Libraries

Machine learning libraries and frameworks forked from josephmisti's awesome machine learning.

- APL
- C
- C++
- Common Lisp
- Clojure
- Elixir
- Erlang
- Go
- Haskell
- Java
- Javascript
- Julia
- Lua
- Matlab
- .NET
- Objective C
- OCaml
- PHP
- Python
- Ruby
- Rust
- R
- SAS
- Scala
- Swift

APL

General-Purpose Machine Learning

- naive-apl - Naive Bayesian Classifier implementation in APL

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 recomender system based upon scikit-learn algorithms.

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 Matlab toolbox

Speech Recognition

- HTK -The Hidden Markov Model Toolkit. HTK is a portable toolkit for building and manipulating hidden Markov models.

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
- OpenCV - OpenCV has C++, C, Python, Java and MATLAB interfaces and supports Windows, Linux, Android and Mac OS.
- VIGRA - VIGRA is a generic cross-platform C++ computer vision and machine learning library for volumes of arbitrary dimensionality with Python bindings.

General-Purpose Machine Learning

- BanditLib - A simple Multi-armed Bandit library.
- Caffe - A deep learning framework developed with cleanliness, readability, and speed in mind. [DEEP LEARNING]
- 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]
- CXXNET - Yet another deep learning framework with less than 1000 lines core code [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.
- Disrtibuted Machine learning Tool Kit (DMTK) 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.
- encog-cpp
- Fido - A highly-modular C++ machine learning library for embedded electronics and robotics.
- igraph - General purpose graph library
- Intel(R) DAAL - 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 framework based on decision tree algorithms, used for ranking, classification and many other machine learning tasks.
- 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
- 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, on both CPU and GPU.
- XGBoost - A parallelized optimized general purpose gradient boosting library.

Natural Language Processing

- BLLIP 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++ for segmenting/labeling sequential data & other Natural Language Processing tasks.
- CRFsuite for labeling sequential data.

- frog - Memory-based NLP suite developed for Dutch: PoS tagger, lemmatiser, dependency parser, NER, shallow parser, morphological analyzer.
- libfolia](https://github.com/LanguageMachines/libfolia) - C++ library for the [FoLiA format
- MeTA](https://github.com/meta-toolkit/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 objected-oriented framework that facilitates the integration of probabilistic models for sequences over a user defined alphabet.

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, Gaussian Processes
- mgl-gpr - Evolutionary algorithms
- cl-libsvm - Wrapper for the libsvm support vector machine library

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

- Touchstone - Clojure A/B testing library
- Clojush - The Push programming language and the PushGP genetic programming system implemented in Clojure

- Infer - Inference and machine learning in clojure
- Clj-ML - A machine learning library for Clojure built on top of Weka and friends
- DL4CLJ - Clojure wrapper for Deeplearning4j
- Encog
- Fungp - A genetic programming library for Clojure
- Statistiker - Basic Machine Learning algorithms in Clojure.
- clortex - General Machine Learning library using Numenta's Cortical Learning Algorithm
- comportex - Functionally composable Machine Learning library using Numenta's Cortical Learning Algorithm
- cortex - Neural networks, regression and feature learning in Clojure.
- lambda-ml - Simple, concise implementations of machine learning techniques and utilities in Clojure.

Data Analysis / Data Visualization

- Incanter - Incanter is a Clojure-based, R-like platform for statistical computing and graphics.
- PigPen - Map-Reduce for Clojure.
- Envision - Clojure Data Visualisation library, based on Statistiker and D3

Elixir

General-Purpose Machine Learning

- Simple Bayes - A Simple Bayes / Naive Bayes implementation in Elixir.

Natural Language Processing

- Stemmer stemming implementation in Elixir.

Erlang

General-Purpose Machine Learning

- Disco - Map Reduce in Erlang

Go

Natural Language Processing

- go-porterstemmer - A native Go clean room implementation of the Porter Stemming algorithm.
- paicehusk - Golang implementation of the Paice/Husk Stemming Algorithm.
- snowball - Snowball Stemmer for Go.
- go-ngram - In-memory n-gram index with compression.

