Advanced Online Learning for Natural Language Processing

Koby Crammer
Department of Computer and Information Science
University of Pennsylvania
Philadelphia, PA 19104
`crammer@cis.upenn.edu`

Introduction: Most research in machine learning has been focused on binary classification, in which the learned classifier outputs one of two possible answers. Important fundamental questions can be analyzed in terms of binary classification, but real-world natural language processing problems often involve richer output spaces. In this tutorial, we will focus on classifiers with a large number of possible outputs with interesting structure. Notable examples include information retrieval, part-of-speech tagging, NP chucking, parsing, entity extraction, and phoneme recognition.

Our algorithmic framework will be that of online learning, for several reasons. First, online algorithms are in general conceptually simple and easy to implement. In particular, online algorithms process one example at a time and thus require little working memory. Second, our example applications have all been treated successfully using online algorithms. Third, the analysis of online algorithms uses simpler mathematical tools than other types of algorithms. Fourth, the online learning framework provides a very general setting which can be applied to a broad setting of problems, where the only machinery assumed is the ability to perform exact inference, which computes a maxima over some score function.

Goals: (1) To provide the audience systematic methods to design, analyze and implement efficiently learning algorithms for their specific complex-output problems: from simple binary classification through multi-class categorization to information extraction, parsing and speech recognition. (2) To introduce new online algorithms which provide state-of-the-art performance in practice backed by interesting theoretical guarantees.

Content: The tutorial is divided into two parts. In the first half we introduce online learning and describe the Perceptron algorithm (Rosenblatt, 1958) and the passive-aggressive framework (Crammer et al., 2006). We then discuss in detail an approach for deriving algorithms for complex natural language processing (Crammer, 2004). In the second half we discuss is detail relevant applications including text classification (Crammer and Singer, 2003), named entity recognition (McDonald et al., 2005), parsing (McDonald, 2006), and other tasks. We also relate the online algorithms to their batch counterparts.

References

K. Crammer and Y. Singer. 2003. A new family of online algorithms for category ranking. *Jornal of Machine Learning Research*, 3:1025–1058.

K. Crammer, O. Dekel, J. Keshet, S. Shalev-Shwartz, and Y. Singer. 2006. Online passive-aggressive algorithms. *JMLR*, 7:551–585.

K. Crammer. 2004. *Online Learning of Complex Categorial Problems*. Ph.D. thesis, Hebrew Universtiy.

R. McDonald, K. Crammer, and F. Pereira. 2005. Flexible text segmentation with structured multilabel classification. In *HLT/EMNLP*.

R. McDonald. 2006. *Discriminative Training and Spanning Tree Algorithms for Dependency Parsing*. Ph.D. thesis, University of Pennsylvania.

F. Rosenblatt. 1958. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65:386–407.