josephmisiti /
awesome-machine-learning

<> **Code** ⊙ Issues � Pull requests ▷ Actions ⊘ Security ⩘ Insights

ℙ master ▾   [awesome-machine-learning](#) / README.md ⧉

⊙ **zechengz** Add PyTorch Frame and PyTorch Geometric   4 days ago ••• ⟲

1825 lines (1587 loc) · 192 KB

Preview   Code   Blame   Raw ⧉ ⭳ ✎ ▾   ☰

Awesome Machine Learning 👓 awesome   👀 Track Awesome List

A curated list of awesome machine learning frameworks, libraries and software (by language). Inspired by `awesome-php`.

If you want to contribute to this list (please do), send me a pull request or contact me [@josephmisiti](#). Also, a listed repository should be deprecated if:

- Repository's owner explicitly says that "this library is not maintained".
- Not committed for a long time (2~3 years).

Further resources:

- For a list of free machine learning books available for download, go [here](#).

- For a list of professional machine learning events, go [here](#).

- For a list of (mostly) free machine learning courses available online, go [here](#).

- For a list of blogs and newsletters on data science and machine learning, go [here](#).

- For a list of free-to-attend meetups and local events, go [here](#).

Table of Contents

Frameworks and Libraries

- [Awesome Machine Learning](#) 👓 awesome
  - [Table of Contents](#)
    - [Frameworks and Libraries](#)
    - [Tools](#)

APL

General-Purpose Machine Learning

- naive-apl - Naive Bayesian Classifier implementation in APL. **[Deprecated]**

C

General-Purpose Machine Learning

- Darknet - Darknet is an open source neural network framework written in C and CUDA. It is fast, easy to install, and supports CPU and GPU computation.
- Recommender - A C library for product recommendations/suggestions using collaborative filtering (CF).
- Hybrid Recommender System - A hybrid recommender system based upon scikit-learn algorithms. **[Deprecated]**
- neonrvm - neonrvm is an open source machine learning library based on RVM technique. It's written in C programming language and comes with Python programming language bindings.
- cONNXr - An `ONNX` runtime written in pure C (99) with zero dependencies focused on small embedded devices. Run inference on your machine learning models no matter which framework you train it with. Easy to install and compiles everywhere, even in very old devices.
- libonnx - A lightweight, portable pure C99 onnx inference engine for embedded devices with hardware acceleration support.

Computer Vision

- CCV - C-based/Cached/Core Computer Vision Library, A Modern Computer Vision Library.

- VLFeat - VLFeat is an open and portable library of computer vision algorithms, which has a Matlab toolbox.

C++

Computer Vision

- DLib - DLib has C++ and Python interfaces for face detection and training general object detectors.
- EBLearn - Eblearn is an object-oriented C++ library that implements various machine learning models **[Deprecated]**
- OpenCV - OpenCV has C++, C, Python, Java and MATLAB interfaces and supports Windows, Linux, Android and Mac OS.
- VIGRA - VIGRA is a genertic cross-platform C++ computer vision and machine learning library for volumes of arbitrary dimensionality with Python bindings.
- Openpose - A real-time multi-person keypoint detection library for body, face, hands, and foot estimation

General-Purpose Machine Learning

- Speedster -Automatically apply SOTA optimization techniques to achieve the maximum inference speed-up on your hardware. [DEEP LEARNING]
- BanditLib - A simple Multi-armed Bandit library. **[Deprecated]**
- Caffe - A deep learning framework developed with cleanliness, readability, and speed in mind. [DEEP LEARNING]
- CatBoost - General purpose gradient boosting on decision trees library with categorical features support out of the box. It is easy to install, contains fast inference implementation and supports CPU and GPU (even multi-GPU) computation.
- CNTK - The Computational Network Toolkit (CNTK) by Microsoft Research, is a unified deep-learning toolkit that describes neural networks as a series of computational steps via a directed graph.
- CUDA - This is a fast C++/CUDA implementation of convolutional [DEEP LEARNING]
- DeepDetect - A machine learning API and server written in C++11. It makes state of the art machine learning easy to work with and integrate into existing applications.
- Distributed Machine learning Tool Kit (DMTK) - A distributed machine learning (parameter server) framework by Microsoft. Enables training models on large data sets across multiple machines. Current tools bundled with it include: LightLDA and Distributed (Multisense) Word Embedding.
- DLib - A suite of ML tools designed to be easy to imbed in other applications.
- DSSTNE - A software library created by Amazon for training and deploying deep neural networks using GPUs which emphasizes speed and scale over experimental flexibility.
- DyNet - A dynamic neural network library working well with networks that have dynamic structures that change for every training instance. Written in C++ with bindings in Python.
- Fido - A highly-modular C++ machine learning library for embedded electronics and robotics.

- [igraph](#) - General purpose graph library.
- [Intel® oneAPI Data Analytics Library](#) - A high performance software library developed by Intel and optimized for Intel's architectures. Library provides algorithmic building blocks for all stages of data analytics and allows to process data in batch, online and distributed modes.
- [LightGBM](#) - Microsoft's fast, distributed, high performance gradient boosting (GBDT, GBRT, GBM or MART) framework based on decision tree algorithms, used for ranking, classification and many other machine learning tasks.
- [libfm](#) - A generic approach that allows to mimic most factorization models by feature engineering.
- [MLDB](#) - The Machine Learning Database is a database designed for machine learning. Send it commands over a RESTful API to store data, explore it using SQL, then train machine learning models and expose them as APIs.
- [mlpack](#) - A scalable C++ machine learning library.
- [MXNet](#) - Lightweight, Portable, Flexible Distributed/Mobile Deep Learning with Dynamic, Mutation-aware Dataflow Dep Scheduler; for Python, R, Julia, Go, JavaScript and more.
- [N2D2](#) - CEA-List's CAD framework for designing and simulating Deep Neural Network, and building full DNN-based applications on embedded platforms
- [oneDNN](#) - An open-source cross-platform performance library for deep learning applications.
- [ParaMonte](#) - A general-purpose library with C/C++ interface for Bayesian data analysis and visualization via serial/parallel Monte Carlo and MCMC simulations. Documentation can be found [here](#).
- [proNet-core](#) - A general-purpose network embedding framework: pair-wise representations optimization Network Edit.
- [PyCaret](#) - An open-source, low-code machine learning library in Python that automates machine learning workflows.
- [PyCUDA](#) - Python interface to CUDA
- [ROOT](#) - A modular scientific software framework. It provides all the functionalities needed to deal with big data processing, statistical analysis, visualization and storage.
- [shark](#) - A fast, modular, feature-rich open-source C++ machine learning library.
- [Shogun](#) - The Shogun Machine Learning Toolbox.
- [sofia-ml](#) - Suite of fast incremental algorithms.
- [Stan](#) - A probabilistic programming language implementing full Bayesian statistical inference with Hamiltonian Monte Carlo sampling.
- [Timbl](#) - A software package/C++ library implementing several memory-based learning algorithms, among which IB1-IG, an implementation of k-nearest neighbor classification, and IGTree, a decision-tree approximation of IB1-IG. Commonly used for NLP.
- [Vowpal Wabbit (VW)](#) - A fast out-of-core learning system.
- [Warp-CTC](#) - A fast parallel implementation of Connectionist Temporal Classification (CTC), on both CPU and GPU.
- [XGBoost](#) - A parallelized optimized general purpose gradient boosting library.
- [ThunderGBM](#) - A fast library for GBDTs and Random Forests on GPUs.
- [ThunderSVM](#) - A fast SVM library on GPUs and CPUs.

- LKYDeepNN - A header-only C++11 Neural Network library. Low dependency, native traditional chinese document.
- xLearn - A high performance, easy-to-use, and scalable machine learning package, which can be used to solve large-scale machine learning problems. xLearn is especially useful for solving machine learning problems on large-scale sparse data, which is very common in Internet services such as online advertising and recommender systems.
- Featuretools - A library for automated feature engineering. It excels at transforming transactional and relational datasets into feature matrices for machine learning using reusable feature engineering "primitives".
- skynet - A library for learning neural networks, has C-interface, net set in JSON. Written in C++ with bindings in Python, C++ and C#.
- Feast - A feature store for the management, discovery, and access of machine learning features. Feast provides a consistent view of feature data for both model training and model serving.
- Hopsworks - A data-intensive platform for AI with the industry's first open-source feature store. The Hopsworks Feature Store provides both a feature warehouse for training and batch based on Apache Hive and a feature serving database, based on MySQL Cluster, for online applications.
- Polyaxon - A platform for reproducible and scalable machine learning and deep learning.
- QuestDB - A relational column-oriented database designed for real-time analytics on time series and event data.
- Phoenix - Uncover insights, surface problems, monitor and fine tune your generative LLM, CV and tabular models.
- XAD - Comprehensive backpropagation tool for C++.
- Truss - An open source framework for packaging and serving ML models.

Natural Language Processing

- BLLIP Parser - BLLIP Natural Language Parser (also known as the Charniak-Johnson parser).
- colibri-core - C++ library, command line tools, and Python binding for extracting and working with basic linguistic constructions such as n-grams and skipgrams in a quick and memory-efficient way.
- CRF++ - Open source implementation of Conditional Random Fields (CRFs) for segmenting/labeling sequential data & other Natural Language Processing tasks. **[Deprecated]**
- CRFsuite - CRFsuite is an implementation of Conditional Random Fields (CRFs) for labeling sequential data. **[Deprecated]**
- frog - Memory-based NLP suite developed for Dutch: PoS tagger, lemmatiser, dependency parser, NER, shallow parser, morphological analyzer.
- libfolia - C++ library for the FoLiA format
- MeTA - MeTA : ModErn Text Analysis is a C++ Data Sciences Toolkit that facilitates mining big text data.
- MIT Information Extraction Toolkit - C, C++, and Python tools for named entity recognition and relation extraction
- ucto - Unicode-aware regular-expression based tokenizer for various languages. Tool and C++ library. Supports FoLiA format.

Speech Recognition

- [Kaldi](#) - Kaldi is a toolkit for speech recognition written in C++ and licensed under the Apache License v2.0. Kaldi is intended for use by speech recognition researchers.

Sequence Analysis

- [ToPS](#) - This is an object-oriented framework that facilitates the integration of probabilistic models for sequences over a user defined alphabet. **[Deprecated]**

Gesture Detection

- [grt](#) - The Gesture Recognition Toolkit (GRT) is a cross-platform, open-source, C++ machine learning library designed for real-time gesture recognition.

Common Lisp

General-Purpose Machine Learning

- [mgl](#) - Neural networks (boltzmann machines, feed-forward and recurrent nets), Gaussian Processes.
- [mgl-gpr](#) - Evolutionary algorithms. **[Deprecated]**
- [cl-libsvm](#) - Wrapper for the libsvm support vector machine library. **[Deprecated]**
- [cl-online-learning](#) - Online learning algorithms (Perceptron, AROW, SCW, Logistic Regression).
- [cl-random-forest](#) - Implementation of Random Forest in Common Lisp.

Clojure

Natural Language Processing

- [Clojure-openNLP](#) - Natural Language Processing in Clojure (opennlp).
- [Infections-clj](#) - Rails-like inflection library for Clojure and ClojureScript.

General-Purpose Machine Learning

- [scicloj.ml](#) - A idiomatic Clojure machine learning library based on tech.ml.dataset with a unique approach for immutable data processing pipelines.
- [clj-ml](#) - A machine learning library for Clojure built on top of Weka and friends.
- [clj-boost](#) - Wrapper for XGBoost
- [Touchstone](#) - Clojure A/B testing library.
- [Clojush](#) - The Push programming language and the PushGP genetic programming system implemented in Clojure.
- [lambda-ml](#) - Simple, concise implementations of machine learning techniques and utilities in Clojure.
- [Infer](#) - Inference and machine learning in Clojure. **[Deprecated]**

- Encog - Clojure wrapper for Encog (v3) (Machine-Learning framework that specializes in neural-nets). **[Deprecated]**
- Fungp - A genetic programming library for Clojure. **[Deprecated]**
- Statistiker - Basic Machine Learning algorithms in Clojure. **[Deprecated]**
- clortex - General Machine Learning library using Numenta's Cortical Learning Algorithm. **[Deprecated]**
- comportex - Functionally composable Machine Learning library using Numenta's Cortical Learning Algorithm. **[Deprecated]**

Deep Learning

- MXNet - Bindings to Apache MXNet - part of the MXNet project
- Deep Diamond - A fast Clojure Tensor & Deep Learning library
- jutsu.ai - Clojure wrapper for deeplearning4j with some added syntactic sugar.
- cortex - Neural networks, regression and feature learning in Clojure.
- Flare - Dynamic Tensor Graph library in Clojure (think PyTorch, DynNet, etc.)
- dl4clj - Clojure wrapper for Deeplearning4j.

Data Analysis

- tech.ml.dataset - Clojure dataframe library and pipeline for data processing and machine learning
- Tablecloth - A dataframe grammar wrapping tech.ml.dataset, inspired by several R libraries
- Panthera - Clojure API wrapping Python's Pandas library
- Incanter - Incanter is a Clojure-based, R-like platform for statistical computing and graphics.
- PigPen - Map-Reduce for Clojure.
- Geni - a Clojure dataframe library that runs on Apache Spark

Data Visualization

- Hanami : Clojure(Script) library and framework for creating interactive visualization applications based in Vega-Lite (VGL) and/or Vega (VG) specifications. Automatic framing and layouts along with a powerful templating system for abstracting visualization specs
- Saite - Clojure(Script) client/server application for dynamic interactive explorations and the creation of live shareable documents capturing them using Vega/Vega-Lite, CodeMirror, markdown, and LaTeX
- Oz - Data visualisation using Vega/Vega-Lite and Hiccup, and a live-reload platform for literate-programming
- Envision - Clojure Data Visualisation library, based on Statistiker and D3.
- Pink Gorilla Notebook - A Clojure/Clojurescript notebook application/-library based on Gorilla-REPL
- clojupyter - A Jupyter kernel for Clojure - run Clojure code in Jupyter Lab, Notebook and Console.
- notespace - Notebook experience in your Clojure namespace
- Delight - A listener that streams your spark events logs to delight, a free and improved spark UI

Interop

- [Java Interop](#) - Clojure has Native Java Interop from which Java's ML ecosystem can be accessed
- [JavaScript Interop](#) - ClojureScript has Native JavaScript Interop from which JavaScript's ML ecosystem can be accessed
- [Libpython-clj](#) - Interop with Python
- [ClojisR](#) - Interop with R and Renjin (R on the JVM)

Misc

- [Neanderthal](#) - Fast Clojure Matrix Library (native CPU, GPU, OpenCL, CUDA)
- [kixistats](#) - A library of statistical distribution sampling and transducing functions
- [fastmath](#) - A collection of functions for mathematical and statistical computing, macine learning, etc., wrapping several JVM libraries
- [matlib](#) - A Clojure library of optimisation and control theory tools and convenience functions based on Neanderthal.

Extra

- [Scicloj](#) - Curated list of ML related resources for Clojure.

Crystal

General-Purpose Machine Learning

- [machine](#) - Simple machine learning algorithm.
- [crystal-fann](#) - FANN (Fast Artificial Neural Network) binding.

Elixir

General-Purpose Machine Learning

- [Simple Bayes](#) - A Simple Bayes / Naive Bayes implementation in Elixir.
- [emel](#) - A simple and functional machine learning library written in Elixir.
- [Tensorflex](#) - Tensorflow bindings for the Elixir programming language.

