

Adaptive Sparse Grid Techniques for Data Mining

H.-J. Bungartz¹, D. Pflüger², and S. Zimmer³

Abstract: It was shown in [1] that the task of classification in data mining can be tackled by employing ansatz functions associated to grid points in the (often high dimensional) feature-space rather than using data-centered ansatz functions. To cope with the curse of dimensionality, sparse grids have been used. The problem obtained by regularisation was solved using the combination technique for sparse grids.

Based on this approach we propose an efficient finite-element-like discretization technique instead of the combination technique for classification. The main goal of our method is to make use of adaptivity to further reduce the number of grid points needed. Employing adaptivity in classification is reasonable as the target function contains flat regions as well as steep ones. We discuss strategies for the adaptive selection of grid points and show that special care has to be taken regarding the boundary values, as adaptive techniques commonly used for solving PDEs are not optimal here. Regarding implementational issues we present the algorithm for the fast multiplication of the vector of unknowns with the matrix resulting from the Laplacian. Results for some typical classification tasks, including problems from the UCI repository, are presented.

References:

- [1] J. Garcke, M. Griebel, and M. Thess. *Data Mining with Sparse Grids*. Computing 67(3), 2001, p. 225–253.

^{1,2,3} Department of Computer Science
Technische Universität München
Boltzmannstrasse 3, 85748 Garching, Germany
bungartz@in.tum.de, pflueged@in.tum.de, zimmer@in.tum.de