学术报告

报 告 人 : 雷理骅助理教授 (斯坦福大学)

报告题目: Learning from a Biased Sample

报告时间: 2023年3月16日, 10:00-11:00

报告地点: 腾讯会议 593 806 120

发布平台: 北京大学统计中心

报告摘要: : The empirical risk minimization approach to data-driven decision making assumes that we can learn a decision rule from training data drawn under the same conditions as the ones we want to deploy it in. However, in a number of settings, we may be concerned that our training sample is biased, and that some groups (characterized by either observable or unobservable attributes) may be under- or over-represented relative to the general population; and in this setting empirical risk minimization over the training set may fail to yield rules that perform well at deployment. We propose a model of sampling bias called Γ-biased sampling, where observed covariates can affect the probability of sample selection arbitrarily much but the amount of unexplained variation in the probability of sample selection is bounded by a constant factor. Applying the distributionally robust optimization framework, we propose a method for learning a decision rule that minimizes the worst-case risk incurred under a family of test distributions that can generate the training distribution under Γ-biased sampling. We apply a result of Rockafellar and Uryasev to show that this problem is equivalent to an augmented convex risk minimization problem. We give statistical guarantees for learning a model that is robust to sampling bias via the method of sieves, and propose a deep learning algorithm whose loss function captures our robust learning target. We empirically validate our proposed method in simulations and a case study on ICU length of stay prediction.

报告人简介 :雷理骅，斯坦福大学商学院经济学组助理教授。2014年毕业于北京大学数学科学学院，2019年于加州大学伯克利分校获得统计学博士学位，师从Peter Bickel教授和Michael Jordan教授。2019-2022年在斯坦福大学与Emmanuel Candès教授完成了博士后工作。主要研究方向是因果推断、计量经济学、conformal inference、多重检验、网络聚类、高维统计推断、随机优化。