1. What is Bias Variance Tradeoff ?

The bias-variance tradeoff is a fundamental concept in machine learning that deals with finding the right balance between two sources of error, bias and variance, when building predictive models. It is a key consideration when developing machine learning algorithms and models because it helps us understand and manage the tradeoff between model complexity and model performance.

Here's an explanation of bias and variance in the context of machine learning:

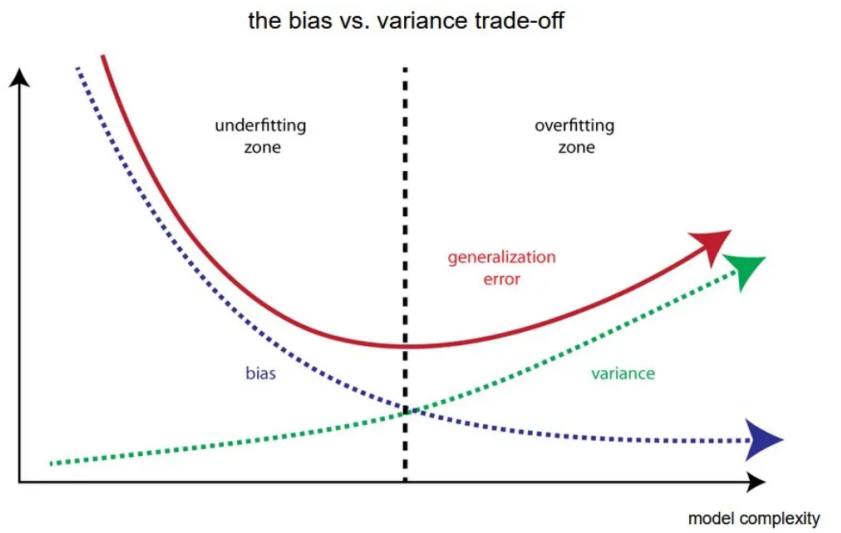
Bias (Underfitting):

Bias represents the error due to overly simplistic assumptions in the learning algorithm. Models with high bias are too simplistic and do not capture the underlying patterns in the data.
High bias often leads to underfitting, where the model is too rigid and performs poorly on both the training data and unseen data. It fails to generalize well.

Variance (Overfitting):

Variance represents the error due to excessive complexity in the learning algorithm. Models with high variance are too flexible and can fit noise or random fluctuations in the training data.
High variance often leads to overfitting, where the model performs exceptionally well on the training data but poorly on unseen data. It fails to generalize because it has essentially memorized the training data.
The tradeoff between bias and variance can be summarized as follows:

If a model is too simple (high bias), it may not capture the underlying patterns in the data and will have poor performance (underfitting).
If a model is too complex (high variance), it may fit the training data very well but will generalize poorly to new, unseen data (overfitting).



The dashed vertical line in the middle shows a sort of "happy medium" (no pun intended) where the model's complexity is just the right amount for the model to be neither underfitting nor overfitting. The vertical axis represents the model's prediction error. The red curve represents the average error on the test data. Image by author.

The goal in machine learning is to find a model that strikes the right balance between bias and variance. This is typically achieved through techniques like hyperparameter tuning, cross-validation, and regularization:

Hyperparameter Tuning: Adjusting hyperparameters (e.g., learning rate, model depth, number of features) to find the optimal tradeoff between bias and variance.

Cross-Validation: Evaluating a model's performance on multiple subsets of the data to estimate its generalization error and choose a model that minimizes this error.

Regularization: Adding constraints or penalties to the model to reduce its complexity and prevent overfitting. Common regularization techniques include L1 (Lasso) and L2 (Ridge) regularization.

In summary, the bias-variance tradeoff is a critical concept in machine learning that emphasizes the importance of finding the right level of model complexity to achieve good generalization on unseen data while avoiding both underfitting and overfitting.

Note: Its very important that our model perform well for new sets of training data and variance among the results of these models should be low. There has to be a right balance between these two. Now if our model is too simplistic (under fitted/high bias) then the variance among the different model performance resulted from the different training sets will be very low and if the model is complex with low bias(over fitted) then the variance among those model performance resulting from different training data sets will be high. Because in this case our model will be very prone to learn the intricate and less important detail of our training set and may not be generalizable for new test data. So we need to optimize these two and hence it’s a tradeoff between bias and variance.