



北京理工大学

数学与统计学院学术报告

Leveraging unlabeled data and model averaging to improve prediction

报告人: 张新雨 中科院数学与系统科学研究院

时间: 2024年5月30日, 16:00-17:00

地点: 北京理工大学良乡校区文萃楼F205

摘要: The available data in semi-supervised learning usually consists of relatively small sized labeled data and much larger sized unlabeled data. How to effectively exploit unlabeled data is the key issue. In this paper, we consider the prediction of the regression function. Write regression function in the form of a copula and marginal distributions, and the unlabeled data can be exploited to improve the estimation of the marginal distributions. The predictions of the different copulas are weighted, where the weights are obtained by minimizing an estimate of the risk. Error-ambiguity decomposition of the risk is performed such that unlabeled data can be exploited to improve the estimation of the risk. We demonstrate the asymptotic normality of copula parameters and regression function estimates of the candidate model under the semi-supervised framework, as well as the asymptotic optimality and weight consistency of model averaging estimates. Our model averaging estimate achieves faster convergence rates of asymptotic optimality and weight consistency than the supervised counterpart. Extensive simulation experiments demonstrate the effectiveness of the proposed method.

个人简介:

张新雨, 中国科学院数学与系统科学研究院预测中心研究员。主要从事统计学和计量经济学的理论和应用研究工作, 具体研究方向包括模型平均、管理统计、经济预测、机器学习等, 多篇论文发表在计量经济学和统计学顶级期刊。担任SCI期刊《JSSC》领域主编和其他5个国内外重要期刊的编委, 是管理科学与工程学会常务理事、国际统计学会当选会员, 先后主持国自科优秀和杰出青年基金项目, 曾获中国青年科技奖。