

DEALING WITH LARGE DIAGONALS IN KERNEL MATRICES

JASON WESTON¹, BERNHARD SCHÖLKOPF¹, ELEAZAR ESKIN², CHRISTINA LESLIE^{2,3} AND
WILLIAM STAFFORD NOBLE⁴

¹*Max-Planck-Institut für biologische Kybernetik, Spemannstr. 38, D-72076 Tübingen, Germany,*
e-mail: jason.weston@tuebingen.mpg.de; bernhard.schoelkopf@tuebingen.mpg.de

²*Department of Computer Science, Columbia University, 450 Computer Science Building, MC 0401,*
1214 Amsterdam Avenue, New York, NY 10027, U.S.A., e-mail: eeskin@cs.columbia.edu;
cleslie@cs.columbia.edu

³*Center for Computational Biology (C2B2), Russ Berrie Pavilion, Columbia University, 1150 St.*
Nicholas Avenue, New York, NY 10032, U.S.A.

⁴*Department of Genome Sciences, University of Washington, Health Sciences Center, Box 357730,*
1705 NE Pacific Street, Seattle, WA 98195, U.S.A., e-mail: noble@gs.washington.edu

(Received June 3, 2002; revised September 9, 2002)

Abstract. In kernel methods, all the information about the training data is contained in the Gram matrix. If this matrix has large diagonal values, which arises for many types of kernels, then kernel methods do not perform well. We propose and test several methods for dealing with this problem by reducing the dynamic range of the matrix while preserving the positive definiteness of the Hessian of the quadratic programming problem that one has to solve when training a Support Vector Machine, which is a common kernel approach for pattern recognition.

Key words and phrases: Kernel methods, Support Vector Machines, pattern recognition, bioinformatics, microarray data analysis, transduction, regularization.