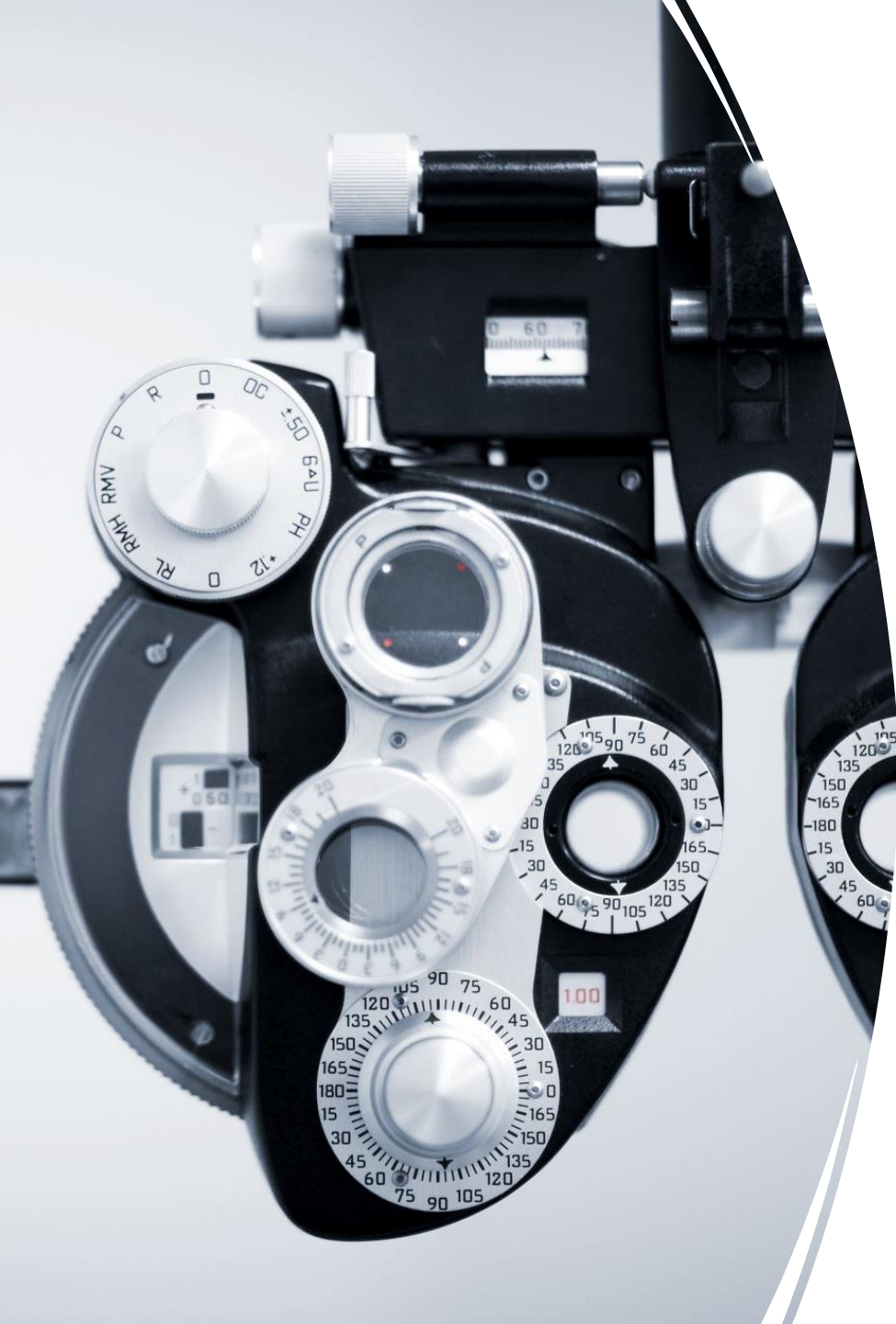

- **Bias and Variance**



Machine Learning

Dr. Jagendra Singh

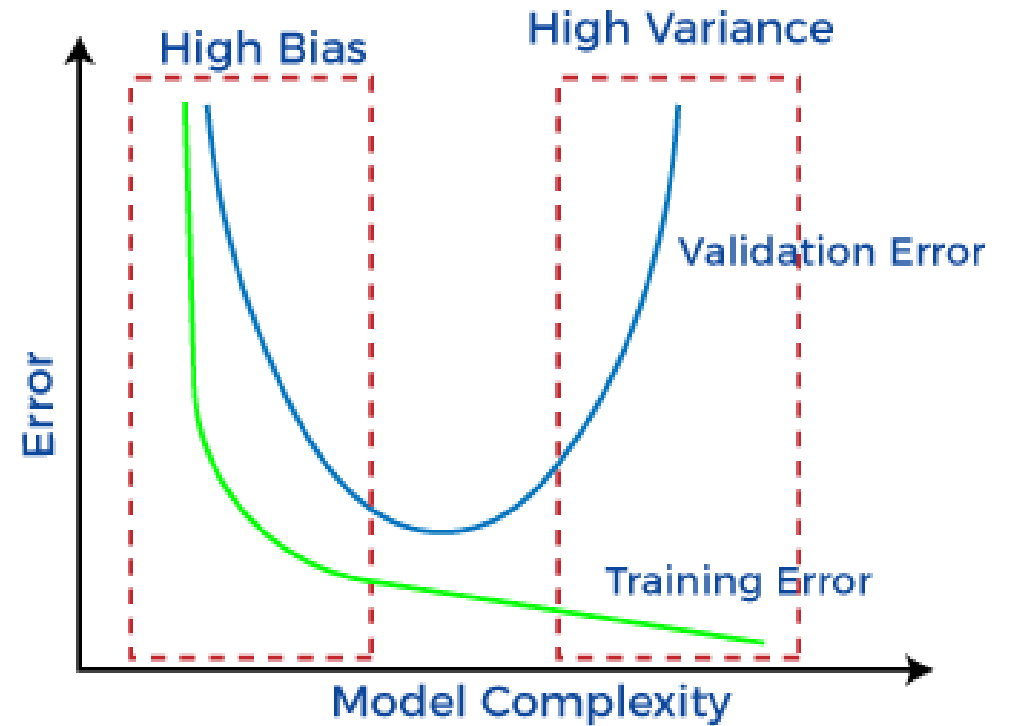


BIAS AND VARIANCE

- Machine learning perform data analysis and make predictions. However, if the machine learning model is not accurate, it can make predictions errors, and these prediction errors are usually known as Bias and Variance.
- In machine learning, these errors will always be present as there is always a slight difference between the model predictions and actual predictions.

BIAS AND VARIANCE

- The main aim of ML/data science analysts is to reduce these errors in order to get more accurate results.
- In this topic, we are going to discuss bias and variance, Bias-variance trade-off, Underfitting and Overfitting. But before starting, let's first understand what errors in Machine learning are?



