Bias Variance Tradeoff

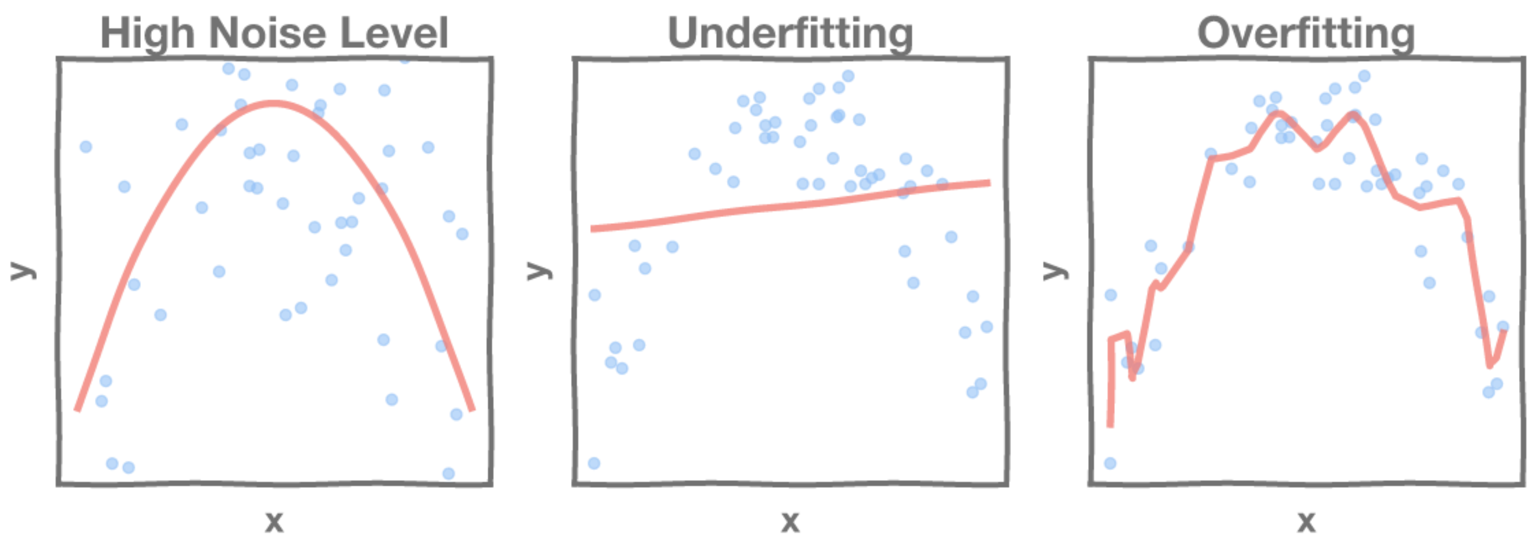
Bookmark this page

Test Error and Generalization

We know to **evaluate** models on both training and test data because models can do well on training data but poorly on new data. When models do well on new data (test data), we call it **generalization**. Models that “generalize well” are good with new data.

> **Reminder**
>
> There are three ways a model can have a high test error: high noise, underfitting, and overfitting.

- **High Noise Level**: If the data is very noisy we will not be able to generalize very well and we will get a high test error due to this noise.
- **Underfitting**: The model is not complex enough to capture the patterns in the data.
- **Overfitting**: The model focuses too much on the training data and does not generalize to the test data.



Let's focus in on noise.

Irreducible and Reducible Error

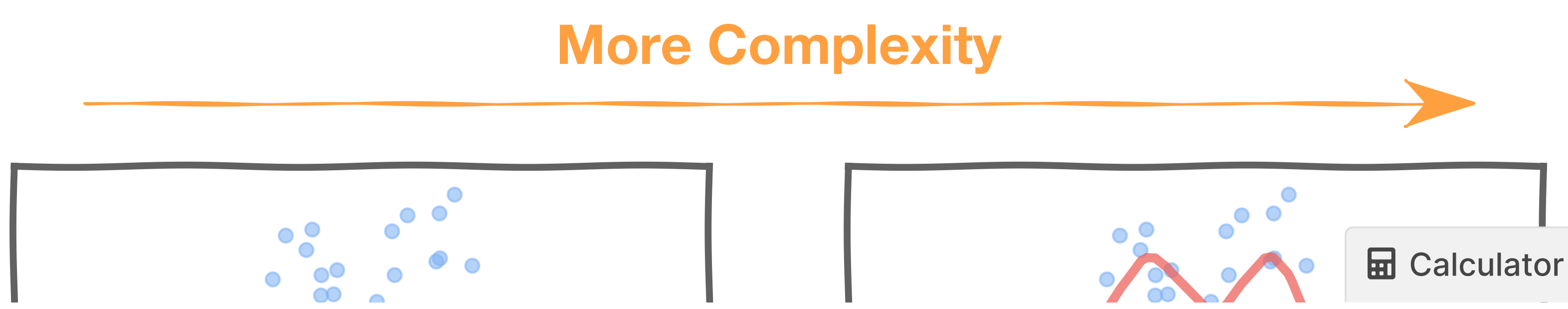
There are two ways that “noise” can contribute to the generalization error:

