

Aplikasi metode bicluster-based robust least square estimation dengan principal components pada imputasi missing value data ekspresi gen = Missing value imputation for microarray data using bicluster-based robust least squares estimation with principal components

Clive Nathaniel, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20516295&lokasi=lokal>

Abstrak

Missing value merupakan masalah yang sering ditemukan pada analisis data ekspresi gen. Salah satu metode yang sering digunakan untuk mengatasi masalah ini adalah dengan melakukan imputasi. Imputasi adalah proses mengganti missing value pada data dengan nilai pengganti yang didapat dari metode tertentu. Pada skripsi ini dibahas mengenai suatu metode imputasi untuk data ekspresi gen yang merupakan pengembangan dari metode imputasi robust least squares estimation dengan principal components (RLSP) dengan menggunakan konsep biclustering. Metode ini dinamakan bicluster-based robust least squares estimation dengan principal components (bi-RLSP). Metode RLSP adalah metode imputasi data ekspresi gen yang menggunakan konsep k-nearest neighbor, principal component analysis, dan regresi kuantil. Konsep biclustering ingin diterapkan untuk menggantikan k-nearest neighbor yang mencari baris yang mirip untuk semua kolom, sehingga dapat dicari baris yang mirip serta kolom yang mirip secara bersamaan. Metode bi-RLSP diaplikasikan untuk mengimputasi missing values pada data ekspresi gen, dimana metode ini ditemukan memiliki akurasi yang lebih tinggi dibandingkan metode RLSP dan metode imputasi row average dengan melihat nilai NRMSE dari estimasi missing value pada metode-metode ini.

.....Missing values are a common problem in gene expression data analysis. One of the methods used to overcome this problem is by performing imputation. Imputation is the process of replacing missing values in data with values obtained from certain methods. This thesis discusses an imputation method for gene expression data which is a development of the robust least squares estimation with principal components (RLSP) imputation method using the biclustering concept. This method is called bicluster-based robust least squares estimation with principal components (bi-RLSP). This RLSP method is a gene expression data imputation method that uses the concept of k-nearest neighbor, principal component analysis, and quantile regression. The concept of biclustering is applied to replace the k-nearest neighbor concept that look for similar rows under all columns, so that it can search for similar rows as well as similar columns simultaneously. The bi-RLSP method was applied to impute missing values in gene expression data, where this method was found to have a higher accuracy than the RLSP method and the row average imputation method by looking at the NRMSE value of the estimated missing values in these methods.