Natural Language Processing

- [Stemmer](#) - An English (Porter2) stemming implementation in Elixir.

Erlang

General-Purpose Machine Learning

- [Disco](#) - Map Reduce in Erlang. **[Deprecated]**

Fortran

General-Purpose Machine Learning

- neural-fortran - A parallel neural net microframework. Read the paper here.

Data Analysis / Data Visualization

- ParaMonte - A general-purpose Fortran library for Bayesian data analysis and visualization via serial/parallel Monte Carlo and MCMC simulations. Documentation can be found here.

Go

Natural Language Processing

- Cybertron - Cybertron: the home planet of the Transformers in Go.
- snowball - Snowball Stemmer for Go.
- word-embedding - Word Embeddings: the full implementation of word2vec, GloVe in Go.
- sentences - Golang implementation of Punkt sentence tokenizer.
- go-ngram - In-memory n-gram index with compression. *[Deprecated]*
- paicehusk - Golang implementation of the Paice/Husk Stemming Algorithm. *[Deprecated]*
- go-porterstemmer - A native Go clean room implementation of the Porter Stemming algorithm. **[Deprecated]**

General-Purpose Machine Learning

- Spago - Self-contained Machine Learning and Natural Language Processing library in Go.
- birdland - A recommendation library in Go.
- eaopt - An evolutionary optimization library.
- leaves - A pure Go implementation of the prediction part of GBRTs, including XGBoost and LightGBM.
- gobrain - Neural Networks written in Go.
- go-featureprocessing - Fast and convenient feature processing for low latency machine learning in Go.
- go-mxnet-predictor - Go binding for MXNet c_predict_api to do inference with a pre-trained model.
- go-ml-benchmarks — benchmarks of machine learning inference for Go.
- go-ml-transpiler - An open source Go transpiler for machine learning models.
- golearn - Machine learning for Go.
- goml - Machine learning library written in pure Go.
- gorgonia - Deep learning in Go.
- goro - A high-level machine learning library in the vein of Keras.
- gorse - An offline recommender system backend based on collaborative filtering written in Go.

- [therfoo](#) - An embedded deep learning library for Go.
- [neat](#) - Plug-and-play, parallel Go framework for NeuroEvolution of Augmenting Topologies (NEAT). **[Deprecated]**
- [go-pr](#) - Pattern recognition package in Go lang. **[Deprecated]**
- [go-ml](#) - Linear / Logistic regression, Neural Networks, Collaborative Filtering and Gaussian Multivariate Distribution. **[Deprecated]**
- [GoNN](#) - GoNN is an implementation of Neural Network in Go Language, which includes BPNN, RBF, PCN. **[Deprecated]**
- [bayesian](#) - Naive Bayesian Classification for Golang. **[Deprecated]**
- [go-galib](#) - Genetic Algorithms library written in Go / Golang. **[Deprecated]**
- [Cloudforest](#) - Ensembles of decision trees in Go/Golang. **[Deprecated]**
- [go-dnn](#) - Deep Neural Networks for Golang (powered by MXNet)

Spatial analysis and geometry

- [go-geom](#) - Go library to handle geometries.
- [gogeo](#) - Spherical geometry in Go.

Data Analysis / Data Visualization

- [dataframe-go](#) - Dataframes for machine-learning and statistics (similar to pandas).
- [gota](#) - Dataframes.
- [gonum/mat](#) - A linear algebra package for Go.
- [gonum/optimize](#) - Implementations of optimization algorithms.
- [gonum/plot](#) - A plotting library.
- [gonum/stat](#) - A statistics library.
- [SVGo](#) - The Go Language library for SVG generation.
- [glot](#) - Glot is a plotting library for Golang built on top of gnuplot.
- [globe](#) - Globe wireframe visualization.
- [gonum/graph](#) - General-purpose graph library.
- [go-graph](#) - Graph library for Go/Golang language. **[Deprecated]**
- [RF](#) - Random forests implementation in Go. **[Deprecated]**

Computer vision

- [GoCV](#) - Package for computer vision using OpenCV 4 and beyond.

Reinforcement learning

- [gold](#) - A reinforcement learning library.
- [stable-baselines3](#) - PyTorch implementations of Stable Baselines (deep) reinforcement learning algorithms.

Haskell

General-Purpose Machine Learning

- haskell-ml - Haskell implementations of various ML algorithms. **[Deprecated]**
- HLearn - a suite of libraries for interpreting machine learning models according to their algebraic structure. **[Deprecated]**
- hnn - Haskell Neural Network library.
- hopfield-networks - Hopfield Networks for unsupervised learning in Haskell. **[Deprecated]**
- DNNGraph - A DSL for deep neural networks. **[Deprecated]**
- LambdaNet - Configurable Neural Networks in Haskell. **[Deprecated]**

Java

Natural Language Processing

- Cortical.io - Retina: an API performing complex NLP operations (disambiguation, classification, streaming text filtering, etc...) as quickly and intuitively as the brain.
- IRIS - Cortical.io's FREE NLP, Retina API Analysis Tool (written in JavaFX!) - See the Tutorial Video.
- CoreNLP - Stanford CoreNLP provides a set of natural language analysis tools which can take raw English language text input and give the base forms of words.
- Stanford Parser - A natural language parser is a program that works out the grammatical structure of sentences.
- Stanford POS Tagger - A Part-Of-Speech Tagger (POS Tagger).
- Stanford Name Entity Recognizer - Stanford NER is a Java implementation of a Named Entity Recognizer.
- Stanford Word Segmenter - Tokenization of raw text is a standard pre-processing step for many NLP tasks.
- Tregex, Tsurgeon and Semgrex - Tregex is a utility for matching patterns in trees, based on tree relationships and regular expression matches on nodes (the name is short for "tree regular expressions").
- Stanford Phrasal: A Phrase-Based Translation System
- Stanford English Tokenizer - Stanford Phrasal is a state-of-the-art statistical phrase-based machine translation system, written in Java.
- Stanford Tokens Regex - A tokenizer divides text into a sequence of tokens, which roughly correspond to "words".
- Stanford Temporal Tagger - SUTime is a library for recognizing and normalizing time expressions.
- Stanford SPIED - Learning entities from unlabeled text starting with seed sets using patterns in an iterative fashion.
- Twitter Text Java - A Java implementation of Twitter's text processing library.
- MALLET - A Java-based package for statistical natural language processing, document classification, clustering, topic modelling, information extraction, and other machine learning

applications to text.

- OpenNLP - A machine learning based toolkit for the processing of natural language text.
- LingPipe - A tool kit for processing text using computational linguistics.
- ClearTK - ClearTK provides a framework for developing statistical natural language processing (NLP) components in Java and is built on top of Apache UIMA. **[Deprecated]**
- Apache cTAKES - Apache Clinical Text Analysis and Knowledge Extraction System (cTAKES) is an open-source natural language processing system for information extraction from electronic medical record clinical free-text.
- NLP4J - The NLP4J project provides software and resources for natural language processing. The project started at the Center for Computational Language and EducAtion Research, and is currently developed by the Center for Language and Information Research at Emory University. **[Deprecated]**
- CogcompNLP - This project collects a number of core libraries for Natural Language Processing (NLP) developed in the University of Illinois' Cognitive Computation Group, for example `illinois-core-utilities` which provides a set of NLP-friendly data structures and a number of NLP-related utilities that support writing NLP applications, running experiments, etc, `illinois-edison` a library for feature extraction from illinois-core-utilities data structures and many other packages.

General-Purpose Machine Learning

- aerosolve - A machine learning library by Airbnb designed from the ground up to be human friendly.
- AMIDST Toolbox - A Java Toolbox for Scalable Probabilistic Machine Learning.
- Chips-n-Salsa - A Java library for genetic algorithms, evolutionary computation, and stochastic local search, with a focus on self-adaptation / self-tuning, as well as parallel execution.
- Datumbox - Machine Learning framework for rapid development of Machine Learning and Statistical applications.
- ELKI - Java toolkit for data mining. (unsupervised: clustering, outlier detection etc.)
- Encog - An advanced neural network and machine learning framework. Encog contains classes to create a wide variety of networks, as well as support classes to normalize and process data for these neural networks. Encog trainings using multithreaded resilient propagation. Encog can also make use of a GPU to further speed processing time. A GUI based workbench is also provided to help model and train neural networks.
- FlinkML in Apache Flink - Distributed machine learning library in Flink.
- H2O - ML engine that supports distributed learning on Hadoop, Spark or your laptop via APIs in R, Python, Scala, REST/JSON.
- htm.java - General Machine Learning library using Numenta's Cortical Learning Algorithm.
- liblinear-java - Java version of liblinear.
- Mahout - Distributed machine learning.
- Meka - An open source implementation of methods for multi-label classification and evaluation (extension to Weka).

- [MLlib in Apache Spark](#) - Distributed machine learning library in Spark.
- [Hydrosphere Mist](#) - a service for deployment Apache Spark MLLib machine learning models as realtime, batch or reactive web services.
- [Neuroph](#) - Neuroph is lightweight Java neural network framework.
- [ORYX](#) - Lambda Architecture Framework using Apache Spark and Apache Kafka with a specialization for real-time large-scale machine learning.
- [Samoa](#) SAMOA is a framework that includes distributed machine learning for data streams with an interface to plug-in different stream processing platforms.
- [RankLib](#) - RankLib is a library of learning to rank algorithms. **[Deprecated]**
- [rapaio](#) - statistics, data mining and machine learning toolbox in Java.
- [RapidMiner](#) - RapidMiner integration into Java code.
- [Stanford Classifier](#) - A classifier is a machine learning tool that will take data items and place them into one of k classes.
- [Smile](#) - Statistical Machine Intelligence & Learning Engine.
- [SystemML](#) - flexible, scalable machine learning (ML) language.
- [Tribou](#) - A machine learning library written in Java by Oracle.
- [Weka](#) - Weka is a collection of machine learning algorithms for data mining tasks.
- [LBJava](#) - Learning Based Java is a modelling language for the rapid development of software systems, offers a convenient, declarative syntax for classifier and constraint definition directly in terms of the objects in the programmer's application.
- [knn-java-library](#) - Just a simple implementation of K-Nearest Neighbors algorithm using with a bunch of similarity measures.

Speech Recognition

- [CMU Sphinx](#) - Open Source Toolkit For Speech Recognition purely based on Java speech recognition library.

Data Analysis / Data Visualization

- [Flink](#) - Open source platform for distributed stream and batch data processing.
- [Hadoop](#) - Hadoop/HDFS.
- [Onyx](#) - Distributed, masterless, high performance, fault tolerant data processing. Written entirely in Clojure.
- [Spark](#) - Spark is a fast and general engine for large-scale data processing.
- [Storm](#) - Storm is a distributed realtime computation system.
- [Impala](#) - Real-time Query for Hadoop.
- [DataMelt](#) - Mathematics software for numeric computation, statistics, symbolic calculations, data analysis and data visualization.
- [Dr. Michael Thomas Flanagan's Java Scientific Library.](#) **[Deprecated]**

Deep Learning

- Deeplearning4j - Scalable deep learning for industry with parallel GPUs.
- Keras Beginner Tutorial - Friendly guide on using Keras to implement a simple Neural Network in Python.
- deepjavalibrary/djl - Deep Java Library (DJL) is an open-source, high-level, engine-agnostic Java framework for deep learning, designed to be easy to get started with and simple to use for Java developers.

JavaScript

Natural Language Processing

- Twitter-text - A JavaScript implementation of Twitter's text processing library.
- natural - General natural language facilities for node.
- Knwl.js - A Natural Language Processor in JS.
- Retext - Extensible system for analyzing and manipulating natural language.
- NLP Compromise - Natural Language processing in the browser.
- nlp.js - An NLP library built in node over Natural, with entity extraction, sentiment analysis, automatic language identify, and so more.

Data Analysis / Data Visualization

- D3.js
- High Charts
- NVD3.js
- dc.js
- chartjs
- dimple
- amCharts
- D3xter - Straight forward plotting built on D3. **[Deprecated]**
- statkit - Statistics kit for JavaScript. **[Deprecated]**
- datakit - A lightweight framework for data analysis in JavaScript
- science.js - Scientific and statistical computing in JavaScript. **[Deprecated]**
- Z3d - Easily make interactive 3d plots built on Three.js **[Deprecated]**
- Sigma.js - JavaScript library dedicated to graph drawing.
- C3.js - customizable library based on D3.js for easy chart drawing.
- Datamaps - Customizable SVG map/geo visualizations using D3.js. **[Deprecated]**
- ZingChart - library written on Vanilla JS for big data visualization.
- cheminfo - Platform for data visualization and analysis, using the visualizer project.
- Learn JS Data
- AnyChart
- FusionCharts

- [Nivo](#) - built on top of the awesome d3 and Reactjs libraries

General-Purpose Machine Learning

- [Auto ML](#) - Automated machine learning, data formatting, ensembling, and hyperparameter optimization for competitions and exploration- just give it a .csv file! **[Deprecated]**
- [Convnet.js](#) - ConvNetJS is a JavaScript library for training Deep Learning models[DEEP LEARNING] **[Deprecated]**
- [Clusterfck](#) - Agglomerative hierarchical clustering implemented in JavaScript for Node.js and the browser. **[Deprecated]**
- [Clustering.js](#) - Clustering algorithms implemented in JavaScript for Node.js and the browser. **[Deprecated]**
- [Decision Trees](#) - NodeJS Implementation of Decision Tree using ID3 Algorithm. **[Deprecated]**
- [DN2A](#) - Digital Neural Networks Architecture. **[Deprecated]**
- [figue](#) - K-means, fuzzy c-means and agglomerative clustering.
- [Gaussian Mixture Model](#) - Unsupervised machine learning with multivariate Gaussian mixture model.
- [Node-fann](#) - FANN (Fast Artificial Neural Network Library) bindings for Node.js **[Deprecated]**
- [Keras.js](#) - Run Keras models in the browser, with GPU support provided by WebGL 2.
- [Kmeans.js](#) - Simple JavaScript implementation of the k-means algorithm, for node.js and the browser. **[Deprecated]**
- [LDA.js](#) - LDA topic modelling for Node.js
- [Learning.js](#) - JavaScript implementation of logistic regression/c4.5 decision tree **[Deprecated]**
- [machinelearn.js](#) - Machine Learning library for the web, Node.js and developers
- [mil-tokyo](#) - List of several machine learning libraries.
- [Node-SVM](#) - Support Vector Machine for Node.js
- [Brain](#) - Neural networks in JavaScript **[Deprecated]**
- [Brain.js](#) - Neural networks in JavaScript - continued community fork of [Brain](#).
- [Bayesian-Bandit](#) - Bayesian bandit implementation for Node and the browser. **[Deprecated]**
- [Synaptic](#) - Architecture-free neural network library for Node.js and the browser.
- [kNear](#) - JavaScript implementation of the k nearest neighbors algorithm for supervised learning.
- [NeuralN](#) - C++ Neural Network library for Node.js. It has advantage on large dataset and multi-threaded training. **[Deprecated]**
- [kalman](#) - Kalman filter for JavaScript. **[Deprecated]**
- [shaman](#) - Node.js library with support for both simple and multiple linear regression. **[Deprecated]**
- [ml.js](#) - Machine learning and numerical analysis tools for Node.js and the Browser!
- [ml5](#) - Friendly machine learning for the web!
- [Pavlov.js](#) - Reinforcement learning using Markov Decision Processes.
- [MXNet](#) - Lightweight, Portable, Flexible Distributed/Mobile Deep Learning with Dynamic, Mutation-aware Dataflow Dep Scheduler; for Python, R, Julia, Go, JavaScript and more.

