

Workshop Cluster Analysis

Workshop Data Mining
6 Nopember 2013
Jurusan Teknologi Informasi
Politeknik Elektronika Negeri Surabaya

Datasets yang dipakai

- Ruspini
- Iris
- Fossil
- Wine
- New Thyroid

Dataset Ruspini*

- Dataset yang sederhana, dengan tipe data yang well-separated
- Cocok untuk dipakai sebagai benchmark problem untuk mengevaluasi tingkat presisi dari algoritma-algoritma clustering
- Jumlah dimensi : 2
- Jumlah data : 75
- Jumlah class : 4

*Incorporated as a built-in data object in both R and S-plus statistics packages

Dataset Iris*

- Data tentang klasifikasi bunga Iris
- Umum dipakai pada pengujian algoritma-algoritma clustering
- Jumlah dimensi : 4
- Jumlah data : 150
- Jumlah class : 3

*Provided by UCI Repository (<http://www.sgi.com/tech/mlc/db/>)

Dataset Fossil*

- Data dari sampel nummulitidae dari bentuk batu kapur kuning Eocene di daerah barat laut Jamaica
- Jumlah dimensi : 6
- Jumlah data : 87
- Jumlah class : 3

*obtained from Chernoff (Yi-tsuu, C., 1978. Interactive Pattern Recognition. Marcel Dekker Inc., New York and Basel

Dataset Wine*

- Berasal dari analisa kimia dari produksi anggur pada suatu daerah di Itali
- Umum dipakai pada pengujian algoritma-algoritma clustering
- Jumlah dimensi : 13
- Jumlah data : 178
- Jumlah class : 3

*Provided by UCI Repository (<http://www.sgi.com/tech/mlc/db/>)

Dataset New Thyroid*

- Data tentang klasifikasi pasien yang terkena penyakit thyroid. Diagnosa didasarkan pada medical record lengkap, termasuk anamnesis.
- Jumlah dimensi : 5
- Jumlah data : 215
- Jumlah class : 3

*Provided by UCI Repository (<http://www.sgi.com/tech/mlc/db/>)

Workshop Guide

(Algoritma: K-means, Dataset: Ruspini)

- Siapkan dataset Ruspini
- Lakukan clustering dengan K-means
- Visualisasikan tampilan hasil clustering
- Hitunglah variance dari hasil clustering
- Hitunglah error ratio dari hasil clustering
- Lakukan percobaan sebanyak n kali dan amati hasilnya
- Catat berapakah rata-rata variance dan rata-rata error ratio dari algoritma K-means terhadap dataset Ruspini

Workshop Guide

(Algoritma: Hierarchical clustering, Dataset: Ruspini)

- Lakukan hal yang sama diatas tetapi dengan memakai algoritma Single Linkage, Centroid Linkage, Complete Linkage dan Average. Lakukan percobaan 1 kali saja untuk masing algoritma Hierarchical clustering
- Catat berapakah rata-rata variance dan rata-rata error ratio dari masing-masing algoritma Hierarchical clustering terhadap dataset Ruspini dan bandingkan
- Bandingkan kinerja dari K-means dan 4 algoritma Hierarchical clustering yang lain.

Workshop Guide

(Algoritma: K-means+Hierarchical clustering,
Dataset: Ruspini, Iris, Fossil, Wine, New Thyroid)

- Lakukan percobaan yang sama tetapi dengan menggunakan datasets Iris, Fossil, Wine dan New Thyroid.
- Cobalah clustering dengan algoritma K-means dan 4 algoritma Hierarchical clustering yang lain
- Catat berapakah rata-rata variance dan rata-rata error ratio dari masing-masing algoritma
- Bandingkan kinerja dari K-means dan 4 algoritma Hierarchical clustering yang lain.