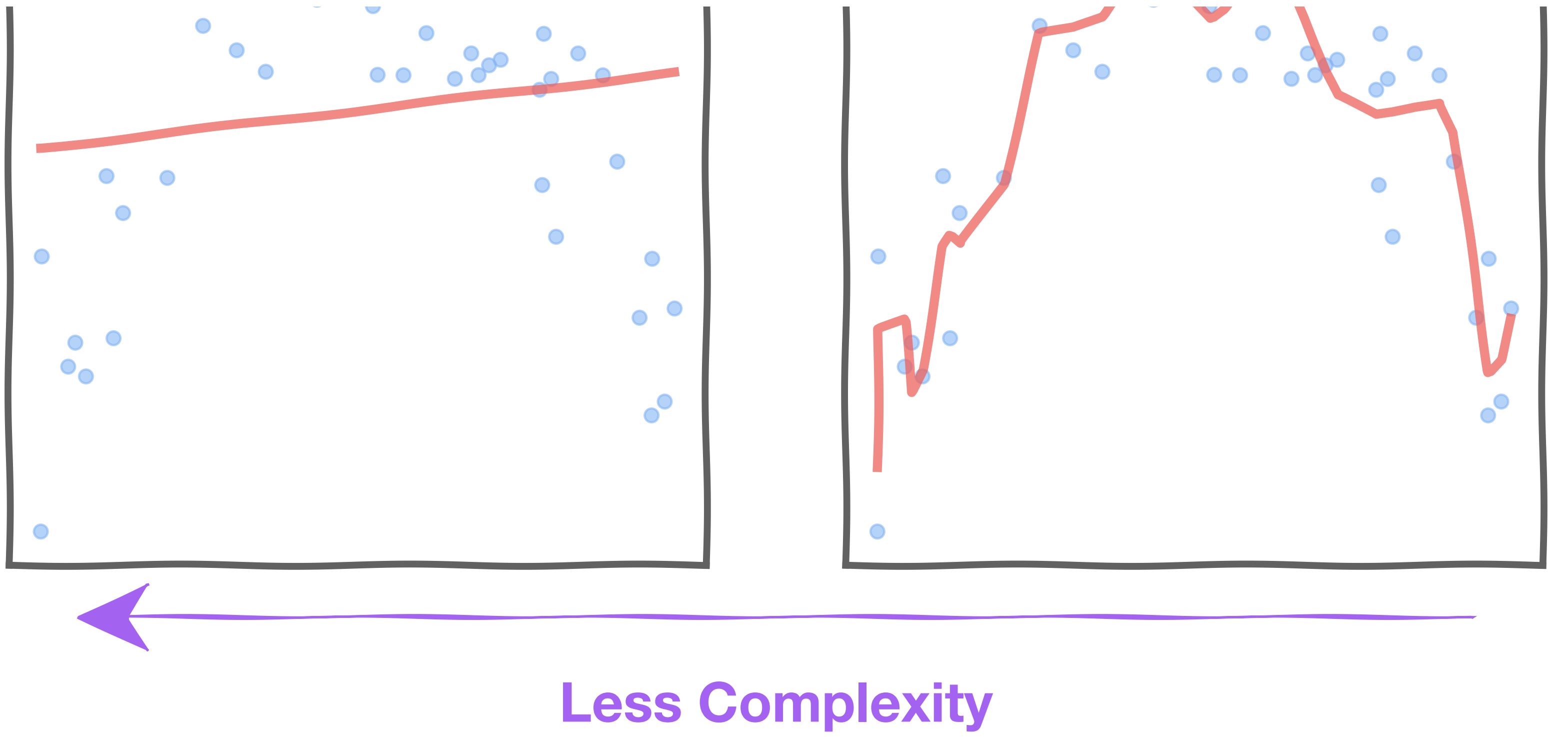
- **Irreducible error**: *This is due to the noise in the data*. We can’t do anything to decrease the this kind of error. This is also known as aleatoric error.
- **Reducible error**: *This is due to the model*. We can decrease the error due to overfitting and underfitting by improving the model. This kind of error is also known as epistemic error.

Model Complexity and Reducible Error

Reducible error comes from either *underfitting* or *overfitting*. Simple, less complex models are more likely to be underfit. However, as the complexity of our models increase, we are more likely to be overfitting.

