Machine Learning Software Development Environments
EXAMPLES OF MACHINE LEARNING SOFTWARE DEVELOPMENT ENVIRONMENTS

MATLAB®

Python

R

Julia
Some Popular Machine Learning / Deep Learning Frameworks
EXAMPLES OF MACHINE LEARNING SOFTWARE FRAMEWORKS

TensorFlow

The Microsoft Cognitive Toolkit

scikit learn

Caffe2

PyTorch

Keras

Dr. Golden’s

SML

*Still under active development
Machine Learning Software
Frameworks
Statistical Machine Learning
Matlab Toolbox (Dr. Golden)

• Software that is still under development which will complement my new book “Statistical Machine Learning”
  – Linear Machine
  – Perceptron
  – Lunar Lander (reinforcement Learning)

• Good Points: Helps you understand nuts and bolts underneath hood. Also useful for the “brain” of your mad scientist creation. Illustrates key ideas from my book...

• Bad Points: In preliminary stages, little documentation, little supporting infrastructure, requires lots of expertise to apply to large-scale problems.
Current Number of Users < 20!
Tensor Flow (https://www.tensorflow.org/)

Get Started with TensorFlow

TensorFlow is an open-source machine learning library for research and production. TensorFlow offers APIs for beginners and experts to develop for desktop, mobile, web, and cloud. See the sections below to get started.

Learn and use ML

The high-level Keras API provides building blocks to create and train deep learning models. Start with these beginner-friendly notebook examples, then read the TensorFlow Keras guide.

1. Basic classification
2. Text classification
3. Regression
4. Overfitting and underfitting
5. Save and load

Research and experimentation

Eager execution provides an imperative, define-by-run interface for advanced operations. Write custom layers, forward passes, and training loops with auto-differentiation. Start with these notebooks, then read the eager execution guide.

ML at production scale

Estimators can train large models on multiple machines in a production environment. TensorFlow provides a collection of pre-made Estimators to implement common ML algorithms. See the Estimators guide.
TensorFlow widely used in Industry....

Machine Learning and Artificial Intelligence

Julia comes ready with Flux, a state-of-the-art framework for machine learning and AI.

As machine learning systems become more complex, researchers increasingly need differentiable languages in which they can simply write their algorithms as code. Julia's mathematical syntax makes it an ideal way to express algorithms just as they are written in papers, while Flux turns code into trainable models with automatic differentiation, GPU acceleration and support for terabytes of data via JuliaDB.

Flux provides a highly intuitive, layer-stacking-based interface for simple models, yet you can easily write your own mathematics for more advanced ones such as variational auto-encoders. Despite how easy it is to use, Flux is the most flexible ML framework on Earth; one can easily integrate it with other Julia libraries like physics simulators or differential equations solvers, and even drop down to writing a quick CUDA kernel via CUDAnative, all from one language, or even from a single script!
Top R packages for Machine Learning


<table>
<thead>
<tr>
<th>Package</th>
<th>CRAN</th>
<th>Stack Overflow</th>
<th>GitHub</th>
</tr>
</thead>
<tbody>
<tr>
<td>caret</td>
<td>1</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>randomForest</td>
<td>2</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>e1071</td>
<td>4</td>
<td>9</td>
<td>8</td>
</tr>
<tr>
<td>nnet</td>
<td>13</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>glmnet</td>
<td>10</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>tree</td>
<td>9</td>
<td>10</td>
<td>8</td>
</tr>
<tr>
<td>party</td>
<td>5</td>
<td>17</td>
<td>17</td>
</tr>
<tr>
<td>arules</td>
<td>8</td>
<td>16</td>
<td>16</td>
</tr>
<tr>
<td>kernlab</td>
<td>14</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>ROCR</td>
<td>12</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>gbm</td>
<td>11</td>
<td>18</td>
<td>18</td>
</tr>
<tr>
<td>RWeka</td>
<td>18</td>
<td>13</td>
<td>13</td>
</tr>
<tr>
<td>rattle</td>
<td>20</td>
<td>14</td>
<td>14</td>
</tr>
<tr>
<td>earth</td>
<td>1</td>
<td>7</td>
<td>7</td>
</tr>
</tbody>
</table>

