The Hong Kong University of Science and Technology (Guangzhou)

Syllabus for DSAA 3073: Theories in Data Science

Course Title: Theories in Data Science

Course Code: DSAA 3073

No. of Credits: 3

Pre-requisites: (DSAA 2011 or DSAA 2012 or AIAA 3111) AND (UFUG2106 or DSAA 2088)

Instructor Name: Wei WANG

Email: weiwcs@hkust-gz.edu.cn

Office Hours: TBD

TA's Name: TBD

Office Hours: TBD

Course Description

This course will cover fundamental topics in Machine Learning and Data Science, including powerful algorithms with provable guarantees for making sense of and generalizing from large amounts of data. The course will include a discussion of the computational complexity of learning and the concepts of convexity and stability; important algorithmic paradigms including stochastic gradient descent, neural networks, and structured output learning; and emerging theoretical concepts such as the PAC-Bayes approach and compression-based bounds. In addressing these and related questions we will make connections to statistics, algorithms, linear algebra, complexity theory, information theory, optimization, and empirical machine learning research.

Intended Learning Outcomes (ILOs)

Upon completion of this course, students are expected to be able to do the following:

	Course ILOs	**Nature of the learning outcomes** *(A - Knowledge/Content Related; B - Academic Skills/Competencies; C - Others)*
1	Explain and exemplify a wide variety of theoretical concepts and techniques of machine learning.	A
2	Explain and exemplify theoretical properties and guaranties of machine learning models and algorithms.	A, B
3	Apply the principles and techniques to a variety of machine learning algorithms.	B
4	Explain and exemplify the rationale and pros and cons of a variety of machine learning algorithms from a theoretical point of view.	A, B, C

5	Evaluate and compare the performance of different machine learning algorithms.	A, B, C

Contribution of Learning Outcomes to Programs of Study

Program of study 1: BSc in Data Science and Big Data Technology Program ILOs	**To be achieved through these course ILOs**
1. Explain data problems arising in different areas of science, technology, and the society. (Knowledge)	CILO-2, CILO-3, CILO-4
2. Model the problem with learned mathematics theories and models. (Execution)	CILO-2, CILO-3, CILO-4
3. Apply different mathematical tools to model data problems in application areas. (Execution)	CILO-3
4. Design and implement efficient algorithms to model the data and solve the problem. (Execution)	CILO-2, CILO-4
5. Evaluate information and make independent judgements through constructing and inferencing with appropriate data models. (Judgement)	CILO-5
6. Communicate effectively about data science to both laymen and experts. (Communication)	CILO-3
7. Demonstrate self-direction in tackling and solving problems and act autonomously in planning and implementing tasks. (Autonomy)	CILO-3, CILO-4
8. Use a global perspective in conjunction with data analytic techniques to address issues of importance in science, technology, and the society. (International outlook)	CILO-4

Weekly schedule & Weekly ILOs

A weekly schedule and corresponding ILOs for clear, week-by-week guidance is attached below.

Week	**Topics**	**Weekly ILOs**
1	Introduction and Maths Preliminaries	CILO-1, CILO-2
2	PAC learning	CILO-1, CILO-2, CILO-3, CILO-4, CILO-5
3	Generalization bounds – I	CILO-1, CILO-2, CILO-3, CILO-4, CILO-5
4	Generalization bounds – II	CILO-1, CILO-2, CILO-3, CILO-4, CILO-5
5	Model Selection	CILO-1, CILO-2, CILO-3, CILO-4, CILO-5
6	SVM	CILO-1, CILO-2, CILO-3, CILO-4, CILO-5
7	Mid-term exam	
8	Kernel Methods	CILO-1, CILO-2, CILO-3, CILO-4, CILO-5
9	Boosting	CILO-1, CILO-2, CILO-3, CILO-4, CILO-5
10	Online learning	CILO-1, CILO-2, CILO-3, CILO-4, CILO-5
11	Dimensionality Reduction	CILO-1, CILO-2, CILO-3, CILO-4, CILO-5
12	Advanced Topics	CILO-1, CILO-2, CILO-3, CILO-4, CILO-5
13	Course Review	

Assessment and Grading

This course will be assessed using criterion-referencing and grades will not be assigned using a curve. Detailed rubrics for each assignment will be provided on the released date, outlining the criteria used for evaluation.

