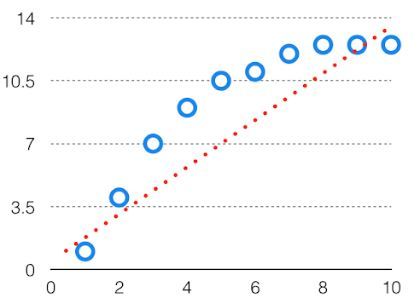
**Bias and Variance in Machine Learning**

There are various ways to evaluate a machine-learning model. We can use MSE ([Mean Squared Error](https://www.geeksforgeeks.org/mean-squared-error/)) for Regression; Precision, Recall, and ROC (Receiver of Characteristics) for a Classification Problem along with Absolute Error. In a similar way, Bias and Variance help us in parameter tuning and deciding better-fitted models among several built.

Bias is one type of error that occurs due to wrong assumptions about data such as assuming data is linear when in reality, data follows a complex function. On the other hand, variance gets introduced with high sensitivity to variations in training data. This also is one type of error since we want to make our model robust against noise. There are two types of error in machine learning. Reducible error and Irreducible error. Bias and Variance come under reducible error.

**What is Bias?**

The bias is known as the difference between the prediction of the values by the [Machine Learning](https://www.geeksforgeeks.org/machine-learning/) model and the correct value. Being high in biasing gives a large error in training as well as testing data. It recommended that an algorithm should always be low-biased to avoid the problem of underfitting. By high bias, the data predicted is in a straight line format, thus not fitting accurately in the data in the data set. Such fitting is known as the [**Underfitting**](https://www.geeksforgeeks.org/underfitting-and-overfitting-in-machine-learning/)**of Data**. This happens when the [hypothesis](https://www.geeksforgeeks.org/understanding-hypothesis-testing/) is too simple or linear in nature. Refer to the graph given below for an example of such a situation.



Bias is simply defined as the inability of the model because of that there is some difference or error occurring between the model’s predicted value and the actual value. These differences between actual or expected values and the predicted values are known as error or bias error or error due to bias. Bias is a systematic error that occurs due to wrong assumptions in the [machine learning](https://www.geeksforgeeks.org/machine-learning/) process.

Let Y*Y* be the true value of a parameter, and let Y^*Y*^ be an estimator of Y*Y* based on a sample of data. Then, the bias of the estimator Y^*Y*^ is given by:

Bias(Y^)=E(Y^)–YBias(*Y*^)=*E*(*Y*^)–*Y*

where E(Y^)  *E*(*Y*^)  is the expected value of the estimator Y^*Y*^. It is the measurement of the model that how well it fits the data.

* **Low Bias:** Low bias value means fewer assumptions are taken to build the target function. In this case, the model will closely match the training dataset.
* **High Bias:** High bias value means more assumptions are taken to build the target function. In this case, the model will not match the training dataset closely.

The high-bias model will not be able to capture the dataset trend. It is considered as the [underfitting](https://www.geeksforgeeks.org/underfitting-and-overfitting-in-machine-learning/) model which has a high error rate. It is due to a very simplified algorithm.

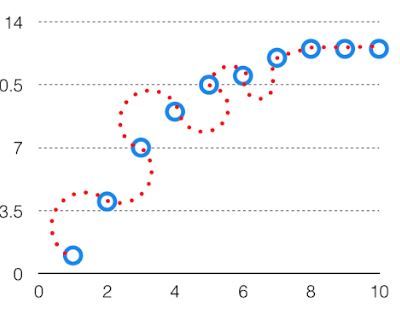
For example, a [linear regression](https://www.geeksforgeeks.org/ml-linear-regression/) model may have a high bias if the data has a non-linear relationship.

