MATH590
Topics in Mathematics: Data Analysis

Course syllabus

Times	TuTh 11:00AM-12:15PM, Phillips 381
Office hours	TuTh 1:15-2:15PM, and by email appointment, Chapman 451
Instructor	Sorin Mitran

A practical, case-based introduction to recent developments within several branches of mathematics to identify patterns within data. Standard numerical methods are based on concepts from mathematical analysis suitable for approximation in $\mathbb{R}^d\,(1 \leqslant d \leqslant 3)$. Contemporary data analysis enlarges the scope of approximation to consider concepts from set theory, topology, stochastic calculus, differential geometry, information theory, and graph theory. Such approaches are introduced through seven two-week long modules that introduce theoretical concepts, simple examples, relevant literature, and conclude with application to a real problem from the physical, life, or social sciences. The focus is on the motivation for choosing a particular mathematical framework for a specific data analysis problem. Coursework introduces software tools used in data analysis, and is suitable for students from a wide variety of backgrounds. Basic familiarity with calculus, linear algebra and computer programming is recommended.

This special topics course is presented more as a research seminar rather than a series of formal lectures. Students are encouraged to engage in independent reading of the bibliography items.

The instructor reserves the right to make changes to the syllabus. Any changes will be announced as early as possible.

Course goals

Upon course completion students:

- will be able to identify a suitable mathematical framework for case-specific data analysis
- will have a basic familiarity with software tools for data analysis
- will be able to place empirical data analysis methods into a proper mathematical framework
- will gain experience in preparation of formal scientific reports resulting from data analysis

Honor Code

Unless explicitly stated otherwise, all work is individual. You may discuss various approaches to homework problems with students, instructors, but must draft your answers by yourself. In joint projects, each student will clearly identify which portions of the work they contributed.

Grading

Required work

- Case studies, submitted as homework: 6 cases x 12 points = 84 points
- Final examination consisting of further work on a case study of student's choice: 28 points
- Extra credit: 2 reading topics x 5 points = 10 points

Mapping of point scores to letter grades

Grade	Points	Grade	Points	Grade	Points	Grade	Points
H+,A *cum laude*	101-110	H-,B+	86-90	P-,C+	71-75	L-,D+	56-60
H+,A	96-100	P+,B	81-85	L+,C	66-70	L–,D-	50-55
H,A-	91-95	P,B-	76-80	L,C-	61-65	F	0-49

Course policies

- Students are free to establish their own schedule; there is no need to inform instructor of absences. Course attendance is highly recommended to gain insight into course topics
- Homework is to be submitted electronically through Sakai

Examinations

A take-home final examination consisting of a more detailed report on a case study of the student's choice is to be submitted before 5:00PM, 04/29/19.

Course materials

Course topics

NUM. Approximation in $\mathbb{R}^d, 1 \leqslant d \leqslant 3$, review of numerical analysis with a focus of where the particular structure of $\mathbb{R}^d$ is used.

SET. Set theory: clustering, sparse data, fuzzy sets, large cardinals.

TOP. Topology: open sets, topological descriptors, homeomorphisms.

STC. Stochastic calculus: Ito, Stratonovich formulations, stochastic processes.

INF. Information theory: Shannon information, information functionals, statistical physics.

DIF. Differential geometry: Manifolds, information metrics.

Textbook

Class notes will be provided, and posted on this website.

Additional references

Entry points into the literature on class topics.

NUM *Numerical Analysis: Mathematics of Scientific Computing*, David Kincaid & Ward Cheney

SET *Data clustering : theory, algorithms, and applications*, Guojun Gan, Chaoqun Ma, Jianhong Wu

TOP P. Bubenik, *Statistical Topological Data Analysis using Persistence Landscapes*

L. Wasserman, *Topological Data Analysis*

Class slides

Class notes will be provided that briefly summarize class discussion topics, and are posted on this website.

Week	Start date	Topic	Tuesday	Thursday
01	01/7	Data analysis	-	Overview
02	01/14	NUM	Theory	Examples
03	01/21		Problem	Analysis
04	01/28	SET	Theory	Examples
05	02/04			
06	02/11	TOP	Theory	Examples
07	02/18			
08	03/19	STC	Theory	
09	03/04			
10	03/18	INF	Theory	
11	03/25			
12	04/01	DIF	Theory	
13	04/08			
14	04/15	DIF		
15	04/22			

Homework

Homework consists of a report on the case study considered in each two-week module. Each report is presented in the form a scientific paper. Templates are provided.

Nr.	Issue Date	Due Date	Topic	Problem	Solution
01	01/14	01/28	NUM	Template	Report
02	02/25	03/04	SET	Template	Report
03	03/22	03/29	TOP	Template	
04	02/25	03/18	STC	Template	
05	03/18	04/01	INF	Template	
06	04/01	04/15	DIF	Template	

Software

Modern software systems allow efficient, productive formulation and solution of mathematical models. A key goal of the course is to familiarize students with these capabilities, using the SciComp@UNC environment in which tools required for data analysis have been preconfigured for immediate use. Follow instructions at SciComp@UNC to install on a laptop with at least 48GB free disk space and that conforms to CCI minimal standards.

Tutorials

Software usage is introduced gradually in each class, so the first resource students should use is careful, active reading of the material posted in class. In particular, carry out small tasks until it becomes clear what the software commands accomplish. Some additional resources:

- Mathematica
  - http://www.wolfram.com/language/fast-introduction-for-math-students/en/
  - http://www.wolfram.com/wolfram-u/catalog/gen005/
  - http://www.wolfram.com/language/fast-introduction-for-programmers/en/
- TeXmacs:
  - http://www.texmacs.org/tmweb/help/tutorial.en.html
  - https://www.youtube.com/watch?v=mlcqGRv7xhc

- Julia:
  - https://julialang.org/learning/
- Scheme:
  - https://www.scheme.com/tspl4/

Course material repository

Course materials are stored in a repository that is accessed through the subversion utility, available on all major operating systems. The URL of the material is http://mitran-lab.amath.unc.edu/courses/MATH590

In the SciComp@UNC virtual machine the initial checkout can be carried out through the terminal commands

```
cd ~/courses
make MATH590
```

Update the course materials before each lecture by:

```
cd ~/courses
svn update
```

Links to course materials will also be posted to this site, but the most up-to-date version is that from the subversion repository, so carry out the svn update procedure prior to each lecture.