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

Course Syllabus: Stochastic Gradient Descent Methods - CS 331

Offering Department	Department of Computer Science
Course Number	CS 331
Course Title	Stochastic Gradient Descent Methods
Academic Semester	Fall 2026/2027
Semester Start Date	08/30/2026
Semester End Date	12/10/2026
Class Schedule (Days & Time)	Stochastic Gradient Descent Methods Lecture CS 331 Sun 14:30 - 17:30

Instructor(s)

Name	Email	Phone	Office Location	Office Hours
Peter Richtarik	PETER.RICHTARIK@KAUST.EDU.SA			by appointment

Teaching Assistant(s)

Name	Email
Majied Ammar Mahran	majiedammar.mahran@kaust.edu.sa

Course Information

Course Description	Stochastic gradient descent (SGD) in one or another of its many variants is the workhorse method for training modern supervised machine learning models. However, the world of SGD methods is vast and expanding, which makes it hard for practitioners and even experts to understand its landscape and inhabitants. This course is a mathematically rigorous and comprehensive introduction to the field, and is based on the latest results and insights. The course develops a convergence and complexity theory for serial, parallel, and distributed variants of SGD, in the strongly convex, convex and nonconvex setup, with randomness coming from sources such as subsampling and compression. Additional topics such as acceleration via Nesterov momentum or curvature information will be covered as well. A substantial part of the course offers a unified analysis of a large family of variants of SGD which have so far required different intuitions, convergence analyses, have different applications, and which have been developed separately in various communities. This framework includes methods with and without the following tricks, and their combinations: variance reduction, data sampling, coordinate sampling, arbitrary sampling, importance sampling, mini-batching, quantization, sketching, dithering and sparsification.
	1. Detailed understanding of the key variants of SGD for training supervised machine

Learning Outcomes	learning models and their differences. 2. Understanding the underlying mathematical theory and of the insights the theory offers for practice. 3. Ability to apply the methodologies to selected applications in machine learning. 4. Ability to code up SGD methods from scratch and verify their theoretical properties through carefully designed experiments. 5. The course offers an ideal background and preparation for original theoretical and applied research in the field of randomized methods for optimization and machine learning.
Textbook/Materials	- Detailed slides including all necessary material, including proofs (these will be handed out before each lecture) - Relevant papers (optional reading)
Method of Assessments	35.00% - Homework /Assignments 30.00% - Midterm exam 30.00% - Final exam 5.00% - Active participation
Nature of the Assignments	14 assignments -- a mix of theoretical and coding problems. Total of 35 points (each assignment worth 2.5 points).
Course Policies	- Attendance at lectures is compulsory - No late submissions will be accepted - Submit all you have done (even if you did not fully complete the assignment) by the deadline
Additional Information	Required Knowledge: - Strong experience with at least one high level computing language (e.g.: Python, Julia, C, MATLAB, ...) - Mathematical maturity (i.e., 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 of expectation, ...)

Tentative Course Schedule

(Time, topic/emphasis & resources)

Week	Lectures	Topic
1	Sun 08/30/2026	Course Organization Introduction to Optimization
2	Sun 09/06/2026	Introduction to Optimization for Machine Learning Broximal Point Method - Convex & Euclidean Theory
3	Sun 09/13/2026	Broximal Point Method - Non-Euclidean and Nonconvex Theory
4	Sun 09/20/2026	Approximate Broximal Point Method (Gradient Descent, Stochastic Gradient Descent, Local LMO, Muon)
5	Sun 09/27/2026	Distributed Broximal Point Method
6	Sun 10/04/2026	Stochastic Broximal Point Method
7	Sun 10/11/2026	Mid-term exam
8	Sun 10/18/2026	Assumptions in Optimization in Machine Learning Convexity and Smoothness 1 Convexity and Smoothness 2
9	Sun 10/25/2026	SGD with Uniform Sampling SGD with Nonuniform Sampling
10	Sun 11/01/2026	SGD with Minibatching General Analysis for SGD with and without Variance Reduction
11	Sun 11/08/2026	SGD with Shift SGD with Learned Shift

Week	Lectures	Topic
12	Sun 11/15/2026	Distributed Training: Gradient Compression CGD (Compressed Gradient Descent) CGD with Sketch Compression: Randomized Coordinate Descent
13	Sun 11/22/2026	CGD with Shift CGD with Learned Shift
14	Sun 11/29/2026	Stochastic Proximal Point Methods
15	Sun 12/06/2026	Further Topics
16	No schedule	Final Exam

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