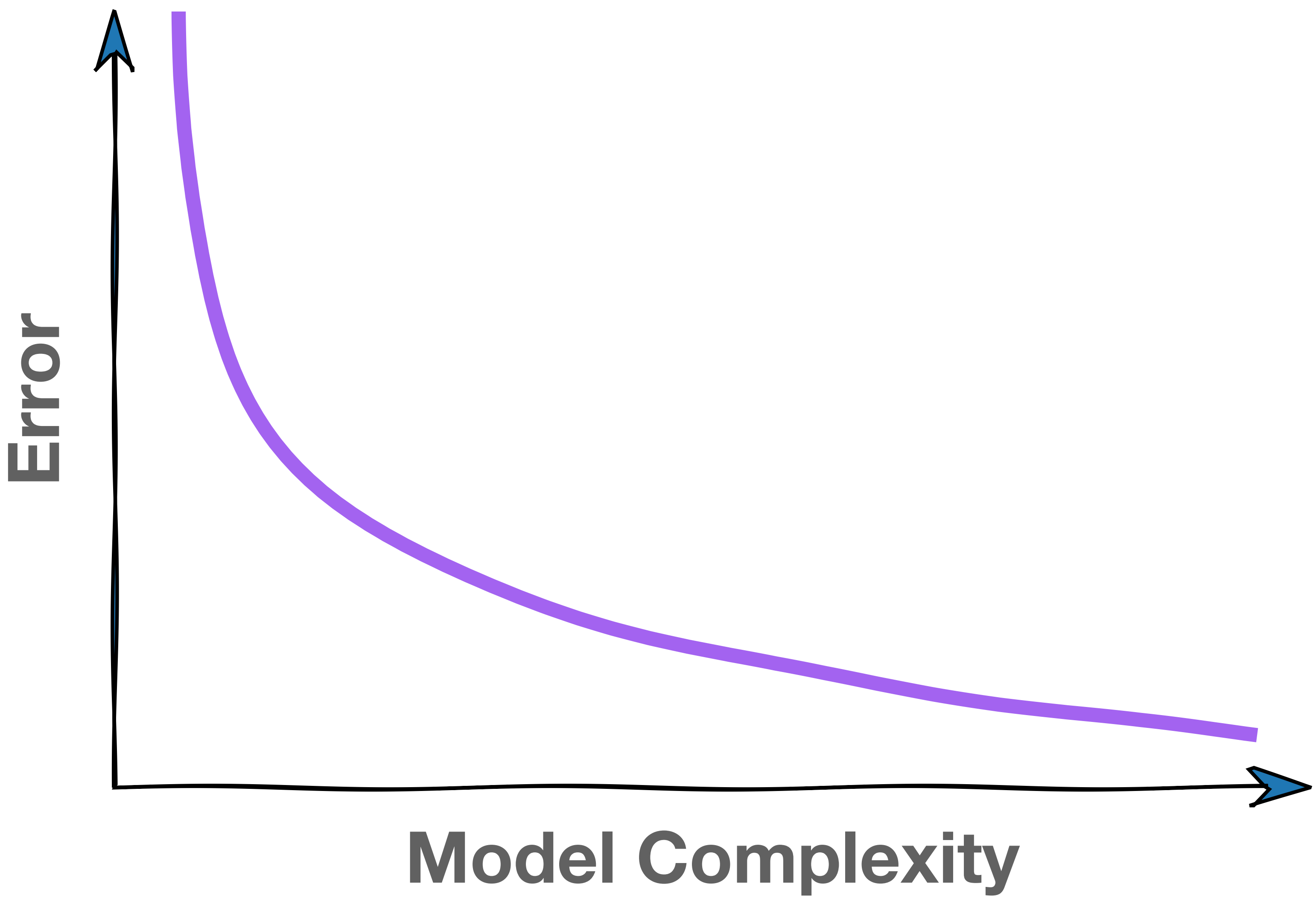
On the left we see a simple linear model, a low complexity model that is underfit to the data. While the data points roughly form a curved parabola, the simple linear model predicts a straight line through the center of the plot. On the right we see a high degree polynomial model, a very complex model that is overfit to the data. The data points form the same U-shaped curve as before. This time, the high degree polynomial model does predict along the center of the curve, but the line is very jagged, shooting up and down at unexpected points.