- TensorFlow.js - A WebGL accelerated, browser based JavaScript library for training and deploying ML models.
- JSMLT - Machine learning toolkit with classification and clustering for Node.js; supports visualization (see visualml.io).
- xgboost-node - Run XGBoost model and make predictions in Node.js.
- Netron - Visualizer for machine learning models.
- tensor-js - A deep learning library for the browser, accelerated by WebGL and WebAssembly.
- WebDNN - Fast Deep Neural Network JavaScript Framework. WebDNN uses next generation JavaScript API, WebGPU for GPU execution, and WebAssembly for CPU execution.

Misc

- stdlib - A standard library for JavaScript and Node.js, with an emphasis on numeric computing. The library provides a collection of robust, high performance libraries for mathematics, statistics, streams, utilities, and more.
- sylvester - Vector and Matrix math for JavaScript. **[Deprecated]**
- simple-statistics - A JavaScript implementation of descriptive, regression, and inference statistics. Implemented in literate JavaScript with no dependencies, designed to work in all modern browsers (including IE) as well as in Node.js.
- regression-js - A javascript library containing a collection of least squares fitting methods for finding a trend in a set of data.
- Lyric - Linear Regression library. **[Deprecated]**
- GreatCircle - Library for calculating great circle distance.
- MLPleaseHelp - MLPleaseHelp is a simple ML resource search engine. You can use this search engine right now at https://jgreenemi.github.io/MLPleaseHelp/, provided via GitHub Pages.
- Pipcook - A JavaScript application framework for machine learning and its engineering.

Demos and Scripts

- The Bot - Example of how the neural network learns to predict the angle between two points created with Synaptic.
- Half Beer - Beer glass classifier created with Synaptic.
- NSFWJS - Indecent content checker with TensorFlow.js
- Rock Paper Scissors - Rock Paper Scissors trained in the browser with TensorFlow.js
- Heroes Wear Masks - A fun TensorFlow.js-based oracle that tells, whether one wears a face mask or not. It can even tell when one wears the mask incorrectly.

Julia

General-Purpose Machine Learning

- MachineLearning - Julia Machine Learning library. **[Deprecated]**
- MLBase - A set of functions to support the development of machine learning algorithms.

- [PGM](#) - A Julia framework for probabilistic graphical models.
- [DA](#) - Julia package for Regularized Discriminant Analysis.
- [Regression](#) - Algorithms for regression analysis (e.g. linear regression and logistic regression). **[Deprecated]**
- [Local Regression](#) - Local regression, so smooooth!
- [Naive Bayes](#) - Simple Naive Bayes implementation in Julia. **[Deprecated]**
- [Mixed Models](#) - A Julia package for fitting (statistical) mixed-effects models.
- [Simple MCMC](#) - basic MCMC sampler implemented in Julia. **[Deprecated]**
- [Distances](#) - Julia module for Distance evaluation.
- [Decision Tree](#) - Decision Tree Classifier and Regressor.
- [Neural](#) - A neural network in Julia.
- [MCMC](#) - MCMC tools for Julia. **[Deprecated]**
- [Mamba](#) - Markov chain Monte Carlo (MCMC) for Bayesian analysis in Julia.
- [GLM](#) - Generalized linear models in Julia.
- [Gaussian Processes](#) - Julia package for Gaussian processes.
- [Online Learning](#) **[Deprecated]**
- [GLMNet](#) - Julia wrapper for fitting Lasso/ElasticNet GLM models using glmnet.
- [Clustering](#) - Basic functions for clustering data: k-means, dp-means, etc.
- [SVM](#) - SVM for Julia. **[Deprecated]**
- [Kernel Density](#) - Kernel density estimators for Julia.
- [MultivariateStats](#) - Methods for dimensionality reduction.
- [NMF](#) - A Julia package for non-negative matrix factorization.
- [ANN](#) - Julia artificial neural networks. **[Deprecated]**
- [Mocha](#) - Deep Learning framework for Julia inspired by Caffe. **[Deprecated]**
- [XGBoost](#) - eXtreme Gradient Boosting Package in Julia.
- [ManifoldLearning](#) - A Julia package for manifold learning and nonlinear dimensionality reduction.
- [MXNet](#) - Lightweight, Portable, Flexible Distributed/Mobile Deep Learning with Dynamic, Mutation-aware Dataflow Dep Scheduler; for Python, R, Julia, Go, JavaScript and more.
- [Merlin](#) - Flexible Deep Learning Framework in Julia.
- [ROCAnalysis](#) - Receiver Operating Characteristics and functions for evaluation probabilistic binary classifiers.
- [GaussianMixtures](#) - Large scale Gaussian Mixture Models.
- [ScikitLearn](#) - Julia implementation of the scikit-learn API.
- [Knet](#) - Koç University Deep Learning Framework.
- [Flux](#) - Relax! Flux is the ML library that doesn't make you tensor
- [MLJ](#) - A Julia machine learning framework.

Natural Language Processing

- [Topic Models](#) - TopicModels for Julia. **[Deprecated]**

- [Text Analysis](#) - Julia package for text analysis.
- [Word Tokenizers](#) - Tokenizers for Natural Language Processing in Julia
- [Corpus Loaders](#) - A Julia package providing a variety of loaders for various NLP corpora.
- [Embeddings](#) - Functions and data dependencies for loading various word embeddings
- [Languages](#) - Julia package for working with various human languages
- [WordNet](#) - A Julia package for Princeton's WordNet

Data Analysis / Data Visualization

- [Graph Layout](#) - Graph layout algorithms in pure Julia.
- [LightGraphs](#) - Graph modelling and analysis.
- [Data Frames Meta](#) - Metaprogramming tools for DataFrames.
- [Julia Data](#) - library for working with tabular data in Julia. **[Deprecated]**
- [Data Read](#) - Read files from Stata, SAS, and SPSS.
- [Hypothesis Tests](#) - Hypothesis tests for Julia.
- [Gadfly](#) - Crafty statistical graphics for Julia.
- [Stats](#) - Statistical tests for Julia.
- [RDataSets](#) - Julia package for loading many of the data sets available in R.
- [DataFrames](#) - library for working with tabular data in Julia.
- [Distributions](#) - A Julia package for probability distributions and associated functions.
- [Data Arrays](#) - Data structures that allow missing values. **[Deprecated]**
- [Time Series](#) - Time series toolkit for Julia.
- [Sampling](#) - Basic sampling algorithms for Julia.

Misc Stuff / Presentations

- [DSP](#) - Digital Signal Processing (filtering, periodograms, spectrograms, window functions).
- [JuliaCon Presentations](#) - Presentations for JuliaCon.
- [SignalProcessing](#) - Signal Processing tools for Julia.
- [Images](#) - An image library for Julia.
- [DataDeps](#) - Reproducible data setup for reproducible science.

Kotlin

Deep Learning

- [KotlinDL](#) - Deep learning framework written in Kotlin.

Lua

General-Purpose Machine Learning

- Torch7
  - cephes - Cephes mathematical functions library, wrapped for Torch. Provides and wraps the 180+ special mathematical functions from the Cephes mathematical library, developed by Stephen L. Moshier. It is used, among many other places, at the heart of SciPy. **[Deprecated]**
  - autograd - Autograd automatically differentiates native Torch code. Inspired by the original Python version.
  - graph - Graph package for Torch. **[Deprecated]**
  - randomkit - Numpy's randomkit, wrapped for Torch. **[Deprecated]**
  - signal - A signal processing toolbox for Torch-7. FFT, DCT, Hilbert, cepstrums, stft.
  - nn - Neural Network package for Torch.
  - torchnet - framework for torch which provides a set of abstractions aiming at encouraging code re-use as well as encouraging modular programming.
  - nngraph - This package provides graphical computation for nn library in Torch7.
  - nnx - A completely unstable and experimental package that extends Torch's builtin nn library.
  - rnn - A Recurrent Neural Network library that extends Torch's nn. RNNs, LSTMs, GRUs, BRNNs, BLSTMs, etc.
  - dpnn - Many useful features that aren't part of the main nn package.
  - dp - A deep learning library designed for streamlining research and development using the Torch7 distribution. It emphasizes flexibility through the elegant use of object-oriented design patterns. **[Deprecated]**
  - optim - An optimization library for Torch. SGD, Adagrad, Conjugate-Gradient, LBFGS, RProp and more.
  - unsup - A package for unsupervised learning in Torch. Provides modules that are compatible with nn (LinearPsd, ConvPsd, AutoEncoder, ...), and self-contained algorithms (k-means, PCA). **[Deprecated]**
  - manifold - A package to manipulate manifolds.
  - svm - Torch-SVM library. **[Deprecated]**
  - lbfgs - FFI Wrapper for liblbfgs. **[Deprecated]**
  - vowpalwabbit - An old vowpalwabbit interface to torch. **[Deprecated]**
  - OpenGM - OpenGM is a C++ library for graphical modelling, and inference. The Lua bindings provide a simple way of describing graphs, from Lua, and then optimizing them with OpenGM. **[Deprecated]**
  - spaghetti - Spaghetti (sparse linear) module for torch7 by @MichaelMathieu **[Deprecated]**
  - LuaSHKit - A Lua wrapper around the Locality sensitive hashing library SHKit **[Deprecated]**
  - kernel smoothing - KNN, kernel-weighted average, local linear regression smoothers. **[Deprecated]**
  - cutorch - Torch CUDA Implementation.
  - cunn - Torch CUDA Neural Network Implementation.
  - imgraph - An image/graph library for Torch. This package provides routines to construct graphs on images, segment them, build trees out of them, and convert them back to images. **[Deprecated]**

- ○ [videograph](#) - A video/graph library for Torch. This package provides routines to construct graphs on videos, segment them, build trees out of them, and convert them back to videos. **[Deprecated]**
  - ○ [saliency](#) - code and tools around integral images. A library for finding interest points based on fast integral histograms. **[Deprecated]**
  - ○ [stitch](#) - allows us to use hugin to stitch images and apply same stitching to a video sequence. **[Deprecated]**
  - ○ [sfm](#) - A bundle adjustment/structure from motion package. **[Deprecated]**
  - ○ [fex](#) - A package for feature extraction in Torch. Provides SIFT and dSIFT modules. **[Deprecated]**
  - ○ [OverFeat](#) - A state-of-the-art generic dense feature extractor. **[Deprecated]**
  - ○ [wav2letter](#) - a simple and efficient end-to-end Automatic Speech Recognition (ASR) system from Facebook AI Research.
- [Numeric Lua](#)
- [Lunatic Python](#)
- [SciLua](#)
- [Lua - Numerical Algorithms](#) **[Deprecated]**
- [Lunum](#) **[Deprecated]**
- [Keras GPT Copilot](#) - A python package that integrates an LLM copilot inside the keras model development workflow.

Demos and Scripts

- [Core torch7 demos repository](#).
  - ○ linear-regression, logistic-regression
  - ○ face detector (training and detection as separate demos)
  - ○ mst-based-segmenter
  - ○ train-a-digit-classifier
  - ○ train-autoencoder
  - ○ optical flow demo
  - ○ train-on-housenumbers
  - ○ train-on-cifar
  - ○ tracking with deep nets
  - ○ kinect demo
  - ○ filter-bank visualization
  - ○ saliency-networks
- [Training a Convnet for the Galaxy-Zoo Kaggle challenge(CUDA demo)](#)
- [torch-datasets](#) - Scripts to load several popular datasets including:
  - ○ BSR 500
  - ○ CIFAR-10
  - ○ COIL

- ○ Street View House Numbers
- ○ MNIST
- ○ NORB
- [Atari2600](#) - Scripts to generate a dataset with static frames from the Arcade Learning Environment.

Matlab

Computer Vision

- [Contourlets](#) - MATLAB source code that implements the contourlet transform and its utility functions.
- [Shearlets](#) - MATLAB code for shearlet transform.
- [Curvelets](#) - The Curvelet transform is a higher dimensional generalization of the Wavelet transform designed to represent images at different scales and different angles.
- [Bandlets](#) - MATLAB code for bandlet transform.
- [mexopencv](#) - Collection and a development kit of MATLAB mex functions for OpenCV library.

Natural Language Processing

- [NLP](#) - A NLP library for Matlab.

General-Purpose Machine Learning

- [Training a deep autoencoder or a classifier on MNIST digits](#) - Training a deep autoencoder or a classifier on MNIST digits[DEEP LEARNING].
- [Convolutional-Recursive Deep Learning for 3D Object Classification](#) - Convolutional-Recursive Deep Learning for 3D Object Classification[DEEP LEARNING].
- [Spider](#) - The spider is intended to be a complete object orientated environment for machine learning in Matlab.
- [LibSVM](#) - A Library for Support Vector Machines.
- [ThunderSVM](#) - An Open-Source SVM Library on GPUs and CPUs
- [LibLinear](#) - A Library for Large Linear Classification.
- [Machine Learning Module](#) - Class on machine w/ PDF, lectures, code
- [Caffe](#) - A deep learning framework developed with cleanliness, readability, and speed in mind.
- [Pattern Recognition Toolbox](#) - A complete object-oriented environment for machine learning in Matlab.
- [Pattern Recognition and Machine Learning](#) - This package contains the matlab implementation of the algorithms described in the book Pattern Recognition and Machine Learning by C. Bishop.
- [Optunity](#) - A library dedicated to automated hyperparameter optimization with a simple, lightweight API to facilitate drop-in replacement of grid search. Optunity is written in Python but interfaces seamlessly with MATLAB.
- [MXNet](#) - Lightweight, Portable, Flexible Distributed/Mobile Deep Learning with Dynamic, Mutation-aware Dataflow Dep Scheduler; for Python, R, Julia, Go, JavaScript and more.

- Machine Learning in MatLab/Octave - Examples of popular machine learning algorithms (neural networks, linear/logistic regressions, K-Means, etc.) with code examples and mathematics behind them being explained.

Data Analysis / Data Visualization

- ParaMonte - A general-purpose MATLAB library for Bayesian data analysis and visualization via serial/parallel Monte Carlo and MCMC simulations. Documentation can be found here.
- matlab_bgl - MatlabBGL is a Matlab package for working with graphs.
- gaimc - Efficient pure-Matlab implementations of graph algorithms to complement MatlabBGL's mex functions.

.NET

Computer Vision

- OpenCVDotNet - A wrapper for the OpenCV project to be used with .NET applications.
- Emgu CV - Cross platform wrapper of OpenCV which can be compiled in Mono to be run on Windows, Linus, Mac OS X, iOS, and Android.
- AForge.NET - Open source C# framework for developers and researchers in the fields of Computer Vision and Artificial Intelligence. Development has now shifted to GitHub.
- Accord.NET - Together with AForge.NET, this library can provide image processing and computer vision algorithms to Windows, Windows RT and Windows Phone. Some components are also available for Java and Android.

Natural Language Processing

- Stanford.NLP for .NET - A full port of Stanford NLP packages to .NET and also available precompiled as a NuGet package.