**Ways to reduce high bias in Machine Learning:**

* **Use a more complex model:**One of the main reasons for high bias is the very simplified model. it will not be able to capture the complexity of the data. In such cases, we can make our mode more complex by increasing the number of hidden layers in the case of a [deep neural network.](https://www.geeksforgeeks.org/introduction-deep-learning/) Or we can use a more complex model like [Polynomial regression](https://www.geeksforgeeks.org/python-implementation-of-polynomial-regression/) for [non-linear datasets](https://www.geeksforgeeks.org/non-linear-regression-examples-ml/), [CNN](https://www.geeksforgeeks.org/introduction-convolution-neural-network/) for [image processing](https://www.geeksforgeeks.org/image-processing/), and [RNN](https://www.geeksforgeeks.org/introduction-to-recurrent-neural-network/) for sequence learning.
* **Increase the number of features:** By adding more features to train the dataset will increase the complexity of the model. And improve its ability to capture the underlying patterns in the data.
* **Reduce**[**Regularization**](https://www.geeksforgeeks.org/regularization-in-machine-learning/)**of the model:**Regularization techniques such as [L1 or L2 regularization](https://www.geeksforgeeks.org/ml-implementing-l1-and-l2-regularization-using-sklearn/) can help to prevent [overfitting](https://www.geeksforgeeks.org/underfitting-and-overfitting-in-machine-learning/) and improve the generalization ability of the model. if the model has a high bias, reducing the strength of regularization or removing it altogether can help to improve its performance.
* **Increase the size of the training data:** Increasing the size of the training data can help to reduce bias by providing the model with more examples to learn from the dataset.

**What is Variance?**

The variability of model prediction for a given data point which tells us the spread of our data is called the variance of the model. The model with high variance has a very complex fit to the training data and thus is not able to fit accurately on the data which it hasn’t seen before. As a result, such models perform very well on training data but have high error rates on test data. When a model is high on variance, it is then said to as **Overfitting of Data**. Overfitting is fitting the training set accurately via complex curve and high order hypothesis but is not the solution as the error with unseen data is high. While training a data model variance should be kept low. The high variance data looks as follows.



Variance is the measure of spread in data from its [mean](https://www.geeksforgeeks.org/mathematics-mean-variance-and-standard-deviation/) position. In machine learning variance is the amount by which the performance of a predictive model changes when it is trained on different subsets of the training data. More specifically, variance is the variability of the model that how much it is sensitive to another subset of the training dataset. i.e. how much it can adjust on the new subset of the training dataset.

Let Y be the actual values of the target variable, and  Y^  *Y*^   be the predicted values of the target variable. Then the [variance](https://www.geeksforgeeks.org/mathematics-mean-variance-and-standard-deviation/) of a model can be measured as the expected value of the square of the difference between predicted values and the expected value of the predicted values.

Variance=E[(Y^–E[Y^])2]Variance=*E*[(*Y*^–*E*[*Y*^])2]

where E[Yˉ]*E*[*Y*ˉ] is the expected value of the predicted values. Here expected value is averaged over all the training data.

Variance errors are either low or high-variance errors.

* **Low variance:** Low variance means that the model is less sensitive to changes in the training data and can produce consistent estimates of the target function with different subsets of data from the same [distribution](https://www.geeksforgeeks.org/introduction-of-statistical-data-distributions/). This is the case of underfitting when the model fails to generalize on both training and test data.
* **High variance:** High variance means that the model is very sensitive to changes in the training data and can result in significant changes in the estimate of the target function when trained on different subsets of data from the same distribution. This is the case of overfitting when the model performs well on the training data but poorly on new, unseen test data. It fits the training data too closely that it fails on the new training dataset.

