

CS724 Topics in Algorithm Theory and Design: Spring 2019

Professor Dan Simovici

Office Hours: M 10:00 – 12:00, W 3:00-4:00

This offering of cs724 is dedicated to a presentation of optimization algorithms in machine learning and data mining and is focused on clustering. This course requires a background in linear algebra, topology, and functional analysis. We will discuss such topics as metric and ultrametric spaces, partitional and hierarchical algorithms, data clusterability, complexity of clustering, etc. This will be supplemented by a presentation of several special topics that you did not see in the typical mathematics course, such as convexity, optimization techniques, etc.

Taking this course may help significantly with your research work; for this reason, the course is intended to attract mainly doctoral students.

Here is a list of topics that I intend to cover:

Partially Ordered Sets	Clustering quality
Dissimilarities, Metrics, and Ultrametrics	Complexity of Clustering
Ultrametrics and Clusterings	Clustering Axiomatization
Partitional Clustering	Clusterability
Density-based Clustering	Spectral Clustering

The main reference for this are slides and several handouts that will be posted on my web page www.cs.umb.edu/~dsim. Bibliography will be indicated at the end of each topic; however, your main sources are these lectures and the slides and handouts that will be posted on my web page.

No late homeworks will be accepted. All homeworks must be typeset using LaTeX.

Download and install **R** and **R Studio** for your operating system version. We will use R as auxiliary software.

Grades in this course will be based on homeworks and class participation. Therefore, the homeworks must be the product of your own efforts. Homework collaboration is strictly prohibited and will be severely sanctioned.