General-Purpose Machine Learning

- Accord-Framework -The Accord.NET Framework is a complete framework for building machine learning, computer vision, computer audition, signal processing and statistical applications.
- Accord.MachineLearning - Support Vector Machines, Decision Trees, Naive Bayesian models, K-means, Gaussian Mixture models and general algorithms such as Ransac, Cross-validation and Grid-Search for machine-learning applications. This package is part of the Accord.NET Framework.
- DiffSharp - An automatic differentiation (AD) library providing exact and efficient derivatives (gradients, Hessians, Jacobians, directional derivatives, and matrix-free Hessian- and Jacobian- vector products) for machine learning and optimization applications. Operations can be nested to any level, meaning that you can compute exact higher-order derivatives and differentiate functions that are internally making use of differentiation, for applications such as hyperparameter optimization.
- Encog - An advanced neural network and machine learning framework. Encog contains classes to create a wide variety of networks, as well as support classes to normalize and process data for

these neural networks. Encog trains using multithreaded resilient propagation. Encog can also make use of a GPU to further speed processing time. A GUI based workbench is also provided to help model and train neural networks.

- GeneticSharp - Multi-platform genetic algorithm library for .NET Core and .NET Framework. The library has several implementations of GA operators, like: selection, crossover, mutation, reinsertion and termination.
- Infer.NET - Infer.NET is a framework for running Bayesian inference in graphical models. One can use Infer.NET to solve many different kinds of machine learning problems, from standard problems like classification, recommendation or clustering through customized solutions to domain-specific problems. Infer.NET has been used in a wide variety of domains including information retrieval, bioinformatics, epidemiology, vision, and many others.
- ML.NET - ML.NET is a cross-platform open-source machine learning framework which makes machine learning accessible to .NET developers. ML.NET was originally developed in Microsoft Research and evolved into a significant framework over the last decade and is used across many product groups in Microsoft like Windows, Bing, PowerPoint, Excel and more.
- Neural Network Designer - DBMS management system and designer for neural networks. The designer application is developed using WPF, and is a user interface which allows you to design your neural network, query the network, create and configure chat bots that are capable of asking questions and learning from your feedback. The chat bots can even scrape the internet for information to return in their output as well as to use for learning.
- Synapses - Neural network library in F#.
- Vulpes - Deep belief and deep learning implementation written in F# and leverages CUDA GPU execution with Alea.cuBase.
- MxNet.Sharp - .NET Standard bindings for Apache MxNet with Imperative, Symbolic and Gluon Interface for developing, training and deploying Machine Learning models in C#. https://mxnet.tech-quantum.com/

Data Analysis / Data Visualization

- numl - numl is a machine learning library intended to ease the use of using standard modelling techniques for both prediction and clustering.
- Math.NET Numerics - Numerical foundation of the Math.NET project, aiming to provide methods and algorithms for numerical computations in science, engineering and everyday use. Supports .Net 4.0, .Net 3.5 and Mono on Windows, Linux and Mac; Silverlight 5, WindowsPhone/SL 8, WindowsPhone 8.1 and Windows 8 with PCL Portable Profiles 47 and 344; Android/iOS with Xamarin.
- Sho - Sho is an interactive environment for data analysis and scientific computing that lets you seamlessly connect scripts (in IronPython) with compiled code (in .NET) to enable fast and flexible prototyping. The environment includes powerful and efficient libraries for linear algebra as well as data visualization that can be used from any .NET language, as well as a feature-rich interactive shell for rapid development.

Objective C

General-Purpose Machine Learning

- [YCML](#) - A Machine Learning framework for Objective-C and Swift (OS X / iOS).
- [MLPNeuralNet](#) - Fast multilayer perceptron neural network library for iOS and Mac OS X. MLPNeuralNet predicts new examples by trained neural networks. It is built on top of the Apple's Accelerate Framework, using vectorized operations and hardware acceleration if available. **[Deprecated]**
- [MAChineLearning](#) - An Objective-C multilayer perceptron library, with full support for training through backpropagation. Implemented using vDSP and vecLib, it's 20 times faster than its Java equivalent. Includes sample code for use from Swift.
- [BPN-NeuralNetwork](#) - It implemented 3 layers of neural networks (Input Layer, Hidden Layer and Output Layer) and it was named Back Propagation Neural Networks (BPN). This network can be used in products recommendation, user behavior analysis, data mining and data analysis. **[Deprecated]**
- [Multi-Perceptron-NeuralNetwork](#) - It implemented multi-perceptrons neural network (ニューラルネットワーク) based on Back Propagation Neural Networks (BPN) and designed unlimited-hidden-layers.
- [KRHebbian-Algorithm](#) - It is a non-supervisory and self-learning algorithm (adjust the weights) in the neural network of Machine Learning. **[Deprecated]**
- [KRKmeans-Algorithm](#) - It implemented K-Means clustering and classification algorithm. It could be used in data mining and image compression. **[Deprecated]**
- [KRFuzzyCMeans-Algorithm](#) - It implemented Fuzzy C-Means (FCM) the fuzzy clustering / classification algorithm on Machine Learning. It could be used in data mining and image compression. **[Deprecated]**

OCaml

General-Purpose Machine Learning

- [Oml](#) - A general statistics and machine learning library.
- [GPR](#) - Efficient Gaussian Process Regression in OCaml.
- [Libra-Tk](#) - Algorithms for learning and inference with discrete probabilistic models.
- [TensorFlow](#) - OCaml bindings for TensorFlow.

OpenCV

OpenSource-Computer-Vision

- [OpenCV](#) - A OpenSource Computer Vision Library

Perl

Data Analysis / Data Visualization

- Perl Data Language, a pluggable architecture for data and image processing, which can be used for machine learning.

General-Purpose Machine Learning

- MXnet for Deep Learning, in Perl, also released in CPAN.
- Perl Data Language, using AWS machine learning platform from Perl.
- Algorithm::SVMLight, implementation of Support Vector Machines with SVMLight under it. [Deprecated]
- Several machine learning and artificial intelligence models are included in the `AI` namespace. For instance, you can find Naïve Bayes.

Perl 6

- Support Vector Machines
- Naïve Bayes

Data Analysis / Data Visualization

- Perl Data Language, a pluggable architecture for data and image processing, which can be used for machine learning.

General-Purpose Machine Learning

PHP

Natural Language Processing

- jieba-php - Chinese Words Segmentation Utilities.

General-Purpose Machine Learning

- PHP-ML - Machine Learning library for PHP. Algorithms, Cross Validation, Neural Network, Preprocessing, Feature Extraction and much more in one library.
- PredictionBuilder - A library for machine learning that builds predictions using a linear regression.
- Rubix ML - A high-level machine learning (ML) library that lets you build programs that learn from data using the PHP language.
- 19 Questions - A machine learning / bayesian inference assigning attributes to objects.

Python

Computer Vision

- Scikit-Image - A collection of algorithms for image processing in Python.

- Scikit-Opt - Swarm Intelligence in Python (Genetic Algorithm, Particle Swarm Optimization, Simulated Annealing, Ant Colony Algorithm, Immune Algorithm, Artificial Fish Swarm Algorithm in Python)
- SimpleCV - An open source computer vision framework that gives access to several high-powered computer vision libraries, such as OpenCV. Written on Python and runs on Mac, Windows, and Ubuntu Linux.
- Vigranumpy - Python bindings for the VIGRA C++ computer vision library.
- OpenFace - Free and open source face recognition with deep neural networks.
- PCV - Open source Python module for computer vision. **[Deprecated]**
- face_recognition - Face recognition library that recognizes and manipulates faces from Python or from the command line.
- deepface - A lightweight face recognition and facial attribute analysis (age, gender, emotion and race) framework for Python covering cutting-edge models such as VGG-Face, FaceNet, OpenFace, DeepFace, DeepID, Dlib and ArcFace.
- retinaface - deep learning based cutting-edge facial detector for Python coming with facial landmarks
- dockerface - Easy to install and use deep learning Faster R-CNN face detection for images and video in a docker container. **[Deprecated]**
- Detectron - FAIR's software system that implements state-of-the-art object detection algorithms, including Mask R-CNN. It is written in Python and powered by the Caffe2 deep learning framework. **[Deprecated]**
- detectron2 - FAIR's next-generation research platform for object detection and segmentation. It is a ground-up rewrite of the previous version, Detectron, and is powered by the PyTorch deep learning framework.
- albumentations - A fast and framework agnostic image augmentation library that implements a diverse set of augmentation techniques. Supports classification, segmentation, detection out of the box. Was used to win a number of Deep Learning competitions at Kaggle, Topcoder and those that were a part of the CVPR workshops.
- pytessarct - Python-tesseract is an optical character recognition (OCR) tool for python. That is, it will recognize and "read" the text embedded in images. Python-tesseract is a wrapper for Google's Tesseract-OCR Engine.
- imutils - A library containing Convenience functions to make basic image processing operations such as translation, rotation, resizing, skeletonization, and displaying Matplotlib images easier with OpenCV and Python.
- PyTorchCV - A PyTorch-Based Framework for Deep Learning in Computer Vision.
- joliGEN - Generative AI Image Toolset with GANs and Diffusion for Real-World Applications.
- Self-supervised learning
- neural-style-pt - A PyTorch implementation of Justin Johnson's neural-style (neural style transfer).
- Detecto - Train and run a computer vision model with 5-10 lines of code.
- neural-dream - A PyTorch implementation of DeepDream.
- Openpose - A real-time multi-person keypoint detection library for body, face, hands, and foot estimation

- [Deep High-Resolution-Net](#) - A PyTorch implementation of CVPR2019 paper "Deep High-Resolution Representation Learning for Human Pose Estimation"
- [TF-GAN](#) - TF-GAN is a lightweight library for training and evaluating Generative Adversarial Networks (GANs).
- [dream-creator](#) - A PyTorch implementation of DeepDream. Allows individuals to quickly and easily train their own custom GoogleNet models with custom datasets for DeepDream.
- [Lucent](#) - Tensorflow and OpenAI Clarity's Lucid adapted for PyTorch.
- [lightly](#) - Lightly is a computer vision framework for self-supervised learning.
- [Learnergy](#) - Energy-based machine learning models built upon PyTorch.
- [OpenVisionAPI](#) - Open source computer vision API based on open source models.
- [IoT Owl](#) - Light face detection and recognition system with huge possibilities, based on Microsoft Face API and TensorFlow made for small IoT devices like raspberry pi.
- [Exadel CompreFace](#) - face recognition system that can be easily integrated into any system without prior machine learning skills. CompreFace provides REST API for face recognition, face verification, face detection, face mask detection, landmark detection, age, and gender recognition and is easily deployed with docker.
- [computer-vision-in-action](#) - as known as `L0CV`, is a new generation of computer vision open source online learning media, a cross-platform interactive learning framework integrating graphics, source code and HTML. the L0CV ecosystem — Notebook, Datasets, Source Code, and from Diving-in to Advanced — as well as the L0CV Hub.
- [timm](#) - PyTorch image models, scripts, pretrained weights -- ResNet, ResNeXT, EfficientNet, EfficientNetV2, NFNet, Vision Transformer, MixNet, MobileNet-V3/V2, RegNet, DPN, CSPNet, and more.
- [segmentation_models.pytorch](#) - A PyTorch-based toolkit that offers pre-trained segmentation models for computer vision tasks. It simplifies the development of image segmentation applications by providing a collection of popular architecture implementations, such as UNet and PSPNet, along with pre-trained weights, making it easier for researchers and developers to achieve high-quality pixel-level object segmentation in images.
- [segmentation_models](#) - A TensorFlow Keras-based toolkit that offers pre-trained segmentation models for computer vision tasks. It simplifies the development of image segmentation applications by providing a collection of popular architecture implementations, such as UNet and PSPNet, along with pre-trained weights, making it easier for researchers and developers to achieve high-quality pixel-level object segmentation in images.
- [MLX](#)- MLX is an array framework for machine learning on Apple silicon, developed by Apple machine learning research.

Natural Language Processing

- [pkuseg-python](#) - A better version of Jieba, developed by Peking University.
- [NLTK](#) - A leading platform for building Python programs to work with human language data.
- [Pattern](#) - A web mining module for the Python programming language. It has tools for natural language processing, machine learning, among others.
- [Quepy](#) - A python framework to transform natural language questions to queries in a database

query language.

- TextBlob - Providing a consistent API for diving into common natural language processing (NLP) tasks. Stands on the giant shoulders of NLTK and Pattern, and plays nicely with both.
- YAlign - A sentence aligner, a friendly tool for extracting parallel sentences from comparable corpora. **[Deprecated]**
- jieba - Chinese Words Segmentation Utilities.
- SnowNLP - A library for processing Chinese text.
- spammy - A library for email Spam filtering built on top of NLTK
- loso - Another Chinese segmentation library. **[Deprecated]**
- genius - A Chinese segment based on Conditional Random Field.
- KoNLPy - A Python package for Korean natural language processing.
- nut - Natural language Understanding Toolkit. **[Deprecated]**
- Rosetta - Text processing tools and wrappers (e.g. Vowpal Wabbit)
- BLLIP Parser - Python bindings for the BLLIP Natural Language Parser (also known as the Charniak-Johnson parser). **[Deprecated]**
- PyNLPl - Python Natural Language Processing Library. General purpose NLP library for Python. Also contains some specific modules for parsing common NLP formats, most notably for FoLiA, but also ARPA language models, Moses phrasetables, GIZA++ alignments.
- PySS3 - Python package that implements a novel white-box machine learning model for text classification, called SS3. Since SS3 has the ability to visually explain its rationale, this package also comes with easy-to-use interactive visualizations tools (online demos).
- python-ucto - Python binding to ucto (a unicode-aware rule-based tokenizer for various languages).
- python-frog - Python binding to Frog, an NLP suite for Dutch. (pos tagging, lemmatisation, dependency parsing, NER)
- python-zpar - Python bindings for ZPar, a statistical part-of-speech-tagger, constituency parser, and dependency parser for English.
- colibri-core - Python binding to C++ library for extracting and working with basic linguistic constructions such as n-grams and skipgrams in a quick and memory-efficient way.
- spaCy - Industrial strength NLP with Python and Cython.
- PyStanfordDependencies - Python interface for converting Penn Treebank trees to Stanford Dependencies.
- Distance - Levenshtein and Hamming distance computation. **[Deprecated]**
- Fuzzy Wuzzy - Fuzzy String Matching in Python.
- Neofuzz - Blazing fast, lightweight and customizable fuzzy and semantic text search in Python with fuzzywuzzy/thefuzz compatible API.
- jellyfish - a python library for doing approximate and phonetic matching of strings.
- editdistance - fast implementation of edit distance.
- textacy - higher-level NLP built on Spacy.
- stanford-corenlp-python - Python wrapper for Stanford CoreNLP **[Deprecated]**

- [CLTK](#) - The Classical Language Toolkit.
- [Rasa](#) - A "machine learning framework to automate text-and voice-based conversations."
- [yase](#) - Transcode sentence (or other sequence) to list of word vector .
- [Polyglot](#) - Multilingual text (NLP) processing toolkit.
- [DrQA](#) - Reading Wikipedia to answer open-domain questions.
- [Dedupe](#) - A python library for accurate and scalable fuzzy matching, record deduplication and entity-resolution.
- [Snips NLU](#) - Natural Language Understanding library for intent classification and entity extraction
- [NeuroNER](#) - Named-entity recognition using neural networks providing state-of-the-art-results
- [DeepPavlov](#) - conversational AI library with many pre-trained Russian NLP models.
- [BigARTM](#) - topic modelling platform.
- [NALP](#) - A Natural Adversarial Language Processing framework built over Tensorflow.
- [DL Translate](#) - A deep learning-based translation library between 50 languages, built with `transformers` .
- [Haystack](#) - A framework for building industrial-strength applications with Transformer models and LLMs.
- [CometLLM](#) - Track, log, visualize and evaluate your LLM prompts and prompt chains.
- [Transformers](#) - A deep learning library containing thousands of pre-trained models on different tasks. The goto place for anything related to Large Language Models.

