SNS COLLEGE OF TECHNOLOGY

Coimbatore-35
An Autonomous Institution

Accredited by NBA – AICTE and Accredited by NAAC – UGC with 'A++' Grade
Approved by AICTE, New Delhi & Affiliated to Anna University, Chennai

DEPARTMENT OF ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

23AMB201 - MACHINE LEARNING

II YEAR IV SEM

UNIT I – INTRODUCTION
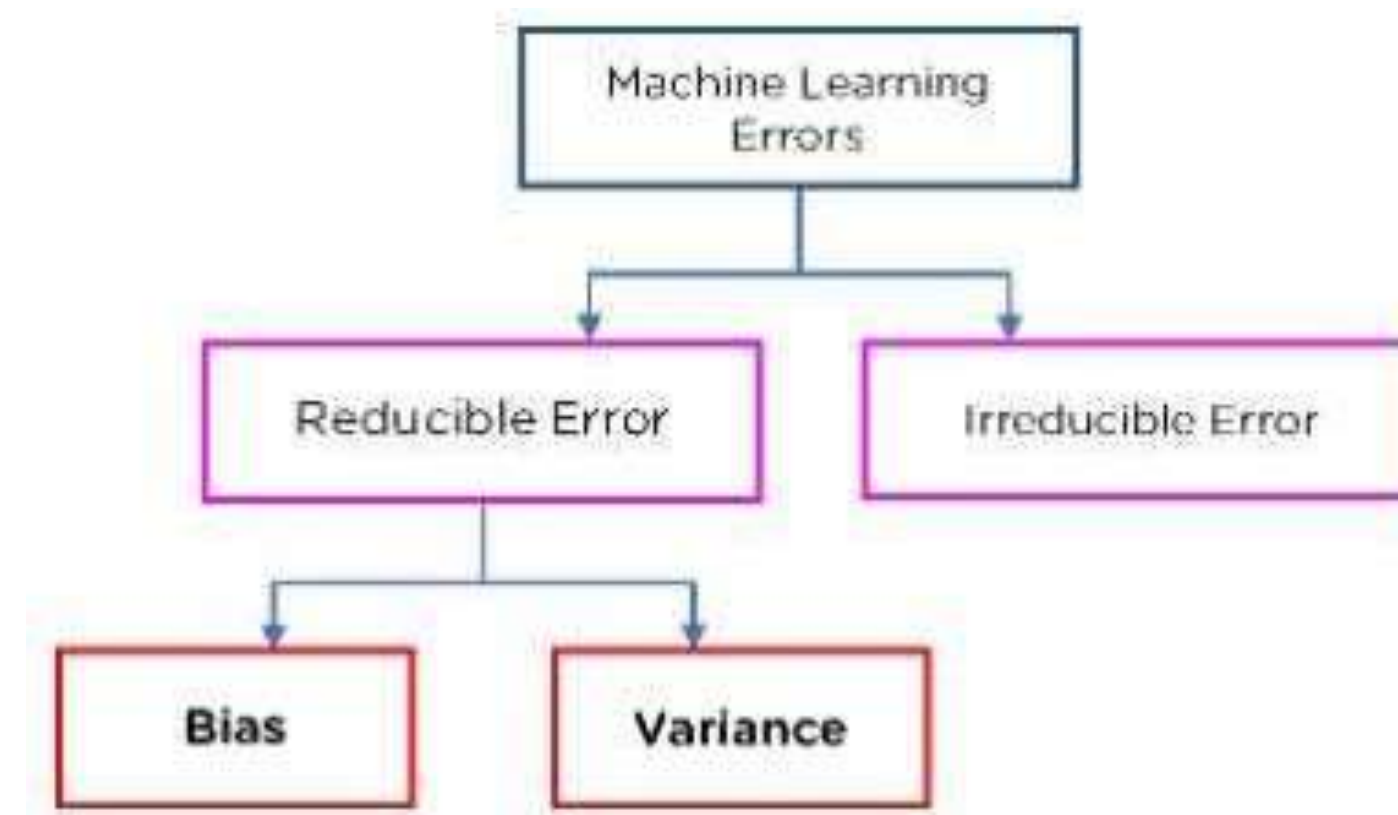
TOPIC 7– Bias, Variance and Tradeoff

Bias

1. Predictive errors (Inaccurate or Wrong)

    1. Bias

    2. Variance

What is Bias?

