

22. Data Science – Machine Learning – Bias Variance Trade Off

Contents

1. Bias.....	2
2. Variance	2
3. Remember	2
4. The goal of supervised algorithm	2
5. Error	2
6. Prediction error	2
7. Bias Error	3
8. Low Bias	3
9. High-Bias	3
10. Examples.....	3
11. Variance Error.....	4
12. Low variance:.....	4
13. High variance	4
14. Examples.....	4
15. Bias-Variance Trade-Off	5
16. Configuring the bias-variance trade-off for specific algorithms.....	5
17. The relationship b/w bias and variance	5
18. In reality.....	5
19. Bias and variance.....	6

22. Data Science – Machine Learning – Bias Variance Trade Off

1. Bias

- ✓ Bias is the error or difference between prediction results and actual values.

2. Variance

- ✓ Variance is the error that occurs due to small changes in the training set.

3. Remember

- ✓ The bias and variance provide the information to understand the performance of machine learning algorithms while prediction.

4. The goal of supervised algorithm

- ✓ In supervised learning technique, an algorithm learns a model from training data.
- ✓ The goal of any supervised learning algorithm is,
 - Find the best mapping function/target function (f) for the output variable (Y) for the given input data (X).

5. Error

- ✓ Error means some wrong calculation.
- ✓ Error can be small error or big error.
- ✓ Small error can be acceptable and big error should be minimize

6. Prediction error

- ✓ Prediction error means, we made some prediction but there are some errors are existing.
- ✓ The prediction error for any machine learning algorithm can be divided into two parts
 - Bias Error
 - Variance Error

7. Bias Error

- ✓ Bias are the simplifying assumptions made by a model to make the target function easy to learn.

8. Low Bias

- ✓ Low bias value suggests **more assumptions** about the form of the target function.

9. High-Bias

- ✓ High bias value suggests **fewer assumptions** about the form of the target function.

10. Examples

- ✓ Low-bias machine learning algorithms are,
 - Decision Trees
 - k-Nearest Neighbors
 - Support Vector Machines.
- ✓ High-bias machine learning algorithms are,
 - Linear Regression
 - Linear Discriminant Analysis
 - Logistic Regression.

11. Variance Error

- ✓ Variance is the error that occurs due to small changes in the training set.
- ✓ The target function is estimated from the training data by a machine learning algorithm, we can expect the algorithm may have some variance.
- ✓ Ideally, it should not change too much from one training dataset to the other.

12. Low variance:

- ✓ Low variance value suggests small changes to the estimate of the target function with changes to the training dataset.

13. High variance

- ✓ High variance value suggests large changes to the estimate of the target function with changes to the training dataset.

14. Examples

- ✓ Low-variance machine learning algorithms are,
 - Linear Regression
 - Linear Discriminant Analysis
 - Logistic Regression.
- ✓ High-variance machine learning algorithms are,
 - Decision Trees
 - k-Nearest Neighbors
 - Support Vector Machines.

15. Bias-Variance Trade-Off

- ✓ The goal of any supervised machine learning algorithm is to achieve the sweet spot of **low bias and low variance**.
- ✓ It means, the algorithm should achieve good prediction performance.

Note 1

- ✓ Parametric or linear machine learning algorithms often have a **high bias** but a **low variance**.
- ✓ Nonparametric or nonlinear machine learning algorithms often have a **low bias** but a **high variance**.

16. Configuring the bias-variance trade-off for specific algorithms

- ✓ Below are two examples of configuring the bias-variance trade-off for specific algorithms
 - The **k-nearest neighbour's 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 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.

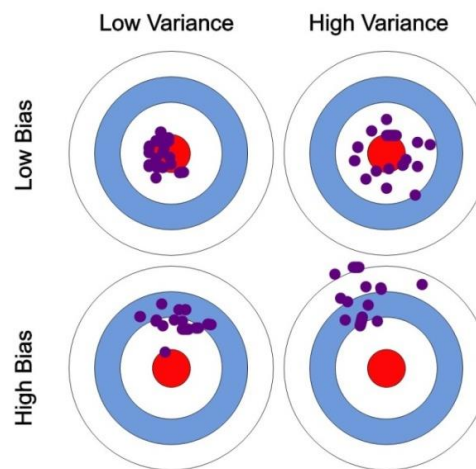
17. The relationship b/w bias and variance

- ✓ Increasing the bias will decrease the variance.
- ✓ Increasing the variance will decrease the bias.

18. In reality

- ✓ In reality we cannot calculate the real bias and variance error terms because we do not know the actual underlying target function.

19. Bias and variance



Error

- ✓ The error is calculated as the difference between predicted and actual value.

Low bias and low variance

- ✓ This is also called as Bias-Variance Tradeoff.
- ✓ If low bias and low variance then the error is very less.
- ✓ So, in this scenario the model is very accurate

Low bias and high variance

- ✓ So, in this scenario the model having lower accuracy.

High bias and low variance

- ✓ So, in this scenario the model having lower accuracy.

High bias and High variance

- ✓ So, in this scenario the model having lower accuracy.

pip install mlxtend

Program Name Calculate bias and variance values
demo1.py

```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from mlxtend.evaluate import bias_variance_decomp

df = pd.read_csv('student_scores.csv')

X = df.iloc[:, :-1].values
y = df.iloc[:, 1].values

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size =
0.2, random_state = 0)

model = LinearRegression()

mse, bias, var = bias_variance_decomp(model, X_train, y_train,
X_test, y_test, loss='mse', num_rounds = 200, random_seed = 1)

print('MSE:', mse)
print('Bias:', bias)
print('Variance:', var)
```

Output

```
MSE: 25.994907724912153
Bias: 22.41211837451609
Variance: 3.5827893503960495
```

- ✓ In this case, we can see that the model has a high bias and a low variance.