Adaptive regularization of weight vectors

Koby Crammer · Alex Kulesza · Mark Dredze

Abstract We present AROW, an online learning algorithm for binary and multiclass problems that combines large margin training, confidence weighting, and the capacity to handle non-separable data. AROW performs adaptive regularization of the prediction function upon seeing each new instance, allowing it to perform especially well in the presence of label noise. We derive mistake bounds for the binary and multiclass settings that are similar in form to the second order perceptron bound. Our bounds do not assume separability. We also relate our algorithm to recent confidence-weighted online learning techniques. Empirical evaluations show that AROW achieves state-of-the-art performance on a wide range of binary and multiclass tasks, as well as robustness in the face of non-separable data.

Keywords Online learning · Supervised learning · Text classification · Adaptive regularization

1 Introduction

Online learning algorithms are fast and simple, make few statistical assumptions, and perform well in a wide variety of settings. The Perceptron algorithm is perhaps the oldest online machine learning algorithm, tracing its origins back to the 1950s. The Perceptron, which

Editor: Shai Shalev-Shwartz.

K. Crammer (✉)
Department of Electrical Engineering, The Technion, Haifa, 32000 Israel
e-mail: koby@ee.technion.ac.il

A. Kulesza
Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48109, USA
e-mail: kulesza@umich.edu

M. Dredze
Human Language Technology Center of Excellence, Johns Hopkins University, Baltimore, MD 21211, USA
e-mail: mdredze@cs.jhu.edu