1. Bias is the difference between our actual and predicted values. Bias is the simple assumptions that our model makes about our data to be able to predict new data.

2. This instance, where the model cannot find patterns in our training set and hence fails for both seen and unseen data, is called Underfitting.

3. High bias is a simple model which pays very little attention to the training data.
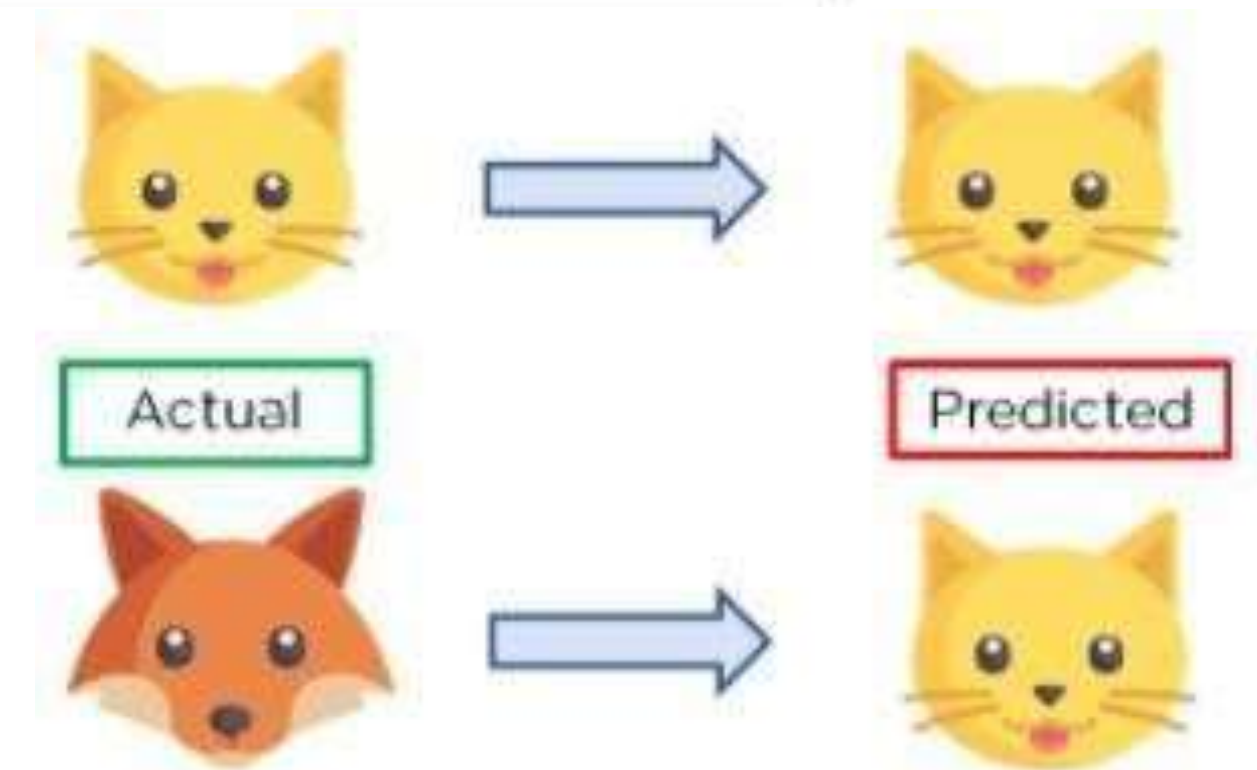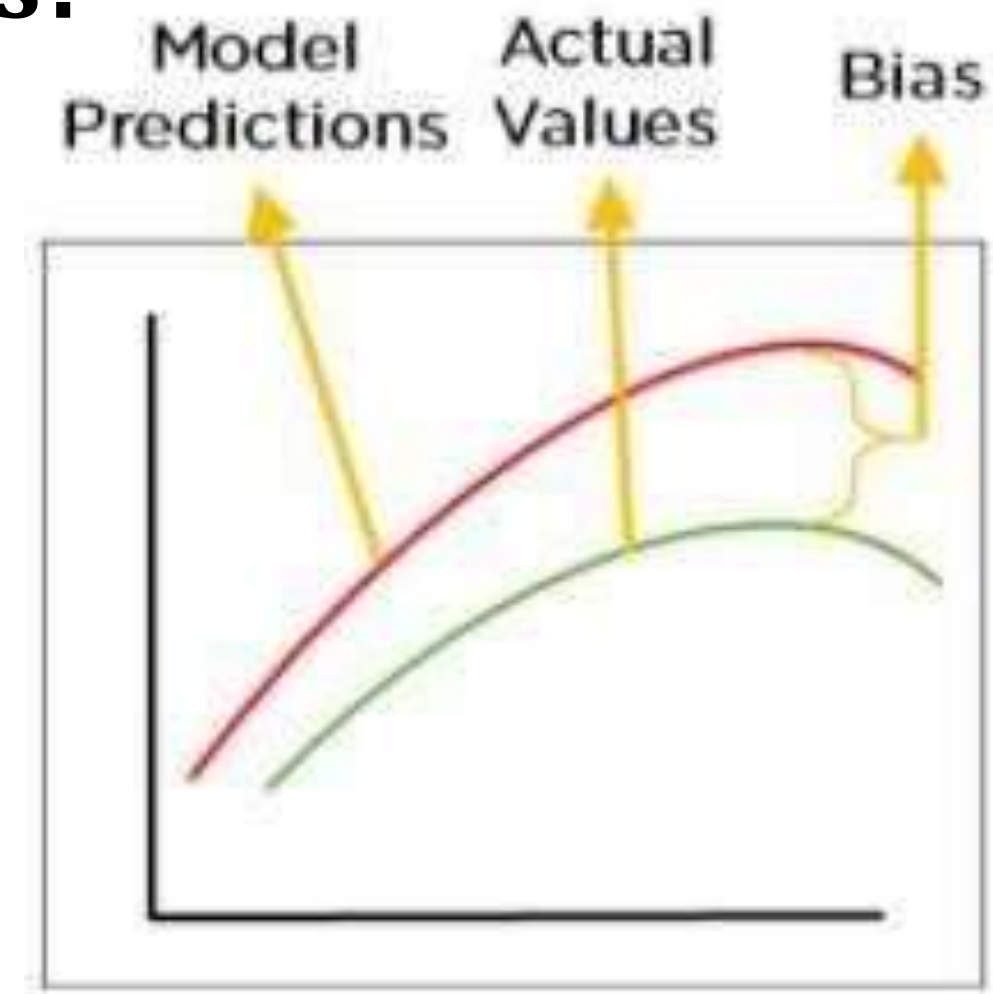
# Characteristics of a high bias model include