General-Purpose Machine Learning

- [XAD](#) -> Fast and easy-to-use backpropagation tool.
- [Aim](#) -> An easy-to-use & supercharged open-source AI metadata tracker.
- [RexMex](#) -> A general purpose recommender metrics library for fair evaluation.
- [ChemicalX](#) -> A PyTorch based deep learning library for drug pair scoring
- [Microsoft ML for Apache Spark](#) -> A distributed machine learning framework Apache Spark
- [Shapley](#) -> A data-driven framework to quantify the value of classifiers in a machine learning ensemble.
- [igel](#) -> A delightful machine learning tool that allows you to train/fit, test and use models **without writing code**
- [ML Model building](#) -> A Repository Containing Classification, Clustering, Regression, Recommender Notebooks with illustration to make them.
- [ML/DL project template](#)
- [PyTorch Frame](#) -> A Modular Framework for Multi-Modal Tabular Learning.
- [PyTorch Geometric](#) -> Graph Neural Network Library for PyTorch.
- [PyTorch Geometric Temporal](#) -> A temporal extension of PyTorch Geometric for dynamic graph representation learning.
- [Little Ball of Fur](#) -> A graph sampling extension library for NetworkX with a Scikit-Learn like API.
- [Karate Club](#) -> An unsupervised machine learning extension library for NetworkX with a Scikit-Learn like API.

- Auto_ViML -> Automatically Build Variant Interpretable ML models fast! Auto_ViML is pronounced "auto vimal", is a comprehensive and scalable Python AutoML toolkit with imbalanced handling, ensembling, stacking and built-in feature selection. Featured in Medium article.
- PyOD -> Python Outlier Detection, comprehensive and scalable Python toolkit for detecting outlying objects in multivariate data. Featured for Advanced models, including Neural Networks/Deep Learning and Outlier Ensembles.
- steppy -> Lightweight, Python library for fast and reproducible machine learning experimentation. Introduces a very simple interface that enables clean machine learning pipeline design.
- steppy-toolkit -> Curated collection of the neural networks, transformers and models that make your machine learning work faster and more effective.
- CNTK - Microsoft Cognitive Toolkit (CNTK), an open source deep-learning toolkit. Documentation can be found here.
- Couler - Unified interface for constructing and managing machine learning workflows on different workflow engines, such as Argo Workflows, Tekton Pipelines, and Apache Airflow.
- auto_ml - Automated machine learning for production and analytics. Lets you focus on the fun parts of ML, while outputting production-ready code, and detailed analytics of your dataset and results. Includes support for NLP, XGBoost, CatBoost, LightGBM, and soon, deep learning.
- dtaidistance - High performance library for time series distances (DTW) and time series clustering.
- einops - Deep learning operations reinvented (for pytorch, tensorflow, jax and others).
- machine learning - automated build consisting of a web-interface, and set of programmatic-interface API, for support vector machines. Corresponding dataset(s) are stored into a SQL database, then generated model(s) used for prediction(s), are stored into a NoSQL datastore.
- XGBoost - Python bindings for eXtreme Gradient Boosting (Tree) Library.
- ChefBoost - a lightweight decision tree framework for Python with categorical feature support covering regular decision tree algorithms such as ID3, C4.5, CART, CHAID and regression tree; also some advanved bagging and boosting techniques such as gradient boosting, random forest and adaboost.
- Apache SINGA - An Apache Incubating project for developing an open source machine learning library.
- Bayesian Methods for Hackers - Book/iPython notebooks on Probabilistic Programming in Python.
- Featureforge A set of tools for creating and testing machine learning features, with a scikit-learn compatible API.
- MLlib in Apache Spark - Distributed machine learning library in Spark
- Hydrosphere Mist - A service for deployment Apache Spark MLLib machine learning models as realtime, batch or reactive web services.
- Towhee - A Python module that encode unstructured data into embeddings.
- scikit-learn - A Python module for machine learning built on top of SciPy.
- metric-learn - A Python module for metric learning.
- OpenMetricLearning - A PyTorch-based framework to train and validate the models producing high-quality embeddings.
- Intel(R) Extension for Scikit-learn - A seamless way to speed up your Scikit-learn applications with

no accuracy loss and code changes.
- SimpleAI Python implementation of many of the artificial intelligence algorithms described in the book "Artificial Intelligence, a Modern Approach". It focuses on providing an easy to use, well documented and tested library.
- astroML - Machine Learning and Data Mining for Astronomy.
- graphlab-create - A library with various machine learning models (regression, clustering, recommender systems, graph analytics, etc.) implemented on top of a disk-backed DataFrame.
- BigML - A library that contacts external servers.
- pattern - Web mining module for Python.
- NuPIC - Numenta Platform for Intelligent Computing.
- Pylearn2 - A Machine Learning library based on Theano. **[Deprecated]**
- keras - High-level neural networks frontend for TensorFlow, CNTK and Theano.
- Lasagne - Lightweight library to build and train neural networks in Theano.
- hebel - GPU-Accelerated Deep Learning Library in Python. **[Deprecated]**
- Chainer - Flexible neural network framework.
- prophet - Fast and automated time series forecasting framework by Facebook.
- gensim - Topic Modelling for Humans.
- tweetopic - Blazing fast short-text-topic-modelling for Python.
- topicwizard - Interactive topic model visualization/interpretation framework.
- topik - Topic modelling toolkit. **[Deprecated]**
- PyBrain - Another Python Machine Learning Library.
- Brainstorm - Fast, flexible and fun neural networks. This is the successor of PyBrain.
- Surprise - A scikit for building and analyzing recommender systems.
- implicit - Fast Python Collaborative Filtering for Implicit Datasets.
- LightFM - A Python implementation of a number of popular recommendation algorithms for both implicit and explicit feedback.
- Crab - A flexible, fast recommender engine. **[Deprecated]**
- python-recsys - A Python library for implementing a Recommender System.
- thinking bayes - Book on Bayesian Analysis.
- Image-to-Image Translation with Conditional Adversarial Networks - Implementation of image to image (pix2pix) translation from the paper by isola et al.[DEEP LEARNING]
- Restricted Boltzmann Machines -Restricted Boltzmann Machines in Python. [DEEP LEARNING]
- Bolt - Bolt Online Learning Toolbox. **[Deprecated]**
- CoverTree - Python implementation of cover trees, near-drop-in replacement for scipy.spatial.kdtree **[Deprecated]**
- nilearn - Machine learning for NeuroImaging in Python.
- neuropredict - Aimed at novice machine learners and non-expert programmers, this package offers easy (no coding needed) and comprehensive machine learning (evaluation and full report of predictive performance WITHOUT requiring you to code) in Python for NeuroImaging and any other type of features. This is aimed at absorbing much of the ML workflow, unlike other packages

like nilearn and pymvpa, which require you to learn their API and code to produce anything useful.

- **imbalanced-learn** - Python module to perform under sampling and oversampling with various techniques.
- **imbalanced-ensemble** - Python toolbox for quick implementation, modification, evaluation, and visualization of ensemble learning algorithms for class-imbalanced data. Supports out-of-the-box multi-class imbalanced (long-tailed) classification.
- **Shogun** - The Shogun Machine Learning Toolbox.
- **Pyevolve** - Genetic algorithm framework. **[Deprecated]**
- **Caffe** - A deep learning framework developed with cleanliness, readability, and speed in mind.
- **breze** - Theano based library for deep and recurrent neural networks.
- **Cortex** - Open source platform for deploying machine learning models in production.
- **pyhsmm** - library for approximate unsupervised inference in Bayesian Hidden Markov Models (HMMs) and explicit-duration Hidden semi-Markov Models (HSMMs), focusing on the Bayesian Nonparametric extensions, the HDP-HMM and HDP-HSMM, mostly with weak-limit approximations.
- **SKLL** - A wrapper around scikit-learn that makes it simpler to conduct experiments.
- **neurolab**
- **Spearmint** - Spearmint is a package to perform Bayesian optimization according to the algorithms outlined in the paper: Practical Bayesian Optimization of Machine Learning Algorithms. Jasper Snoek, Hugo Larochelle and Ryan P. Adams. Advances in Neural Information Processing Systems, 2012. **[Deprecated]**
- **Pebl** - Python Environment for Bayesian Learning. **[Deprecated]**
- **Theano** - Optimizing GPU-meta-programming code generating array oriented optimizing math compiler in Python.
- **TensorFlow** - Open source software library for numerical computation using data flow graphs.
- **pomegranate** - Hidden Markov Models for Python, implemented in Cython for speed and efficiency.
- **python-timbl** - A Python extension module wrapping the full TiMBL C++ programming interface. Timbl is an elaborate k-Nearest Neighbours machine learning toolkit.
- **deap** - Evolutionary algorithm framework.
- **pydeep** - Deep Learning In Python. **[Deprecated]**
- **mlxtend** - A library consisting of useful tools for data science and machine learning tasks.
- **neon** - Nervana's **high-performance** Python-based Deep Learning framework [DEEP LEARNING]. **[Deprecated]**
- **Optunity** - A library dedicated to automated hyperparameter optimization with a simple, lightweight API to facilitate drop-in replacement of grid search.
- **Neural Networks and Deep Learning** - Code samples for my book "Neural Networks and Deep Learning" [DEEP LEARNING].
- **Annoy** - Approximate nearest neighbours implementation.
- **TPOT** - Tool that automatically creates and optimizes machine learning pipelines using genetic programming. Consider it your personal data science assistant, automating a tedious part of

machine learning.

- pgmpy A python library for working with Probabilistic Graphical Models.
- DIGITS - The Deep Learning GPU Training System (DIGITS) is a web application for training deep learning models.
- Orange - Open source data visualization and data analysis for novices and experts.
- MXNet - Lightweight, Portable, Flexible Distributed/Mobile Deep Learning with Dynamic, Mutation-aware Dataflow Dep Scheduler; for Python, R, Julia, Go, JavaScript and more.
- milk - Machine learning toolkit focused on supervised classification. **[Deprecated]**
- TFLearn - Deep learning library featuring a higher-level API for TensorFlow.
- REP - an IPython-based environment for conducting data-driven research in a consistent and reproducible way. REP is not trying to substitute scikit-learn, but extends it and provides better user experience. **[Deprecated]**
- rgf_python - Python bindings for Regularized Greedy Forest (Tree) Library.
- skbayes - Python package for Bayesian Machine Learning with scikit-learn API.
- fuku-ml - Simple machine learning library, including Perceptron, Regression, Support Vector Machine, Decision Tree and more, it's easy to use and easy to learn for beginners.
- Xcessiv - A web-based application for quick, scalable, and automated hyperparameter tuning and stacked ensembling.
- PyTorch - Tensors and Dynamic neural networks in Python with strong GPU acceleration
- PyTorch Lightning - The lightweight PyTorch wrapper for high-performance AI research.
- PyTorch Lightning Bolts - Toolbox of models, callbacks, and datasets for AI/ML researchers.
- skorch - A scikit-learn compatible neural network library that wraps PyTorch.
- ML-From-Scratch - Implementations of Machine Learning models from scratch in Python with a focus on transparency. Aims to showcase the nuts and bolts of ML in an accessible way.
- Edward - A library for probabilistic modelling, inference, and criticism. Built on top of TensorFlow.
- xRBM - A library for Restricted Boltzmann Machine (RBM) and its conditional variants in Tensorflow.
- CatBoost - General purpose gradient boosting on decision trees library with categorical features support out of the box. It is easy to install, well documented and supports CPU and GPU (even multi-GPU) computation.
- stacked_generalization - Implementation of machine learning stacking technique as a handy library in Python.
- modAL - A modular active learning framework for Python, built on top of scikit-learn.
- Cogitare: A Modern, Fast, and Modular Deep Learning and Machine Learning framework for Python.
- Parris - Parris, the automated infrastructure setup tool for machine learning algorithms.
- neonrvm - neonrvm is an open source machine learning library based on RVM technique. It's written in C programming language and comes with Python programming language bindings.
- Turi Create - Machine learning from Apple. Turi Create simplifies the development of custom machine learning models. You don't have to be a machine learning expert to add recommendations, object detection, image classification, image similarity or activity classification

to your app.

- [xLearn](#) - A high performance, easy-to-use, and scalable machine learning package, which can be used to solve large-scale machine learning problems. xLearn is especially useful for solving machine learning problems on large-scale sparse data, which is very common in Internet services such as online advertisement and recommender systems.
- [mlens](#) - A high performance, memory efficient, maximally parallelized ensemble learning, integrated with scikit-learn.
- [Thampi](#) - Machine Learning Prediction System on AWS Lambda
- [MindsDB](#) - Open Source framework to streamline use of neural networks.
- [Microsoft Recommenders](#): Examples and best practices for building recommendation systems, provided as Jupyter notebooks. The repo contains some of the latest state of the art algorithms from Microsoft Research as well as from other companies and institutions.
- [StellarGraph](#): Machine Learning on Graphs, a Python library for machine learning on graph-structured (network-structured) data.
- [BentoML](#): Toolkit for package and deploy machine learning models for serving in production
- [MiraiML](#): An asynchronous engine for continuous & autonomous machine learning, built for real-time usage.
- [numpy-ML](#): Reference implementations of ML models written in numpy
- [Neuraxle](#): A framework providing the right abstractions to ease research, development, and deployment of your ML pipelines.
- [Cornac](#) - A comparative framework for multimodal recommender systems with a focus on models leveraging auxiliary data.
- [JAX](#) - JAX is Autograd and XLA, brought together for high-performance machine learning research.
- [Catalyst](#) - High-level utils for PyTorch DL & RL research. It was developed with a focus on reproducibility, fast experimentation and code/ideas reusing. Being able to research/develop something new, rather than write another regular train loop.
- [Fastai](#) - High-level wrapper built on the top of Pytorch which supports vision, text, tabular data and collaborative filtering.
- [scikit-multiflow](#) - A machine learning framework for multi-output/multi-label and stream data.
- [Lightwood](#) - A Pytorch based framework that breaks down machine learning problems into smaller blocks that can be glued together seamlessly with objective to build predictive models with one line of code.
- [bayeso](#) - A simple, but essential Bayesian optimization package, written in Python.
- [mljar-supervised](#) - An Automated Machine Learning (AutoML) python package for tabular data. It can handle: Binary Classification, MultiClass Classification and Regression. It provides explanations and markdown reports.
- [evostra](#) - A fast Evolution Strategy implementation in Python.
- [Determined](#) - Scalable deep learning training platform, including integrated support for distributed training, hyperparameter tuning, experiment tracking, and model management.
- [PySyft](#) - A Python library for secure and private Deep Learning built on PyTorch and TensorFlow.
- [PyGrid](#) - Peer-to-peer network of data owners and data scientists who can collectively train AI

