

Kriteria GSP pada metode subset selection dalam kasus outlier dan multikolinieritas = The GSP criterion on subset selection method in the presence of outlier and multicollinearity

Agnes Nabila, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20466192&lokasi=lokal>

Abstrak

ABSTRAK

Analisis regresi linier memodelkan hubungan linier antara variabel respon dengan variabel regresor. Analisis regresi linier menghasilkan suatu model yang disebut model regresi linier. Metode yang sering digunakan untuk menaksir parameter pada model regresi linier adalah metode ordinary least square OLS. Metode OLS akan menghasilkan taksiran terbaik ketika semua asumsinya terpenuhi. Namun pada kenyataannya, asumsi-asumsi tersebut seringkali tidak terpenuhi. Asumsi yang seringkali tidak terpenuhi adalah adanya multikolinieritas dan outlier. Multikolinieritas akan menyebabkan variansi dari taksiran parameter regresi membesar, sedangkan outlier akan membuat taksiran parameter menjadi bias. Jika terdapat multikolinieritas dan outlier pada data, digunakan jackknife ridge M-estimator. Jackknife ridge M-estimator adalah taksiran koefisien regresi yang memiliki sifat robust sehingga tidak terpengaruh oleh outlier dan menggunakan metode ridge untuk mengatasi masalah multikolinieritas serta menggunakan metode jackknife untuk mereduksi bias yang dihasilkan metode ridge. Pada regresi linier berganda, model regresi dapat terdiri dari banyak kandidat variabel regresor. Metode subset selection digunakan untuk memilih sebagian kecil saja dari kandidat variabel regresor pada model regresi linier berganda yang paling baik dalam memprediksi nilai dari variabel respon. Kriteria seleksi yang dapat digunakan pada metode subset selection apabila terdapat outlier dan multikolinieritas pada data yaitu kriteria GSp yang merupakan generalized version dari kriteria Sp. Hal tersebut dikarenakan, kriteria GSp didasarkan oleh jackknife ridge M-estimator yang dapat mengatasi masalah outlier dan multikolinieritas. Selanjutnya, kriteria GSp diimplementasikan untuk mendapatkan model terbaik pada data IQ yang memiliki masalah outlier dan multikolinieritas. Berdasarkan analisis data, diperoleh bahwa untuk mengetahui IQ seseorang orang tersebut tidak harus melakukan 5 tes kepribadian berbeda, karena dengan hanya melakukan 3 tes kepribadian saja yaitu tes 1, tes 3 dan tes 5, sudah dapat diketahui besar IQ dari orang tersebut, dimana hal tersebut dapat mengurangi waktu, biaya dan tenaga yang diperlukan.

ABSTRACT

Linear regression analysis model the linear relation between response variable with regressor variables. Linear regression analysis produce a linear regression model. Ordinary least squares OLS method is often utilized to estimate parameters of linear regression model. OLS method will produce the best estimates when all the assumptions are met. However, in real data, those assumptions are often not violated. Such as the multicollinearity and outliers. Multicollinearity will produce a large variance of the regression parameters, while outliers will cause a biased estimates. Jackknife ridge M estimator is recommended to be implemented in the model when these problems are present. Jackknife ridge M estimator is the regression with robust property, which makes it able not to be influenced by the presence of outlier, and ridge method is utilized to overcome multicollinearity, where the jackknife method is utilized to reduce the bias from the

result of the ridge method. In multiple linear regression, regression model could possibly contain many regressor variable candidates. Subset selection method is utilized to select some of those regressor variable candidates on the multiple linear regression model that best predicts the value of the response variable. The selection criterion that can be utilized on the subset selection method when outlier and multicollinearity are present is the GSp criterion which is the generalized version of the Sp criterion. The reason for this is because the GSp criterion is based on jackknife ridge M estimator, which is able to solve the problem of outlier and multicollinearity. Furthermore, GSp criterion is implemented to obtain the best model on IQ data that has outlier and multicollinearity issues. Based on the data analysis, it was found that to know the IQ of a person the person does not need to do 5 different personality tests, because by only doing 3 personality test which is test 1, test 3 and test 5, it can be known the IQ of that person, where it can reduce the time, cost and energy required.