学校编码：10384　　分类号_____密级_____
学号：　　UDC_____

厦门大学

硕　士　学　位　论　文

单指标模型中的变量选择和参数估计：基于最大化距离协方差模型

Variable Selection and Direction Estimation for Single-index Models via Distance Covariance

刘西

指导教师姓名：　马双鸽教授，钟威副教授
专　业　名　称：　统计学
论文提交日期：　2016 年 05 月
论文答辩时间：　2016 年 05 月
学位授予日期：　2016 年 05 月

答辩委员会主席：_____
评　　阅　　人：_____

年　　月

厦门大学学位论文原创性声明

本人呈交的学位论文是本人在导师指导下,独立完成的研究成果。本人在论文写作中参考其他个人或集体已经发表的研究成果，均在文中以适当方式明确标明，并符合法律规范和《厦门大学研究生学术活动规范（试行）》。

另外，该学位论文为（　　）课题（组）的研究成果，获得（　　）课题（组）经费或实验室的资助，在（　　）实验室完成。（请在以上括号内填写课题或课题组负责人或实验室名称，未有此项声明内容的，可以不作特别声明。）

声明人（签名）：

年　　月　　日

厦门大学学位论文著作权使用声明

本人同意厦门大学根据《中华人民共和国学位条例暂行实施办法》等规定保留和使用此学位论文，并向主管部门或其指定机构送交学位论文（包括纸质版和电子版），允许学位论文进入厦门大学图书馆及其数据库被查阅、借阅。本人同意厦门大学将学位论文加入全国博士、硕士学位论文共建单位数据库进行检索，将学位论文的标题和摘要汇编出版，采用影印、缩印或者其它方式合理复制学位论文。

本学位论文属于：

（　　）1.经厦门大学保密委员会审查核定的保密学位论文，于　　年　　月　　日解密，解密后适用上述授权。

（　　）2.不保密，适用上述授权。

（请在以上相应括号内打“√”或填上相应内容。保密学位论文应是已经厦门大学保密委员会审定过的学位论文，未经厦门大学保密委员会审定的学位论文均为公开学位论文。此声明栏不填写的，默认为公开学位论文，均适用上述授权。）

声明人（签名）：

年　　月　　日

摘要

在本文中，我们提出了两种在单指标模型中同时进行变量选择和参数估计的方法：最大化带惩罚项的距离协方差法(PeDcov)和带阀值梯度正则化方法优化距离协方差法(DC-TGDR)。在最大化单指标和响应变量之间距离协方差的过程中，通过惩罚或正则化单指标方向参数，我们可以有效的筛选出重要变量，并排除无关变量。与文献中已有的方法相比，新提出的两种方法从充分降维的角度出发，避免了非参数联系函数的估计，且可以解决响应变量是离散变量时的方向参数估计问题。再者，新提出的DC-TGDR方法鼓励变量成组选择，即它倾向于将高度相关的解释变量同时保留在或者剔除出模型。因为DC-TGDR方法具有正则化性质，当自变量的维数很高但真实模型是稀疏模型时，DC-TGDR方法的估计结果稳健且计算速度快。因此，DC-TGDR方法可以分析自变量维数大于样本容量的高维数据。我们进行了蒙特卡洛模拟实验且分析了现实中的数据集，计算结果表明了新提出的两种方法在有限样本情况下的有效性。

关键词：单指标模型；变量选择；距离协方差

Abstract

In this paper, we propose two new methods, Penalized Distance covariance (PeDcov) method and maximizing Distance Covariance via Threshold Gradient Directed Regularization (DC-TGDR) method, to select significant covariates and estimate the single-index direction simultaneously for single-index models. When utilizing regularization methods to maximize the distance covariance between single index and response variable, we can obtain a sparse estimation of single-index direction. Compared with other methods, our methods keep model-free advantage from the view of sufficient dimension reduction approaches, and avoid estimating the nonparametric link function. In addition, the newly proposed methods can conduct variable selection in single-index models even when the response and covariates are discrete variables. Besides, DC-TGDR method encourages grouping selection, which can keep the highly correlated covariates in or out of the model together. When the dimension of covariates is high and the true model is sparse, DC-TGDR method maintains stability and efficiency due to its regularization property. We examine the finite sample performance of the newly proposed methods by Monte Carlo simulation and real data analysis.

Key Words: Single-index model; Variable selection; Distance covariance.

目 录

Table of Contents