

# THE ART OF ANALYTICS

## Bias

Bias and variance are important parameters that describe prediction errors and accuracy in machine learning.

## Variance

- A machine learning model's performance is based on its correct predictions and how well it is generalized.
- Bias is the difference between the average prediction of our model and the correct target value which model is trying to predict.
- Variance is the amount that the estimate of the target function will change if different training data was used.
- High bias means underfitting of the model on training data, while high variance means overfitting of the model on training data.
- Linear machine learning algorithms often have a high bias but a low variance.
- Nonlinear machine learning algorithms often have a low bias but a high variance.
- Finding the right balance between the bias and variance of the model is called the Bias-Variance trade-off.

- Underfitting/High Bias happens when algorithms are simple, making them fast to learn and easier to understand but generally less flexible to learn complex patterns in data.
- Overfitting happens when algorithms have a high variance and are strongly influenced by the specifics of the training data and try to learn patterns which are noisy and not generalized and only limited to training data set.

# THE ART OF ANALYTICS

Bias refers to the error introduced by the assumptions made in the learning algorithm. It represents the difference between the predicted values by our model and the true values in the underlying dataset. In simpler terms, bias measures how much our model's predictions deviate from the actual values on average.

*Imagine you have a dartboard, and you want to hit the bullseye at the center. However, your aim is slightly off, and you consistently hit the target slightly to the left of the center. In this case, your aim is biased. Similarly, in machine learning, if a model consistently predicts values that are different from the true values, it has a bias problem.*

Variance, measures the variability or spread of the model's predictions. It represents how much the model's predictions vary or fluctuate for different training sets. A high variance indicates that the model is sensitive to the specific training data it was trained on and may not generalize well to unseen data.

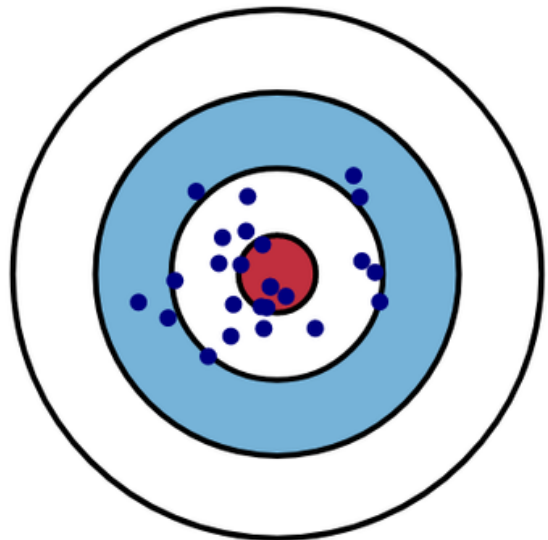
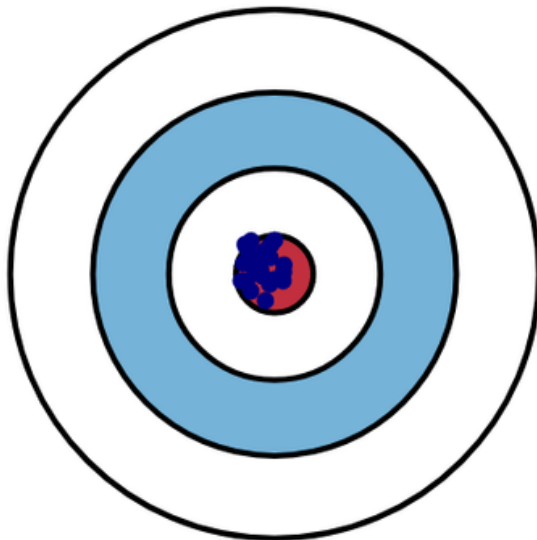
*Continuing with our dartboard example, let's say you have inconsistent aim and your throws sometimes hit the bullseye, but other times they are scattered all over the board. In this case, you have high variance because your throws are highly variable. Similarly, in machine learning, if a model's predictions vary significantly across different training sets, it has a variance problem.*

# THE ART OF ANALYTICS

Low Variance

High Variance

Low Bias



High Bias

