Machine Learning Engineering Career Track
Top Performing Machine Learning Tools At A Glance

Top Machine Learning Libraries

This guide is a one-stop-shop where you can find the documentation for many of the most popular, high-performance machine learning libraries you'll be using in your daily work. No tool is going to work in every situation: some of these libraries are screwdrivers, others are drills, hammers, or wrenches. You'll need to learn how to recognize which problem calls for which tool. Several of the libraries below do accomplish similar things, by playing around with the various offerings you'll be able to find the right library to suit your needs.

We've broken down the libraries by the services they provide like NLP tools or Computer Vision. We'll give you a link to the documentation and a description of each library. Take a look at the documentation on your own time, and return to this guide whenever you are choosing between libraries.

General Purpose Libraries

Scikit-learn

Scikit-learn is the closest thing you'll find to a one-stop-shop for all your machine learning needs. It is a library with tools for classification, regression, clustering, SVMs, random forests, gradient boosting, KNN, and it was designed to use alongside NumPy. As you can see from this list of algorithms, scikit-learn is constantly in a state of evolution, keeping up with trends in ML as they emerge. It is a great starting point for an engineer and essential knowledge for all industry professionals.

Gradient Boosting Libraries

XGBoost

Extreme Gradient Boost is considered the most state of the art, highest-performing gradient boost model on the market. The cream of Kaggle, XGBoost is commonly used in winning solutions on the competition platform. It is worth some time familiarizing yourself with this tool.

lightGBM

There are few boosting algorithms that can rival XGBoost, however, lightGBM is a powerful competition. This model can handle categorical data and boasts faster training and tuning times than XGBoost. There will always be situations where lightGBM is a better tool than XGBoost, so dive into the documentation.

CatBoost

Another strong contender for the gradient boosting throne is CatBoost. Developed by Yandex, one of the most innovative tech companies in Russia, CatBoost has incredible features like native GPU support and lightning-fast training time.

Deep Learning Libraries

PyTorch
PyTorch is a bright beacon of deep learning prowess, loved and lauded by machine learning practitioners worldwide. The framework allows you to create new styles of deep neural networks but also allows you to use out of the box solutions. If you ever hope to implement research papers yourself, you'll need to be familiar with the PyTorch documentation.

TensorFlow
TensorFlow is Google's gift to the world. It is perhaps the most sophisticated deep learning library in the machine learning ecosystem, and many of the world's most cutting edge machine learning applications are built with TensorFlow. We've used it extensively for our course already, but explore more on your own time.

Vowpal Wabbit
Vowpal Wabbit offers a wide range of high-speed algorithms. It is especially useful for online machine learning problems where data is fed in continuously. This library is easy to miss, but will undoubtedly contain tools you'll want to use over the course of your career.

Computer Vision Libraries

OpenCV
The most popular Computer Vision library is OpenCV. This library is an open-source masterpiece which is one of the most high tech free resources available in all of machine learning. If Computer Vision interests you one iota, then you must get used to OpenCV.

fast.ai
Jeremy Howard is one of the towering figures in machine learning. His company fast.ai is aimed at democratizing machine learning, and their fast.ai library has some of the best deep learning tools for NLP, Computer Vision, and more. User testimonial from one student says that they were able to build models with a 91% accuracy for image classification their first-day using fast.ai. What will you do with fast.ai today?

NLP Libraries

spaCy

spaCy is a great starting point for NLP. The production-oriented library is able to handle reams of data and return fantastic results. You can use spaCy both as an the primary processing library, or as part of your pre-processing pipeline!

Flair

Flair is an easy to use framework for multilingual NLP that is built on top of PyTorch. You can use it for tasks like entity and speech recognition, and it has extremely high performance. If NLP interests you, Flair is certainly worth a look.

gensim

gensim is a metalibrary which combines tons of small NLP libraries together into a cohesive and easy to use package. It is a mix of classic and fresh NLP algorithms. Go take a look!

fastText

If there's anyone who knows about analyzing big data amounts of texts in an intelligent way, it's Facebook. And fastText is Facebook's publicly available NLP library. It's perfect if you have a small amount of computing power, like a PC, but a large text dataset. If you want to get an intimate understanding of NLP, there is no better place to start than fastText.

Stanford Core NLP

Stanford is almost synonymous with machine learning excellence. After all, it's where Andrew Ng based himself. The Stanford Core NLP library has more valuable tools than we can count. Every aspiring NLP practitioner should get used to these juicy tools.