General-Purpose Machine Learning

- gago - Multi-population, flexible, parallel genetic algorithm.
- Go Learn - Machine Learning for Go
- go-pr - Pattern recognition package in Go lang.
- go-ml - Linear / Logistic regression, Neural Networks, Collaborative Filtering and Gaussian Multivariate Distribution
- bayesian - Naive Bayesian Classification for Golang.
- go-galib - Genetic Algorithms library written in Go / golang
- Cloudforest - Ensembles of decision trees in go/golang.
- gobrain - Neural Networks written in go
- GoNN - GoNN is an implementation of Neural Network in Go Language, which includes BPNN, RBF, PCN
- MXNet - Lightweight, Portable, Flexible Distributed/Mobile Deep Learning with Dynamic, Mutation-aware Dataflow Dep Scheduler; for Python, R, Julia, Go, Javascript and more.
- go-mxnet-predictor - Go binding for MXNet c_predict_api to do inference with pre-trained model

Data Analysis / Data Visualization

- go-graph - Graph library for Go/golang language.
- SVGo - The Go Language library for SVG generation
- RF - Random forests implementation in Go

Haskell

General-Purpose Machine Learning

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

Java

Natural Language Processing

- Cortical.io as quickly and intuitively as the brain.
- 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.
- 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
- Stanford Topic Modeling Toolbox - Topic modeling tools to social scientists and others who wish to perform analysis on datasets
- 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 modeling, 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 components in Java and is built on top of Apache UIMA.
- Apache cTAKES is an open-source natural language processing system for information extraction from electronic medical record clinical free-text.
- ClearNLP - The ClearNLP 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. This project is under the Apache 2 license.
- CogcompNLP 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.

- Datumbox - Machine Learning framework for rapid development of Machine Learning and Statistical applications
- ELKI
- 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.
- 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
- java-deeplearning - Distributed Deep Learning Platform for Java, Clojure,Scala
- Mahout - Distributed machine learning
- Meka.
- 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
- 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.
- SmileMiner - Statistical Machine Intelligence & Learning Engine
- SystemML language.
- WalnutiQ - object oriented model of the human brain
- Weka - Weka is a collection of machine learning algorithms for data mining tasks
- LBJava - Learning Based Java is a modeling 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.

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
- 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

Deep Learning

- Deeplearning4j - Scalable deep learning for industry with parallel GPUs

Javascript

Natural Language Processing

- Twitter-text - A JavaScript implementation of Twitter's text processing library
- NLP.js - NLP utilities in javascript and coffeescript
- natural - General natural language facilities for node
- Knwl.js - A Natural Language Processor in JS
- Retext - Extensible system for analyzing and manipulating natural language
- TextProcessing - Sentiment analysis, stemming and lemmatization, part-of-speech tagging and chunking, phrase extraction and named entity recognition.
- NLP Compromise - Natural Language processing in the browser

Data Analysis / Data Visualization

- D3.js
- High Charts
- NVD3.js
- dc.js
- chartjs
- dimple
- amCharts
- D3xter - Straight forward plotting built on D3
- statkit - Statistics kit for JavaScript
- datakit - A lightweight framework for data analysis in JavaScript
- science.js - Scientific and statistical computing in JavaScript.
- Z3d - Easily make interactive 3d plots built on Three.js
- 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.

- ZingChart- library written on Vanilla JS for big data visualization.
- cheminfo - Platform for data visualization and analysis, using the visualizer project.

General-Purpose Machine Learning

- Convnet.js - ConvNetJS is a Javascript library for training Deep Learning models[DEEP LEARNING]
- Clusterfck - Agglomerative hierarchical clustering implemented in Javascript for Node.js and the browser
- Clustering.js - Clustering algorithms implemented in Javascript for Node.js and the browser
- Decision Trees - NodeJS Implementation of Decision Tree using ID3 Algorithm
- DN2A - Digital Neural Networks Architecture
- figue - K-means, fuzzy c-means and agglomerative clustering
- Node-fann bindings for Node.js
- Kmeans.js - Simple Javascript implementation of the k-means algorithm, for node.js and the browser
- LDA.js - LDA topic modeling for node.js
- Learning.js - Javascript implementation of logistic regression/c4.5 decision tree
- Machine Learning - Machine learning library for Node.js
- machineJS - Automated machine learning, data formatting, ensembling, and hyperparameter optimization for competitions and exploration- just give it a .csv file!
- mil-tokyo - List of several machine learning libraries
- Node-SVM - Support Vector Machine for nodejs
- Brain - Neural networks in JavaScript **[Deprecated]**
- Bayesian-Bandit - Bayesian bandit implementation for Node and the browser.
- 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.
- kalman - Kalman filter for Javascript.
- shaman - node.js library with support for both simple and multiple linear regression.
- ml.js - Machine learning and numerical analysis tools for Node.js and the Browser!
- 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.

