**BIAS AND VARIANCE :**

**BIAS :**

* 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 values and the predicted values are known as bias error or error due to bias.
* Bias is a systematic error that occurs due to wrong assumptions in the machine learning process.

Formula :



**Low Bias**: Low bias value which means few assumptions are taken to build the target function. In this case, the model will closely match the training

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

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

1. **Use a more complex model:** High bias occurs when a model is overly simplified and cannot capture the data's complexity. To address this, we can increase model complexity, such as adding hidden layers in deep neural networks.
2. **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**.**
3. **Reduce Regularization of the model:** Regularization is the techniques such as L1 or L2 regularization can help to prevent overfitting and improve the generalization ability of the model
4. **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 .

**Advantages:**

* Simpler models are easy to understand and interpret.
* Less likely to overfit (less focus on noise).

**Disadvantages:**

* Can lead to underfitting (missing important patterns).
* Poor performance on both training and new data.

**VARIANCE :**

* 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 extent to which a model's performance or predictions change when it is trained on different subsets of the training data.
* It reflects the model's sensitivity to variations in the training data.

Formula :



**Variance errors** can be categorized as **low variance** or **high variance**:

1. **Low Variance**:
   * The model is less sensitive to changes in the training data and produces consistent predictions across different subsets.
   * However, this can lead to **underfitting**, where the model is too simple and fails to capture patterns in both the training and test data.
2. **High Variance**:
   * The model is very sensitive to changes in the training data, leading to significant variations in predictions across different subsets.
   * This results in **overfitting**, where the model performs well on the training data but poorly on unseen test data.

**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

**Advantages:**

* Can capture complex patterns.
* High accuracy on training data.

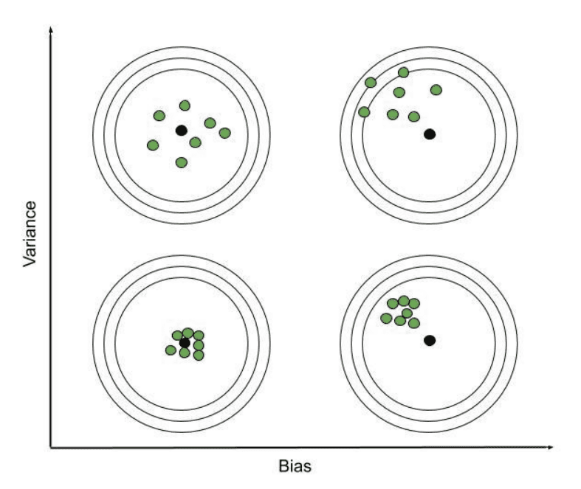
**Disadvantages:**

* Can lead to overfitting (learning noise instead of actual patterns).
* Poor generalization to new data.

**Different Combinations of Bias-Variance**

There can be four combinations between bias and variance.

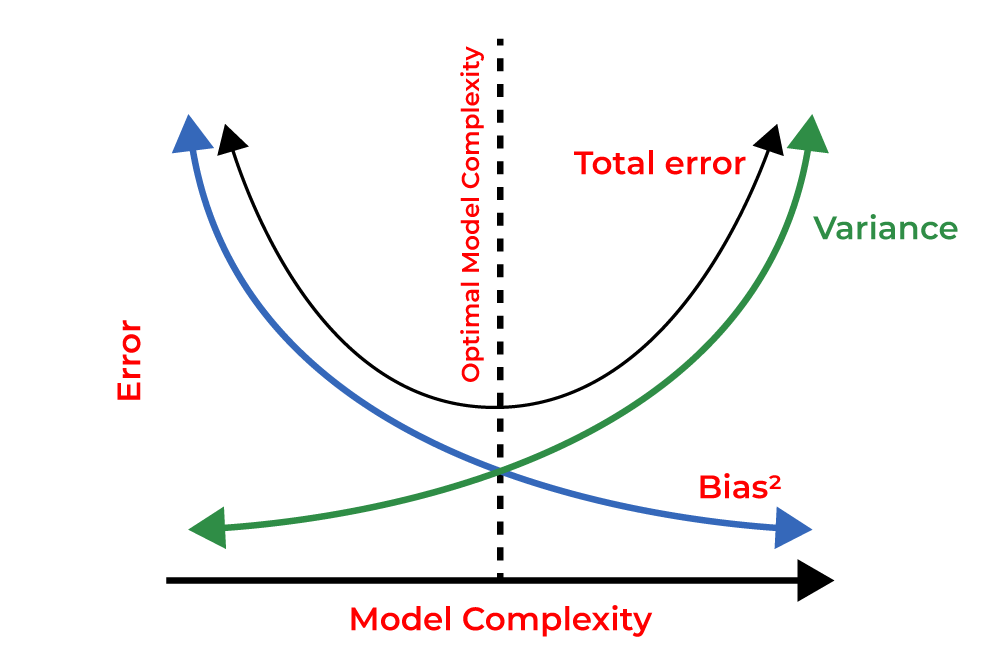
* **High Bias, Low Variance:** A model with high bias and low variance is said to be underfitting.
* **High Variance, Low Bias:**A model with high variance and low bias is said to be overfitting.
* **High-Bias, High-Variance:**A model has both high bias and high variance, which means that the model is 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).
* **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).



**Bias Variance Tradeoff**

If the algorithm is too simple then it may be on high bias and low variance condition and thus is error-prone. If algorithms fit too complex then it may be on high variance and low bias. n 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.



The **tradeoff** lies in finding the right balance where the model is neither too simple nor too complex. At this point, the model achieves a **low total error**, generalizing well on both training and Testing data. A graph of this tradeoff typically shows:

* Bias decreasing as model complexity increases.
* Variance increasing with model complexity.
* The **optimal point** is where the combined error from bias and variance is minimized.

**Balance Between Bias and Variance**

* Striking the right balance is crucial for building effective models.
* Low bias and low variance help in creating models that generalize well to new data.