

## Guest editorial: special issue on learning theory

Sham Kakade · Ping Li

Published online: 24 April 2010  
© The Author(s) 2010

This special issue collects some of the most notable learning theory papers of 2008, which were chosen from the Conference on Learning Theory (COLT). The diversity of problems addressed in these papers reflects the significant progress being made in our theoretical understanding of the foundations of learning at both a computational and statistical level—topics include active learning, ranking, stability, graphical models, computational learning, to name a few.

The best student paper, by Balcan, Hanneke, and Wortman, provides a novel analysis of active learning, showing how in a variety of settings substantial asymptotic improvements are always possible under active learning. Shalev-Shwartz and Singer utilize an elegant minimax interpretation of boosting to provide more efficient and robust boosting algorithms. In a regret minimization setting, Hazan and Kale provide an algorithm which achieves even less regret when the variation in experts is low (thus making another connection to our concentration inequalities, particular those which utilize variance properties of random variables). Ailon and Mohri provide an efficient reduction of ranking with a classification algorithm: the simple algorithm cleverly uses binary classifiers with a QuickSort algorithm. Shamir and Tishby provide a general characterization of the stability of the widely-used  $k$ -means algorithm, whose analysis identifies the factors which influence the stability of clustering algorithms. Kleinberg, Niculescu-Mizil, and Sharma provide online algorithms where the action set (of the experts) is time varying. These algorithms are shown to be optimal in that they match the information theoretic lower bounds. Blais, O'Donnell, and Wimmer

---

S. Kakade (✉)

Department of Statistics, Wharton School, University of Pennsylvania, Philadelphia, USA  
e-mail: [shamkakade@gmail.com](mailto:shamkakade@gmail.com)

P. Li

Department of Statistical Science, Cornell University, Ithaca, USA  
e-mail: [pingli@cornell.edu](mailto:pingli@cornell.edu)

show how to use polynomial regression to perform agnostic learning under arbitrary product distributions and even under limited mixtures of product distributions. Zhou, Lafferty, and Wasserman demonstrate that if the covariance changes smoothly over time, then minimizing an  $\ell_1$ -penalized kernel risk function leads to good estimates of the covariance matrix—this, in turn allows estimation of time varying graphical structure and is also feasible in high dimensions.