Sören Sonnenburg, Gunnar Rätsch, Fabio De Bona

† Fraunhofer FIRST.IDA, Kekuléstr. 7, 12489 Berlin, Germany
‡ Friedrich Miescher Laboratory of the Max Planck Society, Spemannstr. 39, 72076 Tübingen, Germany
Machine Learning Toolbox SHOGUN features algorithms:

- to learn 2-class classification and regression problems
- to train hidden markov models
- toolbox’s focus is on kernel methods esp. Support Vector Machines (SVMs)
- also implements a number of linear methods like Linear Discriminant Analysis (LDA), Linear Programming Machine (LPM), (Kernel) Perceptrons
Support Vector Machine

- given: points \( x_i \in X \ (i = 1, \ldots, N) \) with respective labels \( y_i \in \{-1, +1\} \)

- in training hyperplane that maximizes margin is chosen

Decision function \( f(x) = w \cdot x + b \)
SVM decision function in kernel feature space:

\[ f(x) = \sum_{i=1}^{N} y_i \alpha_i \Phi(x) \cdot \Phi(x_i) + b \]

• SVM decision function in kernel feature space:

\[ f(x) = \sum_{i=1}^{N} y_i \alpha_i \Phi(x) \cdot \Phi(x_i) + b \]  

• find parameters \( \alpha \) by solving quadratic optimization problem
• SHOGUN interfaces to Matlab™, Octave and Python and NEW! R

• provides generic SVM object interfacing to seven different SVM implementations, among them the state-of-the-art LibSVM and SVMlight

• SVMs can be trained using a variety of common kernels (efficient implementations for: Linear, Polynomial, Gaussian and Sigmoid Kernel, recent String Kernels)

• kernels can be combined; weighting can be learned using Multiple Kernel Learning.

• input feature-objects can be dense, sparse or strings and of type int/short/double/char; can be converted into different feature types.

• multiprocessor parallelization ⇒ able train on 10 million examples

... and many more...
• Support Vector Classification
  – Task: separate 2 clouds of gaussian distributed points in 2D

• Support Vector Regression
  – Task: learn a sine function

• Hidden Markov Model
  – Task: 3 loaded dice are drawn 1000 times, find out when which dice was drawn
Summary

- SHOGUN is a large scale machine learning toolbox ⇒ able to train on 10 million examples
- unified SVM framework
- Algorithms: HMM, LDA, LPM, Perceptron, SVM, SVR + many kernels

We need your help:

- Documentation
- Examples
- Testing
- Test Suite

Source Code is freely available under the GPLv2.
http://www.fml.tuebingen.mpg.de/raetsch/projects/shogun