models using PySyft

- sktime - A unified framework for machine learning with time series
- OPFython - A Python-inspired implementation of the Optimum-Path Forest classifier.
- Opytimizer - Python-based meta-heuristic optimization techniques.
- Gradio - A Python library for quickly creating and sharing demos of models. Debug models interactively in your browser, get feedback from collaborators, and generate public links without deploying anything.
- Hub - Fastest unstructured dataset management for TensorFlow/PyTorch. Stream & version-control data. Store even petabyte-scale data in a single numpy-like array on the cloud accessible on any machine. Visit activeloop.ai for more info.
- Synthia - Multidimensional synthetic data generation in Python.
- ByteHub - An easy-to-use, Python-based feature store. Optimized for time-series data.
- Backprop - Backprop makes it simple to use, finetune, and deploy state-of-the-art ML models.
- River: A framework for general purpose online machine learning.
- FEDOT: An AutoML framework for the automated design of composite modelling pipelines. It can handle classification, regression, and time series forecasting tasks on different types of data (including multi-modal datasets).
- Sklearn-genetic-opt: An AutoML package for hyperparameters tuning using evolutionary algorithms, with built-in callbacks, plotting, remote logging and more.
- Evidently: Interactive reports to analyze machine learning models during validation or production monitoring.
- Streamlit: Streamlit is an framework to create beautiful data apps in hours, not weeks.
- Optuna: Optuna is an automatic hyperparameter optimization software framework, particularly designed for machine learning.
- Deepchecks: Validation & testing of machine learning models and data during model development, deployment, and production. This includes checks and suites related to various types of issues, such as model performance, data integrity, distribution mismatches, and more.
- Shapash : Shapash is a Python library that provides several types of visualization that display explicit labels that everyone can understand.
- Eurybia: Eurybia monitors data and model drift over time and securizes model deployment with data validation.
- Colossal-AI: An open-source deep learning system for large-scale model training and inference with high efficiency and low cost.
- dirty_cat - facilitates machine-learning on dirty, non-curated categories. It provides transformers and encoders robust to morphological variants, such as typos.
- Upgini: Free automated data & feature enrichment library for machine learning - automatically searches through thousands of ready-to-use features from public and community shared data sources and enriches your training dataset with only the accuracy improving features.
- AutoML-Implementation-for-Static-and-Dynamic-Data-Analytics: A tutorial to help machine learning researchers to automatically obtain optimized machine learning models with the optimal learning performance on any specific task.

- SKBEL: A Python library for Bayesian Evidential Learning (BEL) in order to estimate the uncertainty of a prediction.
- NannyML: Python library capable of fully capturing the impact of data drift on performance. Allows estimation of post-deployment model performance without access to targets.
- cleanlab: The standard data-centric AI package for data quality and machine learning with messy, real-world data and labels.
- AutoGluon: AutoML for Image, Text, Tabular, Time-Series, and MultiModal Data.
- PyBroker - Algorithmic Trading with Machine Learning.
- Frouros: Frouros is an open source Python library for drift detection in machine learning systems.
- CometML: The best-in-class MLOps platform with experiment tracking, model production monitoring, a model registry, and data lineage from training straight through to production.

Data Analysis / Data Visualization

- DataComPy - A library to compare Pandas, Polars, and Spark data frames. It provides stats and lets users adjust for match accuracy.
- DataVisualization - A GitHub Repository Where you can Learn Datavisualizatoin Basics to Intermediate level.
- Cartopy - Cartopy is a Python package designed for geospatial data processing in order to produce maps and other geospatial data analyses.
- SciPy - A Python-based ecosystem of open-source software for mathematics, science, and engineering.
- NumPy - A fundamental package for scientific computing with Python.
- AutoViz AutoViz performs automatic visualization of any dataset with a single line of Python code. Give it any input file (CSV, txt or JSON) of any size and AutoViz will visualize it. See Medium article.
- Numba - Python JIT (just in time) compiler to LLVM aimed at scientific Python by the developers of Cython and NumPy.
- Mars - A tensor-based framework for large-scale data computation which is often regarded as a parallel and distributed version of NumPy.
- NetworkX - A high-productivity software for complex networks.
- igraph - binding to igraph library - General purpose graph library.
- Pandas - A library providing high-performance, easy-to-use data structures and data analysis tools.
- ParaMonte - A general-purpose Python library for Bayesian data analysis and visualization via serial/parallel Monte Carlo and MCMC simulations. Documentation can be found here.
- Vaex - A high performance Python library for lazy Out-of-Core DataFrames (similar to Pandas), to visualize and explore big tabular datasets. Documentation can be found here.
- Open Mining - Business Intelligence (BI) in Python (Pandas web interface) **[Deprecated]**
- PyMC - Markov Chain Monte Carlo sampling toolkit.
- zipline - A Pythonic algorithmic trading library.
- PyDy - Short for Python Dynamics, used to assist with workflow in the modelling of dynamic motion based around NumPy, SciPy, IPython, and matplotlib.

- SymPy - A Python library for symbolic mathematics.
- statsmodels - Statistical modelling and econometrics in Python.
- astropy - A community Python library for Astronomy.
- matplotlib - A Python 2D plotting library.
- bokeh - Interactive Web Plotting for Python.
- plotly - Collaborative web plotting for Python and matplotlib.
- altair - A Python to Vega translator.
- d3py - A plotting library for Python, based on D3.js.
- PyDexter - Simple plotting for Python. Wrapper for D3xterjs; easily render charts in-browser.
- ggplot - Same API as ggplot2 for R. **[Deprecated]**
- ggfortify - Unified interface to ggplot2 popular R packages.
- Kartograph.py - Rendering beautiful SVG maps in Python.
- pygal - A Python SVG Charts Creator.
- PyQtGraph - A pure-python graphics and GUI library built on PyQt4 / PySide and NumPy.
- pycascading **[Deprecated]**
- Petrel - Tools for writing, submitting, debugging, and monitoring Storm topologies in pure Python.
- Blaze - NumPy and Pandas interface to Big Data.
- emcee - The Python ensemble sampling toolkit for affine-invariant MCMC.
- windML - A Python Framework for Wind Energy Analysis and Prediction.
- vispy - GPU-based high-performance interactive OpenGL 2D/3D data visualization library.
- cerebro2 A web-based visualization and debugging platform for NuPIC. **[Deprecated]**
- NuPIC Studio An all-in-one NuPIC Hierarchical Temporal Memory visualization and debugging super-tool! **[Deprecated]**
- SparklingPandas Pandas on PySpark (POPS).
- Seaborn - A python visualization library based on matplotlib.
- ipychart - The power of Chart.js in Jupyter Notebook.
- bqplot - An API for plotting in Jupyter (IPython).
- pastalog - Simple, realtime visualization of neural network training performance.
- Superset - A data exploration platform designed to be visual, intuitive, and interactive.
- Dora - Tools for exploratory data analysis in Python.
- Ruffus - Computation Pipeline library for python.
- SOMPY - Self Organizing Map written in Python (Uses neural networks for data analysis).
- somoclu Massively parallel self-organizing maps: accelerate training on multicore CPUs, GPUs, and clusters, has python API.
- HDBScan - implementation of the hdbscan algorithm in Python - used for clustering
- visualize_ML - A python package for data exploration and data analysis. **[Deprecated]**
- scikit-plot - A visualization library for quick and easy generation of common plots in data analysis and machine learning.
- Bowtie - A dashboard library for interactive visualizations using flask socketio and react.

- [lime](#) - Lime is about explaining what machine learning classifiers (or models) are doing. It is able to explain any black box classifier, with two or more classes.
- [PyCM](#) - PyCM is a multi-class confusion matrix library written in Python that supports both input data vectors and direct matrix, and a proper tool for post-classification model evaluation that supports most classes and overall statistics parameters
- [Dash](#) - A framework for creating analytical web applications built on top of Plotly.js, React, and Flask
- [Lambdo](#) - A workflow engine for solving machine learning problems by combining in one analysis pipeline (i) feature engineering and machine learning (ii) model training and prediction (iii) table population and column evaluation via user-defined (Python) functions.
- [TensorWatch](#) - Debugging and visualization tool for machine learning and data science. It extensively leverages Jupyter Notebook to show real-time visualizations of data in running processes such as machine learning training.
- [dowel](#) - A little logger for machine learning research. Output any object to the terminal, CSV, TensorBoard, text logs on disk, and more with just one call to `logger.log()`.

Misc Scripts / iPython Notebooks / Codebases

- [MiniGrad](#) – A minimal, educational, Pythonic implementation of autograd (~100 loc).
- [Map/Reduce implementations of common ML algorithms](#): Jupyter notebooks that cover how to implement from scratch different ML algorithms (ordinary least squares, gradient descent, k-means, alternating least squares), using Python NumPy, and how to then make these implementations scalable using Map/Reduce and Spark.
- [BioPy](#) - Biologically-Inspired and Machine Learning Algorithms in Python. **[Deprecated]**
- [CAEs for Data Assimilation](#) - Convolutional autoencoders for 3D image/field compression applied to reduced order [Data Assimilation](#).
- [handsonml](#) - Fundamentals of machine learning in python.
- [SVM Explorer](#) - Interactive SVM Explorer, using Dash and scikit-learn
- [pattern_classification](#)
- [thinking stats 2](#)
- [hyperopt](#)
- [numpic](#)
- [2012-paper-diginorm](#)
- [A gallery of interesting IPython notebooks](#)
- [ipython-notebooks](#)
- [data-science-ipython-notebooks](#) - Continually updated Data Science Python Notebooks: Spark, Hadoop MapReduce, HDFS, AWS, Kaggle, scikit-learn, matplotlib, pandas, NumPy, SciPy, and various command lines.
- [decision-weights](#)
- [Sarah Palin LDA](#) - Topic Modelling the Sarah Palin emails.
- [Diffusion Segmentation](#) - A collection of image segmentation algorithms based on diffusion methods.

- [Scipy Tutorials](#) - SciPy tutorials. This is outdated, check out scipy-lecture-notes.
- [Crab](#) - A recommendation engine library for Python.
- [BayesPy](#) - Bayesian Inference Tools in Python.
- [scikit-learn tutorials](#) - Series of notebooks for learning scikit-learn.
- [sentiment-analyzer](#) - Tweets Sentiment Analyzer
- [sentiment_classifier](#) - Sentiment classifier using word sense disambiguation.
- [group-lasso](#) - Some experiments with the coordinate descent algorithm used in the (Sparse) Group Lasso model.
- [jProcessing](#) - Kanji / Hiragana / Katakana to Romaji Converter. Edict Dictionary & parallel sentences Search. Sentence Similarity between two JP Sentences. Sentiment Analysis of Japanese Text. Run Cabocha(ISO--8859-1 configured) in Python.
- [mne-python-notebooks](#) - IPython notebooks for EEG/MEG data processing using mne-python.
- [Neon Course](#) - IPython notebooks for a complete course around understanding Nervana's Neon.
- [pandas cookbook](#) - Recipes for using Python's pandas library.
- [climin](#) - Optimization library focused on machine learning, pythonic implementations of gradient descent, LBFGS, rmsprop, adadelta and others.
- [Allen Downey's Data Science Course](#) - Code for Data Science at Olin College, Spring 2014.
- [Allen Downey's Think Bayes Code](#) - Code repository for Think Bayes.
- [Allen Downey's Think Complexity Code](#) - Code for Allen Downey's book Think Complexity.
- [Allen Downey's Think OS Code](#) - Text and supporting code for Think OS: A Brief Introduction to Operating Systems.
- [Python Programming for the Humanities](#) - Course for Python programming for the Humanities, assuming no prior knowledge. Heavy focus on text processing / NLP.
- [GreatCircle](#) - Library for calculating great circle distance.
- [Optunity examples](#) - Examples demonstrating how to use Optunity in synergy with machine learning libraries.
- [Dive into Machine Learning with Python Jupyter notebook and scikit-learn](#) - "I learned Python by hacking first, and getting serious *later.* I wanted to do this with Machine Learning. If this is your style, join me in getting a bit ahead of yourself."
- [TDB](#) - TensorDebugger (TDB) is a visual debugger for deep learning. It features interactive, node-by-node debugging and visualization for TensorFlow.
- [Suiron](#) - Machine Learning for RC Cars.
- [Introduction to machine learning with scikit-learn](#) - IPython notebooks from Data School's video tutorials on scikit-learn.
- [Practical XGBoost in Python](#) - comprehensive online course about using XGBoost in Python.
- [Introduction to Machine Learning with Python](#) - Notebooks and code for the book "Introduction to Machine Learning with Python"
- [Pydata book](#) - Materials and IPython notebooks for "Python for Data Analysis" by Wes McKinney, published by O'Reilly Media
- [Homemade Machine Learning](#) - Python examples of popular machine learning algorithms with

interactive Jupyter demos and math being explained
- [Prodmodel](#) - Build tool for data science pipelines.
- [the-elements-of-statistical-learning](#) - This repository contains Jupyter notebooks implementing the algorithms found in the book and summary of the textbook.
- [Hyperparameter-Optimization-of-Machine-Learning-Algorithms](#) - Code for hyperparameter tuning/optimization of machine learning and deep learning algorithms.
- [Heart_Disease-Prediction](#) - Given clinical parameters about a patient, can we predict whether or not they have heart disease?
- [Flight Fare Prediction](#) - This basically to gauge the understanding of Machine Learning Workflow and Regression technique in specific.
- [Keras Tuner](#) - An easy-to-use, scalable hyperparameter optimization framework that solves the pain points of hyperparameter search.

Neural Networks

- [Kinho](#) - Simple API for Neural Network. Better for image processing with CPU/GPU + Transfer Learning.
- [nn_builder](#) - nn_builder is a python package that lets you build neural networks in 1 line
- [NeuralTalk](#) - NeuralTalk is a Python+numpy project for learning Multimodal Recurrent Neural Networks that describe images with sentences.
- [NeuralTalk](#) - NeuralTalk is a Python+numpy project for learning Multimodal Recurrent Neural Networks that describe images with sentences. **[Deprecated]**
- [Neuron](#) - Neuron is simple class for time series predictions. It's utilize LNU (Linear Neural Unit), QNU (Quadratic Neural Unit), RBF (Radial Basis Function), MLP (Multi Layer Perceptron), MLP-ELM (Multi Layer Perceptron - Extreme Learning Machine) neural networks learned with Gradient descent or LeLevenberg–Marquardt algorithm. **[Deprecated]**
- [Data Driven Code](#) - Very simple implementation of neural networks for dummies in python without using any libraries, with detailed comments.
- [Machine Learning, Data Science and Deep Learning with Python](#) - LiveVideo course that covers machine learning, Tensorflow, artificial intelligence, and neural networks.
- [TResNet: High Performance GPU-Dedicated Architecture](#) - TResNet models were designed and optimized to give the best speed-accuracy tradeoff out there on GPUs.
- [TResNet: Simple and powerful neural network library for python](#) - Variety of supported types of Artificial Neural Network and learning algorithms.
- [Jina AI](#) An easier way to build neural search in the cloud. Compatible with Jupyter Notebooks.
- [sequitur](#) PyTorch library for creating and training sequence autoencoders in just two lines of code

Spiking Neural Networks

- [Rockpool](#) - A machine learning library for spiking neural networks. Supports training with both torch and jax pipelines, and deployment to neuromorphic hardware.
- [Sinabs](#) - A deep learning library for spiking neural networks which is based on PyTorch, focuses on fast training and supports inference on neuromorphic hardware.

