جامعة الملك عبدالله
للعلوم والتقنية
King Abdullah University of
Science and Technology

Course Syllabus: Special Topics in Data Sciences - CS 390FF

Division	Computer, Electrical and Mathematical Sciences & Engineering
Course Number	CS 390FF
Course Title	Special Topics in Data Sciences
Academic Semester	Fall
Academic Year	2019/2020
Semester Start Date	08/25/2019
Semester End Date	12/10/2019
Class Schedule (Days & Time)	09:00 AM - 10:30 AM | Sun Tue

Instructor(s)

Name	Email	Phone	Office Location	Office Hours
Peter Richtarik	PETER.RICHTARIK@KAUST.EDU.SA			I will be ready to answer any questions after each lecture (10:30-11:00) in the lecture room. If you want to chat with me during some other time, please send an email to me to ask for an appointment.

Teaching Assistant(s)

Name	Email
Adil Salim	adil.salim@kaust.edu.sa

Course Information

Comprehensive Course Description	**Special Topics in Data Sciences: Big Data Optimization** The course is a mathematically rigorous introduction to the emerging field of big data optimization. The focus is on algorithms and assocaited theory. Randomized/stochastic algorithms play a dominant role. The course is based on a novel and unified approach to recent developments in the field developed by the lecturer. Big data optimization is the study of optimization problems described by big quantities of data, where "big" is loosely defined as large enough for traditional approaches to suffer or not be applicable at all. As we live in a digital age where it is increasingly easier to collect and store data in digital form (e.g., transaction records, YouTube clicks, internet activity, Wikipedia, twitter, customer behaviour databases, government records, image collections), big data problems are becoming ubiquitous. New methods and tools are needed to analyze such vast datasets, and optimization algorithms are at the heart of such efforts, underpinning much of data science, including machine learning, operations research and statistical analysis. Alongside computer science and statistics, optimization is one of the pillars of big data analysis. The course will cover topics such as supervised learning, empirical risk minimization, big data problems, stochastic gradient descent, minibatching, importance sampling, arbitrary sampling, variance reduction, quantization and compression for distributed training, convex feasibility problems, high dimensional problems, randomized coordinate descent, and acceleration.
Course Description from Program Guide	

Goals and Objectives	-Detailed understanding of the role of randomization as a decomposition tool for solving optimization problems of big sizes. -Understanding the underlying mathematical theory. -Ability to apply the methodologies to selected applications in machine learning and data science. -Preparation for original theoretical and applied research in the field.
Required Knowledge	-Strong experience with at least one high level computing language (e.g.: MATLAB, Python, Julia, C, ...) -Mathematical maturity. -Ability to comprehend and generate proofs. -Linear algebra (abstract vector spaces, linear independence, basis, linear operators, quadratic forms, Euclidean spaces, inner product, norm, …) -Matrix theory (matrices, determinants, singular values, eigenvalues, matrix decompositions, …) -Multivariate calculus (gradient, Hessian, Taylor approximation, chain rule, …) -Probability theory (probability spaces, expectation, law of large numbers, tower property, …)
Reference Texts	-Detailed slides (these will be handed out before each lecture) -Relevant papers (optional reading)
Method of evaluation	**25.00**% - Midterm exam **50.00**% - Homework /Assignments **25.00**% - Final exam
Nature of the assignments	2 assignments: computational (2 x 25%) 1 midterm exam: theory (25%) 1 final exam: theory (25%)
Course Policies	-No late submissions will be accepted. Submit all you have done (even if you did not fully complete the assignment) by the deadline. -A no-show at midterm / exam results in a 0 mark. In very exceptional circumstances, **agreed with the lecturer well in advance** (e.g., serious illnes with a document from the doctor; attendance at a conference abroad, if approved by the lecturer in advance), a make-up midterm/exam might be arranged. However, this should be avoided at all costs.
Additional Information	Te MIDTERM covers the first half of the course, and the FINAL EXAM covers the second part. The course this year will have a substantially different content than in the past 2 years; the material was refreshed reflecting some of the latest developments in the field.

Tentative Course Schedule *(Time, topic/emphasis & resources)*		
Week	**Lectures**	**Topic**
1	Sun 08/25/2019 Tue 08/27/2019	Lecture 1 Lecture 2
2	Sun 09/01/2019 Tue 09/03/2019	Lecture 3 Lecture 4
3	Sun 09/08/2019 Tue 09/10/2019	Lecture 5 Lecture 6
4	Sun 09/15/2019 Tue 09/17/2019	Exercise session (Assignment 1 released) Exercise session
5	Sun 09/22/2019 Tue 09/24/2019	university holiday Lecture 7: guest lecture
6	Sun 09/29/2019 Tue 10/01/2019	Lecture 8 (Assignment 1 due) Lecture 9
7	Sun 10/06/2019 Tue 10/08/2019	Lecture 10 Lecture 11
8	Sun 10/13/2019 Tue 10/15/2019	MIDTERM EXAM Lecture 12
9	Sun 10/20/2019 Tue 10/22/2019	Lecture 13: guest lecture Lecture 14: guest lecture
10	Sun 10/27/2019 Tue 10/29/2019	Mid-semester break (27-28.10) no class
11	Sun 11/03/2019 Tue 11/05/2019	Lecture 15 (Assignment 2 released) Lecture 16
12	Sun 11/10/2019 Tue 11/12/2019	Lecture 17 Lecture 18
13	Sun 11/17/2019 Tue 11/19/2019	Lecture 19 (Assignment 2 due) Lecture 20
14	Sun 11/24/2019 Tue 11/26/2019	Lecture 21 Lecture 22
15	Sun 12/01/2019 Tue 12/03/2019	Lecture 23 Lecture 24
16	Sun 12/08/2019 Tue 12/10/2019	FINAL EXAM for the course will be on one of these days: Dec 8, 9, 10. Final exams were not scheduled yet.

Note

The instructor reserves the right to make changes to this syllabus as necessary.