Misc

- sylvester - Vector and Matrix math for JavaScript.
- simple-statistics 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.
- GreatCircle - Library for calculating great circle distance.

Julia

General-Purpose Machine Learning

- MachineLearning - Julia Machine Learning library
- 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
- Local Regression - Local regression, so smooooth!
- Naive Bayes - Simple Naive Bayes implementation in Julia
- Mixed Models mixed-effects models
- Simple MCMC - basic mcmc sampler implemented in Julia
- Distance - Julia module for Distance evaluation
- Decision Tree - Decision Tree Classifier and Regressor
- Neural - A neural network in Julia
- MCMC - MCMC tools for Julia
- Mamba for Bayesian analysis in Julia
- GLM - Generalized linear models in Julia
- Online Learning
- GLMNet - Julia wrapper for fitting Lasso/ElasticNet GLM models using glmnet
- Clustering - Basic functions for clustering data: k-means, dp-means, etc.
- SVM - SVM's for Julia
- Kernal Density - Kernel density estimators for julia
- Dimensionality Reduction - Methods for dimensionality reduction
- NMF - A Julia package for non-negative matrix factorization
- ANN - Julia artificial neural networks
- Mocha - Deep Learning framework for Julia inspired by Caffe
- 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

Natural Language Processing

- Topic Models - TopicModels for Julia
- Text Analysis - Julia package for text analysis

Data Analysis / Data Visualization

- Graph Layout - Graph layout algorithms in pure Julia
- Data Frames Meta - Metaprogramming tools for DataFrames
- Julia Data - library for working with tabular data in Julia
- 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
- Time Series - Time series toolkit for Julia
- Sampling - Basic sampling algorithms for Julia

Misc Stuff / Presentations

- DSP.
- JuliaCon Presentations - Presentations for JuliaCon
- SignalProcessing - Signal Processing tools for Julia
- Images - An image library for Julia

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.
- autograd - Autograd automatically differentiates native Torch code. Inspired by the original Python version.
- graph - Graph package for Torch
- randomkit - Numpy's randomkit, wrapped for Torch
- 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.
- optim - An optimization library for Torch. SGD, Adagrad, Conjugate-Gradient, LBFGS, RProp and more.
- unsup.
- manifold - A package to manipulate manifolds
- svm - Torch-SVM library
- lbfgs - FFI Wrapper for liblbfgs
- vowpalwabbit - An old vowpalwabbit interface to torch.
- OpenGM - OpenGM is a C++ library for graphical modeling, and inference. The Lua bindings provide a simple way of describing graphs, from Lua, and then optimizing them with OpenGM.
- sphagetti module for torch7 by @MichaelMathieu
- LuaSHKit - A lua wrapper around the Locality sensitive hashing library SHKit
- kernel smoothing - KNN, kernel-weighted average, local linear regression smoothers
- 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.
- 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.
- saliency - code and tools around integral images. A library for finding interest points based on fast integral histograms.
- stitch - allows us to use hugin to stitch images and apply same stitching to a video sequence
- sfm - A bundle adjustment/structure from motion package
- fex - A package for feature extraction in Torch. Provides SIFT and dSIFT modules.
- OverFeat - A state-of-the-art generic dense feature extractor
- Numeric Lua
- Lunatic Python
- SciLua
- Lua - Numerical Algorithms
- Lunum

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)
- Music Tagging - Music Tagging scripts for torch7
- 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 - An NLP library for Matlab

General-Purpose Machine Learning

- Training a deep autoencoder or a classifier on MNIST
- Convolutional-Recursive Deep Learning for 3D Object Classification - Convolutional-Recursive Deep Learning for 3D Object Classification[DEEP LEARNING]
- t-Distributed Stochastic Neighbor Embedding technique for dimensionality reduction that is particularly well suited for the visualization of high-dimensional datasets.
- Spider - The spider is intended to be a complete object orientated environment for machine learning in Matlab.
- LibSVM - A Library for Support Vector Machines
- 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.

