition (other general conditions apart from the given specific conditions while training).

This condition is called **over fitting**. And it offers **poor generalizability**.



High Bias or High Variance?

Then what is better, high bias (high generalizability) or high variance (high accuracy on training data) ?

Well, the answer is, “Best of both worlds”. We neither need high bias nor high variance. We would want our algorithm to perform better on training set and also offer best result on unseen data (the test set).

In general, **having high bias reduces the performance of the algorithm on training set while having high variance reduces performance on unseen data.**

This is known as **Bias Variance Trade off.**



Reading Material

MUST READ

Understanding the Bias Variance Tradeoff:

<https://towardsdatascience.com/understanding-the-bias-variance-trade-off-165e6942b229>



Reference

<https://medium.com/@viveksingh.heritage/an-intuitive-explanation-to-bias-variance-tradeoff-ec540fb13e12>



Slide Download Link

- You can download the slides here:

https://docs.google.com/presentation/d/1c_mkhl2ac9gapFwCQdn6tySCi09Ve8jCv8Z1BqbrPG8/edit?usp=sharing



That's it for the day. Thank you!

Feel free to post any queries in the #help channel on Slack



Democratizing Data Science Learning