

Bias and Variance

Tushar B. Kute,
<http://tusharkute.com>

Statistical Modeling

- In predictive modeling algorithm learns a model from training data.
- The goal of any such algorithm is to best estimate the mapping function (f) for the output variable (Y) given the input data (X).
- The mapping function is often called the target function because it is the function that a given supervised machine learning algorithm aims to approximate.

Statistical Modeling

- The prediction error for any machine learning algorithm can be broken down into three parts:
 - Bias Error
 - Variance Error
 - Irreducible Error

Irreducible Error

- The irreducible error cannot be reduced regardless of what algorithm is used.
- It is the error introduced from the chosen framing of the problem and may be caused by factors like unknown variables that influence the mapping of the input variables to the output variable.

Bias Error

- Bias are the simplifying assumptions made by a model to make the target function easier to learn.
- Generally, linear algorithms have a high bias making them fast to learn and easier to understand but generally less flexible.
- In turn, they have lower predictive performance on complex problems that fail to meet the simplifying assumptions of the algorithms bias.

Bias Error

- Low Bias: Suggests less assumptions about the form of the target function.
- High-Bias: Suggests more assumptions about the form of the target function.
- Examples of low-bias machine learning algorithms include: Decision Trees, k-Nearest Neighbors and Support Vector Machines.
- Examples of high-bias machine learning algorithms include: Linear Regression, Linear Discriminant Analysis and Logistic Regression.

Variance Error

- Variance is the amount that the estimate of the target function will change if different training data was used.
- The target function is estimated from the training data by a machine learning algorithm, so we should expect the algorithm to have some variance.
- Ideally, it should not change too much from one training dataset to the next, meaning that the algorithm is good at picking out the hidden underlying mapping between the inputs and the output variables.

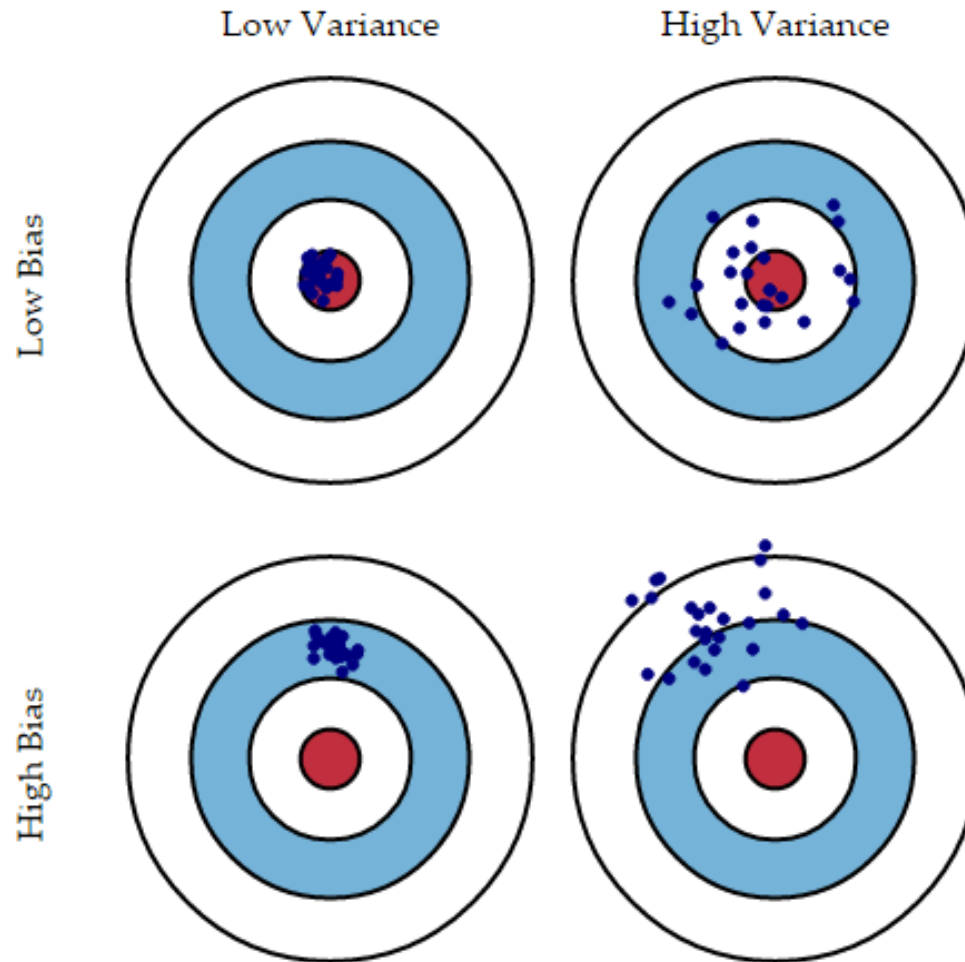
Variance Error

- Machine learning algorithms that have a high variance are strongly influenced by the specifics of the training data.
- This means that the specifics of the training have influences the number and types of parameters used to characterize the mapping function.
 - Low Variance: Suggests small changes to the estimate of the target function with changes to the training dataset.
 - High Variance: Suggests large changes to the estimate of the target function with changes to the training dataset.