Data Analysis / Data Visualization

- matlab_gbl - MatlabBGL is a Matlab package for working with graphs.
- gamic - 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 e 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 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.
- Vulpes - Deep belief and deep learning implementation written in F# and leverages CUDA GPU execution with Alea.cuBase.

- 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.
- 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 feed back. The chat bots can even scrape the internet for information to return in their output as well as to use for learning.
- 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 to customised 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.

Data Analysis / Data Visualization

- numl - numl is a machine learning library intended to ease the use of using standard modeling 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 every day 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 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.
- MLPNeuralNet - Fast multilayer perceptron neural network library for iOS and Mac OS X. MLPNeuralNet predicts new examples by trained neural network. It is built on top of the Apple's Accelerate Framework, using vectorized operations and hardware acceleration if available.
- 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. This network can be used in products recommendation, user behavior analysis, data mining and data analysis.

- Multi-Perceptron-NeuralNetwork and designed unlimited-hidden-layers.
- KRHebbian-Algorithm in neural network of Machine Learning.
- KRKmeans-Algorithm - It implemented K-Means the clustering and classification algorithm. It could be used in data mining and image compression.
- KRFuzzyCMeans-Algorithm the fuzzy clustering / classification algorithm on Machine Learning. It could be used in data mining and image compression.

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.

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.

Python

Computer Vision

- Scikit-Image - A collection of algorithms for image processing 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

Natural Language Processing

- 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 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.
- 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.
- genius - A Chinese segment base on Conditional Random Field.
- KoNLPy - A Python package for Korean natural language processing.
- nut - Natural language Understanding Toolkit
- Rosetta
- BLLIP Parser
- PyNLPl](https://github.com/proycon/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.
- python-ucto
- python-frog
- python-zpar](https://github.com/EducationalTestingService/python-zpar) - Python bindings for [ZPar, a statistical part-of-speech-tagger, constiuency parser, and dependency parser for English.
- colibri-core - Python binding to C++ library for extracting and working with 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
- Fuzzy Wuzzy - Fuzzy String Matching in Python
- 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](https://github.com/dasmith/stanford-corenlp-python) - Python wrapper for [Stanford CoreNLP

General-Purpose Machine Learning

- 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, LightGBM, and soon, deep learning.
- machine learning](https://github.com/jeff1evesque/machine-learning) - automated build consisting of a [web-interface](https://github.com/jeff1evesque/machine-learning#web-interface), and set of [programmatic-interface, are stored into a NoSQL datastore.
- XGBoost 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.
- scikit-learn - A Python module for machine learning built on top of SciPy.
- metric-learn - A Python module for metric learning.
- SimpleAI Python implementation of many of the artificial intelligence algorithms described on 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 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](https://github.com/lisa-lab/pylearn2) - A Machine Learning library based on [Theano.
- keras](https://github.com/fchollet/keras) - Modular neural network library based on [Theano.
- Lasagne - Lightweight library to build and train neural networks in Theano.
- hebel - GPU-Accelerated Deep Learning Library in Python.
- Chainer - Flexible neural network framework
- prohpet - Fast and automated time series forecasting framework by Facebook.
- gensim - Topic Modelling for Humans.
- topik - Topic modelling toolkit
- PyBrain - Another Python Machine Learning Library.
- Brainstorm - Fast, flexible and fun neural networks. This is the successor of PyBrain.
- Crab - A flexible, fast recommender engine.
- python-recsys - A Python library for implementing a Recommender System.
- thinking bayes - Book on Bayesian Analysis
- Image-to-Image Translation with Conditional Adversarial Networks](https://github.com/williamFalcon/pix2pix-keras) - 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

