ML
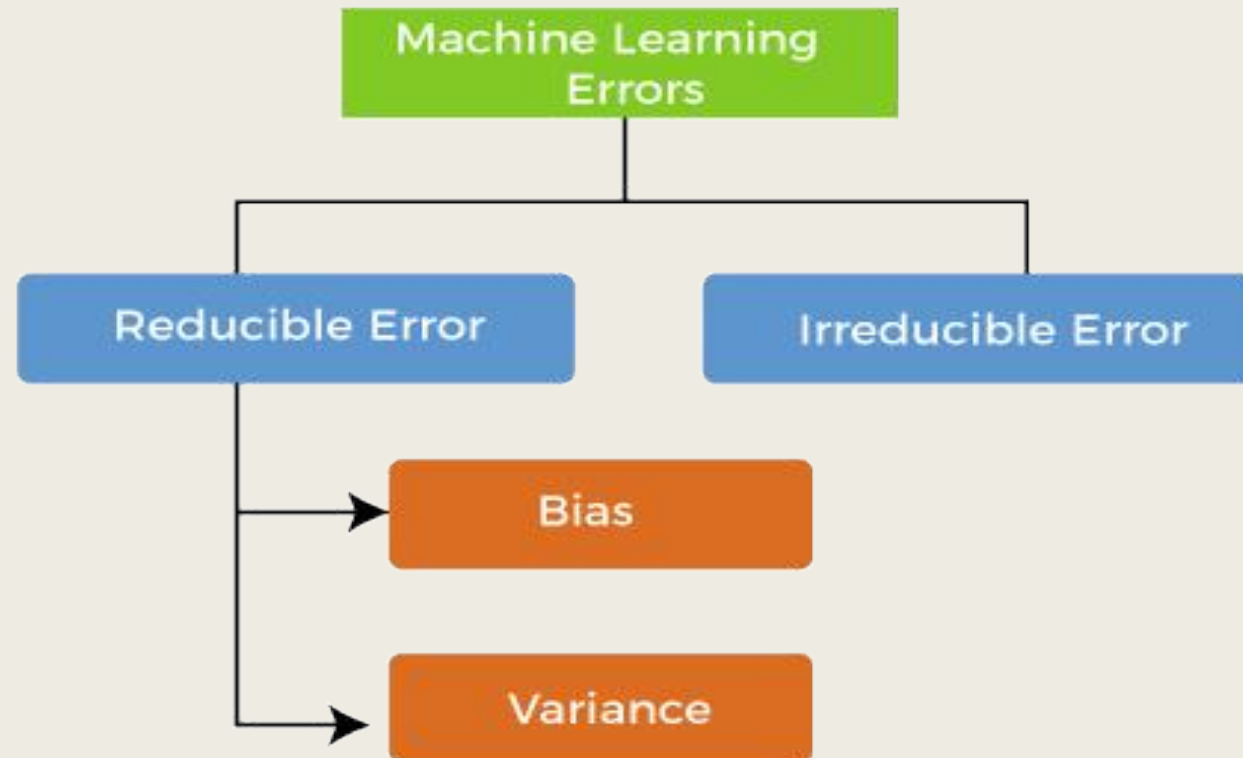
Errors in Machine Learning

❖ In machine learning, an **error is a measure of how accurately an algorithm can make predictions for the previously unknown dataset**.

❖ On the basis of these errors, the machine learning model is selected that can perform best on the particular dataset. There are mainly two types of errors in machine learning, which are:

❖ **Reducible errors:** These errors **can be reduced to improve the model accuracy**. Such errors can further be classified into bias and Variance.

❖ **Irreducible errors:** These errors **will always be present in the model.**

What is Bias?

❖ While making predictions, **a difference occurs between prediction values made by the model and actual values, and this difference is known as bias errors** or Errors due to bias.

❖ It can be defined as **an inability of machine learning algorithms such as Linear Regression to capture the true relationship between the data points.**

❖ Each algorithm begins with some amount of bias because bias occurs from assumptions in the model, which makes the target function simple to learn. A model has either:

❖ **Low Bias:** A low bias model **will make fewer assumptions about the form of the target function.**

❖ **High Bias:** A model with a high bias **makes more assumptions, and the model becomes unable to capture the important features of our dataset.** A high bias model also cannot perform well on new data.

