IE 452/552: Algebraic and Geometric Methods in Data Analytics Spring 2021

- **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 15:30-17:20 (EA-Z03) and Fridays 10:30 - 12:20 (EB-103)

  In some weeks, all **four** reserved slots will be used which shall be announced during the course.

- **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 Algortihms 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 Study-Set discussions.

- **MidTerms:** There will be **One (1)** scheduled **Midterm** some time during the Mid Term period. 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 circumstances (such as serious health or family problems) as approved and mandated by University Rules. The student should contact the instructor as early as possible and provide the instructor with proper documentation (such as a Medical Report issued/certified by Bilkent University's Health Center).

- **Project:** There shall be two project assignments (one due sometime during the middle (around Mid-March 2021) and other due sometime during the end of the semester (around End-April 2021). These will be group projects 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 17:00 on Friday, February 12, 2021. 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 during its Report Submission.
- **Grading Policy:** Your overall score will be computed based on 15% X 2 Projects = 30%, 20% MidTerm, 20% Homeworks (2 HWs 10% each) and 30% 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
textsf13	Applications to Growth Models and Small Worlds
14	Summarization and Discussion

The instructor reserves the right to change this syllabus as necessary. Students will be duly notified.

- **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. **No make-up will be given for Projects/HWs under any circumstances.**
3. **Cheating will not be tolerated and will be severely penalized. Disciplinary action will be taken.**
4. 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. Such behavior might also reflect negatively on the course grade of the student.