THE UNIVERSITY OF CHICAGO

Department of Statistics

MASTER'S THESIS PRESENTATION

ERIN LIPMAN

Department of Statistics
The University of Chicago

Beyond Accuracy: Diversification in a Classification Setting

MONDAY, FEBRUARY 18, 2019, at 1:00 PM
Jones 304, 5747 S. Ellis Avenue

ABSTRACT

Recent work in such settings as recommender systems and information retrieval has studied the problem of trading off between selecting a set of relevant items and selecting a diverse set of items to present to a user. This paper brings this work into a classification setting in which, given items {x_1,..., x_n}, unknown labels y_i in {0,1}, and probabilities p_i=P(y_i=1 | x_i), we wish to select a set of (small, fixed) size n of items with high probabilities p_i, such that the set is spread out within the sample space.

We formulate this problem as a maximization of the expectation of a score function capturing set diversity on the subset of selected points that have y_i=1. In particular, we propose three possible score functions and explore the implications of the choice of score functions for the tradeoff between relevance and diversity for the selected subset. Finally, we demonstrate the use of our method on both simulated data and on MNIST digit classification.

For information about building access for persons with disabilities, please contact Keisha Prowoznik at 773.702-0541 or send an email to kprowoznik@galton.uchicago.edu. If you wish to subscribe to our email list, please visit the following web site: https://lists.uchicago.edu/web/subscribe/statseminars.