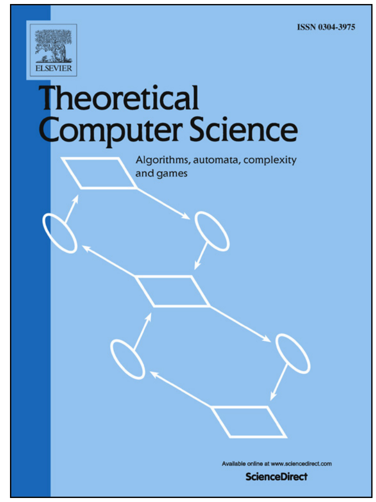
Accepted Manuscript



Large width nearest prototype classification on general distance spaces

Martin Anthony, Joel Ratsaby

PII: S0304-3975(18)30295-0
DOI: https://doi.org/10.1016/j.tcs.2018.04.045
Reference: TCS 11575

To appear in: *Theoretical Computer Science*

Received date: 27 June 2017
Revised date: 20 December 2017
Accepted date: 24 April 2018

Please cite this article in press as: M. Anthony, J. Ratsaby, Large width nearest prototype classification on general distance spaces, *Theoret. Comput. Sci.* (2018), https://doi.org/10.1016/j.tcs.2018.04.045

Large width nearest prototype classification on general distance spaces

Martin Anthony, Joel Ratsaby

Department of Mathematics, London School of Economics and Political Science, Houghton Street, London WC2A2AE, U.K.

Department of Electrical and Electronics Engineering, Ariel University of Samaria, ARIEL 40700, ISRAEL

Abstract

In this paper we consider the problem of learning nearest-prototype classifiers in any finite distance space; that is, in any finite set equipped with a distance function. An important advantage of a distance space over a metric space is that the triangle inequality need not be satisfied, which makes our results potentially very useful in practice. We consider a family of binary classifiers for learning nearest-prototype classification on distance spaces, building on the concept of large-width learning which we introduced and studied in earlier works. Nearest-prototype is a more general version of the ubiquitous nearest-neighbor classifier: a prototype may or may not be a sample point. One advantage in the approach taken in this paper is that the error bounds depend on a 'width' parameter, which can be sample-dependent and thereby yield a tighter bound.

1. Introduction

Learning Vector Quantization (LVQ) and its various extensions introduced by Kohonen [22] are used successfully in many machine learning tools and applications. Learning pattern classification by LVQ is based on adapting a fixed set of labeled prototypes in Euclidean space and using the resulting set of prototypes in a nearest-prototype rule (winner-take-all) to classify any point in the input space. As [21] mentions, LVQ fails if Euclidean representation is not well-suited for the data; and there have been extensions of LVQ to try to allow different metrics [21, 26] and take advantage of samples for which a more confident (or a large margin) classification can be obtained. Generalization error bounds with