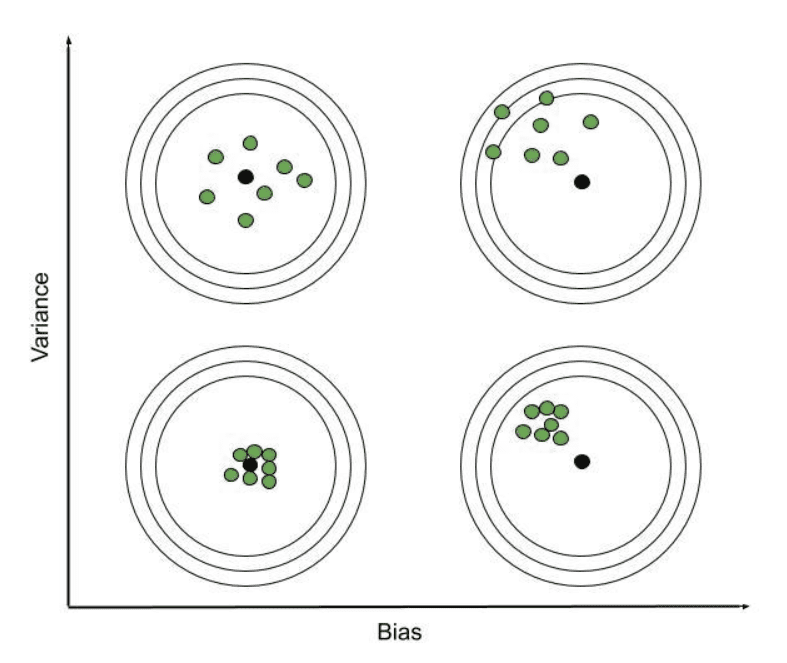
**Ways to Reduce the reduce Variance in Machine Learning:**

* [**Cross-validation**](https://www.geeksforgeeks.org/cross-validation-machine-learning/)**:** By splitting the data into training and testing sets multiple times, cross-validation can help identify if a model is overfitting or underfitting and can be used to tune hyperparameters to reduce variance.
* [**Feature selection:**](https://www.geeksforgeeks.org/feature-selection-techniques-in-machine-learning/)By choosing the only relevant feature will decrease the model’s complexity. and it can reduce the variance error.
* [**Regularization**](https://www.geeksforgeeks.org/regularization-in-machine-learning/)**:** We can use L1 or L2 regularization to reduce variance in machine learning models
* [**Ensemble methods**](https://www.geeksforgeeks.org/ensemble-classifier-data-mining/)**:** It will combine multiple models to improve generalization performance. [Bagging, boosting](https://www.geeksforgeeks.org/bagging-vs-boosting-in-machine-learning/), and stacking are common ensemble methods that can help reduce variance and improve generalization performance.
* **Simplifying the model:**Reducing the complexity of the model, such as decreasing the number of parameters or layers in a neural network, can also help reduce variance and improve generalization performance.
* [**Early stopping**](https://www.geeksforgeeks.org/choose-optimal-number-of-epochs-to-train-a-neural-network-in-keras/)**:** Early stopping is a technique used to prevent overfitting by stopping the training of the deep learning model when the performance on the validation set stops improving.

**Different Combinations of Bias-Variance**

There can be four combinations between bias and variance.

* **High Bias, Low Variance:** A model with high bias and low variance is said to be underfitting.
* **High Variance, Low Bias:**A model with high variance and low bias is said to be overfitting.
* **High-Bias, High-Variance:**A model has both high bias and high variance, which means that the model is not able to capture the underlying patterns in the data (high bias) and is also too sensitive to changes in the training data (high variance). As a result, the model will produce inconsistent and inaccurate predictions on average.
* **Low Bias, Low Variance:** A model that has low bias and low variance means that the model is able to capture the underlying patterns in the data (low bias) and is not too sensitive to changes in the training data (low variance). This is the ideal scenario for a machine learning model, as it is able to generalize well to new, unseen data and produce consistent and accurate predictions. But in practice, it’s not possible.



*Bias-Variance Combinations*

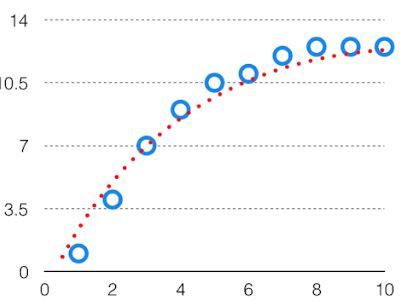
Now we know that the ideal case will be **Low Bias and Low variance**, but in practice, it is not possible. So, we trade off between Bias and variance to achieve a balanced bias and variance.

