

# The SHOGUN Machine Learning Toolbox - Feature Comparison

Sören Sonnenburg, Gunnar Rätsch, Sebastian Henschel,  
 Christian Widmer, Jonas Behr, Alexander Zien, Fabio de Bona,  
 Alexander Binder, Christian Gehl, and Vojtěch Franc

Table 1: A comparison of shogun with the popular machine learning toolboxes weka, kernlab, dlib, nieme, orange, java-ml, pyML, mlp, pybrain, torch3. A '?' denotes unknown, '-' feature is missing. This table is available as a google spreadsheet from <http://spreadsheets.google.com/ccc?key=0Aunb9cCVAP6NdDVBmZY1TjdPcmx4ei1EeUZNNgtKUHc&hl=en>.

Feature		shogun	weka	kernlab	dlib	nieme	orange	java-ml	pyML	mlp	pybrain	torch3
General Features	Graphical User Interface	-	•	-	•	•	•	-	-	-	•	•
	One Class Classification	•	•	•	•	-	-	-	•	-	-	-
	Classification	•	•	•	•	•	•	•	•	•	•	•
	Multiclass classification	•	•	•	-	•	-	•	•	•	•	•
	Regression	•	•	•	•	•	•	-	•	-	•	•
	Structured Output Learning	•	-	-	-	•	-	-	-	-	-	-
	Pre-Processing	•	•	•	•	•	•	•	•	•	-	•
	Built-in Model Selection Strategies	-	•	•	•	-	•	•	•	-	-	-
	Visualization	-	•	-	-	•	•	-	•	•	•	•
	Test Framework	•	•	-	•	•	?	•	-	-	-	-
	Large Scale Learning	•	-	-	•	•	-	-	-	•	-	-
	Semi-supervised Learning	-	-	-	-	-	-	-	-	-	-	-
	Multitask Learning	•	-	-	-	-	-	-	-	-	-	-
	Domain Adaptation	•	-	-	-	-	-	-	-	-	-	-
	Serialization	•	•	•	•	•	•	•	•	•	•	-
	Parallelized Code	•	•	-	•	-	-	-	-	-	-	-
	Performance Measures (auROC etc)	•	•	-	•	•	•	•	•	•	•	•
	Image Processing	-	-	-	•	-	-	-	-	-	-	-
Supported Operating Systems	Linux	•	•	•	•	•	•	•	•	•	•	•
	Windows	•	•	•	•	•	•	•	-	•	•	•
	Mac OSX	•	•	•	•	•	•	•	•	•	-	•
	Other Unix	•	•	•	•	•	•	•	-	•	-	•
Language Bindings	Python	•	-	-	-	•	•	-	•	•	•	-
	R	•	-	•	-	-	-	-	-	-	-	-

Continued on next page...

Feature		shogun	weka	kernlab	dlib	nieme	orange	java-ml	pyML	mlpy	pybrain	torch3
	Matlab	•	-	-	-	-	-	-	-	-	-	-
	Octave	•	-	-	-	-	-	-	-	-	-	-
	C/C++	•	-	-	•	•	-	-	-	-	-	•
	Command Line	•	-	-	-	-	-	-	-	•	•	•
	Java	-	•	-	-	•	-	•	-	-	-	-
SVM Solvers	SVMLight	•	•	-	-	-	-	-	-	-	-	-
	LibSVM	•	•	•	•	•	•	•	•	-	•	-
	SVM Ocas	•	-	-	•	-	-	-	-	-	-	-
	LibLinear	•	•	-	-	-	-	-	-	-	-	-
	BMRM	•	-	-	-	-	-	-	-	-	-	-
	LaRank	•	-	-	-	-	-	-	-	-	-	-
	SVM Pegasos	-	•	-	•	•	-	-	-	-	-	-
	SVM SGD	•	-	-	-	-	-	-	-	-	-	-
	other	•	-	•	-	-	-	-	•	•	-	•
Regression	Kernel Ridge Regression	•	-	-	-	-	-	-	•	-	-	-
	Support Vector Regression	•	•	•	-	-	-	-	•	-	-	•
	Gaussian Processes	-	•	•	-	-	-	-	-	-	-	-
	Relevance Vector Machine	-	•	•	•	-	-	-	-	-	-	-
Multiple Kernel Learning	MKL	•	-	-	-	-	-	-	-	-	-	-
	q-norm MKL	•	-	-	-	-	-	-	-	-	-	-
Classifiers	Naive Bayes	•	•	-	-	-	•	-	-	-	•	•
	Bayesian Networks	-	•	-	•	-	-	-	-	-	•	-
	Multi Layer Perceptron	-	•	-	•	•	-	-	-	-	•	•
	RBF Networks	-	•	-	•	-	-	-	-	-	•	-
	Logistic Regression	•	•	?	-	•	•	-	-	-	-	-
	LASSO	-	-	?	-	•	-	-	-	-	-	-
	Decision Trees	-	•	-	-	-	•	•	-	-	-	-
	k-NN	•	•	•	•	-	•	•	•	•	•	•
Linear Classifiers	Linear Programming Machine	•	-	-	-	-	-	-	-	-	-	-
	LDA	•	-	-	-	-	-	-	-	•	-	-
Distributions	Markov Chains	•	-	-	-	-	-	•	-	-	-	-
	Hidden Markov Models	•	-	-	-	-	-	-	-	-	-	•
Kernels	Linear	•	•	•	•	•	•	•	•	•	•	•
	Gaussian	•	•	•	•	-	•	•	•	•	•	•
	Polynomial	•	•	•	•	-	•	•	•	•	•	•
	String Kernels	•	•	•	-	-	-	-	•	-	-	-
	Sigmoid Kernel	•	•	-	•	-	•	-	-	-	-	-
	Kernel Normalizer	•	?	•	-	-	-	-	•	-	-	-
Feature Selection	Forward	-	•	-	?	-	•	•	•	•	-	-