- CoverTree - Python implementation of cover trees, near-drop-in replacement for scipy.spatial.kdtree
- nilearn - Machine learning for NeuroImaging in Python
- imbalanced-learn - Python module to perform under sampling and over sampling with various techniques.
- Shogun - The Shogun Machine Learning Toolbox
- Pyevolve - Genetic algorithm framework.
- Caffe - A deep learning framework developed with cleanliness, readability, and speed in mind.
- breze - Theano based library for deep and recurrent neural networks
- pyhsmm, focusing on the Bayesian Nonparametric extensions, the HDP-HMM and HDP-HSMM, mostly with weak-limit approximations.
- mrjob - A library to let Python program run on Hadoop.
- SKLL - A wrapper around scikit-learn that makes it simpler to conduct experiments.
- neurolab - https://github.com/zueve/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.
- Pebl - Python Environment for Bayesian Learning
- 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
- yahmm - 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
- mlxtend - A library consisting of useful tools for data science and machine learning tasks.
- neon](https://github.com/NervanaSystems/neon) - Nervana's [high-performance Python-based Deep Learning framework [DEEP LEARNING]
- 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
- skflow - Simplified interface for TensorFlow, mimicking Scikit Learn.
- 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 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.
- 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.
- rgf_python Library.
- gym - OpenAI Gym is a toolkit for developing and comparing reinforcement learning algorithms.
- 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.

Data Analysis / Data Visualization

- SciPy - A Python-based ecosystem of open-source software for mathematics, science, and engineering.
- NumPy - A fundamental package for scientific computing with Python.
- Numba complier to LLVM aimed at scientific Python by the developers of Cython and 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.
- Open Mining
- 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 modeling of dynamic motion based around NumPy, SciPy, IPython, and matplotlib.
- SymPy - A Python library for symbolic mathematics.
- statsmodels - Statistical modeling 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.
- vincent - A Python to Vega translator.
- d3py](https://github.com/mikedewar/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.
- 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
- 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.
- NuPIC Studio An all-in-one NuPIC Hierarchical Temporal Memory visualization and debugging super-tool!
- SparklingPandas
- Seaborn - A python visualization library based on matplotlib
- bqplot
- pastalog - Simple, realtime visualization of neural network training performance.
- caravel - 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.
- 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.
- scikit-plot - A visualization library for quick and easy generation of common plots in data analysis and machine learning.

Neural networks

- Neural networks - NeuralTalk is a Python+numpy project for learning Multimodal Recurrent Neural Networks that describe images with sentences.
- Neuron neural networks learned with Gradient descent or LeLevenberg–Marquardt algorithm.
- Data Driven Code - Very simple implementation of neural networks for dummies in python without using any libraries, with detailed comments.

Ruby

Natural Language Processing

- Treat - Text REtrieval and Annotation Toolkit, definitely the most comprehensive toolkit I've encountered so far for Ruby

- Ruby Linguistics - Linguistics is a framework for building linguistic utilities for Ruby objects in any language. It includes a generic language-independent front end, a module for mapping language codes into language names, and a module which contains various English-language utilities.
- Stemmer - Expose libstemmer_c to Ruby
- Ruby Wordnet - This library is a Ruby interface to WordNet
- Raspel - raspell is an interface binding for ruby
- 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

- Ruby Machine Learning - Some Machine Learning algorithms, implemented in Ruby
- Machine Learning Ruby
- jRuby Mahout - JRuby Mahout is a gem that unleashes the power of Apache Mahout in the world of JRuby.
- 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
- Random Forester - Creates Random Forest classifiers from PMML files

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
- ruby-plot - gnuplot wrapper for ruby, especially for plotting roc curves into svg files
- plot-rb - A plotting library in Ruby built on top of Vega and D3.
- scruffy - A beautiful graphing toolkit for Ruby
- SciRuby
- Glean - A data management tool for humans
- Bioruby
- Arel

Misc

- Big Data For Chimps
- Listof](https://github.com/kevincobain2000/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

- 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](https://github.com/autumnai/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.

