

Open in app ↗



Search Medium



This is your last free member-only story this month. [Upgrade](#) for unlimited access.

★ Member-only story

Bias, Variance and How they are related to Underfitting, Overfitting

I came across the terms bias, variance, underfitting and overfitting while doing a course. The terms seemed daunting and articles online didn't help either. Although concepts related to them are complex, the terms themselves are pretty simple. Below I will give a brief overview of the above-mentioned terms and Bias-Variance Tradeoff in an easy to understand manner.



Rahul Banerjee · [Follow](#)

Published in Towards Data Science

5 min read · Sep 16, 2020



Listen



Share



More

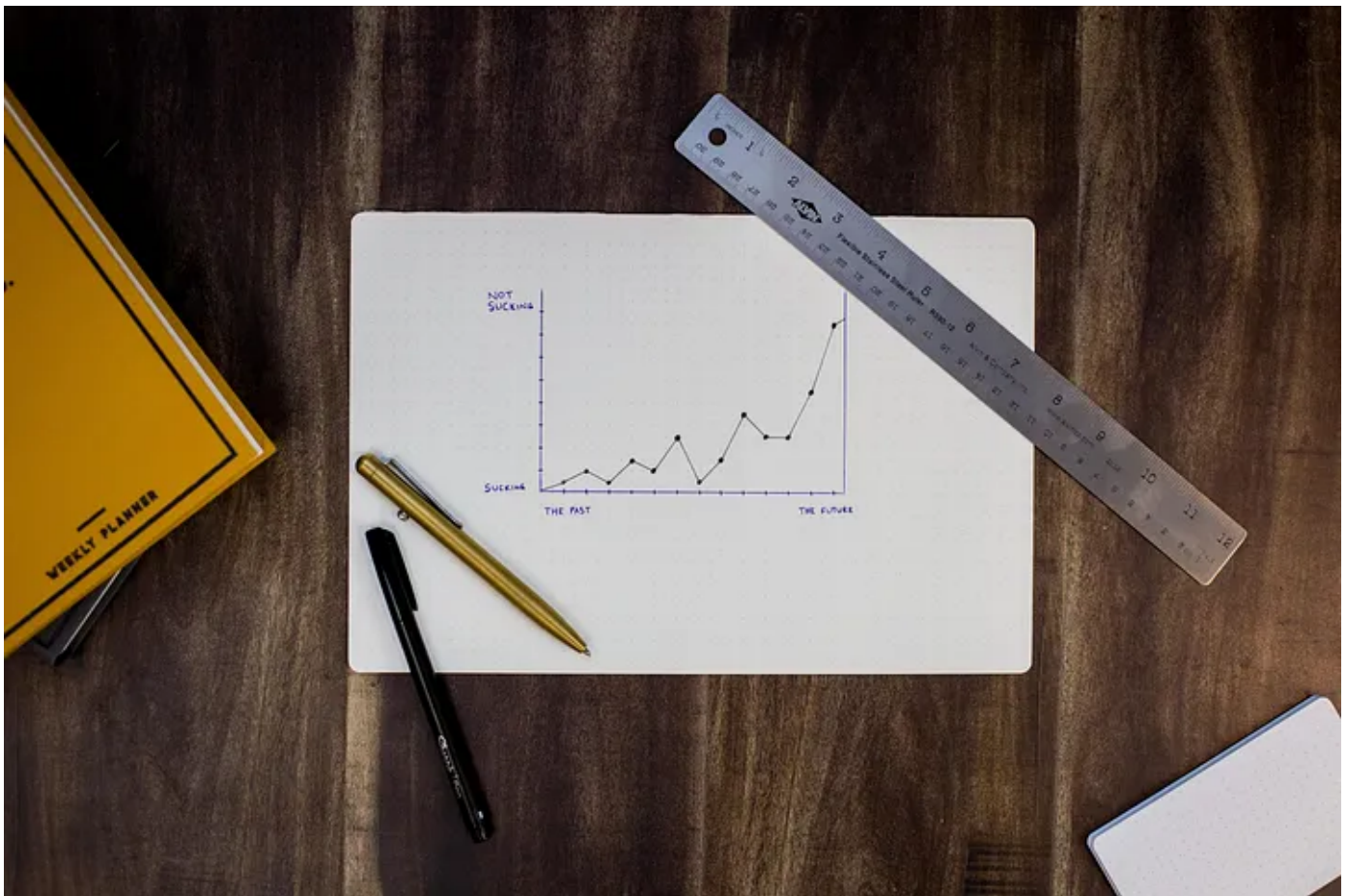


Photo by [Isaac Smith](#) on [Unsplash](#)

Assume you have a classification model, training data and testing data

```
x_train , y_train // This is the training data
x_test , y_test // This is the testing data
y_predicted // the values predicted by the model given an input
```

The error rate is the average error of value predicted by the model and the correct value.

