

XGBoost

 [_ \(https://github.com/learn-co-curriculum/dsc-xgboost\)](https://github.com/learn-co-curriculum/dsc-xgboost)  [_ \(https://github.com/learn-co-curriculum/dsc-xgboost/issues/new/choose\)](https://github.com/learn-co-curriculum/dsc-xgboost/issues/new/choose)

Introduction

Now that you are familiar with gradient boosting, you'll learn about the top gradient boosting algorithm currently in use -- XGBoost!

Objectives

You will be able to:

- Compare XGBoost to other boosting algorithms

What is XGBoost?

Gradient boosting is one of the most powerful concepts in machine learning right now. As you've seen, the term *gradient boosting* refers to a class of algorithms, rather than any single one. The version with the highest performance right now is **XGBoost**, which is short for **eXtreme Gradient Boosting**.



XGBoost is a stand-alone library that implements popular gradient boosting algorithms in the fastest, most performant way possible. There are many under-the-hood optimizations that allow XGBoost to train more quickly than any other library implementations of gradient boosting algorithms. For instance, XGBoost is configured in such a way that it parallelizes the construction of trees across all your computer's CPU cores during the training phase. It also allows for more advanced use cases, such as distributing training across a cluster of computers, which is often a technique used to speed up computation. The algorithm even automatically handles missing values!

Installing **XGBoost**

XGBoost is an independent library that provides implementations in C++, Python, and other languages. Luckily, the open-source community has had the good sense to make the Python API for **XGBoost** mirror that of **sklearn**, so using **XGBoost** feels no different than using any other supervised learning algorithm from **sklearn**. The only downside is that it does not come packaged with **sklearn**, so we must install it ourselves. **conda** makes this quite easy.

All you need to do is run the command `conda install py-xgboost` in your terminal, and **conda** will take care of the rest!

Use cases

XGBoost has risen to prominence by being the go-to algorithm for winning competitions on [Kaggle](https://www.kaggle.com/) , a competitive data science platform. It is so common to see XGBoost cited as an algorithm used by the winners of Kaggle competitions that it has become a bit of a running joke in the community. [This page](https://github.com/dmlc/xgboost/tree/master/demo#machine-learning-challenge-winning-solutions)  [\(https://github.com/dmlc/xgboost/tree/master/demo#machine-learning-challenge-winning-solutions\)](https://github.com/dmlc/xgboost/tree/master/demo#machine-learning-challenge-winning-solutions) contains an (incomplete) list of all the recent competitions with place winners that used XGBoost for their solution!

XGBoost is a great choice for classification tasks. It provides best-in-class performance compared to other classification algorithms (with the exception of Deep Learning, which we'll learn more about soon).

Takeaways

When approaching a supervised learning problem, you should always use multiple algorithms, and compare the performances of the various models. There will always be use cases where some classes of models tend to outperform others. However, there are some models that generally outperform all the others -- XGBoost is at the top of this list! Make sure that this is an algorithm you're familiar with, as there are many situations where you'll find it quite useful!

You can find the full documentation for XGBoost [here](https://xgboost.readthedocs.io/en/latest/)  [\(https://xgboost.readthedocs.io/en/latest/\)](https://xgboost.readthedocs.io/en/latest/).

Summary

In this lesson, we learned about what XGBoost is, and why it is so powerful and useful to Data Scientists!

How do you feel about this lesson?



Have specific feedback?

[Tell us here!](https://github.com/learn-co-curriculum/dsc-xgboost/issues/new/choose) [\(https://github.com/learn-co-curriculum/dsc-xgboost/issues/new/choose\)](https://github.com/learn-co-curriculum/dsc-xgboost/issues/new/choose)