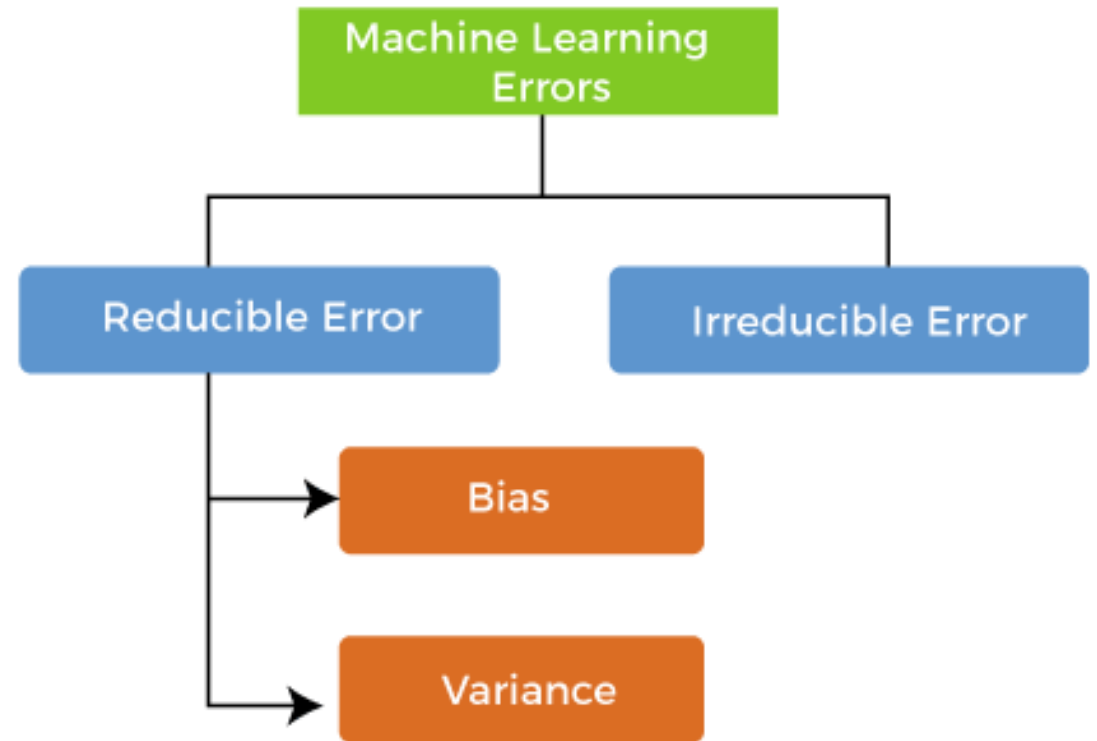
ERRORS IN MACHINE LEARNING

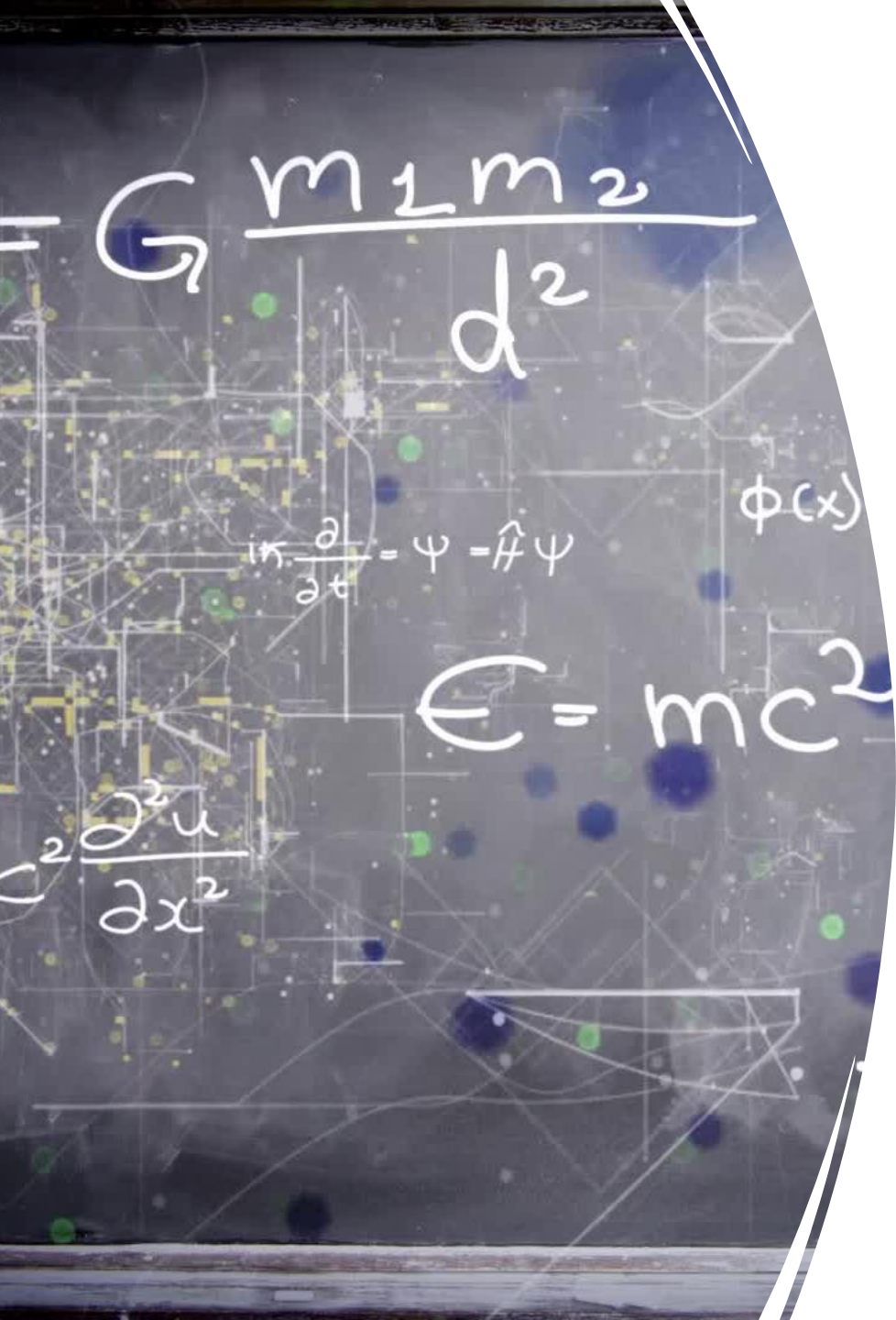
- In machine learning, an error is a measure of how accurately an algorithm can make predictions for the previously unknown dataset.
- On the basis of these errors, the machine learning model is selected that can perform best on the particular dataset. There are mainly two types of errors in machine 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 regardless of which algorithm has been used. The cause of these errors is unknown variables whose value can't be reduced.

