SF2956 Topological Data Analysis 7.5 credits

Topologisk dataanalys

This is a translation of the Swedish, legally binding, course syllabus.

Establishment

Course syllabus for SF2956 valid from Spring 2017

Grading scale

A, B, C, D, E, FX, F

Education cycle

Second cycle

Main field of study

Mathematics

Specific prerequisites

Courses in liner algebra, calculus in one and several variables, numerical methods, modern methods in statistical learning, computer intensive methods.

Language of instruction

The language of instruction is specified in the course offering information in the course catalogue.

Intended learning outcomes

In this course we present topology-based constructions of some stable signatures (called bar codes) of data sets and how to use these signatures to extract information out of outcomes of various hierarchical clustering schemes. We also present how to use these methods as a preprocessing step transforming possibly complex data into objects, which might be more suitable for statistical analysis. We also illustrate how these methods can help in visualizing certain aspects of the data. Part of the course is to learn how to extract homology out of data based simplicial complexes. The course addresses both theoretical and practical aspects of topological data analysis techniques. Computer-aided project work will be part of the learning activity.

To pass the course, the student should:

- explain the concepts of a metric space and an ultrametric space
- be able to visualise ultra metrics as dendrograms
- extract bar codes from dendrograms
- be able to calculate bottle neck distances between bar codes
- use bar codes to look for homological features in data sets
- be familiar with basic methods of extracting homology of data based simplical complexes
- be familiar with stability properties of classical hierarchical clustering schemes

To receive the highest grade, the student should in addition be able to:

- apply the discussed methods such as hierarchical clustering and bar coding of various homology groups to visualise properties of different type of data

Course contents

The course contains the following topics:

- Kleinberg theorem about impossibility of clustering,
- metric spaces and dendrograms,
- classical hierarchical clustering schemes (single, complete, average, and Haussdorff linkage),
- elements of simplicial complexes,
- transforming data into simplicial complexes via Chech and Vietoris -Rips constructions
- extracting homology out of data based simplicial complexes
- persistence modules, barcoding , and feature visualization

Disposition

Lectures, presentations, work with computer-aided data analysis.

Course literature

Various books and lecture notes presented on the course web page.

Examination

• TEN1 - Examination, 7.5 credits, grading scale: A, B, C, D, E, FX, F

Based on recommendation from KTH's coordinator for disabilities, the examiner will decide how to adapt an examination for students with documented disability.

The examiner may apply another examination format when re-examining individual students.

If the course is discontinued, students may request to be examined during the following two academic years.

Written exam.

Other requirements for final grade

Passed final exam.

Ethical approach

• All members of a group are responsible for the group's work.

• In any assessment, every student shall honestly disclose any help received and sources used.

• In an oral assessment, every student shall be able to present and answer questions about the entire assignment and solution.