

# Efficient cross-validation for kernelized least-squares regression with sparse basis expansions

Tapio Pahikkala · Hanna Suominen · Jorma Boberg

Received: 3 December 2010 / Accepted: 21 March 2012 / Published online: 13 April 2012  
© The Author(s) 2012

**Abstract** We propose an efficient algorithm for calculating hold-out and cross-validation (CV) type of estimates for sparse regularized least-squares predictors. Holding out  $H$  data points with our method requires  $O(\min(H^2n, Hn^2))$  time provided that a predictor with  $n$  basis vectors is already trained. In addition to holding out training examples, also some of the basis vectors used to train the sparse regularized least-squares predictor with the whole training set can be removed from the basis vector set used in the hold-out computation. In our experiments, we demonstrate the speed improvements provided by our algorithm in practice, and we empirically show the benefits of removing some of the basis vectors during the CV rounds.

**Keywords** Hold-out · Cross-validation · Regularized least-squares · Least-squares support vector machine · Kernel methods · Sparse basis expansions

## 1 Introduction

This paper considers using the regularized least-squares (RLS) algorithm (Rifkin et al. 2003; Poggio and Smale 2003), a kernel-based learning algorithm that is also known as the kernel

---

Editor: Paolo Frasconi.

---

T. Pahikkala (✉) · J. Boberg  
Department of Information Technology and Turku Centre for Computer Science, University of Turku,  
20014, Turku, Finland  
e-mail: [tapio.pahikkala@utu.fi](mailto:tapio.pahikkala@utu.fi)  
Fax: +358-2-3338600

J. Boberg  
e-mail: [boberg@utu.fi](mailto:boberg@utu.fi)

H. Suominen  
NICTA Canberra Research Laboratory, Canberra, Australia  
e-mail: [Hanna.Suominen@nicta.com.au](mailto:Hanna.Suominen@nicta.com.au)

H. Suominen  
College of Engineering and Computer Science, Australian National University, Canberra, Australia