26. Data Science – Machine Learning – Underfitting and Overfitting

Contents

26. Data Science – Machine Learning – Underfitting and Overfitting

1. Overfitting & Underfitting

- ✓ The main goal of every machine learning model is to generalize well.
- ✓ A generalized model provides a suitable output on unknown dataset
  - This means after providing training on the dataset, it can produce reliable and accurate output.
- ✓ Overfitting and Underfitting are the two main problems that occur in machine learning, because of these ML model performances may impact to reduce.

2. Noise

- ✓ Noise means irrelevant data; it reduces the performance of the model.

3. Bias

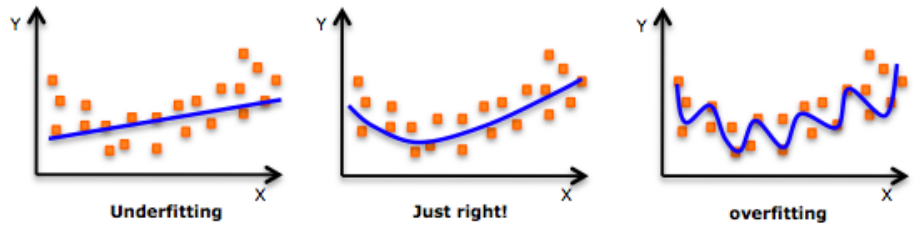
- ✓ Difference between the predicted values and the actual values.

4. Variance

- ✓ Machine learning model performs well with the training dataset, but does not perform well with the test dataset.

5. Overfitting

- ✓ Overfitting is the scenario where a machine learning model cannot generalize or fit well on unseen dataset.
- ✓ The over fitted model has low bias and high variance.



Note:

- ✓ We can avoid the Overfitting in Model by using cross-Validation, Training with more data, removing features, regularization, Ensemble

6. Underfitting

- ✓ During training if model unable to learn properly then this is called as Underfitting.
- ✓ So it reduces the accuracy and produces unreliable predictions.
- ✓ The under fitted model has high bias and low variance.

7. Good fit model

- ✓ If model predicts well on training dataset and unseen dataset then it's called as good fit model

8. Good example

- ✓ Assuming that there are three students have prepared for a mathematics examination.

- ✓ First student:
  - Prepared only Addition operations and skipped other math operations from textbook[**X**]

- ✓ Second student
  - Prepared all math operations from textbook[**X**]

- ✓ Third student
  - Prepared all math operations from textbook[**X**].
  - Practiced more on new topics from other math text books[**Y**, **Z**]

During exam

- ✓ First student:
  - He can answer only for addition related questions.

- ✓ Second student
  - He can answer to the questions which are from only textbook[**X**]

- ✓ Third student
  - He can answer to the questions which are from textbook[**X**, **Y**, **Z**]

Comparison

- ✓ First student - Under fitting
- ✓ Second student - Over fitting
- ✓ Third student - Good fit