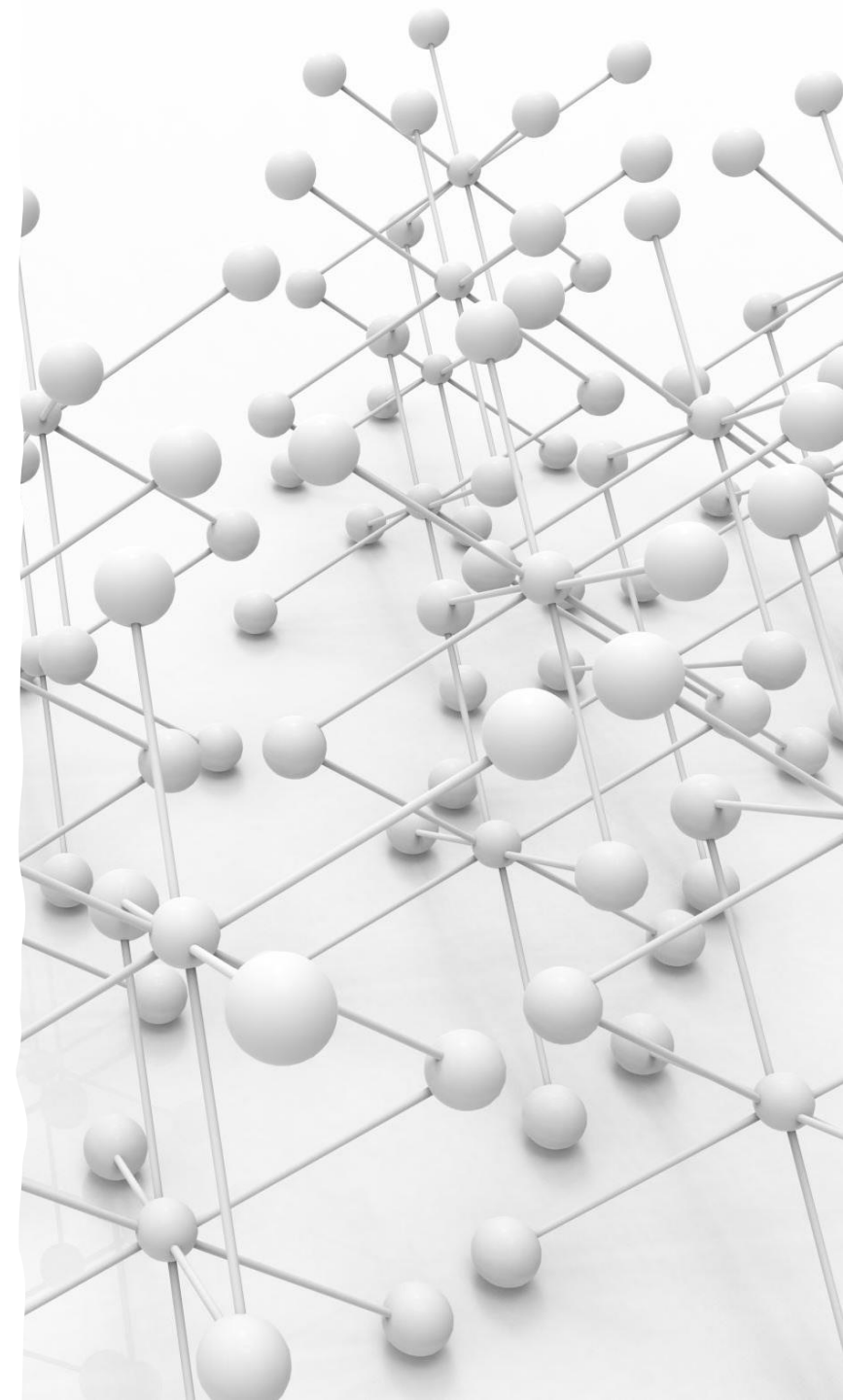


WHAT IS BIAS

- In general, a machine learning model analyses the data, find patterns in it and make predictions. While training, the model learns these patterns in the dataset and applies them to test data for prediction.
- ***While making predictions, a difference occurs between prediction values made by the model and actual values/expected values, and this difference is known as bias errors or Errors due to bias.***