1. Not able to capture proper data trends

2. Trained over noise also. So giving less accurate results

3. Suffers from underfitting

4. A more general or simple model

Ms.S.Rajarjeswari/AP/AIML/SNSCT

What is Variance?

1. Variance is the opposite of Bias. Variance is also an error that measures the randomness of the predicted value from the actual value.

2. Variance can be defined as the model's sensitivity to fluctuations in the data. if we model is allowed to view the data too many times, it will learn very well for only that data.

3. It will capture most patterns in the data, but it will also learn from the unnecessary data present, or from the noise.
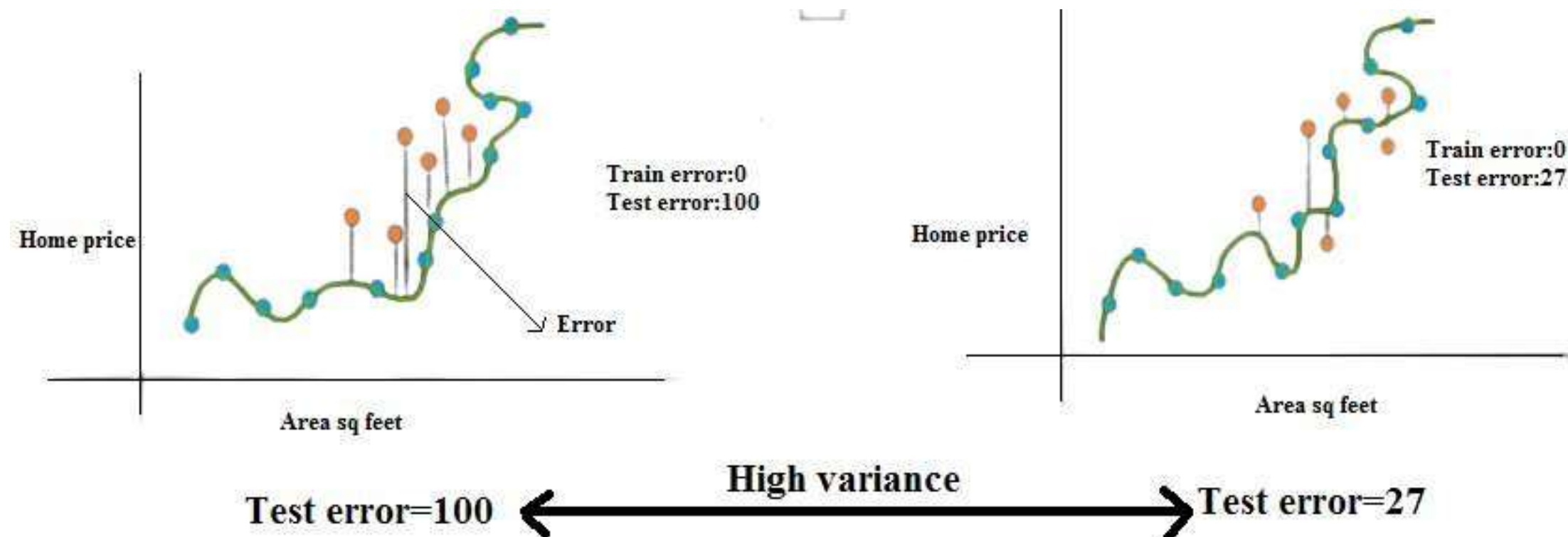
Examples of bias and variance

1. Some machine learning algorithms with low bias **are**

   k-Nearest Neighbours, Decision Trees, and

   Support Vector Machines.

2. At the same time, some machine learning algorithms

   that have high bias are **Linear Regression and**

   Logistic Regression.

Summary

Bias——>Underfitting—->High train and test error

Variance—->Overfitting——>High test error

Ms.S.Rajarjeswari/AP/AIML/SNSCT

Summary

1.Low-Bias, Low-Variance:
The combination is an ideal machine learning model. However, it is not possible practically.

2.Low-Bias, High-Variance: This is a case of **overfitting** where model predictions are inconsistent and accurate on average. The predicted values will be accurate(average) but will be scattered.

3.High-Bias, Low-Variance: This is a case of **underfitting** where predictions are consistent but inaccurate on average. The predicted values will be inaccurate but will be not scattered.