Bias

Let's assume we have trained the model and are trying to predict values with input 'x_train'. The predicted values are y_predicted. Bias is the error rate of y_predicted and y_train.

In simple terms, think of bias as the error rate of the training data.

When the error rate is high, we call it High Bias and when the error rate is low, we call it Low Bias

Variance

Let's assume we have trained the model and this time we are trying to predict values with input 'x_test'. Again, the predicted values are y_predicted. Variance is the error rate of the y_predicted and y_test

In simple terms, think of variance as the error rate of the testing data.

When the error rate is high, we call it High Variance and when the error rate is low, we call it Low Variance

Underfitting

When the model has a high error rate in the training data, we can say the model is underfitting. This usually occurs when the number of training samples is too low. Since our model performs badly on the training data, it consequently performs badly on the testing data as well.

A high error rate in training data implies a High Bias, therefore

In simple terms, High Bias implies underfitting

OverFitting

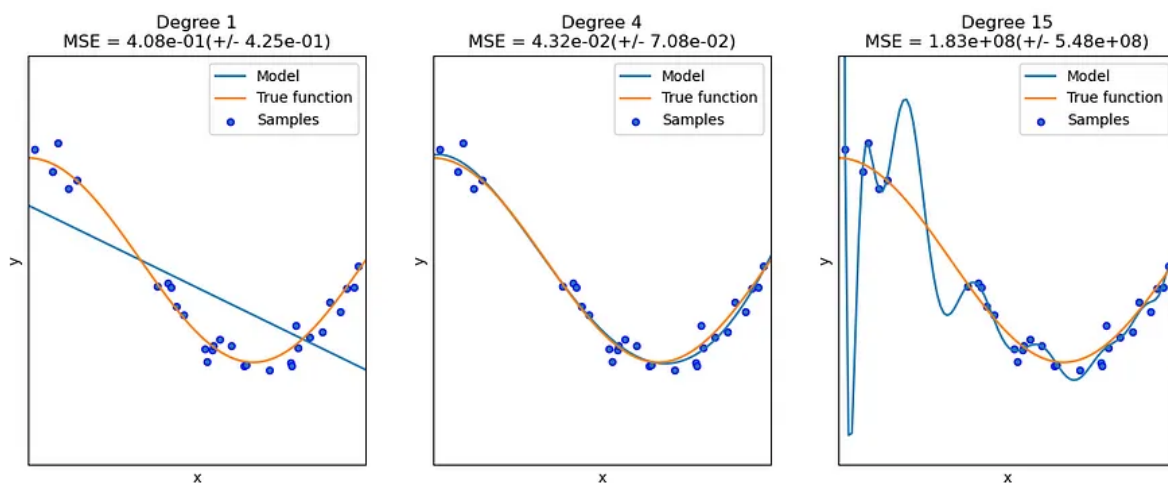
When the model has a low error rate in training data but a high error rate in testing data, we can say the model is overfitting. This usually occurs when the number of training samples is too high or the hyperparameters have been tuned to produce a low error rate on the training data.

Think of a student who studied a certain set of questions and then gave a mock exam which contains those exact questions they studied. They might do well on the mock exam but on the real exam, which contains unseen questions, they might not necessarily do well. If the student gets a 95% in the mock exam but a 50% in the real exam, we can call it overfitting.

A low error rate in training data implies Low Bias whereas a high error rate in testing data implies a High Variance, therefore

In simple terms, Low Bias and High Variance implies overfitting

Overfitting, Underfitting in Regression



Source: https://scikit-learn.org/stable/auto_examples/model_selection/plot_underfitting_overfitting.html

In the first image, we try to fit the data using a linear equation. The model is rigid and not at all flexible. Due to the low flexibility of a linear equation, it is not able to predict the samples (training data), therefore the error rate is high and it has a High Bias which in turn means it's **underfitting**. This model won't perform well on unseen data.

In the second image, we use an equation with degree 4. The model is flexible enough to predict most of the samples correctly but rigid enough to avoid overfitting. In this case, our model will be able to do well on the testing data therefore this is an **ideal model**.

In the third image, we use an equation with degree 15 to predict the samples. Although it's able to predict almost all the samples, it has too much flexibility and will not be able to perform well on unseen data. As a result, it will have a high error rate in testing data. Since it has a low error rate in training data (Low Bias) and high error rate in training data (High Variance), it's **overfitting**.

Overfitting, Underfitting in Classification

Assume we have three models (Model A , Model B , Model C) with the following error rates on training and testing data.

