**Underfitting in Machine Learning**

A [statistical model](https://www.geeksforgeeks.org/difference-between-statistical-model-and-machine-learning/) or a machine learning algorithm is said to have underfitting when a model is too simple to capture data complexities. It represents the inability of the model to learn the training data effectively result in poor performance both on the training and testing data. In simple terms, an underfit model’s are inaccurate, especially when applied to new, unseen examples. It mainly happens when we uses very simple model with overly simplified assumptions. To address underfitting problem of the model, we need to use more complex models, with enhanced feature representation, and less regularization.

**Note: The underfitting model has High bias and low variance.**

**Reasons for Underfitting**

1. The model is too simple, So it may be not capable to represent the complexities in the data.
2. The input features which is used to train the model is not the adequate representations of underlying factors influencing the target variable.
3. The size of the training dataset used is not enough.
4. Excessive regularization are used to prevent the overfitting, which constraint the model to capture the data well.
5. Features are not scaled.

**Techniques to Reduce Underfitting**

1. Increase model complexity.
2. Increase the number of features, performing [feature engineering](https://www.geeksforgeeks.org/what-is-feature-engineering/).
3. Remove noise from the data.
4. Increase the number of [epochs](https://www.geeksforgeeks.org/epoch-in-machine-learning/) or increase the duration of training to get better results.

**Overfitting in Machine Learning**

A [statistical model](https://www.geeksforgeeks.org/difference-between-statistical-model-and-machine-learning/) is said to be overfitted when the model does not make accurate predictions on testing data. When a model gets trained with so much data, it starts learning from the noise and inaccurate data entries in our data set. And when testing with test data results in High variance. Then the model does not categorize the data correctly, because of too many details and noise. The causes of overfitting are the non-parametric and non-linear methods because these types of machine learning algorithms have more freedom in building the model based on the dataset and therefore they can really build unrealistic models. A solution to avoid overfitting is using a linear algorithm if we have linear data or using the parameters like the maximal depth if we are using decision trees.

In a nutshell, [Overfitting](https://www.geeksforgeeks.org/underfitting-and-overfitting-in-machine-learning/) is a problem where the evaluation of machine learning algorithms on training data is different from unseen data.

**Reasons for Overfitting:**

1. High variance and low bias.
2. The model is too complex.
3. The size of the training data.

**Techniques to Reduce Overfitting**

1. Improving the quality of training data reduces overfitting by focusing on meaningful patterns, mitigate the risk of fitting the noise or irrelevant features.
2. Increase the training data can improve the model’s ability to generalize to unseen data and reduce the likelihood of overfitting.
3. Reduce model complexity.
4. [Early stopping](https://www.geeksforgeeks.org/regularization-by-early-stopping/) during the training phase (have an eye over the loss over the training period as soon as loss begins to increase stop training).
5. [Ridge Regularization](https://www.geeksforgeeks.org/lasso-vs-ridge-vs-elastic-net-ml/) and [Lasso Regularization](https://www.geeksforgeeks.org/implementation-of-lasso-regression-from-scratch-using-python/).
6. Use [dropout](https://www.geeksforgeeks.org/dropout-in-neural-networks/) for [neural networks](https://www.geeksforgeeks.org/neural-networks-a-beginners-guide/) to tackle overfitting.



*Underfitting and Overfitting*

**Good Fit in a Statistical Model**

Ideally, the case when the model makes the predictions with 0 error, is said to have a good fit on the data. This situation is achievable at a spot between overfitting and underfitting. In order to understand it, we will have to look at the performance of our model with the passage of time, while it is learning from the training dataset.

With the passage of time, our model will keep on learning, and thus the error for the model on the training and testing data will keep on decreasing. If it will learn for too long, the model will become more prone to overfitting due to the presence of noise and less useful details. Hence the performance of our model will decrease. In order to get a good fit, we will stop at a point just before where the error starts increasing. At this point, the model is said to have good skills in training datasets as well as our unseen testing dataset.

Top of Form

Bottom of Form

* [Online Degrees](https://www.coursera.org/degrees)
* [Find your New Career](https://www.coursera.org/career-academy?trk_ref=globalnav)
* [Log In](https://www.coursera.org/articles/overfitting-vs-underfitting?authMode=login&msockid=340670e60b6e6b8b0fb262d40ac86a51)
* [**Join for Free**](https://www.coursera.org/articles/overfitting-vs-underfitting?authMode=signup&msockid=340670e60b6e6b8b0fb262d40ac86a51)

1. [Data](https://www.coursera.org/articles/category/data)
2. [AI and Machine Learning](https://www.coursera.org/articles/category/ai-and-machine-learning)
3. Overfitting vs. Underfitting: What’s the Difference?

**Indicators of overfitting and underfitting: Bias and variance**

Being aware of bias and variance can help you assess the reliability of a machine learning model. Here’s what they mean:

Bias represents how far off, on average, the model's predictions are from the real outcomes. A high bias suggests that the model may be too simplistic, missing out on essential patterns in the data.

Variance, on the other hand, pertains to the fluctuations in a model's behavior when tested on different sections of the training data set. A high variance model can accommodate diverse data sets but can result in very dissimilar models for each instance. Complex models can exhibit high variance.

**A more succinct comparison of bias and variance:**

* High-bias models oversimplify data.
* High variance models over-adapt to data.

High bias and low variance signify underfitting, while low bias and high variance indicate overfitting. Note that bias and variance exhibit an inverse correlation. As you continue training a model, bias decreases while variance grows, so you are attempting to balance bias and variance somewhat. Still, your ML model may function properly even with a higher variance.

Now that you understand the bias-variance trade-off, let's explore the steps to adjust an ML model so that it is neither overfitted nor underfitted.