TensorFlow

Theano

Torch

scikit-learn

Caffe

Choosing an Open Source Machine Learning Library:

TensorFlow, Theano, Torch, scikit-learn, Caffe

TensorFlow

Theano

Torch

scikit-learn

Caffe

- From healthcare and security to marketing personalization, despite being at the early stages of development, machine learning has been changing the way we use technology to solve business challenges and everyday tasks. This potential has prompted companies to start looking at machine learning as a relevant opportunity rather than a distant, unattainable virtue.

Why Open Source Machine Learning?

TensorFlow

Theano

Torch

scikit-learn

Caffe

- The term open source software refers to a tool with a source code available via the Internet for free. Proprietary (closed source) software code is private and distributed via licensed rights. For a business that's just starting its ML initiative, using open source tools can be a great way to practice data science gratis before deciding on enterprise level tools like Microsoft Azure or Amazon Machine Learning.

- The benefits of using open source tools don't stop at their availability. Generally, such projects have a vast community of data engineers and data scientists eager to share datasets and pre-trained models. For instance, instead of building image recognition from scratch, you can use classification models trained on the data from ImageNet, or build your own using this dataset. Open source ML tools also let you leverage transfer learning, meaning solving machine learning problems by applying knowledge gained after working with a problem from a related or even distant domain. So, you can transfer some capacities form the model that has learned to recognize cars to the model aimed at trucks recognition.

TensorFlow

Theano

Torch

scikit-learn

Caffe

- Depending on the task you're working with, pre-trained models and open datasets may not be as accurate as custom ones, but they will save a substantial amount of effort and time, and they don't require you to gather datasets. According to Andrew Ng, former chief scientist at Baidu and professor at Stanford, the concept of reusing open source models and datasets will be the second biggest driver of the commercial ML success after supervised learning.
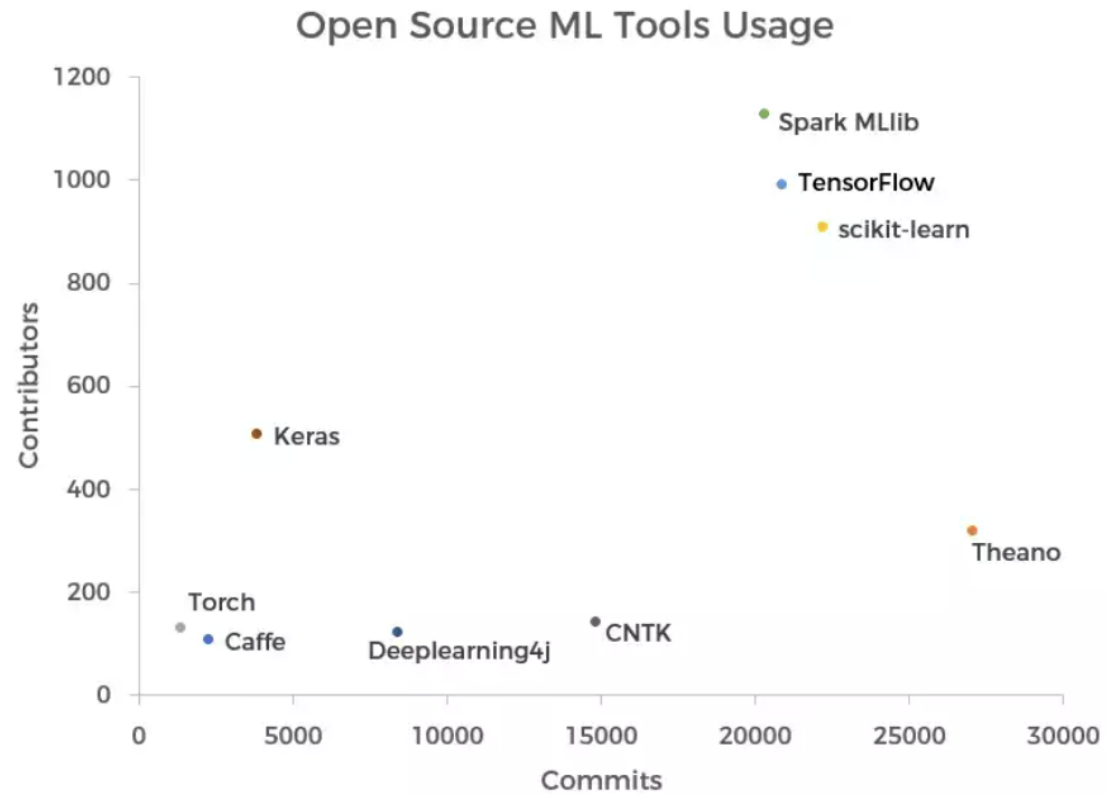
TensorFlow

