

VARIABLE SELECTION USING
LEAST ABSOLUTE SHRINKAGE AND SELECTION OPERATOR

RAHAINI BINTI MOHD SAID

A thesis submitted in partial fulfillment of
the requirements for the award of the degree of
Master of Science (Mathematics)

Faculty of Science
Universiti Teknologi Malaysia

MAY 2011

**Special dedicated to
My beloved family and friends**

ACKNOWLEDGEMENT

First of all, I would like to express my sincere appreciation to my supervisor, Assoc. Prof. Dr Ismail Mohamad for his encouragement and guidance as well as suggestions.

Furthermore, I am grateful to Universiti Teknologi Malaysia for funding my master study, the librarians at Sultanah Zanariah Library (PSZ), UTM plays an important role in supplying the relevant literatures and resources used in this research.

I would also like to say thank all my friends and colleagues that helps me to complete my thesis.

ABSTRACT

Least Absolute Shrinkage and Selection Operator (LASSO) and Forward Selection are variable selection methods that are implemented in this study. The objectives of this study are to apply forward selection method in variable selection for a regression model, to apply LASSO method in variable selection for a regression model using quadratic programming and leave one out cross validation and choosing the better model obtained from forward selection and LASSO method using least mean square error. The forward selection method implemented in the statistical package for social sciences (SPSS). Quadratic programming technique and leave one out cross validation from MATLAB software is applied to solve LASSO. The analyzed result showed forward selection and LASSO are chosen the same variable that should be included in the model. However the coefficient of the regression for both models differ. To choose between the two models, MSE is used as the criteria where the model with the smallest MSE is taken as the best model. The MSE for forward selection and LASSO are 0.4959 and 0.4765 respectively. Thus, LASSO is the better model compared to forward selection model.

ABSTRAK

“Least absolute shrinkage and selection operator” (LASSO) dan “forward selection” merupakan pilihan kaedah pembolehubah dalam kajian ini. Tujuan kajian ini adalah untuk melaksanakan kaedah “forward selection” dalam pemilihan pembolehubah untuk model regresi, untuk melaksanakan kaedah LASSO dalam pemilihan pembolehubah untuk model regresi menggunakan pengaturcaraan kuadratik dan “leave one-out cross validation”, memilih model yang lebih baik daripada “forward selection” dan LASSO dengan menggunakan memilih purata ralat kuasa dua. Kaedah regresi “forward selection” dilaksanakan dalam pakej statistik untuk ilmu sosial (SPSS) dan teknik pengaturcaraan kuadrat dan “leave one out cross validation” daripada perisian MATLAB diterapkan untuk menyelesaikan LASSO. Keputusan analisis menunjukkan “forward selection” dan LASSO adalah memilih pembolehubah yang sama yang harus dimasukkan ke dalam model. Namun pekali regresi untuk kedua model adalah berbeza. Untuk memilih antara dua model, purata ralat kuasa dua digunakan sebagai kriteria yang mana model dengan purata ralat kuasa dua terkecil diambil sebagai model terbaik. Purata ralat kuasa dua untuk “forward selection” dan LASSO adalah 0.4959 dan 0.4765 masing-masing. Dengan demikian, LASSO adalah model lebih baik berbanding dengan “forward selection”.