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

Ms.S.Rajarjeswari/AP/AIML/SNSCT
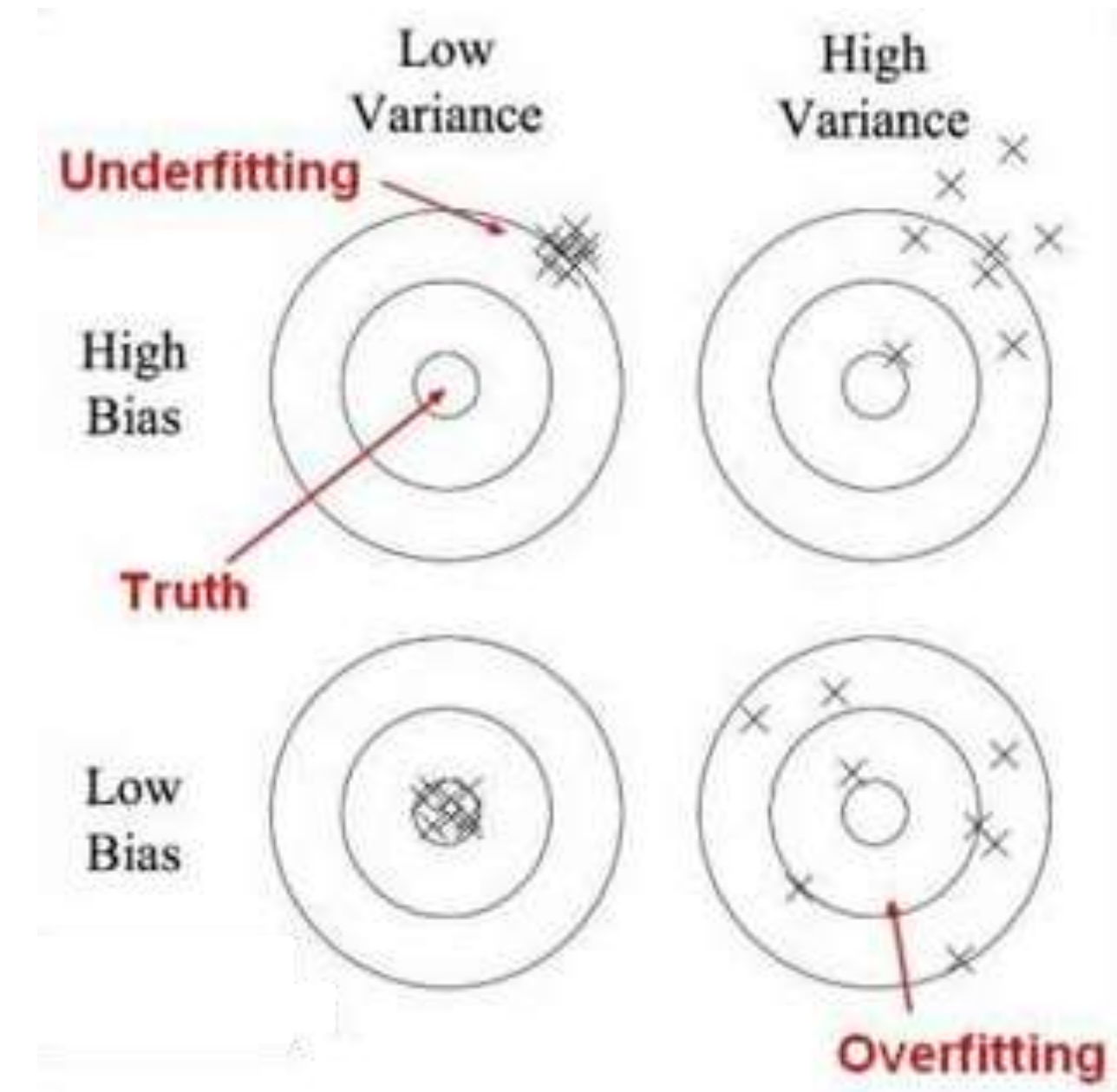
Bias variance Tradeoff

If the model is having fewer parameters, it may have low variance and high bias. Whereas, if the model is complex with a large number of parameters, it will have high variance and low bias.