Theano

Torch

scikit-learn

Caffe

Open Source ML Tools Usage

Spark MLlib
TensorFlow
scikit-learn
Keras
Theano
Torch
Caffe
Deeplearning4j
CNTK

Comparing GitHub commits and contributors for different open source tools

TensorFlow

Theano

Torch

scikit-learn

Caffe

TensorFlow: Profound and Favored Tool from Google

- Originally built by Google for internal use, TensorFlow was released under an Apache 2.0 open source license in 2015. The library is still used by the corporation for a number of services, such as speech recognition, Photo Search, and automatic responses for Gmail's Inbox. Google's reputation and useful flowgraphs to construct models have attracted a massive number of contributors to TensorFlow. This resulted in public access to exhaustive documentation and tutorials allowing for an easy entrance point into the world of neural networks applications.

- TensorFlow is a great Python tool for both deep neural networks research and complex mathematical computations, and it can even support reinforcement learning. The uniqueness of TensorFlow also lies in dataflow graphs – structures that consist of nodes (mathematical operations) and edges (numerical arrays or tensors).

TensorFlow

Theano

Torch

scikit-learn

Caffe

TensorFlow: Datasets and models

The flexibility of TensorFlow is based on the possibility of using it both for research and recurring machine learning tasks. Thus, you can use the low level API called TensorFlow Core. It allows you to have full control over models and train them using your own dataset. But there are also public and official pre-trained models to build higher level APIs on top of TensorFlow Core. Some of the popular models you can apply are MNIST, a traditional dataset helping identify handwritten digits on an image, or Medicare Data, a dataset by Google used to predict charges for medical services among others.

TensorFlow

Theano

Torch

scikit-learn

Caffe

TensorFlow: Audience and learning

- For someone exploring machine learning for the first time, TensorFlow's variety of functions may be a bit of a struggle. Some even argue that the library doesn't try to accelerate a machine learning curve, instead making it even steeper. TensorFlow is a low-level library that requires ample code writing and a good understanding of data science specifics to start successfully working with the product. Consequently, it may not be your first choice if your data science team is IT-centric and there: There are simpler alternatives we'll be discussing.

TensorFlow

Theano

Torch

scikit-learn

Caffe

TensorFlow: curve Use cases

- Considering its complexity, the use cases for TensorFlow mostly include solutions by large companies with access to machine learning specialists. For example, the British online supermarket Ocado applied TensorFlow to prioritize emails coming to their contact center and improve demand forecasting. Also, the global insurance company Axa used the library to predict large-loss car incidents involving their clients.

TensorFlow

Theano

Torch

scikit-learn

Caffe

Theano: Mature Library with Extended Possibilities

- Theano is a low-level library for scientific computing based on Python, which is used to target deep learning tasks related to defining, optimizing, and evaluating mathematical expressions. While it has an impressive computing performance, users complain about an inaccessible interface and unhelpful error messages. For these reasons, Theano is mainly applied in combination with more user-friendly wrappers, such as Keras, Lasagne, and Blocks – three high-level frameworks aimed at fast prototyping and model testing.

Theano: Datasets and models

- There are public models for Theano, but each framework used on top also has plenty of tutorials and pre-trained datasets to choose from. Keras, for instance, stores available models and detailed usage tutorials in its documentation.

TensorFlow

Theano

Torch

scikit-learn

Caffe

TensorFlow

Theano

Torch

scikit-learn

Caffe

Theano: Audience and learning curve

- If you use Lasagne or Keras as high-level wrappers on top of Theano, again you'll have a multitude of tutorials and pre-trained datasets at your fingertips. Moreover, Keras is considered one of the easiest libraries to start with at early stages of deep learning exploration.

- Since TensorFlow was designed to replace Theano, a big part of its fanbase left. But there are still a lot of advantages that many data scientists find compelling enough keep them with an outdated version. Theano's simplicity and maturation are serious points to consider when making this choice.

