

Abstract

Dr. Yasser A Yakout Shehata

A Randomization Method to Correctly Estimate Overall Significance in Best Subsets Regression

Best subsets regression is often used to identify a good regression model. The standard approach to assess statistical significance for a best subsets regression model is flawed. A computationally intensive randomization algorithm which corrects the problem in all instances is outlined and implemented. Simulation studies show that under the null hypothesis the mean difference in p-values between the two methods is in excess of 0.3 for a small number of predictors and typically increases with increasing number of predictors. The proposed method is shown to retain good comparable power in a non-null situation.