❖ Generally, a linear algorithm has a high bias, as it makes them learn fast. **The simpler the algorithm, the higher the bias it has likely to be introduced. Whereas a nonlinear algorithm often has low bias.**

❖ Some examples of machine learning algorithms with **low bias are Decision Trees, k-Nearest Neighbours and Support Vector Machines.**

❖ At the same time, **an algorithm with high bias is Linear Regression, Linear Discriminant Analysis and Logistic Regression.**

Ways to reduce High Bias

❖ High bias mainly **occurs due to a much simple model.**

❖ Below are some ways to reduce the high bias:

❖ **Increase the input features** as the model is underfitted.

❖ **Decrease the regularization term.**

❖ **Use more complex models, such as including some polynomial features**

What is a Variance Error?

❖ The variance specify the **amount of variation in the prediction if the different training data was used.**

❖ **Ideally, a model should not vary too much from one training dataset to another, which means the algorithm should be good in understanding the hidden mapping between inputs and output variables.**

❖ Variance errors are either of low variance or high variance.

❖ **Low variance means there is a small variation in the prediction of the target function with changes in the training data set.**

❖ At the same time, **High variance shows a large variation in the prediction of the target function with changes in the training dataset.**

❖ A model that shows high variance learns a lot and perform well with the training dataset, and does not generalize well with the unseen dataset.

❖ As a result, such a model gives good results with the training dataset but shows high error rates on the test dataset.

What is a Variance Error?

❖ **With high variance, the model learns too much from the dataset, it leads to overfitting of the model.**

❖ A model with high variance has the below problems:

■ A high variance model **leads to overfitting.**

■ **Increase model complexities.**

■ Usually, **nonlinear algorithms have a lot of flexibility to fit the model, have high variance.**

Underfitting   Optimal Balance   Overfitting
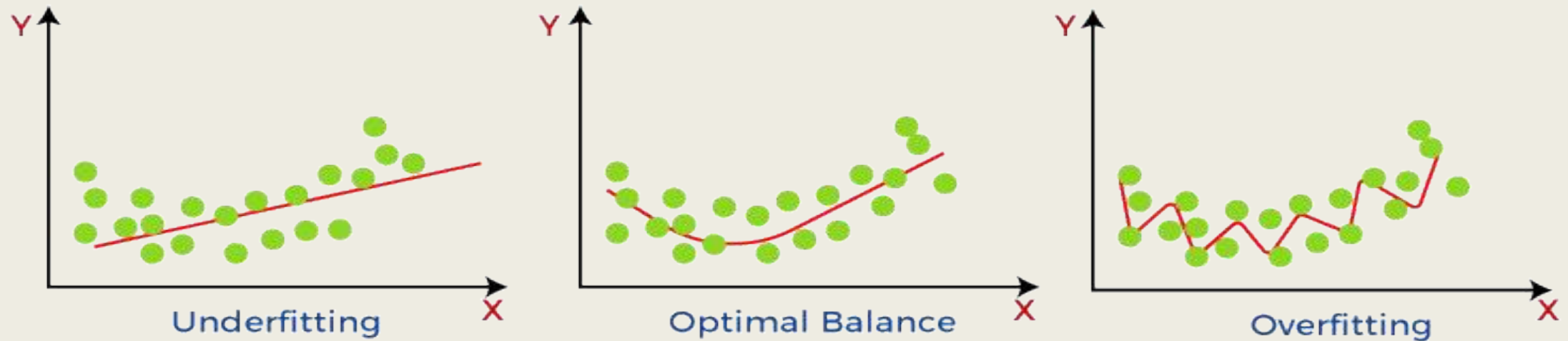
❖ Some examples of machine learning **algorithms with low variance are, Linear Regression, Logistic Regression, and Linear discriminant analysis.**

❖ At the same time, **algorithms with high variance are decision tree, Support Vector Machine, and K-nearest neighbours.**

Ways to Reduce High Variance

- **Reduce the input features or number of parameters** as a model is overfitted.