TensorFlow

Theano

Torch

scikit-learn

Caffe

Theano: Use cases

- Considered an industry standard for deep learning research and development, Theano was originally designed to implement state-of-the-art deep learning algorithms. However, considering that you probably won't use Theano directly, its numerous uses expand as you use it as foundation for other libraries: digit and image recognition, object localization, and even chatbots.

TensorFlow

Theano

Torch

scikit-learn

Caffe

Torch: Facebook-Backed Framework Powered by Lua Scripting Language

- Torch is often called the easiest deep learning tool for beginners. It has a simple scripting language, Lua, and a helpful community sharing an impressive array of tutorials and packages for almost any deep learning purpose. Despite using a less common language than Python, it's widely adopted – Facebook, Google, and Twitter are known for using it in their AI projects.

TensorFlow

Theano

Torch

scikit-learn

Caffe

Torch: Datasets and models

- You can find a list of popular datasets to be loaded for use in Torch on its GitHub cheatsheet page. Moreover, Facebook released an official code for Deep Residual Networks (ResNets) implementation with pre-trained models with instructions for fine-tuning your own datasets.

TensorFlow

Theano

Torch

scikit-learn

Caffe

Torch: Audience and learning curve

- Regardless of the differences and similarities, the choice will always come down to the language. The market population of experienced Lua engineers will always be smaller than that of Python. However, Lua is significantly easier to read, which is reflected in the simple syntax of Torch. The active Torch contributors swear by Lua, so it's a framework of choice for both novices and those wishing to expand their toolset.

TensorFlow

Theano

Torch

scikit-learn

Caffe

scikit-learn: Accessible and Robust Framework from the Python Ecosystem

- In November 2016, scikit-learn became a number one open source machine learning project for Python, according to KDNuggets.

- scikit-learn is a high level framework designed for supervised and unsupervised machine learning algorithms. Being one of the components of the Python scientific ecosystem, it's built on top of NumPy and SciPy libraries, each responsible for lower-level data science tasks. While NumPy sits on Python and deals with numerical computing, the SciPy library covers more specific numerical routines such as optimization and interpolation. Subsequently, scikit-learn was built precisely for machine learning. The relationship between the three along with other tools in the Python ecosystem reflects different levels in the data science field: The higher you go, the more specific the problems you can solve.
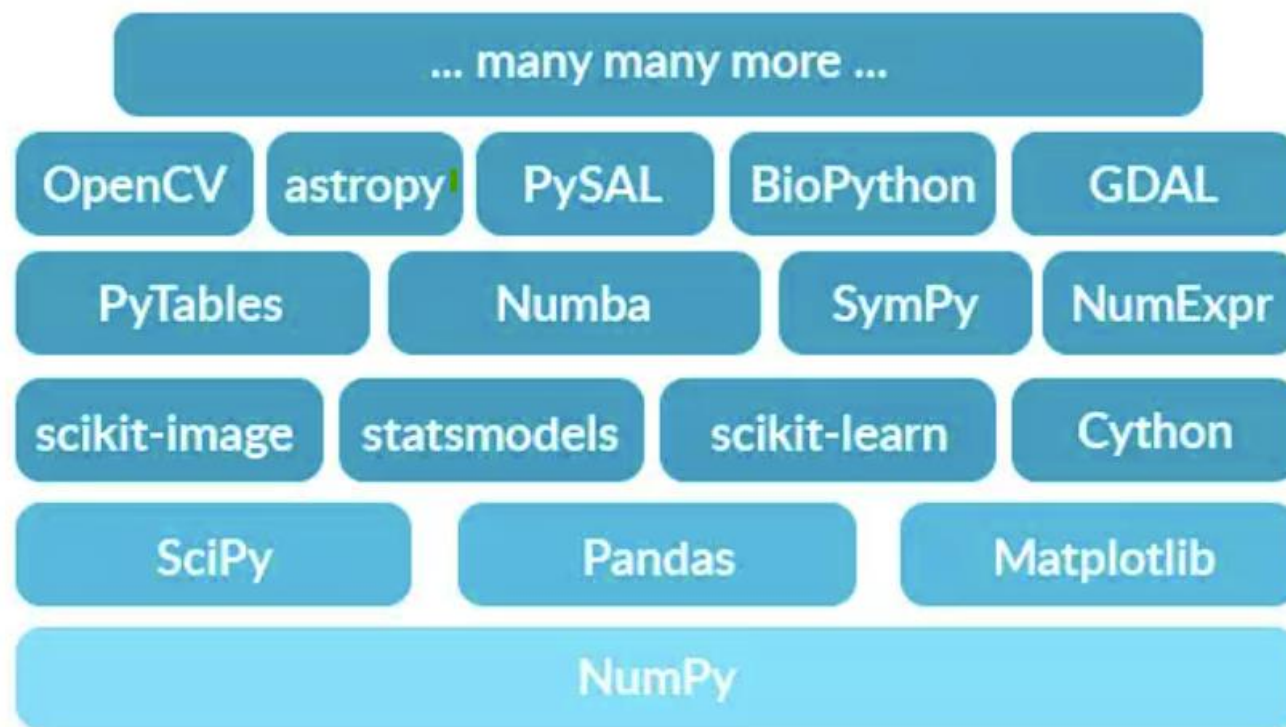
scikit-learn

