

IE 452 / 552: Algebraic and Geometric Methods in Data Analytics Spring 2020

- **Instructor:**

Arnab Basu, Ext: 3343, Office: EA329, E-mail: arnab@bilkent.edu.tr

- **Course Web Page:** TBA

- **Teaching Assistant:**

Aleyna Kof, Office: EA-307, Office Hours: TBA, E-mail: aleyna.kof@bilkent.edu.tr

- **Class hours:**

Tuesdays 10:40-12:30 (EA-Z03) and Fridays 08:40 - 10:30 (EA-Z03)

In a typical week, all **four** reserved slots will be used. Please, see the detailed schedule given at the end of the syllabus.

- **Recitations:** During lecture hours prior to each midterm and the final examination.

- **Office hours:**

By appointment.

- **Prerequisites:** (MATH 220 or 225 or 241) AND (MATH 250 or 255)

- **Recommended Textbooks:** No required textbooks as class notes would be sufficient. However, the following supplementary materials may be used:

- Foundations of Data Science by Blum, Hopcroft and Kannan (CUP - To be released on March 31, 2020)
- On the Mathematical Foundations of Learning by Cucker and Samle (Bulletin of the AMS 39(1) - pp. 1-49, 2001)
- Spectral Algorithms by Kannan and Vempala (NOW Foundations and Trends in TCS 4(3-4) - pp. 157-288, 2008)
- Introduction to Semi-Supervised Learning by Zhu and Goldberg (Morgan and Claypool Synthesis Lectures on AI and ML 6, 2009)
- Learning with Kernels by Schölkopf and Smola (MIT Press, 2002)
- Algorithms for Reinforcement Learning by Szepesvári (Morgan and Claypool Synthesis Lectures on AI and ML 9, 2010)

- **Course Objectives:** This course is designed to develop an understanding of the mathematical foundations of Data Analytics in general and in particular enable the students to:

1. Create a mathematical formulation of any given large high-dimensional data-set
2. Develop algebraic and geometric insights into the structure of such data-sets
3. Understand how to handle noise and sparsity in such data-sets
4. Use several critical algorithms in these domains to extract relevant information from such data-sets
5. Leverage the concepts and techniques in the above points to model and solve actual industrial problems involving such data-sets

- **Study Sets:** Study sets will be assigned prior to the examinations. Though they will not be collected or graded, it is strongly recommended that you work on these problems on your own since they will serve as an excellent educational tool and preparation for the examinations. The TA will solve these problems during recitations.

- **MidTerms:** There will be **One (1)** scheduled Midterm during evening hours. The exact date and place will be announced during the semester.

- **Final Examination:** There will be **One (1) Comprehensive** Final Exam some time during the final exam period. The exact date and place will be announced during the semester.

- **Make-up Policy:** A make-up examination will only be given under highly unusual circumstances (such as serious health or family problems). The student should contact the instructor as early as possible and provide the instructor with proper documentation (such as a medical note certified by Bilkent University's Health Center).
- **Class Participation:** Attendance will be taken at all lectures due to new regulations. Note that attendance will be taken with student signatures. Grades obtained for this component shall depend on both actual attendance as well as active participation in classroom discussions. Any violation of this process (i.e. forging signatures) is a direct violation of academic integrity!
- **Project:** The project assignment (due sometime towards the end of the semester) will be a group project involving **at most 3** students in one group. Please form your project groups as soon as possible and e-mail the names of your group members to the TA by 12:00 on Friday, February 28, 2020. Students who do not belong to any group by the deadline will solely be responsible from the project. The details of the project will be subsequently announced during the semester. You shall be needing some software like Python, Matlab etc. to implement the project and need to be on your own in learning how to use the software. **All** the group members are responsible from every step of the project. You will be called in for an oral exam to present the knowledge on the project.
- **Grading Policy:** Your overall score will be computed based on 10 % Class Participation, 20% X 2 Projects = 40%, 15% Midterm, 10% Take-home Assignments (HWs 5% each) and 25% Final Examination.
- **FZ Policy:** There is **no** FZ grade, everyone is entitled for the final exam.
- **Tentative Course Outline:**

Week	Topics
01	Geometric Introduction to Data Clouds
02	Algebraic Introduction to Vector Models of Data
03	Applications to Data Query Problems
04	Geometric Introduction to Data Dimensionality Reduction
05	Applications to Centering, Ranking and Clustering
06	Applications to Discrete Optimization Problems
07	Algebraic Introduction to Markov Chains and Random Walks
08	Applications to Physical Networks
09	Applications to the Internet
10	Algebraic Introduction to Massive Data Sets
11	Applications to Streaming, Sketching and Sampling
12	Algebraic Introduction to Random Graphs
13	Applications to Growth Models and Small Worlds
14	Summarization and Discussion

The instructor reserves the right to update any part of this syllabus as necessary. Students will be notified of any changes.

	Section 1	
Week of	Tuesdays	Fridays
February 03	Lecture (2 hrs)	Lecture (2 hrs)
February 10	Lecture (2 hrs)	Lecture (2 hrs)
February 17	Lecture (2 hrs)	Lecture (2 hrs)
February 24	Lecture (2 hrs)	Lecture (2 hrs)
March 02	Lecture (2 hrs)	Lecture (2 hrs)
March 09 (MidTerm 1)	Lecture (2 hrs) / Recitation (2 hrs)	Lecture (2 hrs)
March 16	Spring break	No class
March 23	Lecture (2 hrs)	Lecture (2 hrs)
March 30	Lecture (2 hrs)	Lecture (2 hrs)
April 06	Lecture (2 hrs)	Lecture (1 hr)
April 13	Lecture (2 hrs)	Lecture (1 hr)
April 20	Lecture (2 hrs)	No class
April 27	Lecture (2 hrs)	Holiday
May 04	No class	No class
May 11	No class	No class
May 18	TBA	

• **Rules:**

1. Please make sure that you have a STARS password and a valid e-mail in the STARS system. All of our communications will be conducted through the STARS system. Please check your **e-mail** regularly for announcements.
2. A **make-up** examination will only be given under extenuating circumstances (such as serious health or family problems). The student should contact the instructor as early as possible and provide the instructor with proper documentation (such as a medical note certified by Bilkent University's Health Center). The (comprehensive) make-up exam will be given during or right after the final exam period.
3. **No make-up will be given for quizzes under any circumstances.**
4. **Cheating will not be tolerated and will be severely penalized. Disciplinary action will be taken.**
5. **As a courtesy, please turn off your cell phones during class.**
6. Every student is expected to respect the instructor's right to teach and other students' right to learn. Any behavior which distracts or disturbs the other students or the instructor, or disrupts class in any way is unacceptable and will not be tolerated. Any student engaging in inappropriate behavior will be asked to leave. Such behavior might also reflect negatively on the course grade of the student.