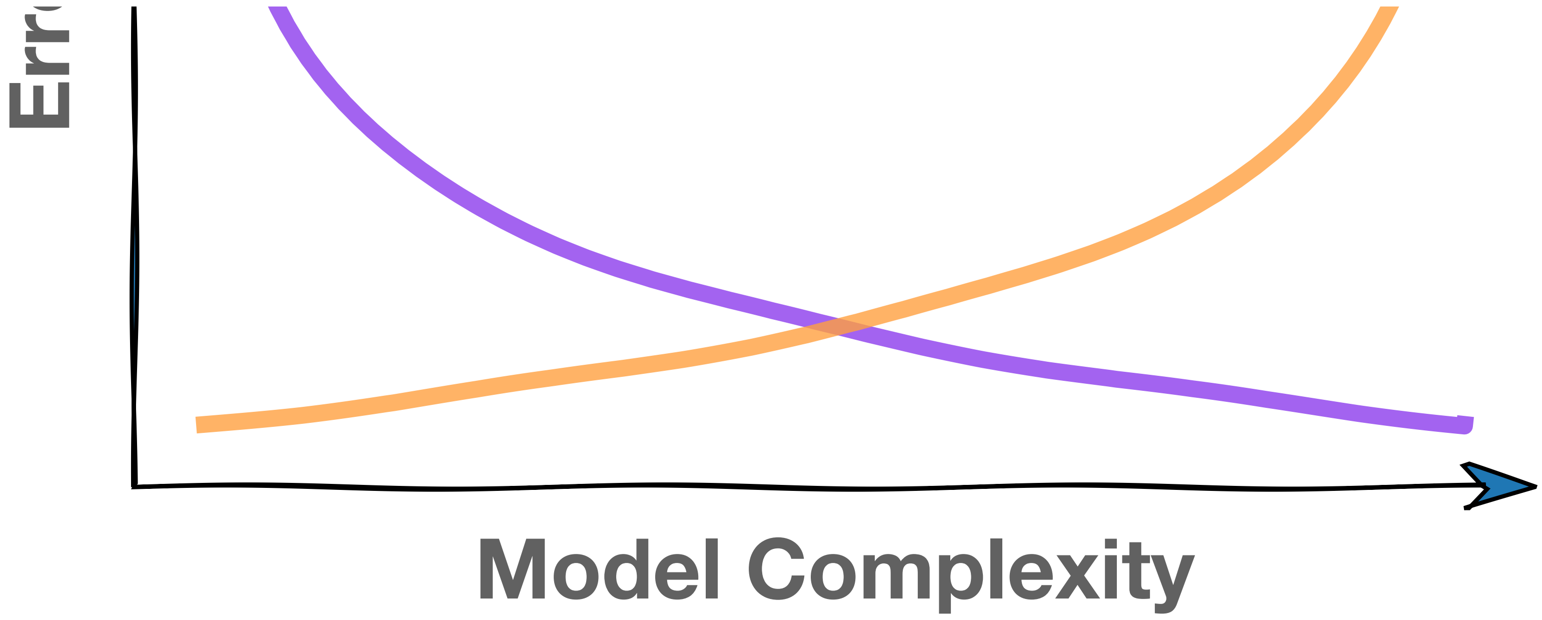
Bias and Variance

As the model complexity increases, the error on our training set will decrease. We call this the **bias**. Said another way, the bias will decrease as the model complexity increases. A low bias model will have it's prediction around the true value according to the training set.



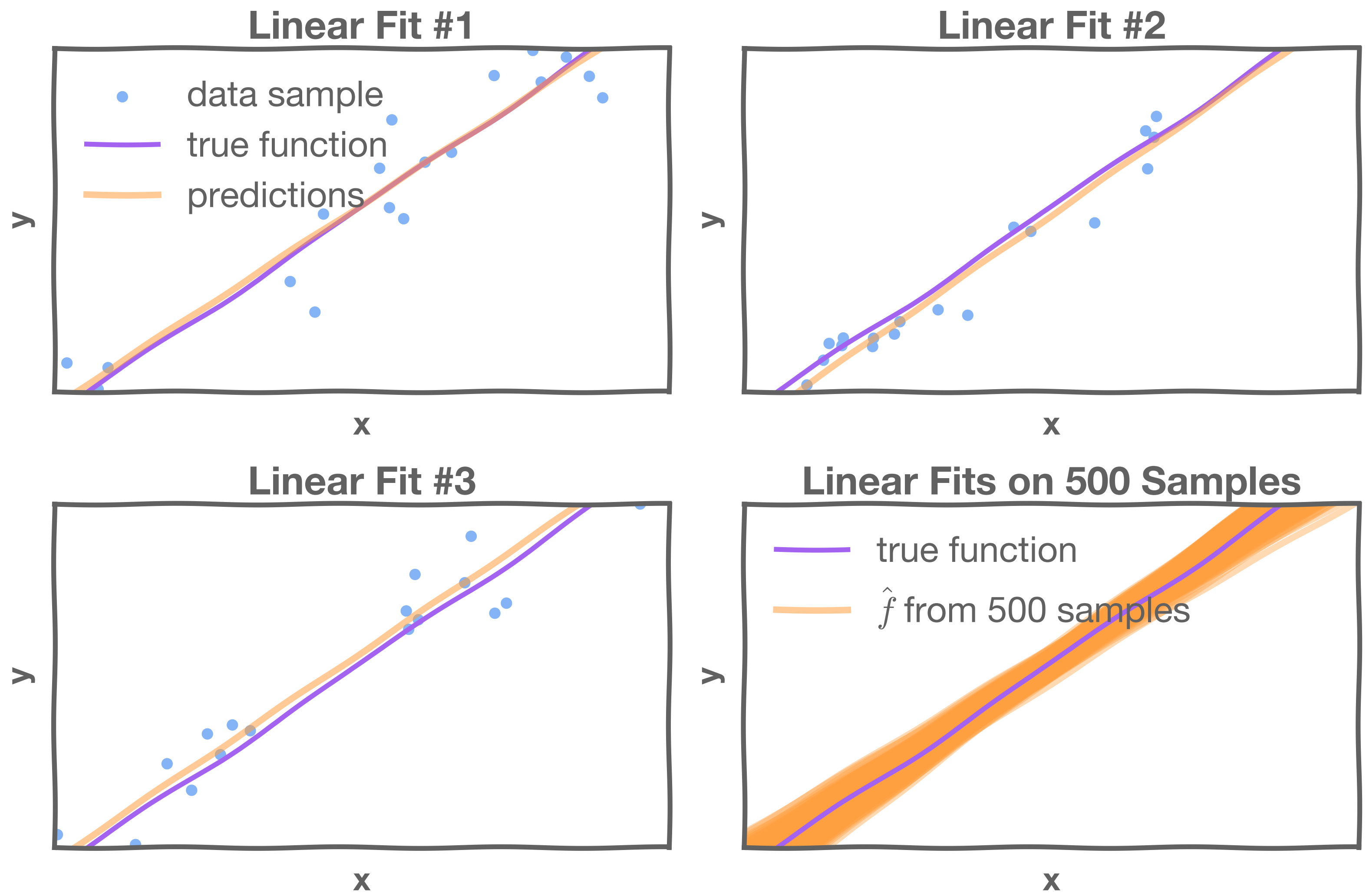
Model **variance**, on the other hand, is the variability among multiple fits of the *same* model on *different* training sets. You can think of variance as indicating how sensitive a model is to changes in its training data. More complex model are more sensitive and so variance will increase as the model complexity increases.

