DASC 3230 Statistical Modeling and Machine Learning, Fall 2023

1. Instructor

Yu HU MATH and LIFS Email: mahy@ust.hk Office: 3443
Yi LIAO LIFS Email: liaoy@ust.hk Office: 5459
Office Hour: TBA

2. Teaching assistants

Jihong TANG Email: jtangbd@connect.ust.hk
Zongchao MO Email: zmoad@connect.ust.hk
Jiabao Li Email: jligm@connect.ust.hk
Weihao Mai Email: wmai@connect.ust.hk

3. Meeting Time and Venue

Lectures: Wed & Fri, 1:30 pm to 2:50 pm, Rm 4502, Lift 25-26

Tutorial: Fri, 3:00 pm to 3:50 pm, Rm 4402, Lift 17-18

4. Course description

This course will provide an introduction to the fundamental concepts and methodologies routinely used in statistical modeling and machine learning in the context of analyzing scientific data. Various predictive and explanatory techniques designed for different data types will be introduced. Specific topics to be covered include linear regression, model selection, linear classification, support vector machines, neural networks, dimension reduction, clustering, graphical model and causal inference, ensemble models. Students will also gain hands-on experience with data analysis and visualization using Python.

Prerequisite: DASC 2220

The student should be familiar with basic linear algebra and calculus, and is also comfortable with basic programming and plotting in Python.

Exclusion: None

5. Intended Learning Outcomes

Upon successful completion of this course, students should:
1. Develop the ability to formulate scientific questions into appropriate statistical models and data analyses
2. Possess the knowledge to select the appropriate modeling and analytic approach(es) based on the nature of questions and the types of data
3. Be able to perform modeling and analysis using Python
4. Be able to interpret modeling outcomes and draw conclusions from analysis results

5. Develop a mechanistic understanding of the theories, potential pitfalls, and limitations of various modeling and analytic techniques covered in the course.

6. Assessment Scheme

Credit points: 3

Assessment	Course ILOs
Homework 40%	1,2,3,4,5
In-class discussions and participation 10%	1,2,4,5
Midterm exam 25%	1,2,4,5
Final exam 25%	1,2,4,5

Homework includes derivations, proofs, and numerical and programming problems.
Exam format: closed-book, but one A4 paper formula sheet is allowed

We will use Canvas to distribute and collect the homework. We will also send course announcements via Canvas. So please turn on the notification or check regularly.
When you have a question, you are encouraged to post it under the Discussion tab on Canvas, so others can see it and it may also help to answer. The instructor will also check and answer the Discussion tab regularly.

Important: You can and are encouraged to discuss and collaborate with classmates on homework. However, each person should write and submit their own answers (copy-and-paste is not allowed).

7. Student Learning Resources

Lecture slides are the primary reference materials and will be uploaded to Canvas before the lecture (usually the evening before).

Optional textbook/reference:
An Introduction to Statistical Learning by James et al. [available online from HKUST library]
Pattern recognition and machine learning by Bishop [reserved at HKUST library]
Machine learning a probabilistic perspective by Murphy [available online from HKUST library]

8. Teaching and Learning Activities

Scheduled weekly activities: 4 hours

Teaching Activities	Course ILOs
Lecture	1,2,3,4,5
Tutorial	3,4,5

9. Tentative Course Schedule

Week 1: Introduction
Week 2: Linear regression
Week 3: Model selection
Week 4: Linear classification
Week 5: Linear classification (cont'd)
Week 6: Kernel method and support vector machines
Week 7: Neural network, review for midterm
Week 8: Dimension reduction
Week 9: Dimension reduction (cont'd), Clustering
Week 10: Clustering (cont'd)
Week 11: Graphical models
Week 12: Graphical models (cont'd), Combining models
Week 13: Combining models (cont'd), review for final