World Academy of Science, Engineering and Technology
International Journal of Mathematical and Computational Sciences
Vol:10, No:01, 2016

Generalized Additive Model for Estimating Propensity Score

Authors : Tahmidul Islam

Abstract : Propensity Score Matching (PSM) technique has been widely used for estimating causal effect of treatment in observational studies. One major step of implementing PSM is estimating the propensity score (PS). Logistic regression model with additive linear terms of covariates is most used technique in many studies. Logistics regression model is also used with cubic splines for retaining flexibility in the model. However, choosing the functional form of the logistic regression model has been a question since the effectiveness of PSM depends on how accurately the PS been estimated. In many situations, the linearity assumption of linear logistic regression may not hold and non-linear relation between the logit and the covariates may be appropriate. One can estimate PS using machine learning techniques such as random forest, neural network etc for more accuracy in non-linear situation. In this study, an attempt has been made to compare the efficacy of Generalized Additive Model (GAM) in various linear and non-linear settings and compare its performance with usual logistic regression. GAM is a non-parametric technique where functional form of the covariates can be unspecified and a flexible regression model can be fitted. In this study various simple and complex models have been considered for treatment under several situations (small/large sample, low/high number of treatment units) and examined which method leads to more covariate balance in the matched dataset. It is found that logistic regression model is impressively robust against inclusion quadratic and interaction terms and reduces mean difference in treatment and control set equally efficiently as GAM does. GAM provided no significantly better covariate balance than logistic regression in both simple and complex models. The analysis also suggests that larger proportion of controls than treatment units leads to better balance for both of the methods.

Keywords : accuracy, covariate balances, generalized additive model, logistic regression, non-linearity, propensity score matching
Conference Title : ICCSDA 2016 : International Conference on Computational Statistics and Data Analysis
Conference Location : Istanbul, Türkiye
Conference Dates : January 25-26, 2016