

Algoritma Nonlinear Iterative Partial Least Squares (NIPALS) untuk Reduksi Dimensi pada Data dengan Missing Value = Nonlinear Iterative Partial Least Squares (NIPALS) Algorithm for Dimensionality Reduction in Data with Missing Value

Rizqa Fatika Fajrianti, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=9999920530560&lokasi=lokal>

Abstrak

Prinsip parsimoni adalah prinsip yang menyatakan bahwa jika terdapat beberapa penjelasan untuk suatu fenomena, maka penjelasan paling sederhanalah yang harus dipilih. Prinsip ini digunakan dalam analisis data untuk memilih model yang paling efisien dalam menjelaskan variabilitas data dengan parameter seminimal mungkin. Namun pada beberapa kondisi, data bisa saja melibatkan pengukuran atau variabel yang cukup banyak. Data berdimensi tinggi dapat menyebabkan kompleksitas dan kesulitan dalam analisis, sehingga reduksi dimensi pada data penting untuk dilakukan. Principal Component Analysis (PCA) adalah salah satu metode yang dapat digunakan untuk melakukan reduksi dimensi, dengan mengekstraksi variabel baru dan mengurangi pengaruh dari variabel yang tidak relevan. Namun, metode PCA tidak toleran terhadap missing value, sehingga algoritma Nonlinear Iterative Partial Least Squares (NIPALS) dapat digunakan dalam mengatasi data yang mengandung missing value. Performa dari algoritma NIPALS dievaluasi menggunakan nilai normalized root mean square error (NRMSE) dan koefisien korelasi Pearson. Kemudian, performa dari algoritma ini dibandingkan dengan dua metode lain, meliputi Probabilistic Principal Component Analysis (PPCA) dan SVDImpute. Setelah dilakukan percobaan sebanyak seratus kali pada data survei COVIDiSTRESS, didapatkan hasil bahwa algoritma NIPALS memiliki performa yang lebih baik dan stabil dalam melakukan reduksi dimensi dibandingkan SVDImpute dan PPCA pada data dengan missing value sebesar 1% hingga 15%.

.....The principle of parsimony, states that if there are multiple explanations for a phenomenon, the simplest explanation should be chosen. This principle is applied in data analysis to select the most efficient model that explains the variability of the data with minimal parameters. However, in some cases, the data may involve a large number of measurements or variables. High-dimensional data can lead to complexity and difficulties in analysis, therefore dimensionality reduction is important. Principal Component Analysis (PCA) is one method that can be used for dimensionality reduction by extracting new variables and reducing the influence of irrelevant variables. However, PCA is not tolerant to missing values, so the Nonlinear Iterative Partial Least Squares (NIPALS) algorithm can be used to handle data with missing values. The performance of the NIPALS algorithm is evaluated using the normalized root mean square error (NRMSE) and Pearson correlation coefficient. Subsequently, the performance of this algorithm is compared with two other methods, including Probabilistic Principal Component Analysis (PPCA) and SVDImpute. After conducting a hundred trials on the COVIDiSTRESS survey data, it was found that the NIPALS algorithm performed better and was more stable in dimension reduction compared to SVDImpute and PPCA algorithms on data with missing values ranging from 1% to 15%.