R

General-Purpose Machine Learning

- ahaz - ahaz: Regularization for semiparametric additive hazards regression
- arules - arules: Mining Association Rules and Frequent Itemsets
- biglasso - biglasso: Extending Lasso Model Fitting to Big Data in R
- bigrf - bigrf: Big Random Forests: Classification and Regression Forests for Large Data Sets
- `bigRR <http://cran.r-project.org/web/packages/bigRR/index.html) - bigRR: Generalized Ridge Regression (with special advantage for p >> n cases>`__
- 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.
- 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
- Cubist - Cubist: Rule- and Instance-Based Regression Modeling
- 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
- GAMBoost - GAMBoost: Generalized linear and additive models by likelihood based boosting
- 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
- 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
- Introduction to Statistical Learning
- ipred - ipred: Improved Predictors
- kernlab - kernlab: Kernel-based Machine Learning Lab
- klaR - klaR: Classification and visualization
- lars - lars: Least Angle Regression, Lasso and Forward Stagewise
- 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
- mboost - mboost: Model-Based Boosting
- medley - medley: Blending regression models, using a greedy stepwise approach
- mlr - mlr: Machine Learning in R
- mvpart - mvpart: Multivariate partitioning
- ncvreg - ncvreg: Regularization paths for SCAD- and MCP-penalized regression models
- nnet - nnet: Feed-forward Neural Networks and Multinomial Log-Linear Models
- oblique.tree - oblique.tree: Oblique Trees for Classification Data
- pamr - pamr: Pam: prediction analysis for microarrays
- party - party: A Laboratory for Recursive Partytioning
- partykit - partykit: A Toolkit for Recursive Partytioning
- penalized penalized estimation in GLMs and in the Cox model
- penalizedLDA - penalizedLDA: Penalized classification using Fisher's linear discriminant

- 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
- rattle - rattle: Graphical user interface for data mining in R
- rda - rda: Shrunken Centroids Regularized Discriminant Analysis
- rdetools in Feature Spaces
- REEMtree Data
- relaxo - relaxo: Relaxed Lasso
- rgenoud - rgenoud: R version of GENetic Optimization Using Derivatives
- rgp - rgp: R genetic programming framework
- Rmalschains in R
- rminer in classification and regression
- ROCR - ROCR: Visualizing the performance of scoring classifiers
- RoughSets - RoughSets: Data Analysis Using Rough Set and Fuzzy Rough Set Theories
- rpart - rpart: Recursive Partitioning and Regression Trees
- RPMM - RPMM: Recursively Partitioned Mixture Model
- RSNNS
- 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
- SDDA - SDDA: Stepwise Diagonal Discriminant Analysis
- SuperLearner](https://github.com/ecpolley/SuperLearner) and [subsemble - Multi-algorithm ensemble learning packages.
- svmpath - svmpath: svmpath: the SVM Path algorithm
- tgp - tgp: Bayesian treed Gaussian process models
- tree - tree: Classification and regression trees
- varSelRF - varSelRF: Variable selection using random forests
- XGBoost.R 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.

Data Analysis / Data Visualization

- ggplot2 - A data visualization package based on the grammar of graphics.

SAS

General-Purpose Machine Learning

- 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.

High Performance Machine Learning

- High Performance Data Mining - Data mining and machine learning that creates deployable models using a GUI or code in an MPP environment, including Hadoop.
- High Performance Text Mining - Text mining using a GUI or code in an MPP environment, including Hadoop.

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.
- FACTORIE - FACTORIE is a toolkit for deployable probabilistic modeling, 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.

Data Analysis / Data Visualization

- 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
- simmer - Reduce your data. A unix filter for algebird-powered aggregation.
- 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.
- Wolfe Declarative Machine Learning
- 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

- Conjecture - Scalable Machine Learning in Scalding
- brushfire - Distributed decision tree ensemble learning in Scala
- ganitha - scalding powered machine learning
- 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.

Swift

General-Purpose Machine Learning

- Swift AI - Highly optimized artificial intelligence and machine learning library written in Swift.
- 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.
- DeepLearningKit an Open Source Deep Learning Framework for Apple's iOS, OS X and tvOS. It currently allows using deep convolutional neural network models trained in Caffe on Apple operating systems.
- 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..