BIAS

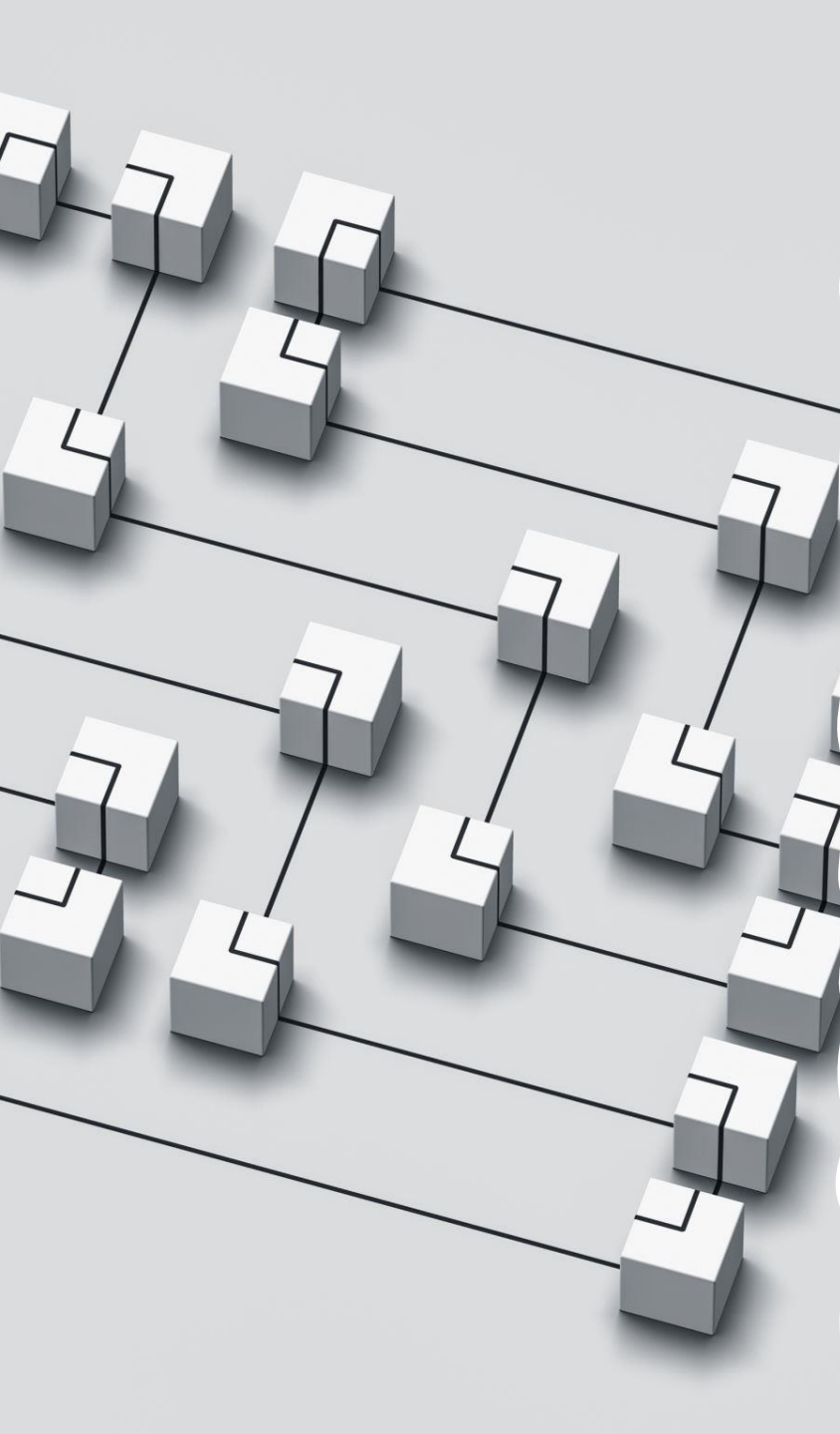
- It can be defined as an inability of machine learning algorithms such as Linear Regression to capture the true relationship between the data points.
- Each algorithm begins with some amount of bias because bias occurs from assumptions in the model, which makes the target function simple to learn. A model has either:





BIAS TYPE

- **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.**



BIAS

- 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



Increase the input features as the model is underfitted.



Decrease the regularization term.



Use more complex models, such as including some polynomial features.

VARIANCE ERROR



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



In simple words, variance tells that how much a random variable is different from its expected value.



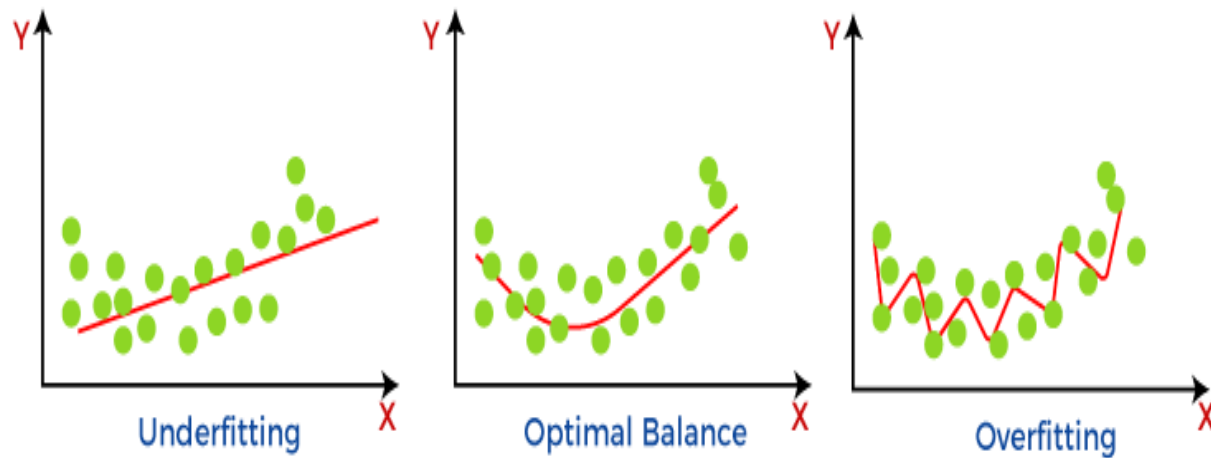
Ideally, a model should not vary too much from one training dataset to another, which means the algorithm should be good in understanding the hidden mapping between inputs and output variables. Variance errors are either of low variance or high variance.