The Python NumPy-based ecosystem includes tools for array-oriented computing

TensorFlow

Theano

Torch

scikit-learn

Caffe

TensorFlow

Theano

Torch

scikit-learn

Caffe

scikit-learn: Datasets and models

- The library already includes a few standard datasets for classification and regression despite their being too small to represent real-life situations. However, the diabetes dataset for measuring disease progression or the iris plants dataset for pattern recognition are good for illustrating how machine learning algorithms in scikit behave. Moreover, the library provides information about loading datasets from external sources, includes sample generators for tasks like multiclass classification and decomposition, and offers recommendations about popular datasets usage.

TensorFlow

Theano

Torch

scikit-learn

Caffe

Audience and learning curve

- Despite being a robust library, scikit-learn focuses on ease of use and documentation. Considering its simplicity and numerous well-described examples, it's an accessible tool for non-experts and neophyte engineers, enabling quick application of machine learning algorithms to data. According to testimonials by software shops AWeber and Yhat, scikit is well-suited for production characterized by limited time and human resources.

TensorFlow

Theano

Torch

scikit-learn

Caffe

Use cases

- scikit-learn has been adopted by a plethora of successful brands like Spotify, Evernote, e-commerce giant Birchbox, and Booking.com, for product recommendations and customer service. However, you don't have to be an expert to explore data science with the library. Thus, a technology and engineering school Télécom ParisTech uses the library for its machine learning courses to allow students to quickly solve interesting problems.

TensorFlow

Theano

Torch

scikit-learn

Caffe

Caffe/Caffe2: Easy to Learn Tool with Abundance of Pre-Trained Models

- While Theano and Torch are designed for research, Caffe isn't fit for text, sound, or time series data. It's a special-purpose machine learning library for image classification. The support from Facebook and the recently open sourced Caffe2 have made the library a popular tool with 248 GitHub contributors.

- Despite being criticized for slow development, Caffe's successor Caffe2 has been eliminating the existing problems of the original technology by adding flexibility, weightlessness, and support for mobile deployment.

TensorFlow

Theano

Torch

scikit-learn

Caffe

Caffe/Caffe2: Datasets and models

- Caffe encourages users to get familiar with datasets provided by both the industry and other users. The team fosters collaboration and links to the most popular datasets that have already been trained with Caffe. One of the biggest benefits of the framework is Model Zoo – a vast reservoir of pre-trained models created by developers and researchers, which allow you to use, or combine a model, or just learn to train a model of your own.

TensorFlow

Theano

Torch

scikit-learn

Caffe

Caffe/Caffe2: Audience and learning curve

- The Caffe team claims that you can skip the learning part and start exploring deep learning using the existing models straightaway. The library is targeted at developers who want to experience deep learning first hand and offers resources that promise to be expanded as the community develops.

TensorFlow

Theano

Torch

scikit-learn

Caffe

Caffe/Caffe2: Use cases

- By using the state-of-the-art Convolutional Neural Networks (CNNs) – deep neural networks successfully applied for visual imagery analysis and even powering vision in self-driving cars – Caffe allowed Facebook to develop its real-time video filtering tool for applying famous artistic styles on videos. Pinterest also used Caffe to expand a visual search function and allow users to discover specific objects on a picture.