

Clustering Methods

Exercises 5/7, 9.2.2012

1. Run the GA software with different parameters (especially population size and number of iterations) for one of the Birch sets. Find out the most competitive parameters and show time-distortion comparison with the Random Swap.
2. Study the results of F-ratio for S1-S4 data sets, and try to develop a method detecting the knee point instead of the minimum point. Use either the first derivative, normalized first derivative by dividing standard deviation, or any idea of your own. Compare the results to that of the minimum point. Use the pre-calculated scores in the files found here: <http://cs.joensuu.fi/pages/franti/cluster/notes.html>
3. Prepare a simple but helpful visualisation to demonstrate the *Rand index* and *Adjusted Rand index*.
4. Show first draft of your course project work.
5. Discuss how to compile your project work software, and help others with their potential problems.