The Chinese University of Hong Kong, Shenzhen

Course Outline

1. Course identity

A. Course as listed at CUHK-Shenzhen

The information in this block should be exactly as approved by CUHK Senate. In case there are any differences, please explain in the table below.

Course code	AIE6003
Course title (English)	Optimization for Machine Learning
Course title (Chinese)	机器学习中的优化
Units	3
Description (English)	This course delves into the theoretical foundations of optimization in machine learning, with a focus on nonlinear programming. Topics include gradient-based methods, adaptive gradient methods, and stochastic optimization. It will also cover emerging topics at the intersection of machine learning and optimization, such as sparse optimization, min-max optimization, bilevel optimization, and black-box optimization. This course is designed for students who are looking to deepen their understanding of advanced optimization techniques and apply them to complex machine learning problems.
Description (Chinese)	本课程深入探讨机器学习领域中优化的理论基础，重点关注非线性规划。课程内容包括基于梯度的方法、自适应梯度方法以及随机优化；同时还将介绍机器学习与优化交叉的前沿话题，如稀疏优化、极小极大优化、双层优化和黑箱优化。本课程面向希望深入掌握高级优化技术并将其应用于复杂机器学习问题的学生。

B. Corresponding course at CUHK

Please give details of the *closest* corresponding course at CUHK (as approved by CUHK Senate and listed in course list). If the course in Shenzhen maps to more than one course at CUHK, please make multiple copies of the block below.

Course code	SEEM5350
Course title (English)	Numerical Optimization
Course title (Chinese)	数值优化
Units	3

Description (English)	This course is to teach students modern numerical optimization methods for large scale systems. Topics covered in this course include gradient method, subgradient method, proximal gradient method, Nesterov's acceleration technique, alternating direction method of multipliers, coordinate descent method, and stochastic / randomized algorithms. Applications of these optimization methods for solving problems in contemporary applications arising from big data analytics, machine learning, statistics, signal processing etc. will also be discussed.
Description (Chinese)	本课程旨在教授面向大规模系统的现代数值优化方法，内容涵盖梯度法、次梯度法、近端梯度法、Nesterov 加速技术、交替方向乘子法（ADMM）、坐标下降法以及随机算法，并将探讨这些方法在大数据分析、机器学习、统计学和信号处理等当代应用中的实际问题求解。

2. Course requisites

Please state prerequisites and co-requisites, in terms of courses at CUHK-Shenzhen or any other requirements.

A. Prerequisites

Undergraduate level of linear algebra, analysis, and probability & statistics are required.

B. Co-requisites

None.

C. Others

Prior courses in linear and nonlinear optimization (including duality, optimality conditions, etc.) and machine learning (including empirical risk minimization, neural network, etc.) are encouraged but not required.

3. Learning outcomes

This course helps students

- get familiar with the main classes of optimization methods used in classical and modern machine learning;
- understand why each algorithm emerged, the problems they address, their strengths and weaknesses, and how subsequent techniques overcome those limitations;
- be able to state and sketch proofs of key convergence guarantees, iteration-complexity bounds, etc.; and

- learn to tailor or innovate optimization algorithms by exploiting problems to achieve faster convergence or better empirical performance.

4. Course syllabus

Please highlight fundamental concepts involved in each topic to help students better understand what is covered in the course.

Gradient descent, heavy ball and accelerated gradient descent, conjugate gradient method, stochastic gradient method, adaptive gradient methods; proximal gradient descent; subgradient method and mirror descent; low-rank optimization and conditional gradient method; empirical risk minimization and variance reduction; minimax optimization and extragradient method; sparse optimization and coordinate descent; black-box optimization and zeroth order methods.

5. Assessment scheme

Component/ method	% weight
Assignments	20
Midterm	30
Final	50

6. Grade descriptor

The grade descriptor should align with specific course learning outcomes, not a generic description applicable to all courses.

Grade	Description
A	Demonstrates the ability to achieve all the learning outcomes in this course and be able to apply the principles in solving problems in novel situations, in a manner that would surpass the normal expectation at this level, and typical of standards that may be common at higher levels of study or research. Has the ability to express the synthesis of ideas or application in a clear and cogent manner.
A-	Demonstrates the ability to state and apply the principles or subject matter learnt in the course to familiar and standard situations in a manner that is logical and comprehensive. Has the ability to express the knowledge or application with clarity.
B	Demonstrates the ability to state and partially apply the principles or subject matter learnt in the course to most (but not all) familiar and standard situations in a manner that is usually logically persuasive. Has the ability to express the knowledge or application in a satisfactory and unambiguous way.
C	Demonstrates the ability to state and apply the principles or subject matter learnt in the course to most (but not all) familiar and standard situations in a manner that is not incorrect but is somewhat fragmented. Has the ability to express the separate pieces of knowledge in

	an unambiguous way.
D	Demonstrates the ability to state and sometimes apply the principles or subject matter learnt in the course to some simple and familiar situations in a manner that is broadly correct in its essentials. Has the ability to state the knowledge or application in simple terms.
F	Demonstrates little or no ability to state or apply the principles or subject matter learnt in the course, even in simple and familiar situations. The knowledge shown is fragmentary, largely incorrect, or absent, and the expression of ideas or application is unclear or misleading.

7. Feedback for evaluation

Please specify forms of evaluation to generate feedback on classes from students.

CTE

Feedback from conversation in my office and e-mail correspondence with students after class.

8. Readings

A. Required

None.

B. Recommended

[1] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press.
[2] Y. Nesterov, Lectures on Convex Optimization, Springer.
[3] Beck, A. *First-order methods in optimization*. Society for Industrial and Applied Mathematics.
[4] Wright, S. J., & Recht, B. *Optimization for data analysis*. Cambridge University Press.
[5] S. Sra, S. Nowozin, and S. J. Wright, *Optimization for Machine Learning*, MIT Press.

9. Course components

Activity	Hours/week
Lectures	3
Tutorial	1
Readings & Assignments	4

10. Indicative teaching plan

Week (Dates)	Content
1 (Jan 05/07)	introduction, basic concepts (projection, optimality condition, convexity, smoothness)
2 (Jan 12/14)	gradient descent and conjugate gradient method
3 (Jan 19/21)	heavy ball and Nesterov's accelerated gradient descent
4 (Jan 26/28)	subgradient methods and mirror descent
5 (Feb 02/04)	proximal operator, proximal gradient descent, FISTA
6 (Mar 02/04)	stochastic gradient descent, variance reduction
7 (Mar 09/11)	stochastic gradient descent with momentum, adaptive (stochastic) gradient methods
8 (Mar 16/18)	midterm review, midterm exam
9 (Mar 23/25)	duality, Lagrangian duality
10 (Mar 30/Apr 01)	Fenchel duality, dual first-order algorithm, ALM, ADMM
11 (Apr 08/13)	low-rank optimization (including conditional gradient methods)
12 (Apr 15/20)	sparse optimization (including coordinate descent)
13 (Apr 22/27)	minimax optimization (including extragradient methods)
14 (Apr 29/May 6)	black-box optimization with zeroth-order methods

11. Academic honesty and plagiarism

A course outline may include subject-specific requirements on plagiarism, in addition to the University established policies.

Students must complete all coursework and assessments independently, unless AI tool use is explicitly permitted. Unauthorized use of AI may constitute academic dishonesty and harm learning. Misuse includes submitting AI-generated work as one's own, using AI to cheat, or employing AI unethically. Always obtain permission before using AI in academic work and be sure to acknowledge any AI use in submissions.