Calculator



Bias vs Variance: Variance of a Simple model

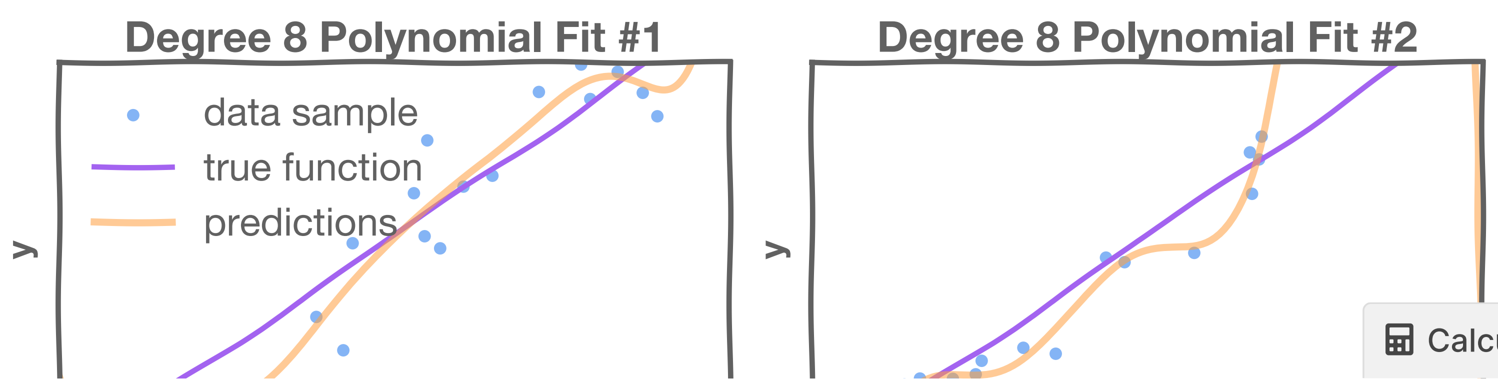
To visualize variance, lets look at a simple model. Each orange line is the same simple linear model but trained using a different split of the training data. Note that there is not much variation between prediction lines. Even when we fit on 500 different samples, all prediction lines are pretty close to each other.