VARIANCE

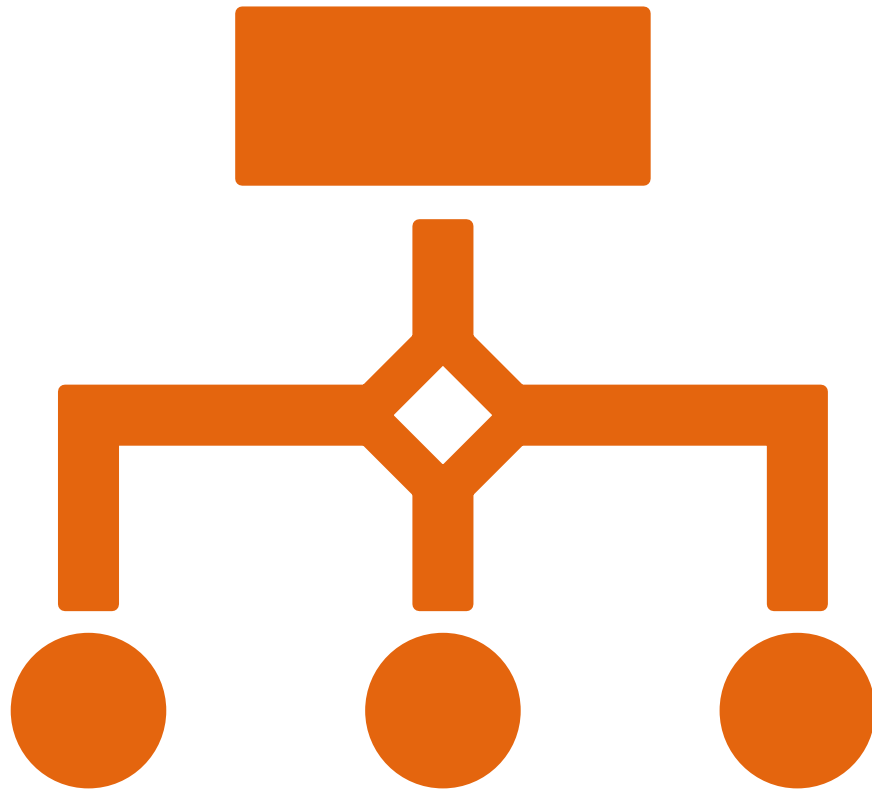
- **Low variance** means there is a small variation in the prediction of the target function with changes in the training data set.
- At the same time, **High variance** shows a large variation in the prediction of the target function with changes in the training dataset.
- Since, with high variance, the model learns too much from the dataset, it leads to overfitting of the model. A model with high variance has the below problems:



