EFFICIENCY AND ROBUSTNESS IN SUPERVISED LEARNING

Anand Vidyashankar[1], Fengnan Deng[1], Giacomo Francisci[1], and Xiaoran Jiang[1]

[1] Department of Statistics, George Mason University, College of Engineering and Computing, Faitfax, VA 22030, (e-mail: avidyash@gmu.edu, fdeng2@gmu.edu, gfranci@gmu.edu, xjiang21@gmu.edu)

ABSTRACT: In recent years, there has been an increasing interest in building machine-learning systems that perform adequately when the training and test data differ. In the context of supervised learning, this problem has been addressed within the distributionally robust framework wherein the ambiguity set for the test distributions is allowed to vary within a neighborhood of the training distribution. While such methods are useful, the tradeoff between statistical efficiency and robustness remains unclear. Focusing on the out-of-distribution generalization problem, in this presentation, we describe a precise notion of statistical efficiency and relate the loss of efficiency to the gain in robustness in these contexts. We illustrate our ideas with examples from label shift estimation arising in diagnostic problems, privacy and utility in healthcare, and generalized adversarial networks.