Machine Learning for Cities
CUSP-GX-5006-002

Prof. Luis Gustavo Nonato

Assignment IV

Release date: Apr. 18th, 2017 Due date: May 3rd, 2017

Spectral Clustering

Motivation

Clustering is major task in machine learning and many different techniques are available. K-means, Hierarchical and Spectral clustering are among the most important techniques, being widely used. This assignment aims to compare those three classes of techniques in a real application scenario, making clear their effectiveness, issues, and computational demand.

Description

For this assignment students should:

1. Perform clustering using K-means, Hierarchical and Spectral techniques. In particular, Hierarchical and Spectral methods should be extensively compared.

2. The accuracy of each clustering model should be assessed via test set, for instance using the concentric noisy circles data presented during the lab class;

3. Comparisons between the three clustering schemes should be accomplished. It is highly encouraged to perform a numerical study as to the impact of parameters in the quality of the clustering, mainly regarding spectral clustering, which demands several parameters.

4. Results should be presented as a report (template in NYU Classes) using a combination of tables and graphics.

Data Set

The noisy circles synthetic data set presented during the lab class can be used. However, we strongly encourage students to look for real data set to perform the comparisons. If other data set is used in the experiments, a clear description of its attributes (and the chosen dependent variable) should be included in the assignment report.