VARIANCE



- 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

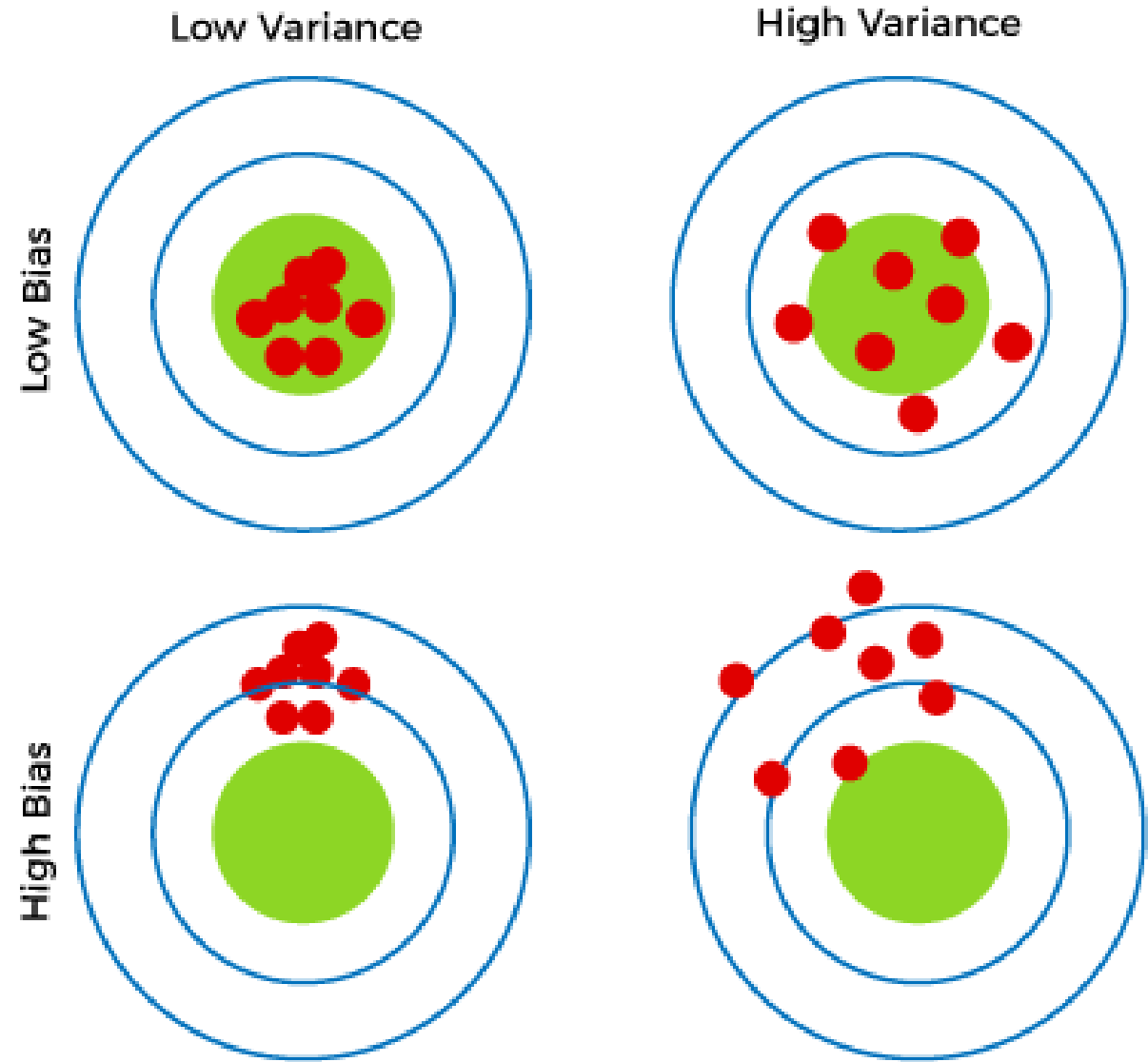
- Reduce the input features or number of parameters as a model is overfitted.
- Do not use a much complex model.
- Increase the training data.
- Increase the Regularization term.



DIFFERENT COMBINATIONS OF BIAS-VARIANCE

- There are four possible combinations of bias and variances, which are represented by the below diagram:
- 1. Low-Bias, Low-Variance:**
The combination of low bias and low variance shows an ideal machine learning model. However, it is not possible practically.
 - 2. Low-Bias, High-Variance:** With low bias and high variance, model predictions are inconsistent and accurate on average.
 - 3. High-Bias, Low-Variance:** With High bias and low variance, predictions are consistent but inaccurate on average.
 - 4. High-Bias, High-Variance:**
With high bias and high variance, predictions are inconsistent and also inaccurate on average.

DIFFERENT COMBINATIONS OF BIAS-VARIANCE





THANK YOU
