

Course 10	Unsupervised Pattern Recognition
Programme	<ul style="list-style-type: none"> 0. Introduction to clustering 1. Data Exploration and preparation <ul style="list-style-type: none"> 1.1 Types of features 1.2 Feature extraction 1.3 Graphical examination 1.4 Missing Data and outlayer removal 1.5 Principal component analysis 1.6 Kernel functions 1.7 Data reduction 1.8 Distance measures 2. Prototype-based clustering <ul style="list-style-type: none"> 2.1. K-Means 2.2. ISODATA 2.3. Fuzzy K-means 2.4. Partitioning Around Medoids (PAM) 2.5. Mixture models (EM algorithm) 2.6. Self-Organizing Maps (SOM) 2.7. Other prototype-based algorithms 3. Density-based clustering <ul style="list-style-type: none"> 3.1. Density Based Spatial Clustering 3.2. Grid Clustering 3.3. DENCLUE (DENsity CLUstEring) 3.4. Other density-based clustering 5. Graph-based clustering <ul style="list-style-type: none"> 5.1. Hierarchical clustering: Introduction 5.2. Hierarchical clustering Locally optimal algorithm 5.3. Hierarchical clustering Linking comparison 5.4. Chameleon 5.5. Hybrid Graph-Density based clustering: SNN-DBSCAN 5.6. Other graph-based clustering 6. Cluster evaluation <ul style="list-style-type: none"> 6.1. Clustering tendency 6.2. Unsupervised cluster evaluation 6.3. Supervised cluster evaluation 6.4. Criteria to determine the number of clusters 7. Miscellanea <ul style="list-style-type: none"> 7.1 Subspace clustering 7.2 Ensemble/Consensus clustering 7.2 Dynamic clustering 7.3 Semisupervised clustering 7.4 Clustering with obstacles 7.5 Biclustering, Coclustering, Two-mode clustering 7.6 Turning a supervised classification into a clustering algorithm 8. Conclusions and final advise
Bibliography	<p>Rui Xu, Don Wunsch, 2009. Clustering. IEEE Press Series on Computational Intelligence. ISBN: 9780470276808.</p> <p>Guojun Gan, Chaoqun Ma, Jianhong WuData, 2007. Clustering: Theory, Algorithms, and Applications (ASA-SIAM Series on Statistics and Applied Probability). ISBN: 9780898716238</p>
Prerequisites	<p>Basic knowledge of programming is desirable, but not essential, to follow the course. Students must bring their own laptop with R installed.http://www.r-project.org/. Ideally, the last day of the course the student should work on his/her own dataset; if this were not possible, there will be several standard data sets to choose from.</p>
Readings before the course	<p>The student will benefit more from the course if before attending she reads (these are not compulsory, only advisable):</p>

	Anil K. Jain. Data Clustering: 50 Years Beyond K-Means. Pattern Recognition Letters, 2009
--	---