Variance Error

- Generally, nonlinear machine learning algorithms that have a lot of flexibility have a high variance. For example, decision trees have a high variance, that is even higher if the trees are not pruned before use.
- Examples of low-variance machine learning algorithms include: Linear Regression, Linear Discriminant Analysis and Logistic Regression.
- Examples of high-variance machine learning algorithms include: Decision Trees, k-Nearest Neighbors and Support Vector Machines.

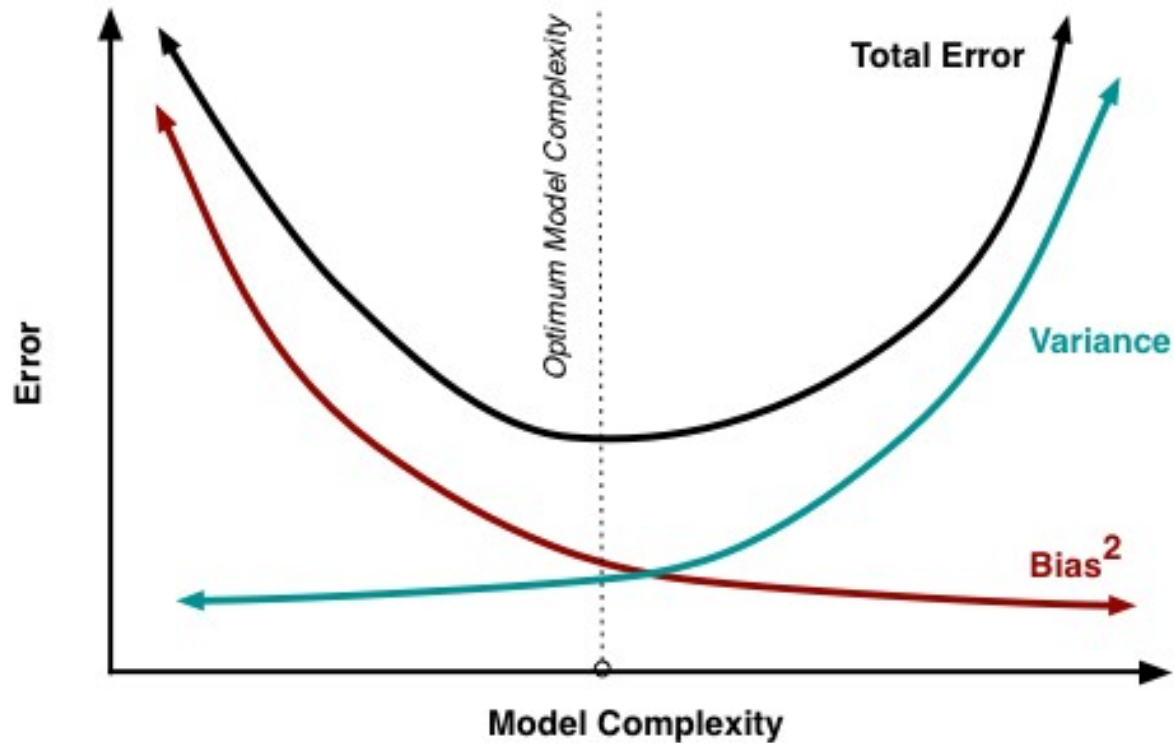
Bias and Variance



Bias-Variance Tradeoff

- The goal of any supervised machine learning algorithm is to achieve low bias and low variance. In turn the algorithm should achieve good prediction performance.
- You can see a general trend in the examples above:
 - 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.
- The parameterization of machine learning algorithms is often a battle to balance out bias and variance.

Bias – Variance Tradeoff



Bias-Variance Tradeoff

- Below are two examples of configuring the bias-variance trade-off for specific algorithms:
 - The k-nearest neighbors algorithm has low bias and high variance, but the trade-off can be changed by increasing the value of k which increases the number of neighbors that contribute to the prediction and in turn increases the bias of the model.
 - The support vector machine algorithm has low bias and high variance, but the trade-off can be changed by increasing the C parameter that influences the number of violations of the margin allowed in the training data which increases the bias but decreases the variance.

Bias-Variance Tradeoff

- There is no escaping the relationship between bias and variance in machine learning.
 - Increasing the bias will decrease the variance.
 - Increasing the variance will decrease the bias.
- There is a trade-off at play between these two concerns and the algorithms you choose and the way you choose to configure them are finding different balances in this trade-off for your problem

Thank you

This presentation is created using LibreOffice Impress 5.1.6.2, can be used freely as per GNU General Public License



@mitu_skillologies



/MITuSkillologies



@mitu_group



/company/mitu-
skillologies



c/MITUSkillologies

Web Resources

<https://mitu.co.in>

<http://tusharkute.com>

contact@mitu.co.in

tushar@tusharkute.com