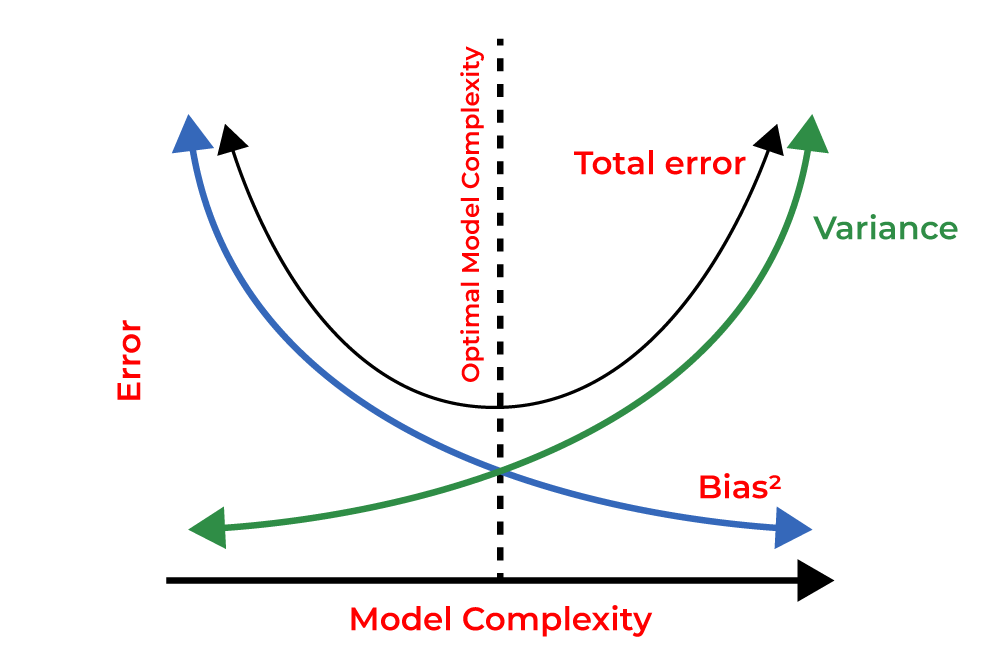
A model with balanced bias and variance is said to have optimal generalization performance. This means that the model is able to capture the underlying patterns in the data without overfitting or underfitting. The model is likely to be just complex enough to capture the complexity of the data, but not too complex to overfit the training data. This can happen when the model has been carefully tuned to achieve a good balance between bias and variance, by adjusting the hyperparameters and selecting an appropriate model architecture.

| **Machine Learning Algorithm** | **Bias** | **Variance** |
| --- | --- | --- |
| [Linear Regression](https://www.geeksforgeeks.org/ml-linear-regression/) | High Bias | Less Variance |
| [Decision Tree](https://www.geeksforgeeks.org/decision-tree/) | Low Bias | High Variance |
| [Random Forest](https://www.geeksforgeeks.org/random-forest-regression-in-python/) | Low Bias | High Variance |
| [Bagging](https://www.geeksforgeeks.org/bagging-vs-boosting-in-machine-learning/) | Low Bias | High Variance |

**Bias Variance Tradeoff**

If the algorithm is too simple (hypothesis with linear equation) then it may be on high bias and low variance condition and thus is error-prone. If algorithms fit too complex (hypothesis with high degree equation) then it may be on high variance and low bias. In the latter condition, the new entries will not perform well. Well, there is something between both of these conditions, known as a Trade-off or Bias Variance Trade-off. This tradeoff in complexity is why there is a tradeoff between bias and variance. An algorithm can’t be more complex and less complex at the same time. For the graph, the perfect tradeoff will be like this.





*Bias-Variance Tradeoff*

The technique by which we analyze the performance of the machine learning model is known as Bias Variance Decomposition. Now we give 1-1 example of Bias Variance Decomposition for [classification](https://www.geeksforgeeks.org/getting-started-with-classification/) and [regression](https://www.geeksforgeeks.org/types-of-regression-techniques/).

**Bias Variance Decomposition for Classification and Regression**

As per the formula, we have derived total error as the sum of Bias squares and variance. We try to make sure that the bias and the variance are comparable and one does not exceed the other by too much difference.

[Bias Variance Tradeoff - Clearly Explained | Machine Learning Plus](https://www.machinelearningplus.com/machine-learning/bias-variance-tradeoff/)