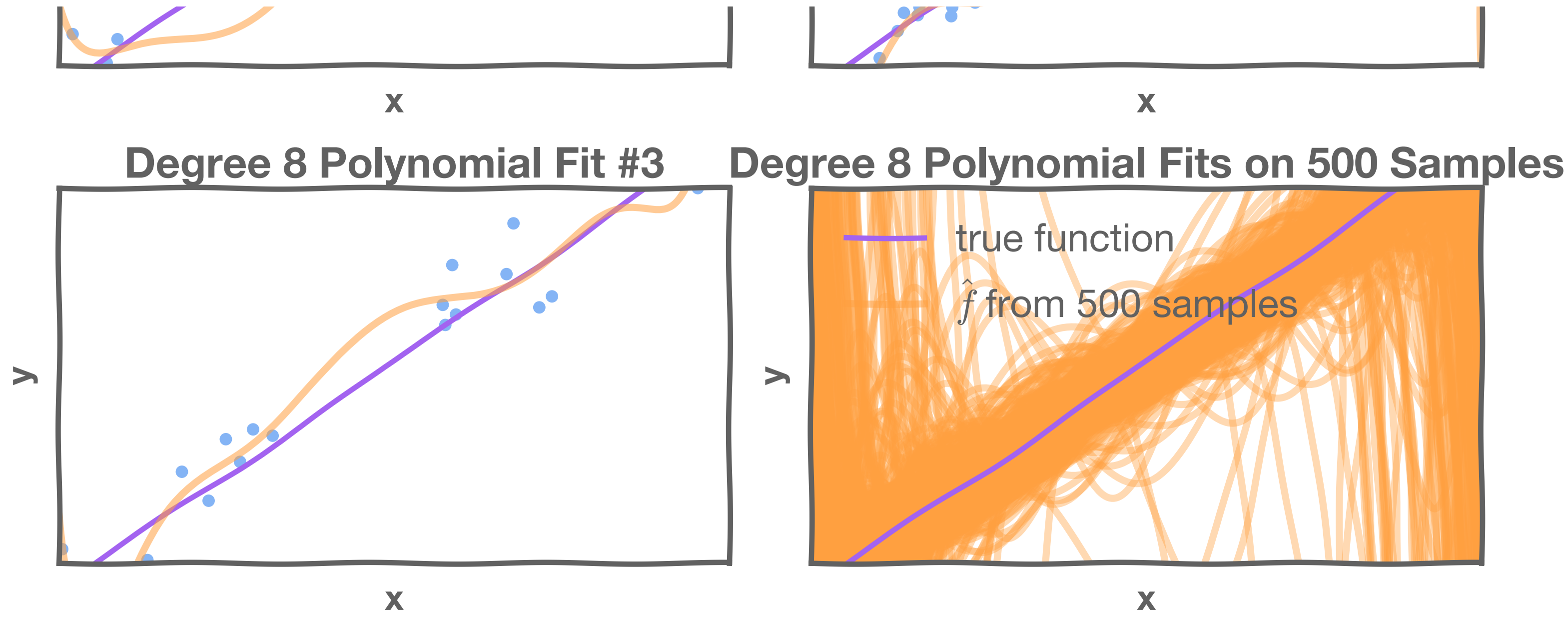


Bias vs Variance: Variance of a Complex model

Now let's do the same thing but using a more complex degree-8 polynomial model. Observe how the predictions now vary wildly between samples used to fit the model. When we plot the predictions of 500 different sample fits, the predictions may generally center around the line of the true function but the variation between predictions creates a mess of lines far above and below the true function line.

You may also notice that you can almost pick out the data points selected for the training sample in the first three plots. This is because the prediction line passes through those points exactly. This is indication of the low bias of the mode.

