BMEG3105 Fall 2025

Data analytics for personalized genomics and precision medicine

Course introduction

Lecturer: YuLI(李煜)fromCSE

Liyu95.com,liyu@cse.cuhk.edu.hk **Wednesday, 3rd September 2025**

Today's agenda

- To understand our collective background and expectations.
- To explain how we've structured the course to meet those expectations.
- To cover essential course logistics.
- To introduce the core theme: The role of DATA in personalized genomics and precision medicine.

Expect Outcomes:

- Make good use of the resulting huge datasets to understand basic biology and medical conditions
- Apply AI and machine learning in clinical applications in the system level
- Data analytics + Personalized genomics + Precision medicine

Teaching team

Instructor	**TAs**
Yu Li liyu@cse.cuhk.edu.hk Office hour: 3pm-5pm, Friday Location: SHB-106	Ziqian Lin linziqian@link.cuhk.edu.hk Office hour: 3pm-5pm, Tuesday Location: SHB-904 Xinyuan Liu 1155246738@link.cuhk.edu.hk Office hour: 3pm-5pm, Thursday Location: SHB-904

About Yu Li:

- 2020-Present, Assistant Professor at CSE, CUHK

  ➢Vice-Chancellor's Exemplary Teaching Award CUHK, 2024

  ➢MIT Technology Review, Innovators Under 35 (TR35) China, 2024

  ➢Asia-Pacific Leaders Under 30, 2024

  ➢Forbes 30 Under 30 Asia list, Class of 2022, Healthcare & Science

- 2022-2024 Summer, Visiting Assistant Professor, MIT/Harvard
- 2022-2024 Summer, Associated Scientist, Broad Institute of MIT and Harvard
- 2017–2020, Ph.D. in Computer Science, KAUST

  ➢2022, KAUST Alumni Change Makers Awards

- 2015–2016, M.S. in Computer Science, KAUST
- 2011-2015, B.S. (First- class honorary) in Biosciences, USTC

  ➢Graduated at 20

Pre-course survey results

Primary motivations	**Background strengths**:
Interest in biological, data analytics and health application Learn data analytics and machine learning techniques	Biology Probability and statistics
Background weaknesses	**Student concerns**
Machine Learning Algorithms Programming	Programming skills Mathematical understanding Biological context

(ESTR mirror of BMEG3105 is ESRE3605)

Course Logistics

Lecture time:

- Wednesday 9.30am-11.15am, SC L4

- Friday: 9.30am-10.15am, MMW703

Tutorial:

- Friday 10.30am-11.15am, MMW703

No video recording anymore according to the school policy

TA support will mainly focus on Python programming

Most of students choose the following grading breakdown:

- Homework (20%)
- Scribing (10%)
- In-class quizzes (10%)
- Midterm (20%)
- Project (20%)
- Final (20%)

Bonus points

- One bonus question in Midterm (2%)
- One additional scribing (1%)
- Pre-course survey + Post-lecture survey: 0.5% for each, and the maximum (3%)

AI tools are allowed with explicit acknowledgement

Late policy: will be 6 late days total and 2 max for any assignments or project.

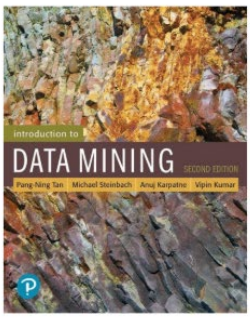
Assignments, Exams, scribing and Projects

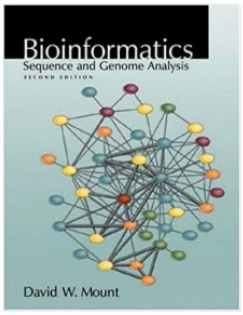
- Assignments: will be total 4 assignments (programming assignment 0 is special which no need to submit anything)
- Scribing: Summarize one of the lectures Submit it within one week after the lecture (Grades will be deducted by 25% for each additional late day)
- Midterm: Date will be October 17 (Friday) which is in -class. Open book exam.
- Final: Date is TBD and it is also open book exam. Centralized exam.
- Projects: A personalized project which provide potential projects.
  - There will be milestone report on November 7th which is for 5% of the total course.
  - The report will be due December 2^{nd,} which is for 5% of the total course
  - Presentation will be held on November 21st and November 28th and the presentation will be 7 minutes.

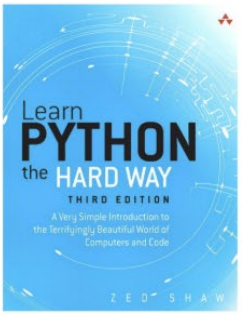
Course Schedule

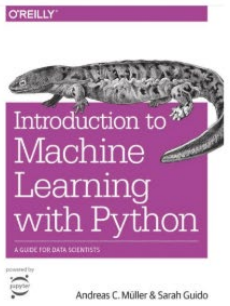
Lecture	Date	Location	Topic	Slides	Notes	Reading	Important dates (All due at 11:59 pm)
1	Sep 3 (Wed)	SC L4	Introduction				
2	Sep 5 (Fri)	MMW703	Data & Python				PA0 posted
3	Sep 10 (Wed)	SC L4	Data & Python & Sequence				Scribing preference registration due
4	Sep 12 (Fri)	MMW703	DP				A1 posted
5	Sep 17 (Wed)	SC L4	Assembly & Mapping				PA0 due
6	Sep 19 (Fri)	MMW703	Data exploration				
7	Sep 24 (Wed)	SC L4	Clustering				A1 due
8	Sep 26 (Fri)	MMW703	Classification				
9	Oct 3 (Fri)	MMW703	Classification & Perf evaluation				A2 posted
10	Oct 8 (Wed)	SC L4	Perf evaluation				
11	Oct 10 (Fri)	MMW703	Dim reduction				
12	Oct 15 (Wed)	SC L4	Midterm review				Quiz, A2 due
13	Oct 17 (Fri)	MMW703	Midterm				8:30am-11:15am, Midterm exam
			Module 2 start				
14	Oct 22 (Wed)	SC L4	Multi-omics overview				
15	Oct 24 (Fri)	MMW703	Cancer genomics overview				PA1 posted
16	Oct 31 (Fri)	MMW703	Genomics data analysis				
17	Nov 5 (Wed)	SC L4	Single cell genomics				
18	Nov 7 (Fri)	MMW703	Data visualization				Project M-report (Proposal) due
			Module 3 start				
19	Nov 12 (Wed)	SC L4	Deep learning				A3 posted
20	Nov 14 (Fri)	MMW703	CNN				PA1 due
21	Nov 19 (Wed)	SC L4	EHRs & Text				
22	Nov 21 (Fri)	MMW703	Project Presentation				A3 due
23	Nov 26 (Wed)	SC L4	Course review				Quiz
24	Nov 28 (Fri)	MMW703	Project Presentation				Project report due on 2 Dec

Reference Book









Brief overview of DATA in personalized genomics and precision medicine

He using different questions to help us to overview.

Why data analytics?

- A more efficient and cost benefit way for solving massive amounts of biological and health data.

What data can measure a person?

Can be concluded in the following part:

- **Molecular:** Genome, Transcriptome, Proteome, Metabolome, Microbiome.
- **Clinical & Physiological:** Medical Imaging (MRI, X-ray), Blood Tests, ECG, EHRs.
- **Lifestyle & Environmental:** Demographics, Diet, Exercise, Social Media, Pollution.