Continued on next page...

Feature		shogun	weka	kernlab	dlib	nieme	orange	java-ml	pyML	mlpy	pybrain	torch3
	Wrapper methods	-	•	-	?	-	?	•	•	•	-	-
	Recursive Feature Selection	-	•	-	•	-	?	•	•	•	-	-
Missing Features	Mean value imputation	-	•	-	-	-	•	•	-	•	-	-
	EM-based/model based imputation	-	•	-	-	-	•	-	-	-	-	-
Clustering	Hierarchical Clustering	•	•	-	-	-	•	-	-	•	-	-
	k-means	•	•	•	•	-	•	•	•	•	•	•
Optimization	BFGS	-	•	-	•	•	-	-	-	-	-	-
	conjugate gradient	-	-	-	•	-	-	-	-	-	-	-
	gradient descent	•	•	•	-	•	-	-	-	•	•	•
	bindings to CPLEX	•	-	-	-	-	-	-	-	-	-	-
	bindings to Mosek	-	-	-	-	-	-	-	-	-	-	-
	bindings to other solver	•	-	•	-	-	•	-	•	-	-	-
Supported File Formats	Binary	•	•	-	-	-	-	-	-	-	•	-
	Arff	-	•	-	-	-	-	•	-	-	-	-
	HDF5	•	-	•	-	-	-	-	-	-	-	-
	CSV	-	•	•	-	-	•	•	•	•	-	•
	libSVM/ SVMLight format	•	•	-	•	•	-	-	•	-	•	-
	Excel	-	-	•	-	-	•	-	-	-	-	-
Supported Data Types	Sparse Data Representation	•	•	-	•	•	•	•	•	•	•	-
	Dense Matrices	•	•	•	•	-	•	•	•	•	•	•
	Strings	•	•	•	•	-	-	-	-	-	-	•
	Support for native (e.g. C) types (char, signed and unsigned int8, int16, int32, int64, float, double, long double)	•	-	-	•	-	-	-	-	•	-	-

Table 2: General overview of popular machine learning toolboxes

	shogun	weka	kernlab	dlib	nieme	orange	java-ml	pyML	mlpy	pybrain	torch3
created	1999	1997	04-2004	2006	09-2006	06-2004	08-2008	08-2004	02-2008	10-2008	01-2002
last updated	03-2010	01-2010	10-2009	03-2010	03-2009	03-2010	08-2009	01-2009	11-2009	11-2009	11-2004
Main Language	C++	java	R	C++	C++	python	java	C++; python	python	python	C++
Main Focus	"Large Scale Kernel Methods"	String Features	SVMs"	General Purpose ML package	Kernel Based Classification/Dimensionality Reduction	"Portability"	Correctness"	Linear Regression; Ranking; Classification	Visual Data Analysis	Feature Selection	Kernel Methods