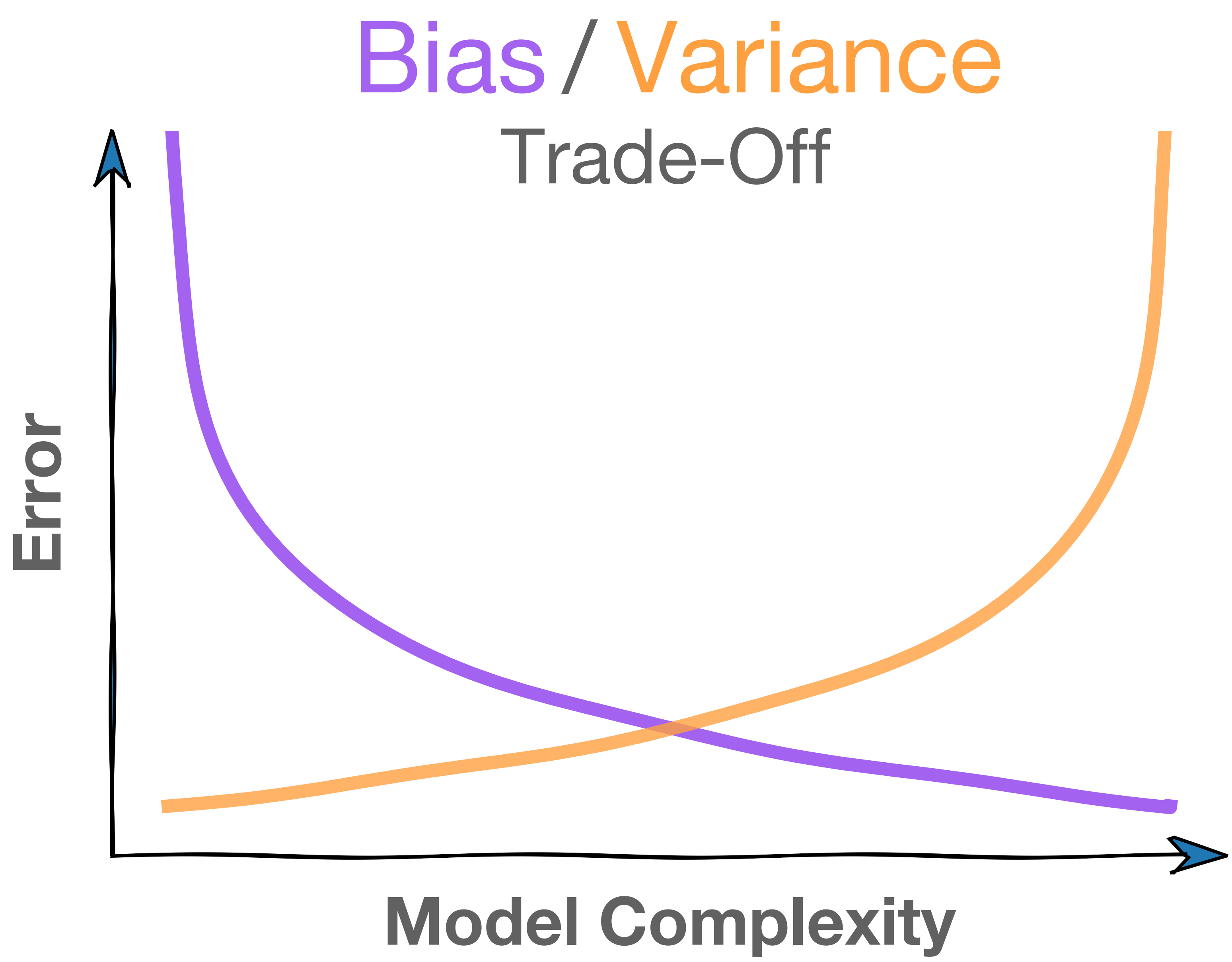




Bias vs Variance: the Trade-off

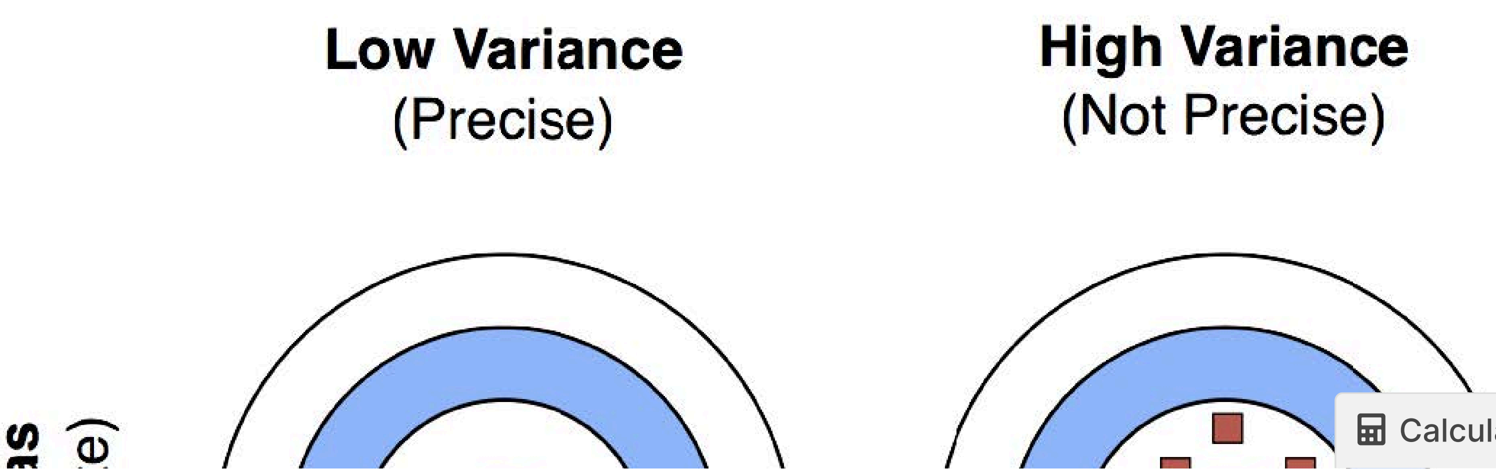
Now we see that the error of the bias decreases as our model complexity increases but the variance of the model increases. This is where we see the trade-off: we cannot lower one without increasing the other. We want to find balance between the two to get the lowest bias and variance possible.