Assessments

Assessment Task	Contribution to Overall Course grade (%)	Due date
Mid-Term Test	20%	TBD *
Written Assignment	30%	TBD *
Final examination	50%	TBD *

* Assessment marks for individual assessed tasks will be released within two weeks of the due date.

Mapping of Course ILOs to Assessment Tasks

Assessed Task	Mapped ILOs
Written Assignments [30%]	CILO-1, CILO-2, CILO-3, CILO-4, CILO-5
Midterm test [20%]	CILO-1, CILO-2, CILO-3, CILO-4, CILO-5
Final exam [50%]	CILO-1, CILO-2, CILO-3, CILO-4, CILO-5

Grading Rubrics

Detailed rubrics for each assignment will be provided. These rubrics clearly outline the criteria used for evaluation. Students can refer to these rubrics to understand how their work will be assessed.

Final Grade Descriptors

As appropriate to the course and aligned with university standards

Grades	Short Description	Elaboration on subject grading description
A	Excellent Performance	Demonstrates a comprehensive grasp of subject matter, expertise in problem-solving, and significant creativity in thinking. Exhibits a high capacity for scholarship and collaboration, going beyond core requirements to achieve learning goals.
B	Good Performance	Shows good knowledge and understanding of the main subject matter, competence in problem-solving, and the ability to analyze and evaluate issues. Displays high motivation to learn and the ability to work effectively with others.
C	Satisfactory Performance	Possesses adequate knowledge of core subject matter, competence in dealing with familiar problems, and some capacity for analysis and critical thinking. Shows persistence and effort to achieve broadly defined learning goals.
D	Marginal Pass	Has threshold knowledge of core subject matter, potential to achieve key professional skills, and the ability to make basic judgments. Benefits from the course and has the potential to develop in the discipline.
F	Fail	Demonstrates insufficient understanding of the subject matter and lacks the necessary problem-solving skills. Shows limited ability to think critically or analytically and exhibits minimal effort towards achieving learning goals. Does not meet the threshold requirements for professional practice or development in the discipline.

Course AI Policy

Students **may** use generative AI tools (e.g., ChatGPT, Copilot) for:

- Clarifying course contents, including concepts, proofs, and algorithms.
- Getting help debugging small, specific pieces of code.
- Brainstorming or outlining ideas, followed by the student's own development and verification of the final content.

Students **should not** (unless explicitly required):

- Use generative AI to produce full or near-complete assignment solutions, project reports, or exam answers.
- Submit AI-generated content without meaningful modification and understanding.
- Use AI tools to circumvent learning or misrepresent authorship.

Any use of generative AI must be openly acknowledged in an appendix or comment (e.g., "Used GPT5 to clarify concept X; prompt: …"). Work that appears to rely substantially on AI without understanding may be treated as academic misconduct in accordance with university policy. The teaching staffs reserve the right to reach out to individual students to check on their understanding of the submitted assessment items.

Communication and Feedback

Assessment marks for individual assessed tasks will be communicated via Canvas within two weeks of submission. Feedback on assignments will include [specific details, e.g., strengths, areas for improvement]. Students who have further questions about the feedback including marks should consult the instructor within five working days after the feedback is received.

Resubmission & Late Day Policy

Each student has a total of **3 late days** for use on assignments. They extend the due date by **24 hours** and a maximum of **2 late days** can be used towards any individual assignment.

Grading Written and Programming assignments will be hand graded. If you have used up your total late days, you will be penalized 25% per day, up to two days max, with no credit after two days. In cases of medical or other emergencies which interfere with your work, or resubmission needs, contact the instructor **before** the deadline for approval is required.

Required Texts and Materials

Reading materials will be available in each week's lecture slides.

References:

- Foundations of Machine Learning. Mehryar Mohri, Afshin Rostamizadeh, and Ameet Talwalkar. MIT Press, Second Edition, 2018. (Errata: https://cs.nyu.edu/~mohri/mlbook/errata_ed2_p1.html)

Academic Integrity

Students are expected to adhere to the university's academic integrity policy. Students are expected to uphold HKUST(GZ)'s Academic Honor Code and to maintain the highest standards of academic integrity. The University has zero tolerance of academic misconduct. Please refer to Regulations for Academic Integrity and Student Conduct for the University's definition of plagiarism and ways to avoid cheating and plagiarism.