The ranking is based on average rank of CRAN (The Comprehensive R Archive Network) downloads and Stack Overflow activity (full ranking here [CSV]). CRAN downloads are from the past year. Stack Overflow activity is from the past 30 days. Note: OneR: 1 (SO); mlr: 2 (GitHub); ranger: 4 (GitHub); SuperLearner: 5 (GitHub).
PYTORCH (https://pytorch.org/)
H2O(https://www.h2o.ai/products/h2o/)

H2O is an open source, in-memory, distributed, fast, and scalable machine learning platform that allows you to build machine learning models on big data and provides easy deployment of models in an enterprise environment. H2O can be used from R, Python, Scala, Java and H2O Flow – Interactive notebook.

Open Source, Distributed Machine Learning for Everyone
Industrial Strength Natural Language Processing...

https://spacy.io/

spaCy

Industrial-Strength Natural Language Processing

IN PYTHON

Fastest in the world

spaCy excels at large-scale information extraction tasks. It's written from the ground up in carefully memory-managed Cython. Independent research has confirmed that spaCy is the fastest in the world. If your application needs to process entire web dumps, spaCy is the library you want to be using.

FACTS & FIGURES

Get things done

spaCy is designed to help you do real work — to build real products, or gather real insights. The library respects your time, and tries to avoid wasting it. It's easy to install, and its API is simple and productive. We like to think of spaCy as the Ruby on Rails of Natural Language Processing.
1) CVonline is a free WWW-based set of introductions to topics in computer vision.

http://homepages.inf.ed.ac.uk/rbf/CVonline/

Because of the improvements in the content available in Wikipedia, it is now possible to find content for more than 50% of CVonline's 2000 topics. CVonline groups together the topics into a sensible topic hierarchy, but tries to exploit the advancing quality and breadth of Wikipedia's content.

2) CVonline has a variety of supplemental information useful to students and researchers, namely lists of:

- 1000+ online and hardcopy computer vision and image processing books: http://homepages.inf.ed.ac.uk/rbf/CVonline/books.htm
- 1000+ datasets for research and student projects: http://homepages.inf.ed.ac.uk/rbf/CVonline/Imagedbase.htm
- 175+ useful software packages: http://homepages.inf.ed.ac.uk/rbf/CVonline/SWEnvironments.htm
- List of over 300 different image analysis application areas: http://homepages.inf.ed.ac.uk/rbf/CVonline/applic.htm
Open AI Gym (https://gym.openai.com/)

Gym

Gym is a toolkit for developing and comparing reinforcement learning algorithms. It supports teaching agents everything from walking to playing games like Pong or Pinball.

View documentation >
View on GitHub >

Open source interface to reinforcement learning tasks.

The gym library provides an easy-to-use suite of reinforcement learning tasks.

```python
import gym
env = gym.make("LunarLander-v2")
observation = env.reset()
for _ in range(1000):
    env.render()
    action = env.action_space.sample() # your agent here (this takes random actions)
    observation, reward, done, info = env.step(action)
```
SCI-KIT LEARN
(http://scikit-learn.org/stable/)

Machine Learning in Python

- Simple and efficient tools for data mining and data analysis
- Accessible to everybody, and reusable in various contexts
- Built on NumPy, SciPy, and matplotlib
- Open source, commercially usable - BSD license

Classification
Identifying to which category an object belongs to.
Applications: Spam detection, Image recognition.
Algorithms: SVM, nearest neighbors, random forest, ...

Regression
Predicting a continuous-valued attribute associated with an object.
Applications: Drug response, Stock prices.
Algorithms: SVR, ridge regression, Lasso, ...

