

INTISARI

SELEKSI VARIABEL REGRESI MENGGUNAKAN BOOTSTRAP LASSO

Oleh

Illuminata Wynn timer
14/363533/PA/15848

Regresi logistik digunakan untuk memprediksi variabel respon yang biner dengan satu set variabel prediktor. Dalam regresi logistik, asumsi yang harus dipenuhi adalah tidak ada multikolinearitas dan jika tidak terpenuhi maka estimasi parameternya akan jauh dari nilai seharusnya dan membuat model regresi yang diperoleh menjadi tidak efisien lagi karena nilai standar error koefisien regresi menjadi sangat besar (*overestimate*).

Metode Least Absolute Shrinkage and Selection Operator (LASSO) dikenal sebagai metode yang digunakan untuk mengatasi data yang memiliki multikolinearitas. LASSO akan menyusutkan koefisien (parameter β) yang berkorelasi, menjadi nol atau mendekati nol, dan sekaligus melakukan seleksi variabel regresi. Namun, LASSO cenderung akan memilih variabel yang relevan dan tidak relevan. Pada skripsi ini, diusulkan metode alternatif yaitu Bootstrap LASSO untuk menyeleksi variabel regresi yang relevan sehingga dapat digunakan untuk mengatasi multikolinearitas. Pada akhirnya, performa model regresi logistik Bootstrap LASSO akan dibandingkan dengan model regresi logistik LASSO. Dengan melihat nilai proporsi konkordansi, metode Bootstrap LASSO dianggap lebih baik daripada regresi logistik LASSO karena memiliki nilai proporsi konkordansi yang lebih tinggi.

Kata kunci : Bootstrap LASSO, LASSO, multikolinearitas, regresi logistik, *Coordinate Descent*, *Generalized Linear Model*

ABSTRACT

VARIABLE SELECTION OF REGRESSION USING BOOTSTRAP LASSO

By

ILLUMINATA WYNNIE
14/363533/PA/15848

Logistic regression is used to predict binary response variables with a set of predictor variables. In logistic regression, the assumption that must be met is that there is no multicollinearity and if not met then the parameter estimation will be far from the value and make the regression model obtained becomes inefficient because the standard error value of the regression coefficient becomes overestimate.

The Least Absolute Shrinkage and Selection Operator (LASSO) method is known as the method used to overcome data that has multicollinearity. LASSO will shrink the correlated coefficients (parameters), to zero or near zero, and simultaneously perform the selection of regression variables. Therefore, LASSO will produce a more representative end model. In this thesis, it is proposed an alternative method of Bootstrap method applied to LASSO estimation which is equally used to overcome multicollinearity in regression variable selection. Ultimately, the performance of the LASSO Bootstrap logistic regression model will be compared with the LASSO logistic regression model. By looking at the concordance proportion value, the LASSO Bootstrap method is considered better than the LASSO logistic regression because it has a higher concordance proportion value.

Keywords : *Bootstrap LASSO, LASSO, multicollinearity, logistic regression, Coordinate Descent, Generalized Linear Model*