

Reading instructions

- Data mining introduction
 - course book 2nd edition: 1.1-6, 2.1 / course book 3rd edition: 1.1-6, 3.1
- Clustering - introduction and distances for different types of data objects
 - course book 2nd edition: 7.1-3 / course book 3rd edition: 2.1, 2.4, 10.1
- Clustering - Partitioning Methods
 - course book 2nd edition: 7.4 / course book 3rd edition: 10.2
 - Raymond T Ng, Jiawei Han. [Efficient and Effective Clustering Methods for Spatial Data Mining](#), VLDB 94, 144--155, 1994. (CLARANS, also introduction to PAM and CLARA)
- Clustering - Hierarchical Methods
 - course book 2nd edition: 7.5 / course book 3rd edition: 10.3 (except 10.3.5) (Obs: ROCK not in 3rd edition)
 - Tian Zhang, Raghu Ramakrishnan, and Miron Livny. [BIRCH : an efficient data clustering method for very large databases](#), SIGMOD 96, 103-114, 1996.
 - Sudipto Guha, Rejeev Rastogi, and Kyuseok Shim. [ROCK: A robust clustering algorithm for categorical attributes](#), *Information Systems* 25(5):345-366, 2000.
 - George Karypis, Eui-Hong Han, and Vipin Kumar. [CHAMELEON: A Hierarchical Clustering Algorithm Using Dynamic Modeling](#), *COMPUTER* 32(8): 68-75, 1999.
- Clustering - Density-Based Methods
 - course book 2nd edition: 7.6 / course book 3rd edition: 10.4
 - Mihael Ankerst, Markus M Breunig, Hans-Peter Kriegel, Jörg Sander. [Optics: Ordering points to identify the clustering structure](#), SIGMOD 99, 49-60, 1999. (also introduction to DBSCAN)
 - Alexander Hinneburg, Daniel A. Keim. [An Efficient Approach to Clustering in Large Multimedia Databases with Noise](#), KDD 98, 58-65, 1998. (DENCLUE)
- Association analysis - introduction
 - course book 2nd edition: 5.1 / course book 3rd edition: 6.1
- Association analysis - Apriori algorithm
 - course book 2nd edition: 5.2.1-2, 5.4 / course book 3rd edition: 6.2.1-2, 6.4
 - R. Agrawal and R. Srikant. [Fast Algorithms for Mining Association Rules](#), Proc. of the 20th Int. Conf. on Very Large Databases, Santiago, Chile, September 1994. Expanded version available as IBM Research Report RJ9839, June 1994.
- Association analysis - FP grow algorithm
 - course book 2nd edition: 5.2.4 / course book 3rd edition: 6.2.4
 - J. Han, J. Pei, and Y. Yin. [Mining Frequent Patterns without Candidate Generation](#), Proc. 2000 ACM-SIGMOD Int. Conf. on Management of Data (SIGMOD'00), Dallas, TX, May 2000.
- Association analysis - Constraints
 - course book 2nd edition: 5.5 / course book 3rd edition: 7.3
 - J. Pei and J. Han. [Can We Push More Constraints into Frequent Pattern Mining?](#), Proc. 2000 Int. Conf. on Knowledge Discovery and Data Mining (KDD'00), Boston, MA, August 2000.