Clustering
Automatic grouping of similar objects into sets.
Applications: Customer segmentation, Grouping experiments outcomes
Algorithms: k-Means, spectral clustering, mean-shift, ...

Dimensionality reduction
Reducing the number of random variables to consider.
Applications: Visualization, Increased efficiency.
Algorithms: PCA, feature selection, non-negative matrix factorization.

Model selection
Comparing, validating and choosing parameters and models.
Goal: Improved accuracy via parameter tuning.
Modules: grid search, cross validation, metrics.

Preprocessing
Feature extraction and normalization.
Application: Transforming input data such as text for use with machine learning algorithms.
Modules: preprocessing, feature extraction.

News
On-going development: What's new (Changelog)

Community
About us See authors and contributing
More Machine Learning Find related projects

Who uses scikit-learn?
Caffe2 (https://caffe2.ai/)

A New Lightweight, Modular, and Scalable Deep Learning Framework

CODE ONCE, RUN ANYWHERE
Your favorite deep learning technology, now from zero to scale, cloud to mobile.

WHAT IS CAFFE2?
Caffe2 aims to provide an easy and straightforward way for you to experiment with deep learning and leverage community contributions of new models and algorithms. You can bring your creations to scale using the power of GPUs in the cloud or to the masses on mobile with Caffe2’s cross-platform libraries.

HOW DO YOU USE IT?
Try out our quickstart tutorials or jump in and start developing. Caffe2 comes with Python & C++ APIs so you can prototype now, easily optimize later. Use cloud services, Docker, or install on Mac, Windows, or Ubuntu. It integrates with Visual Studio, Android Studio, and Xcode for mobile development.
Keras (https://keras.io/)

Keras: The Python Deep Learning library

You have just found Keras.

Keras is a high-level neural networks API, written in Python and capable of running on top of TensorFlow, CNTK, or Theano. It was developed with a focus on enabling fast experimentation. Being able to go from idea to result with the least possible delay is key to doing good research.

Use Keras if you need a deep learning library that:

- Allows for easy and fast prototyping (through user friendliness, modularity, and extensibility).
- Supports both convolutional networks and recurrent networks, as well as combinations of the two.
- Runs seamlessly on CPU and GPU.

Read the documentation at Keras.io.

Keras is compatible with: Python 2.7-3.6.

Guiding principles

- User friendliness. Keras is an API designed for human beings, not machines. It puts user experience front and center. Keras follows best practices for reducing cognitive load: it offers consistent & simple APIs, it minimizes the number of user actions required for common use cases, and it provides clear and actionable feedback upon user error.
- Modularity. A model is understood as a sequence or a graph of standalone, fully-configurable modules that can
Unlock deeper learning with the new Microsoft Cognitive Toolkit

The Microsoft Cognitive Toolkit—previously known as CNTK—empowers you to harness the intelligence within massive datasets through deep learning by providing uncompromised scaling, speed, and accuracy with commercial-grade quality and compatibility with the programming languages and algorithms you already use. Hear about the team that developed the Cognitive Toolkit, or read more below.
Theano 1.0 (deeplearning.net/software/Theano)

Welcome

Theano is a Python library that allows you to define, optimize, and evaluate mathematical expressions involving multi-dimensional arrays efficiently. Theano features:

- **tight integration with NumPy** – Use `numpy.ndarray` in Theano-compiled functions.
- **transparent use of a GPU** – Perform data-intensive computations much faster than on a CPU.
- **efficient symbolic differentiation** – Theano does your derivatives for functions with one or many inputs.
- **speed and stability optimizations** – Get the right answer for \( \log(1+x) \) even when \( x \) is really tiny.
- **dynamic C code generation** – Evaluate expressions faster.
- **extensive unit-testing and self-verification** – Detect and diagnose many types of errors.

Theano has been powering large-scale computationally intensive scientific investigations since 2007. But it is also approachable enough to be used in the classroom (University of Montreal’s deep learning/machine learning classes).