- Tonic - A library that makes downloading publicly available neuromorphic datasets a breeze and provides event-based data transformation/augmentation pipelines.

Python Survival Analysis

- lifelines - lifelines is a complete survival analysis library, written in pure Python
- Scikit-Survival - scikit-survival is a Python module for survival analysis built on top of scikit-learn. It allows doing survival analysis while utilizing the power of scikit-learn, e.g., for pre-processing or doing cross-validation.

Federated Learning

- Flower - A unified approach to federated learning, analytics, and evaluation. Federate any workload, any ML framework, and any programming language.
- PySyft - A Python library for secure and private Deep Learning.
- Tensorflow-Federated A federated learning framework for machine learning and other computations on decentralized data.

Kaggle Competition Source Code

- open-solution-home-credit -> source code and experiments results for Home Credit Default Risk.
- open-solution-googleai-object-detection -> source code and experiments results for Google AI Open Images - Object Detection Track.
- open-solution-salt-identification -> source code and experiments results for TGS Salt Identification Challenge.
- open-solution-ship-detection -> source code and experiments results for Airbus Ship Detection Challenge.
- open-solution-data-science-bowl-2018 -> source code and experiments results for 2018 Data Science Bowl.
- open-solution-value-prediction -> source code and experiments results for Santander Value Prediction Challenge.
- open-solution-toxic-comments -> source code for Toxic Comment Classification Challenge.
- wiki challenge - An implementation of Dell Zhang's solution to Wikipedia's Participation Challenge on Kaggle.
- kaggle insults - Kaggle Submission for "Detecting Insults in Social Commentary".
- kaggle_acquire-valued-shoppers-challenge - Code for the Kaggle acquire valued shoppers challenge.
- kaggle-cifar - Code for the CIFAR-10 competition at Kaggle, uses cuda-convnet.
- kaggle-blackbox - Deep learning made easy.
- kaggle-accelerometer - Code for Accelerometer Biometric Competition at Kaggle.
- kaggle-advertised-salaries - Predicting job salaries from ads - a Kaggle competition.
- kaggle amazon - Amazon access control challenge.
- kaggle-bestbuy_big - Code for the Best Buy competition at Kaggle.

- kaggle-bestbuy_small
- Kaggle Dogs vs. Cats - Code for Kaggle Dogs vs. Cats competition.
- Kaggle Galaxy Challenge - Winning solution for the Galaxy Challenge on Kaggle.
- Kaggle Gender - A Kaggle competition: discriminate gender based on handwriting.
- Kaggle Merck - Merck challenge at Kaggle.
- Kaggle Stackoverflow - Predicting closed questions on Stack Overflow.
- kaggle_acquire-valued-shoppers-challenge - Code for the Kaggle acquire valued shoppers challenge.
- wine-quality - Predicting wine quality.

Reinforcement Learning

- DeepMind Lab - DeepMind Lab is a 3D learning environment based on id Software's Quake III Arena via ioquake3 and other open source software. Its primary purpose is to act as a testbed for research in artificial intelligence, especially deep reinforcement learning.
- Gymnasium - A library for developing and comparing reinforcement learning algorithms (successor of [gym])(https://github.com/openai/gym).
- Serpent.AI - Serpent.AI is a game agent framework that allows you to turn any video game you own into a sandbox to develop AI and machine learning experiments. For both researchers and hobbyists.
- ViZDoom - ViZDoom allows developing AI bots that play Doom using only the visual information (the screen buffer). It is primarily intended for research in machine visual learning, and deep reinforcement learning, in particular.
- Roboschool - Open-source software for robot simulation, integrated with OpenAI Gym.
- Retro - Retro Games in Gym
- SLM Lab - Modular Deep Reinforcement Learning framework in PyTorch.
- Coach - Reinforcement Learning Coach by Intel® AI Lab enables easy experimentation with state of the art Reinforcement Learning algorithms
- garage - A toolkit for reproducible reinforcement learning research
- metaworld - An open source robotics benchmark for meta- and multi-task reinforcement learning
- acme - An Open Source Distributed Framework for Reinforcement Learning that makes build and train your agents easily.
- Spinning Up - An educational resource designed to let anyone learn to become a skilled practitioner in deep reinforcement learning
- Maze - Application-oriented deep reinforcement learning framework addressing real-world decision problems.
- RLlib - RLlib is an industry level, highly scalable RL library for tf and torch, based on Ray. It's used by companies like Amazon and Microsoft to solve real-world decision making problems at scale.
- DI-engine - DI-engine is a generalized Decision Intelligence engine. It supports most basic deep reinforcement learning (DRL) algorithms, such as DQN, PPO, SAC, and domain-specific algorithms like QMIX in multi-agent RL, GAIL in inverse RL, and RND in exploration problems.

Ruby

Natural Language Processing

- [Awesome NLP with Ruby](#) - Curated link list for practical natural language processing in Ruby.
- [Treat](#) - Text Retrieval and Annotation Toolkit, definitely the most comprehensive toolkit I've encountered so far for Ruby.
- [Stemmer](#) - Expose libstemmer_c to Ruby. **[Deprecated]**
- [Raspell](#) - raspell is an interface binding for ruby. **[Deprecated]**
- [UEA Stemmer](#) - Ruby port of UEALite Stemmer - a conservative stemmer for search and indexing.
- [Twitter-text-rb](#) - A library that does auto linking and extraction of usernames, lists and hashtags in tweets.

General-Purpose Machine Learning

- [Awesome Machine Learning with Ruby](#) - Curated list of ML related resources for Ruby.
- [Ruby Machine Learning](#) - Some Machine Learning algorithms, implemented in Ruby. **[Deprecated]**
- [Machine Learning Ruby](#) **[Deprecated]**
- [jRuby Mahout](#) - JRuby Mahout is a gem that unleashes the power of Apache Mahout in the world of JRuby. **[Deprecated]**
- [CardMagic-Classifier](#) - A general classifier module to allow Bayesian and other types of classifications.
- [rb-libsvm](#) - Ruby language bindings for LIBSVM which is a Library for Support Vector Machines.
- [Scoruby](#) - Creates Random Forest classifiers from PMML files.
- [rumale](#) - Rumale is a machine learning library in Ruby

Data Analysis / Data Visualization

- [rsruby](#) - Ruby - R bridge.
- [data-visualization-ruby](#) - Source code and supporting content for my Ruby Manor presentation on Data Visualisation with Ruby. **[Deprecated]**
- [ruby-plot](#) - gnuplot wrapper for Ruby, especially for plotting ROC curves into SVG files. **[Deprecated]**
- [plot-rb](#) - A plotting library in Ruby built on top of Vega and D3. **[Deprecated]**
- [scruffy](#) - A beautiful graphing toolkit for Ruby.
- [SciRuby](#)
- [Glean](#) - A data management tool for humans. **[Deprecated]**
- [Bioruby](#)
- [Arel](#) **[Deprecated]**

Misc

- [Big Data For Chimps](#)

- [Listof](#) - Community based data collection, packed in gem. Get list of pretty much anything (stop words, countries, non words) in txt, JSON or hash. [Demo/Search for a list](#)

Rust

General-Purpose Machine Learning

- [smartcore](#) - "The Most Advanced Machine Learning Library In Rust."
- [linfa](#) - a comprehensive toolkit to build Machine Learning applications with Rust
- [deeplearn-rs](#) - deeplearn-rs provides simple networks that use matrix multiplication, addition, and ReLU under the MIT license.
- [rustlearn](#) - a machine learning framework featuring logistic regression, support vector machines, decision trees and random forests.
- [rusty-machine](#) - a pure-rust machine learning library.
- [leaf](#) - open source framework for machine intelligence, sharing concepts from TensorFlow and Caffe. Available under the MIT license. **[Deprecated]**
- [RustNN](#) - RustNN is a feedforward neural network library. **[Deprecated]**
- [RusticSOM](#) - A Rust library for Self Organising Maps (SOM).
- [candle](#) - Candle is a minimalist ML framework for Rust with a focus on performance (including GPU support) and ease of use.
- [linfa](#) - `linfa` aims to provide a comprehensive toolkit to build Machine Learning applications with Rust

Deep Learning

- [tch-rs](#) - Rust bindings for the C++ API of PyTorch
- [dfdx](#) - Deep learning in Rust, with shape checked tensors and neural networks
- [burn](#) - Burn is a new comprehensive dynamic Deep Learning Framework built using Rust with extreme flexibility, compute efficiency and portability as its primary goals

Natural Language Processing

- [huggingface/tokenizers](#) - Fast State-of-the-Art Tokenizers optimized for Research and Production
- [rust-bert](#) - Rust native ready-to-use NLP pipelines and transformer-based models (BERT, DistilBERT, GPT2,...)

R

General-Purpose Machine Learning

- [ahaz](#) - ahaz: Regularization for semiparametric additive hazards regression. **[Deprecated]**
- [arules](#) - arules: Mining Association Rules and Frequent Itemsets
- [biglasso](#) - biglasso: Extending Lasso Model Fitting to Big Data in R.

- [bmrm](#) - bmrm: Bundle Methods for Regularized Risk Minimization Package.
- [Boruta](#) - Boruta: A wrapper algorithm for all-relevant feature selection.
- [bst](#) - bst: Gradient Boosting.
- [C50](#) - C50: C5.0 Decision Trees and Rule-Based Models.
- [caret](#) - Classification and Regression Training: Unified interface to ~150 ML algorithms in R.
- [caretEnsemble](#) - caretEnsemble: Framework for fitting multiple caret models as well as creating ensembles of such models. **[Deprecated]**
- [CatBoost](#) - General purpose gradient boosting on decision trees library with categorical features support out of the box for R.
- [Clever Algorithms For Machine Learning](#)
- [CORElearn](#) - CORElearn: Classification, regression, feature evaluation and ordinal evaluation. -* [CoxBoost](#) - CoxBoost: Cox models by likelihood based boosting for a single survival endpoint or competing risks **[Deprecated]**
- [Cubist](#) - Cubist: Rule- and Instance-Based Regression Modelling.
- [e1071](#) - e1071: Misc Functions of the Department of Statistics (e1071), TU Wien
- [earth](#) - earth: Multivariate Adaptive Regression Spline Models
- [elasticnet](#) - elasticnet: Elastic-Net for Sparse Estimation and Sparse PCA.
- [ElemStatLearn](#) - ElemStatLearn: Data sets, functions and examples from the book: "The Elements of Statistical Learning, Data Mining, Inference, and Prediction" by Trevor Hastie, Robert Tibshirani and Jerome Friedman Prediction" by Trevor Hastie, Robert Tibshirani and Jerome Friedman.
- [evtree](#) - evtree: Evolutionary Learning of Globally Optimal Trees.
- [forecast](#) - forecast: Timeseries forecasting using ARIMA, ETS, STLM, TBATS, and neural network models.
- [forecastHybrid](#) - forecastHybrid: Automatic ensemble and cross validation of ARIMA, ETS, STLM, TBATS, and neural network models from the "forecast" package.
- [fpc](#) - fpc: Flexible procedures for clustering.
- [frbs](#) - frbs: Fuzzy Rule-based Systems for Classification and Regression Tasks. **[Deprecated]**
- [GAMBoost](#) - GAMBoost: Generalized linear and additive models by likelihood based boosting. **[Deprecated]**
- [gamboostLSS](#) - gamboostLSS: Boosting Methods for GAMLSS.
- [gbm](#) - gbm: Generalized Boosted Regression Models.
- [glmnet](#) - glmnet: Lasso and elastic-net regularized generalized linear models.
- [glmpath](#) - glmpath: L1 Regularization Path for Generalized Linear Models and Cox Proportional Hazards Model.
- [GMMBoost](#) - GMMBoost: Likelihood-based Boosting for Generalized mixed models. **[Deprecated]**
- [grplasso](#) - grplasso: Fitting user specified models with Group Lasso penalty.
- [grpreg](#) - grpreg: Regularization paths for regression models with grouped covariates.
- [h2o](#) - A framework for fast, parallel, and distributed machine learning algorithms at scale -- Deeplearning, Random forests, GBM, KMeans, PCA, GLM.
- [hda](#) - hda: Heteroscedastic Discriminant Analysis. **[Deprecated]**

- [Introduction to Statistical Learning](#)
- [ipred](#) - ipred: Improved Predictors.
- [kernlab](#) - kernlab: Kernel-based Machine Learning Lab.
- [klaR](#) - klaR: Classification and visualization.
- [L0Learn](#) - L0Learn: Fast algorithms for best subset selection.
- [lars](#) - lars: Least Angle Regression, Lasso and Forward Stagewise. **[Deprecated]**
- [lasso2](#) - lasso2: L1 constrained estimation aka 'lasso'.
- [LiblineaR](#) - LiblineaR: Linear Predictive Models Based On The Liblinear C/C++ Library.
- [LogicReg](#) - LogicReg: Logic Regression.
- [Machine Learning For Hackers](#)
- [maptree](#) - maptree: Mapping, pruning, and graphing tree models. **[Deprecated]**
- [mboost](#) - mboost: Model-Based Boosting.
- [medley](#) - medley: Blending regression models, using a greedy stepwise approach.
- [mlr](#) - mlr: Machine Learning in R.
- [ncvreg](#) - ncvreg: Regularization paths for SCAD- and MCP-penalized regression models.
- [nnet](#) - nnet: Feed-forward Neural Networks and Multinomial Log-Linear Models. **[Deprecated]**
- [pamr](#) - pamr: Pam: prediction analysis for microarrays. **[Deprecated]**
- [party](#) - party: A Laboratory for Recursive Partitioning
- [partykit](#) - partykit: A Toolkit for Recursive Partitioning.
- [penalized](#) - penalized: L1 (lasso and fused lasso) and L2 (ridge) penalized estimation in GLMs and in the Cox model.
- [penalizedLDA](#) - penalizedLDA: Penalized classification using Fisher's linear discriminant. **[Deprecated]**
- [penalizedSVM](#) - penalizedSVM: Feature Selection SVM using penalty functions.
- [quantregForest](#) - quantregForest: Quantile Regression Forests.
- [randomForest](#) - randomForest: Breiman and Cutler's random forests for classification and regression.
- [randomForestSRC](#) - randomForestSRC: Random Forests for Survival, Regression and Classification (RF-SRC).
- [rattle](#) - rattle: Graphical user interface for data mining in R.
- [rda](#) - rda: Shrunken Centroids Regularized Discriminant Analysis.
- [rdetools](#) - rdetools: Relevant Dimension Estimation (RDE) in Feature Spaces. **[Deprecated]**
- [REEMtree](#) - REEMtree: Regression Trees with Random Effects for Longitudinal (Panel) Data. **[Deprecated]**
- [relaxo](#) - relaxo: Relaxed Lasso. **[Deprecated]**
- [rgenoud](#) - rgenoud: R version of GENetic Optimization Using Derivatives
- [Rmalschains](#) - Rmalschains: Continuous Optimization using Memetic Algorithms with Local Search Chains (MA-LS-Chains) in R.
- [rminer](#) - rminer: Simpler use of data mining methods (e.g. NN and SVM) in classification and

regression. **[Deprecated]**

- ROCR - ROCR: Visualizing the performance of scoring classifiers. **[Deprecated]**
- RoughSets - RoughSets: Data Analysis Using Rough Set and Fuzzy Rough Set Theories. **[Deprecated]**
- rpart - rpart: Recursive Partitioning and Regression Trees.
- RPMM - RPMM: Recursively Partitioned Mixture Model.
- RSNNS - RSNNS: Neural Networks in R using the Stuttgart Neural Network Simulator (SNNS).
- RWeka - RWeka: R/Weka interface.
- RXshrink - RXshrink: Maximum Likelihood Shrinkage via Generalized Ridge or Least Angle Regression.
- sda - sda: Shrinkage Discriminant Analysis and CAT Score Variable Selection. **[Deprecated]**
- spectralGraphTopology - spectralGraphTopology: Learning Graphs from Data via Spectral Constraints.
- SuperLearner - Multi-algorithm ensemble learning packages.
- svmpath - svmpath: svmpath: the SVM Path algorithm. **[Deprecated]**
- tgp - tgp: Bayesian treed Gaussian process models. **[Deprecated]**
- tree - tree: Classification and regression trees.
- varSelRF - varSelRF: Variable selection using random forests.
- XGBoost.R - R binding for eXtreme Gradient Boosting (Tree) Library.
- Optunity - A library dedicated to automated hyperparameter optimization with a simple, lightweight API to facilitate drop-in replacement of grid search. Optunity is written in Python but interfaces seamlessly to R.
- igraph - binding to igraph library - General purpose graph library.
- MXNet - Lightweight, Portable, Flexible Distributed/Mobile Deep Learning with Dynamic, Mutation-aware Dataflow Dep Scheduler; for Python, R, Julia, Go, JavaScript and more.
- TDSP-Utilities - Two data science utilities in R from Microsoft: 1) Interactive Data Exploration, Analysis, and Reporting (IDEAR) ; 2) Automated Modelling and Reporting (AMR).