- **Do not use a much complex model.**

- **Increase the training data.**

- **Increase the Regularization term.**

Different Combinations of Bias-Variance

❖ **There are four possible combinations of bias and variances:-**

■ **Low-Bias, Low-Variance:** The combination of low bias and low variance shows an **ideal machine learning model. However, it is not possible practically.**

■ **Low-Bias, High-Variance:** With low bias and high variance, model **predictions are inconsistent and accurate on average.** This case occurs when the **model learns with a large number of parameters and hence leads to an overfitting**

■ **High-Bias, Low-Variance:** With High bias and low variance, **predictions are consistent but inaccurate on average.** This case occurs when a **model does not learn well with the training dataset or uses few numbers of the parameter. It leads to underfitting problems in the model.**

■ **High-Bias, High-Variance:** With high bias and high variance, **predictions are inconsistent and also inaccurate on average.**

How to identify High variance or High Bias?

❖ High variance can be identified if the model has:

❖ **Low training error and high test error.**

❖ High Bias can be identified if the model has:

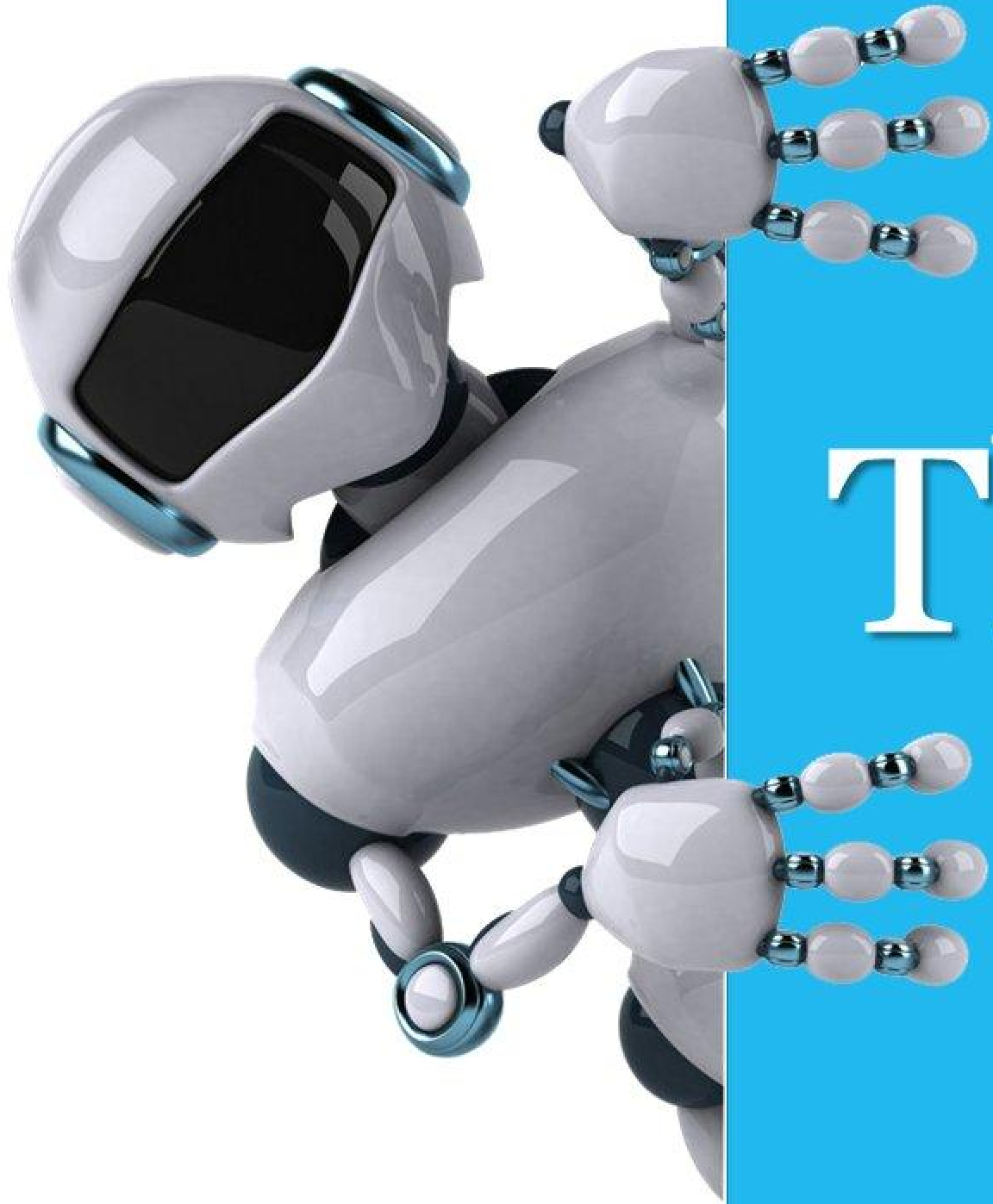❖ **High training error and the test error is almost similar to training error.**
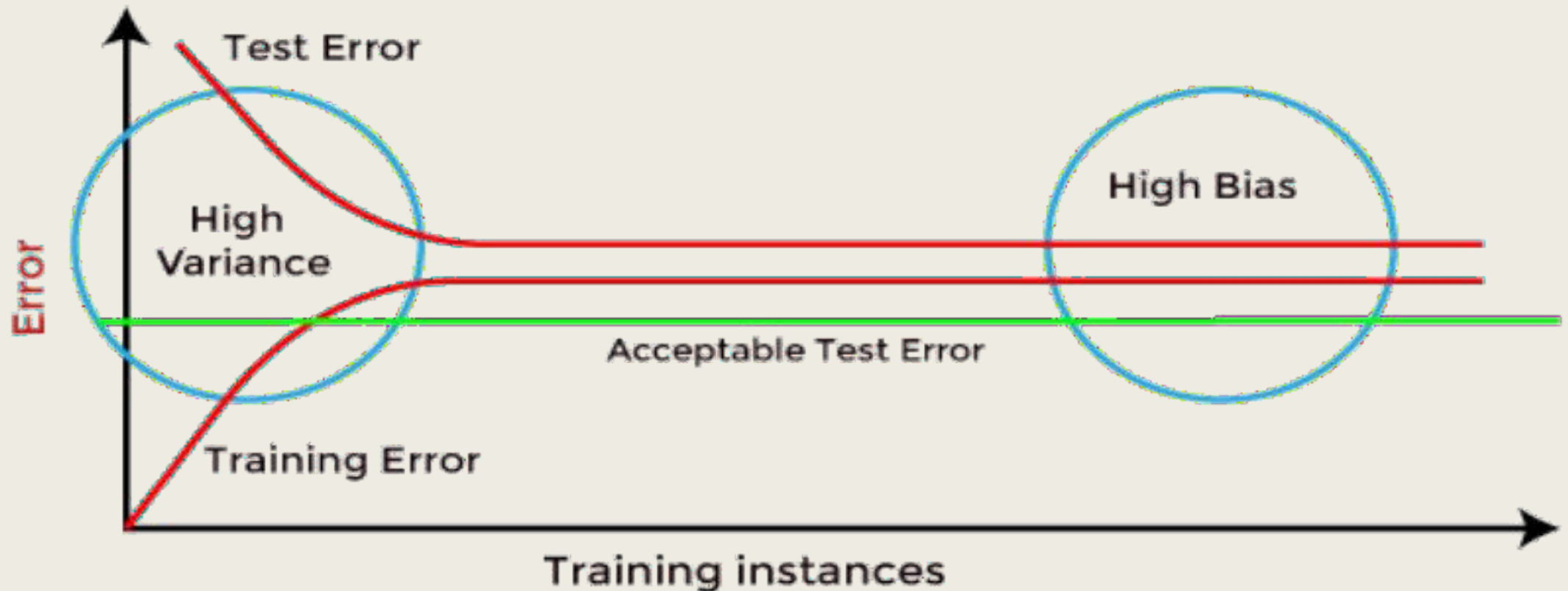
Bias-Variance Trade-Off

❖ While building the machine learning model, it is really important to take care of bias and variance in order to avoid overfitting and underfitting in the model.

❖ If the model is very simple with fewer parameters, it may have low variance and high bias.