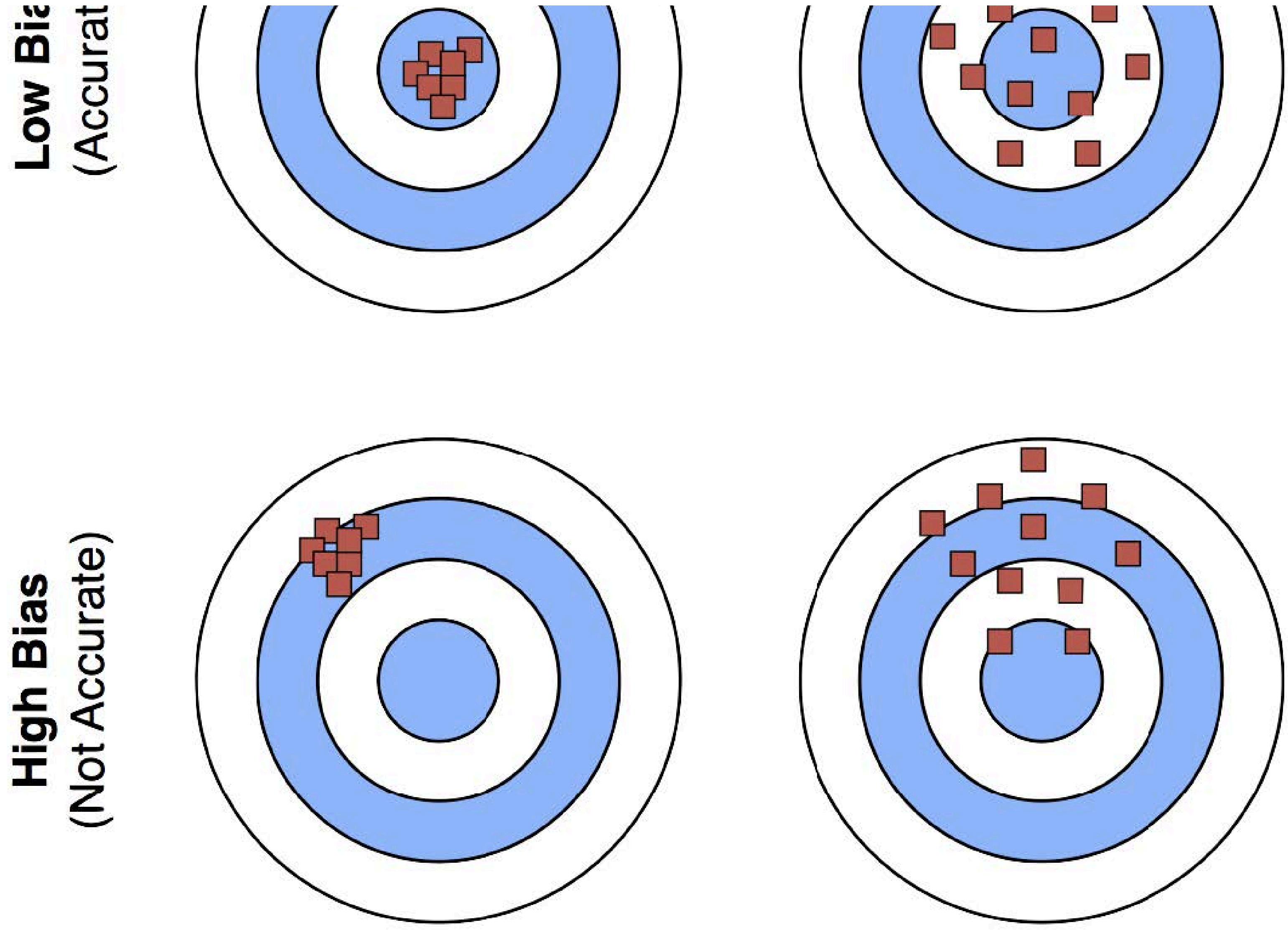
Based on the image below, the point where the lines intersect would be a good balance between bias and variance.



You can roughly think of **bias** as how *accurate* a model is and **variance** as how *precise* it is.

Consider the diagram below:





In the case of **high bias** and **low variance** the model shows very similar predictions regardless of what training data is used to fit it. However, these predictions are systematically off; predictions are wrong but all wrong in roughly the same way. It is *precise* but not *accurate*.

If we have **low bias** and **high variance** then the model is very complex and so very sensitive to changes in the training data. The predictions across the different fits are correct on average, but there is large variability or "spread" in the predictions. It is *accurate* but not *precise*.

We ideally want **low bias** and **low variance**. This would be a model whose predictions when fit on different training sets would be both *accurate* (centered on the true value) and *precise* (low spread).

If we have **high bias** and **high variance**, then model is neither *accurate* nor *precise*. So predictions are systematically off and there is also a lot of spread. If this is the case then we have a very poor model indeed. We would do well to reassess the steps and assumptions of our modeling approach in this situation.

Discussion Board (External resource)

Haga clic en Aceptar para que su nombre de usuario y dirección de correo electrónico se envíen a una aplicación de terceros.

Aceptar

‹ Previous

Next ›

Calculator