Data Manipulation | Data Analysis | Data Visualization

- dplyr - A data manipulation package that helps to solve the most common data manipulation problems.
- ggplot2 - A data visualization package based on the grammar of graphics.
- tmap for visualizing geospatial data with static maps and leaflet for interactive maps
- tm and quanteda are the main packages for managing, analyzing, and visualizing textual data.
- shiny is the basis for truly interactive displays and dashboards in R. However, some measure of interactivity can be achieved with htmlwidgets bringing javascript libraries to R. These include, plotly, dygraphs, highcharter, and several others.

SAS

General-Purpose Machine Learning

- [Visual Data Mining and Machine Learning](#) - Interactive, automated, and programmatic modelling with the latest machine learning algorithms in and end-to-end analytics environment, from data prep to deployment. Free trial available.
- [Enterprise Miner](#) - Data mining and machine learning that creates deployable models using a GUI or code.
- [Factory Miner](#) - Automatically creates deployable machine learning models across numerous market or customer segments using a GUI.

Data Analysis / Data Visualization

- [SAS/STAT](#) - For conducting advanced statistical analysis.
- [University Edition](#) - FREE! Includes all SAS packages necessary for data analysis and visualization, and includes online SAS courses.

Natural Language Processing

- [Contextual Analysis](#) - Add structure to unstructured text using a GUI.
- [Sentiment Analysis](#) - Extract sentiment from text using a GUI.
- [Text Miner](#) - Text mining using a GUI or code.

Demos and Scripts

- [ML_Tables](#) - Concise cheat sheets containing machine learning best practices.
- [enlighten-apply](#) - Example code and materials that illustrate applications of SAS machine learning techniques.
- [enlighten-integration](#) - Example code and materials that illustrate techniques for integrating SAS with other analytics technologies in Java, PMML, Python and R.
- [enlighten-deep](#) - Example code and materials that illustrate using neural networks with several hidden layers in SAS.
- [dm-flow](#) - Library of SAS Enterprise Miner process flow diagrams to help you learn by example about specific data mining topics.

Scala

Natural Language Processing

- [ScalaNLP](#) - ScalaNLP is a suite of machine learning and numerical computing libraries.
- [Breeze](#) - Breeze is a numerical processing library for Scala.
- [Chalk](#) - Chalk is a natural language processing library. **[Deprecated]**
- [FACTORIE](#) - FACTORIE is a toolkit for deployable probabilistic modelling, implemented as a software library in Scala. It provides its users with a succinct language for creating relational factor graphs, estimating parameters and performing inference.
- [Montague](#) - Montague is a semantic parsing library for Scala with an easy-to-use DSL.

- **Spark NLP** - Natural language processing library built on top of Apache Spark ML to provide simple, performant, and accurate NLP annotations for machine learning pipelines, that scale easily in a distributed environment.

Data Analysis / Data Visualization

- **NDScala** - N-dimensional arrays in Scala 3. Think NumPy ndarray, but with compile-time type-checking/inference over shapes, tensor/axis labels & numeric data types
- **MLlib in Apache Spark** - Distributed machine learning library in Spark
- **Hydrosphere Mist** - a service for deployment Apache Spark MLLib machine learning models as realtime, batch or reactive web services.
- **Scalding** - A Scala API for Cascading.
- **Summing Bird** - Streaming MapReduce with Scalding and Storm.
- **Algebird** - Abstract Algebra for Scala.
- **xerial** - Data management utilities for Scala. **[Deprecated]**
- **PredictionIO** - PredictionIO, a machine learning server for software developers and data engineers.
- **BIDMat** - CPU and GPU-accelerated matrix library intended to support large-scale exploratory data analysis.
- **Flink** - Open source platform for distributed stream and batch data processing.
- **Spark Notebook** - Interactive and Reactive Data Science using Scala and Spark.

General-Purpose Machine Learning

- **Microsoft ML for Apache Spark** -> A distributed machine learning framework Apache Spark
- **ONNX-Scala** - An ONNX (Open Neural Network eXchange) API and backend for typeful, functional deep learning in Scala (3).
- **DeepLearning.scala** - Creating statically typed dynamic neural networks from object-oriented & functional programming constructs.
- **Conjecture** - Scalable Machine Learning in Scalding.
- **brushfire** - Distributed decision tree ensemble learning in Scala.
- **ganitha** - Scalding powered machine learning. **[Deprecated]**
- **adam** - A genomics processing engine and specialized file format built using Apache Avro, Apache Spark and Parquet. Apache 2 licensed.
- **bioscala** - Bioinformatics for the Scala programming language
- **BIDMach** - CPU and GPU-accelerated Machine Learning Library.
- **Figaro** - a Scala library for constructing probabilistic models.
- **H2O Sparkling Water** - H2O and Spark interoperability.
- **FlinkML in Apache Flink** - Distributed machine learning library in Flink.
- **DynaML** - Scala Library/REPL for Machine Learning Research.
- **Saul** - Flexible Declarative Learning-Based Programming.
- **SwiftLearner** - Simply written algorithms to help study ML or write your own implementations.
- **Smile** - Statistical Machine Intelligence and Learning Engine.

- doddle-model - An in-memory machine learning library built on top of Breeze. It provides immutable objects and exposes its functionality through a scikit-learn-like API.
- TensorFlow Scala - Strongly-typed Scala API for TensorFlow.

Scheme

Neural Networks

- layer - Neural network inference from the command line, implemented in CHICKEN Scheme.

Swift

General-Purpose Machine Learning

- Bender - Fast Neural Networks framework built on top of Metal. Supports TensorFlow models.
- Swift AI - Highly optimized artificial intelligence and machine learning library written in Swift.
- Swift for Tensorflow - a next-generation platform for machine learning, incorporating the latest research across machine learning, compilers, differentiable programming, systems design, and beyond.
- BrainCore - The iOS and OS X neural network framework.
- swix - A bare bones library that includes a general matrix language and wraps some OpenCV for iOS development. **[Deprecated]**
- AIToolbox - A toolbox framework of AI modules written in Swift: Graphs/Trees, Linear Regression, Support Vector Machines, Neural Networks, PCA, KMeans, Genetic Algorithms, MDP, Mixture of Gaussians.
- MLKit - A simple Machine Learning Framework written in Swift. Currently features Simple Linear Regression, Polynomial Regression, and Ridge Regression.
- Swift Brain - The first neural network / machine learning library written in Swift. This is a project for AI algorithms in Swift for iOS and OS X development. This project includes algorithms focused on Bayes theorem, neural networks, SVMs, Matrices, etc...
- Perfect TensorFlow - Swift Language Bindings of TensorFlow. Using native TensorFlow models on both macOS / Linux.
- PredictionBuilder - A library for machine learning that builds predictions using a linear regression.
- Awesome CoreML - A curated list of pretrained CoreML models.
- Awesome Core ML Models - A curated list of machine learning models in CoreML format.

TensorFlow

General-Purpose Machine Learning

- Awesome Keras - A curated list of awesome Keras projects, libraries and resources.
- Awesome TensorFlow - A list of all things related to TensorFlow.

- [Golden TensorFlow](#) - A page of content on TensorFlow, including academic papers and links to related topics.

Tools

Neural Networks

- [layer](#) - Neural network inference from the command line

Misc

- [Infinity](#) - The AI-native database built for LLM applications, providing incredibly fast vector and full-text search. Developed using C++20
- [Synthical](#) - AI-powered collaborative research environment. You can use it to get recommendations of articles based on reading history, simplify papers, find out what articles are trending, search articles by meaning (not just keywords), create and share folders of articles, see lists of articles from specific companies and universities, and add highlights.
- [Humanloop](#) – Humanloop is a platform for prompt experimentation, finetuning models for better performance, cost optimization, and collecting model generated data and user feedback.
- [Qdrant](#) – Qdrant is [open source](#) vector similarity search engine with extended filtering support, written in Rust.
- [milvus](#) – Milvus is [open source](#) vector database for production AI, written in Go and C++, scalable and blazing fast for billions of embedding vectors.
- [Weaviate](#) – Weaviate is an [open source](#) vector search engine and vector database. Weaviate uses machine learning to vectorize and store data, and to find answers to natural language queries. With Weaviate you can also bring your custom ML models to production scale.
- [txtai](#) - Build semantic search applications and workflows.
- [MLReef](#) - MLReef is an end-to-end development platform using the power of git to give structure and deep collaboration possibilities to the ML development process.
- [Chroma](#) - Chroma - the AI-native open-source embedding database
- [Pinecone](#) - Vector database for applications that require real-time, scalable vector embedding and similarity search.
- [CatalyzeX](#) - Browser extension ([Chrome](#) and [Firefox](#)) that automatically finds and shows code implementations for machine learning papers anywhere: Google, Twitter, Arxiv, Scholar, etc.
- [ML Workspace](#) - All-in-one web-based IDE for machine learning and data science. The workspace is deployed as a docker container and is preloaded with a variety of popular data science libraries (e.g., Tensorflow, PyTorch) and dev tools (e.g., Jupyter, VS Code).
- [Notebooks](#) - A starter kit for Jupyter notebooks and machine learning. Companion docker images consist of all combinations of python versions, machine learning frameworks (Keras, PyTorch and Tensorflow) and CPU/CUDA versions.
- [DVC](#) - Data Science Version Control is an open-source version control system for machine learning projects with pipelines support. It makes ML projects reproducible and shareable.
- [DVCLive](#) - Python library for experiment metrics logging into simply formatted local files.

- VDP - open source visual data ETL to streamline the end-to-end visual data processing pipeline: extract unstructured visual data from pre-built data sources, transform it into analysable structured insights by Vision AI models imported from various ML platforms, and load the insights into warehouses or applications.
- Kedro - Kedro is a data and development workflow framework that implements best practices for data pipelines with an eye towards productionizing machine learning models.
- guild.ai - Tool to log, analyze, compare and "optimize" experiments. It's cross-platform and framework independent, and provided integrated visualizers such as tensorboard.
- Sacred - Python tool to help you configure, organize, log and reproduce experiments. Like a notebook lab in the context of Chemistry/Biology. The community has built multiple add-ons leveraging the proposed standard.
- Comet - ML platform for tracking experiments, hyper-parameters, artifacts and more. It's deeply integrated with over 15+ deep learning frameworks and orchestration tools. Users can also use the platform to monitor their models in production.
- MLFlow - platform to manage the ML lifecycle, including experimentation, reproducibility and deployment. Framework and language agnostic, take a look at all the built-in integrations.
- Weights & Biases - Machine learning experiment tracking, dataset versioning, hyperparameter search, visualization, and collaboration
- More tools to improve the ML lifecycle: Catalyst, PachydermIO. The following are GitHub-alike and targeting teams Weights & Biases, Neptune.ai, Comet.ml, Valohai.ai, DAGsHub.
- Arize AI - Model validaiton and performance monitoring, drift detection, explainability, visualization across structured and unstructured data
- MachineLearningWithTensorFlow2ed - a book on general purpose machine learning techniques regression, classification, unsupervised clustering, reinforcement learning, auto encoders, convolutional neural networks, RNNs, LSTMs, using TensorFlow 1.14.1.
- m2cgen - A tool that allows the conversion of ML models into native code (Java, C, Python, Go, JavaScript, Visual Basic, C#, R, PowerShell, PHP, Dart) with zero dependencies.
- CML - A library for doing continuous integration with ML projects. Use GitHub Actions & GitLab CI to train and evaluate models in production like environments and automatically generate visual reports with metrics and graphs in pull/merge requests. Framework & language agnostic.
- Pythonizr - An online tool to generate boilerplate machine learning code that uses scikit-learn.
- Flyte - Flyte makes it easy to create concurrent, scalable, and maintainable workflows for machine learning and data processing.
- Chaos Genius - ML powered analytics engine for outlier/anomaly detection and root cause analysis.
- MLEM - Version and deploy your ML models following GitOps principles
- DockerDL - Ready to use deeplearning docker images.
- Aqueduct - Aqueduct enables you to easily define, run, and manage AI & ML tasks on any cloud infrastructure.
- Ambrosia - Ambrosia helps you clean up your LLM datasets using *other* LLMs.

Books

- [Distributed Machine Learning Patterns](#) - This book teaches you how to take machine learning models from your personal laptop to large distributed clusters. You'll explore key concepts and patterns behind successful distributed machine learning systems, and learn technologies like TensorFlow, Kubernetes, Kubeflow, and Argo Workflows directly from a key maintainer and contributor, with real-world scenarios and hands-on projects.
- [Grokking Machine Learning](#) - Grokking Machine Learning teaches you how to apply ML to your projects using only standard Python code and high school-level math.
- [Machine Learning Bookcamp](#) - Learn the essentials of machine learning by completing a carefully designed set of real-world projects.
- [Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow](#) - Through a recent series of breakthroughs, deep learning has boosted the entire field of machine learning. Now, even programmers who know close to nothing about this technology can use simple, efficient tools to implement programs capable of learning from data. This bestselling book uses concrete examples, minimal theory, and production-ready Python frameworks (Scikit-Learn, Keras, and TensorFlow) to help you gain an intuitive understanding of the concepts and tools for building intelligent systems.

- [Netron](#) - An opensource viewer for neural network, deep learning and machine learning models
- [Teachable Machine](#) - Train Machine Learning models on the fly to recognize your own images, sounds, & poses.
- [Model Zoo](#) - Discover open source deep learning code and pretrained models.

Credits

- Some of the python libraries were cut-and-pasted from [vinta](#)
- References for Go were mostly cut-and-pasted from [gopherdata](#)