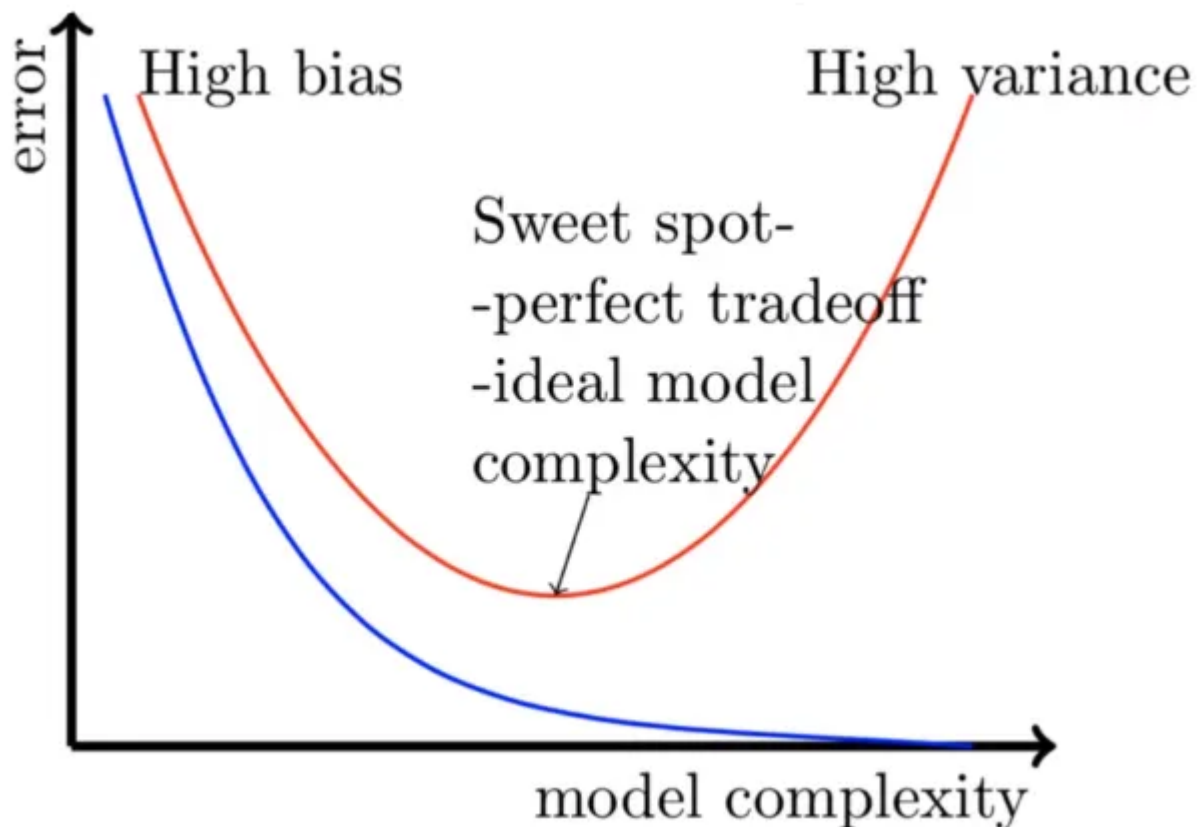
Error Rate	Model A	Model B	Model C
Training Data	30%	6%	1%
Testing Data	45%	8%	25%

For Model A, The error rate of training data is too high as a result of which the error rate of Testing data is too high as well. It has a High Bias and a High Variance, therefore it's underfit. This model won't perform well on unseen data.

For Model B, The error rate of training data is low and the error rate of Testing data is low as well. It has a Low Bias and a Low Variance, therefore it's an **ideal model**. This model will perform well on unseen data.

For Model C, The error rate of training data is too low. However, the error rate of Testing data is too high as well. It has a Low Bias and a High Variance, therefore it's **overfit**. This model won't perform well on unseen data.

Bias-Variance Tradeoff



Source: <https://medium.com/@prvnk10/bias-variance-tradeoff-ebf13adcea42>

When the model's complexity is too low, i.e a simple model, the model won't be able to perform well on the training data nor the testing data, therefore it's underfit

At the sweet spot, the model has a low error rate on the training data as well as the testing data, therefore, that's the ideal model

As the complexity of the model increases, the model performs well on the training data but it doesn't perform well on the testing data and therefore it's overfit

Thank You for reading the article. Please let me know if I made any mistakes or have any misconceptions. Always happy to receive feedback :)

I recently created a blog using WordPress, I would love it if you check it out 😊

Python Project Tutorials — Improve your CV/Portfolio with these Python Project Tutorials.