kernlab: Machine Learning with kernels in R

Alexandros Karatzoglou
joint work with Alex Smola and Kurt Hornik

1 INSA de Rouen
LITIS
France

December 12, 2008
Outline

1. kernlab Introduction
2. Demo
**kernlab**

**R package:**
- Contains a wide range of kernel-based Machine Learning methods
- Uses modern Open Source R environment
- Extensible by exploiting the inherent modularity of kernel methods
- GPL 2
- Environment for statistical data analysis, inference and visualization.
- Ports for Unix, Windows and MacOSX
- Highly extensible through user-defined functions
- Generic functions and conventions for standard operations like plot, predict etc.
- $\sim$ 1200 add-on packages contributed by developers from all over the world
- e.g. Multivariate Statistics, Machine Learning, Natural Language Processing, Bioinformatics (Bioconductor).
- Interfaces to C, C++, Fortran, Java
9 kernel functions, 4 kernel expression functions
18 kernel-based ML methods:
- Regression, Quantile regression, Classification and Novelty detection (SVM, RVM, Gaussian Processes)
- Clustering (kernel $k$-means, Spectral Clustering)
- Ranking and dimensionality reduction (kernel PCA, kernel CCA, etc.) kernel-based two sample test (KMMD)

Utility Functions:
- Quadratic Problem solver
- Incomplete Cholesky Decomposition
Some kernel functions in kernlab

- Linear kernel
- Gaussian radial basis kernel
- Hyperbolic tangent kernel
- String kernel
- Polynomial kernel
- ...

Inputs

- kernlab methods take kernel function and data or kernel matrix as input
- Function parameters are given as options

Outputs

- Return objects that contain model parameters.
- Objects can be used with generic functions such as predict, plot etc.
Performance

- Nystrom Method for eigendecomposition based methods (e.g. spectral clustering). Eigenvectors can be calculated using only part of the kernel matrix yielding speedups and better scaling $O(n^3) + O(nN)$.
- Build in kernel function computations are vectorized
- C++ on performance bottlenecks
- Use of scalable algorithms
- ...
Features

Most function implement additional features such as e.g.

- Cross-validation (svm, GP’s)
- Model-selection in particular hyper-parameter estimation
- Data Scaling
- Extensive documentation of every function and options