❖ Whereas, if the model has a large number of parameters, it will have high variance and low bias.

❖ So, **it is required to make a balance between bias and variance errors, and this balance between the bias error and variance error is known as the Bias-Variance trade-off.**
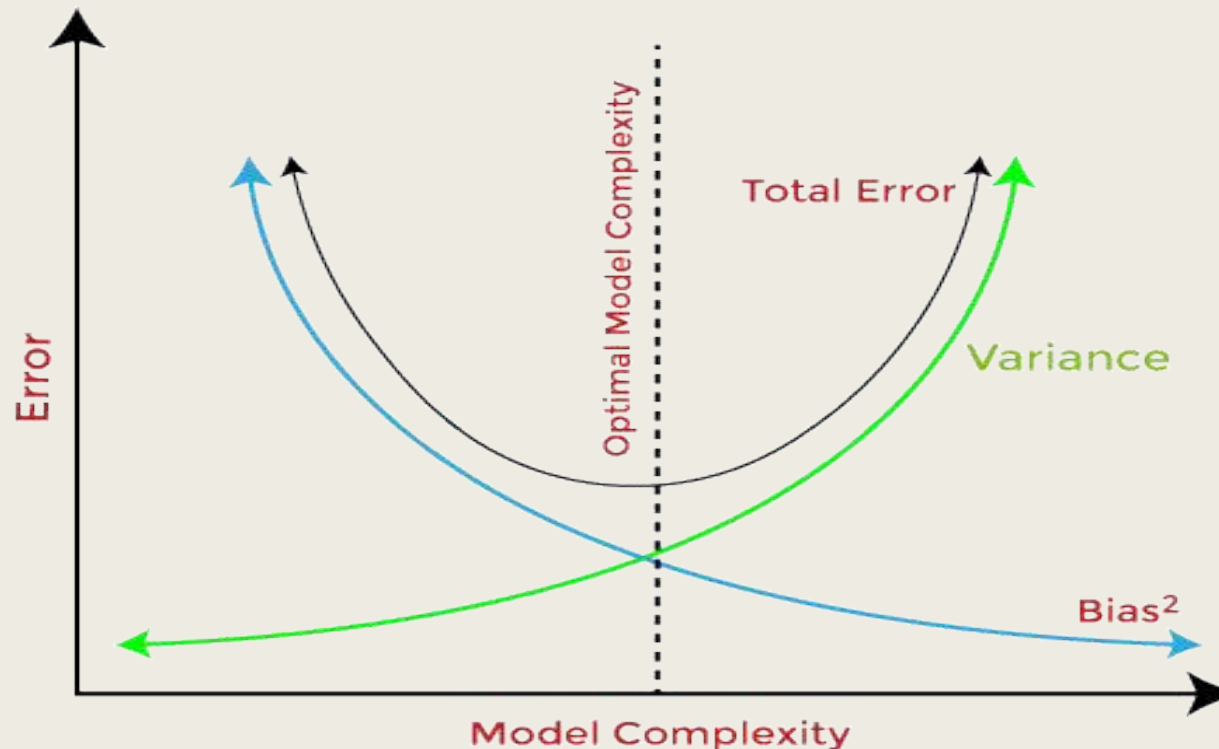
Bias-Variance Trade-Off

❖ For an accurate prediction of the model, **algorithms need a low variance and low bias. But this is not possible because bias and variance are related to each other:**

    ■ **If we decrease the variance, it will increase the bias.**

    ■ **If we decrease the bias, it will increase the variance.**

❖ Bias-Variance trade-off is a central issue in supervised learning.

❖ Ideally, **we need a model that accurately captures the regularities in training data and simultaneously generalizes well with the unseen dataset.**

❖ Unfortunately, doing this is not possible simultaneously. Because a high variance algorithm may perform well with training data, but it may lead to overfitting to noisy data.

❖ Whereas, high bias algorithm generates a much simple model that may not even capture important regularities in the data. So, we need to find a sweet spot between bias and variance to make an optimal model.

❖ **Hence, the Bias-Variance trade-off is about finding the sweet spot to make a balance between bias and variance errors.**

Thank you