

PENDUGAAN KOEFISIEN REGRESI UNTUK DATA MENGANDUNG MULTIKOLINEARITAS MENGGUNAKAN PLS

Yuliani Setia Dewi

Jurusan Matematika Fakultas MIPA Universitas Jember

Abstract : Correlation between predictor variables (multicollinearity) become a problem in regression analysis. There are some methods to solve the problem. Each method has complexity. One of those methods is PLS. This research aims to describe the procedur of Estimation of PLS regression. PLS combine principal component analysis and multiple linear regression. PLS method estimates regression coefficient by selecting the number of component that is used in the model that has optimum Mean Square Error (minimum MSE). Optimal MSE is taken from cross validation.

Keywords: PLS, multicollinearity, MSE, cross validation.

I. PENDAHULUAN

Dalam analisis regresi , terkadang kita jumpai kondisi terdapatnya korelasi antar variabel bebas (variabel prediktor) atau yang biasa disebut dengan istilah multikolinearitas. Multikolinearitas menjadi suatu masalah dalam analisis regresi, terutama dalam regresi linear standar (OLS). Adanya multikolinearitas yang tinggi tidak memungkinkan melihat pengaruh variabel bebas terhadap variabel respon secara terpisah (Gujarati, 1992).

Terdapat beberapa metode untuk mengatasi masalah multikolinearitas ini. Masing-masing metode mempunyai kekomplekan. Salah satu metode untuk mengatasi masalah tersebut adalah PLS.

PLS dapat digunakan untuk pemodelan yang mengandung sejumlah besar regressor/varilabel bebas. PLS pertama kali populer penerapannya dalam bidang kemometrik (Geladi, 1992). Kemudian berkembang dan digunakan dalam bidang-bidang lain. Datta(2001) menggunakan PLS untuk konteks data microarray. Namun demikian, meskipun metode ini sudah lama diperkenalkan (tahun 1960an) sifat-sifat statistiknya relatif baru dipelajari (Frank & Friedman, 1993). Datta et al (2007) menggunakan metode PLS dan LASSO untuk memodelkan waktu daya tahan hidup pasien dalam konteks data microarray tersensor.

Regresi PLS merupakan teknik baru yang menjeneralisasi dan mengkombinasikan analisis komponen utama dan regresi berganda (Abdi, 2006). PLS mereduksi dimensi variabel-variabel penjelas asal melalui pembentukan variabel-variabel laten dengan