So, there is a need to strike 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.
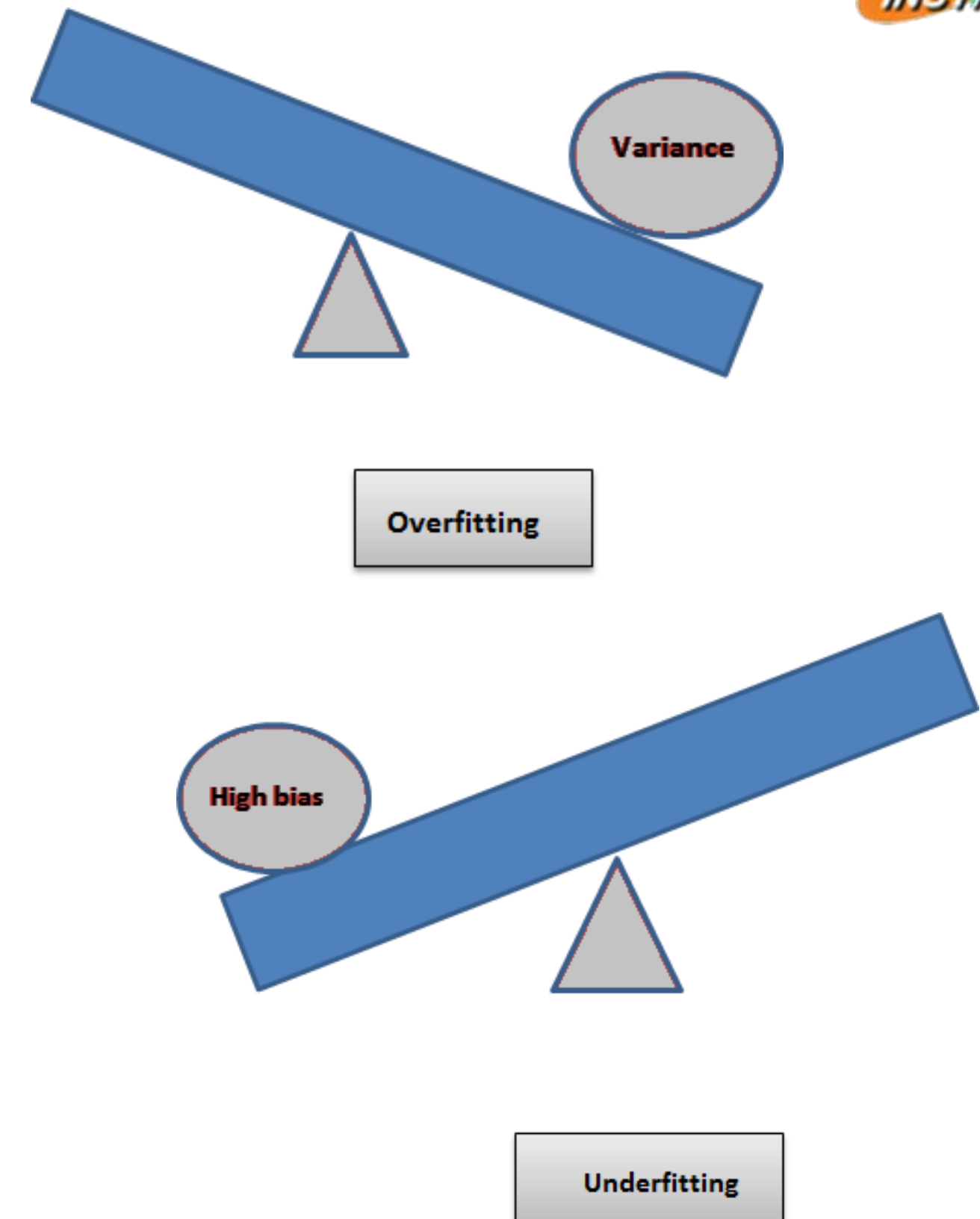
References

1. Aurélien Géron "Hands-On Machine Learning with Scikit-Learn and TensorFlow" Publisher(s): O'Reilly Media, Inc 2017.

2. https://www.geeksforgeeks.org/probability-theory/

3. https://www.analyticsvidhya.com/blog/2021/04/probability-theory-basics-in-machine-learning/

4. https://www.shiksha.com/online-courses/articles/bias-and-variance/