ABSTRAK

ANALISIS KOMPONEN UTAMA

Analisis komponen utama adalah salah satu teknik dalam statistik multivariat. Tujuan umumnya adalah untuk mereduksi data dan intepretasi. Analisis komponen utama merupakan teknik membentuk variabel baru (yang disebut komponen utama) yang merupakan kombinasi linear dari variabel asal. Banyaknya variabel baru yang dibentuk sama dengan banyaknya variabel asal, dan variabel baru tidak berkorelasi satu dengan yang lain. Model dari komponen utama adalah sebagai berikut :

$$\xi_1 = w_{11}x_1 + w_{12}x_2 + \ldots.. + w_{1p}x_p$$

$$\xi_2 = w_{21}x_1 + w_{22}x_2 + \ldots. + w_{2p}x_p$$

$$.$$

$$.$$

$$\xi_p = w_{p1}x_1 + w_{2p}x_2 + \ldots. + w_{pp}x_p$$

di mana $\xi_1, \xi_2, \ldots, \xi_p$ adalah p komponen utama dan w_{ij} adalah bobot variabel ke j *untuk komponen utama* ke i.

Komponen utama *pertama menerangkan variansi maksimum dari data*, komponen utama kedua menerangkan variansi maksimum dari data yang tidak diterangkan oleh komponen utama pertama dan seterusnya. Jika total variansi sejumlah besar data diterangkan oleh beberapa komponen utama atau variabel baru tersebut, maka komponen utama tersebut dapat digunakan untuk tujuan intepretasi atau analisa lanjutan terhadap data dari p variabel asal. Jumlah data yang tereduksi akan cukup besar jika nilai p besar. Analisis komponen utama dapat dikerjakan dengan dua cara yaitu mencari eigenstruktur matriks variansi kovariansi dari data asal dan juga mencari dekomposisi nilai singular dari data matriks. Dekomposisi nilai singular lebih mudah dikerjakan karena skor komponen utama dapat dicari dengan perkalian matriks dekomposisinya.

Analisis komponen utama dapat diterapkan dalam analisis regresi untuk menyelesaikan masalah multikolinearitas dengan cara menggantikan variabel asal dengan komponen utama. Penduga paramenter regresi dapat dicari dengan menggantikan variabel bebasnya dengan komponen utama dan persamaan akhirnya diperoleh dari subsitusi variabel asal yang telah distandarisasi ke komponen utama.

ABSTRACT

PRINCIPAL COMPONENT ANALYSIS

Principal component analysis is one of the multivariate statistical analysis techniques. Its general objectives are data reduction and intrepretation. Principal component analysis is a technique to form new variables (called principal components), each of which is a linear combination of the original variables. The maximum number of the new variables that can be formed is equal to the number of the original variables. The new variables are uncorrelated with each other. The principal component model is as follows:

$$\xi_1 = w_{11}x_1 + w_{12}x_2 + \ldots.. + w_{1p}x_p$$

$$\xi_2 = w_{21}x_1 + w_{22}x_2 + \ldots. + w_{2p}x_p$$

$$\vdots$$

$$\xi_p = w_{p1}x_1 + w_{2p}x_2 + \ldots. + w_{pp}x_p$$

where $\xi_1, \xi_2, \ldots, \xi_p$ are the p principal components and w_{ij} is the weight of the j th variable for i th principal component.

The first principal component, ξ_1, accounts for the maximum variance in the data, the second principal component, ξ_2, accounts for the maximum variance that has not been accounted for by the first principal component, and so on. If a substansial amount of the total variance in the data is accounted for by a few principal components or new variables, then we can use these few principal components for intrepretational purporse or further data analysis of the original p variables. This would result in a substansial amount of data reduction if the value of p is large. Principal component analysis can be done by finding the eigenstructure of the covariance matrix of the original data, as well as by finding the singular value decomposition (SVD) of the data matrix. Singular value decomposition is easier to use because the principal components'scores are given by the product of its matrix decomposition.

In regression, principal component analysis can be used in solving the multicolinearity problem. by substituting the original variables by the principal components. The regression parameter estimates can be found by substituting its independent variables by the principal components, and the last regression equation is obtained by substituting the standardized original variables to the principal components.