COMP7990 Principles and Practices of Data Analytics

Lecturer:

Dr. Zhang Lu, Eric

Lab Instructors:

Mr. Jintian Jiang (labs and quiz grading)

Teaching Assistants:

Ding Yi, ZOU Bohao

About Me

- Dr. Zhang Lu Eric
- Office: DLB 629
- Email: ericluzhang@comp.hkbu.edu.hk
- Research Interest: Deep learning in genomics, Complex disease prediction, AI in drug discovery

- Served:
  - Stanford University (Postdoctoral scholar)
  - CityU (PhD) and HKU (Mphil)

Contact Information

- Dr. Zhang Lu, Eric
  - Office: DLB629
  - Email: ericluzhang@comp.hkbu.edu.hk
- Mr. Jintian Jiang (For labs and quiz grading)
  - Office: RRS637
  - Email: jintian-jiang@comp.hkbu.edu.hk
- Mr. Ding Yi
  - Email: dylan1021@comp.hkbu.edu.hk
- Mr. ZOU Bohao
  - Email: csbhzou@comp.hkbu.edu.hk

Outline

Timetable

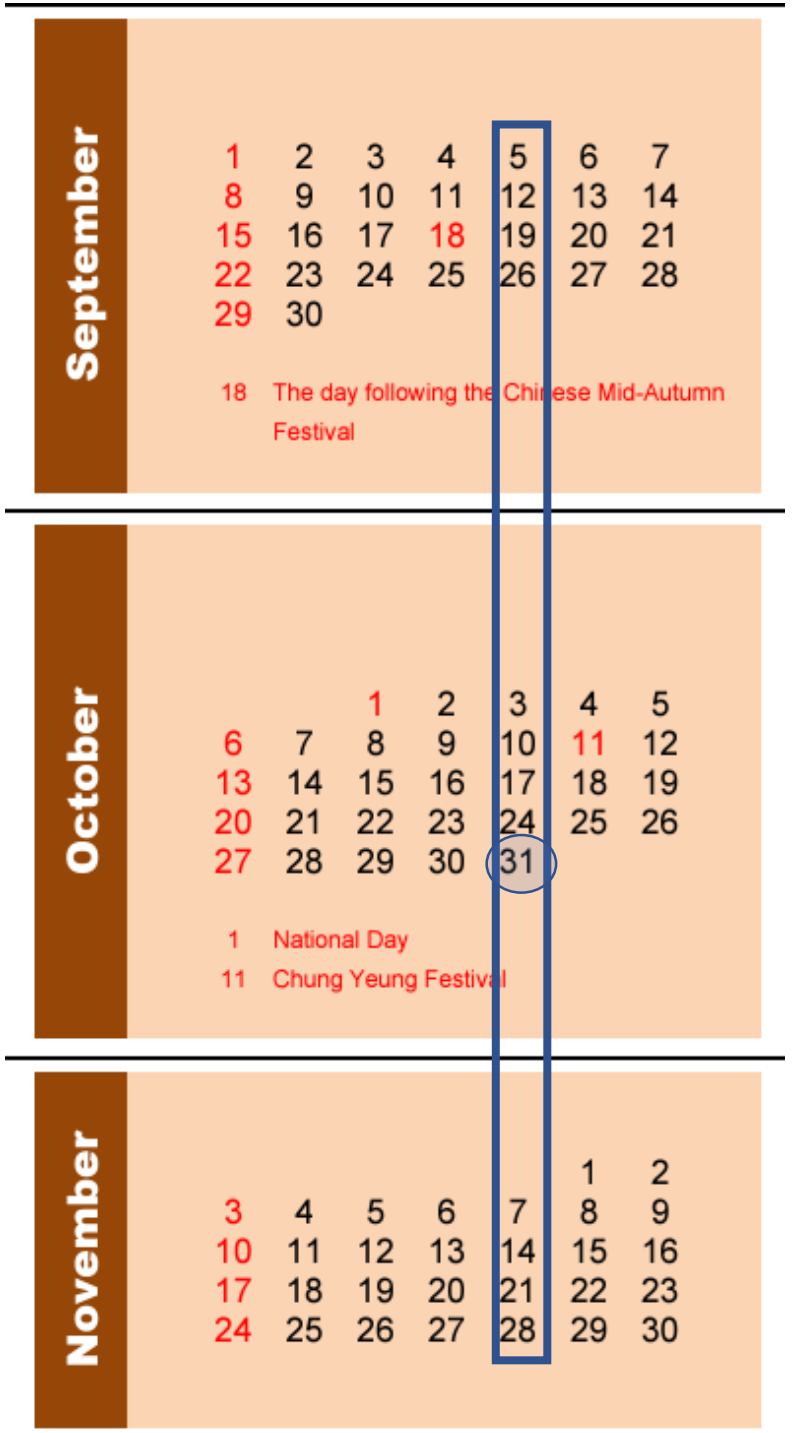
Course Contents

Learning Outcomes

Assessment Method

Timetable

- Instructor:
  - Dr. Zhang Lu, Eric
- Time of our classes
  - 13 weeks from Sep 5 to Nov 28
  - Time: 18:30~21:20 (Thursday)
  - Lecture room: WLB 103
- Quiz:
  - Scope: Lectures 1 to 6
  - Tentative Time: Oct 31st
  - Venue: WLB 103 (Seat arrangement will be released later)



Course Content

- Data Mining – Data preparation; Data mining algorithms (classification; clustering)
- Data Analytics – Background statistics; hypothesis testing, statistical analysis techniques
- Data Management – Database system concepts; Relational data model; SQL
- Data Visualization – Concepts of data visualization; charts, maps and infographics
- Data Security and Privacy – Concepts of data security and privacy; privacy protection principles

Tentative Class Schedule

Week	Topic
1	Basic Statistic
2	Inferential Statistics
3	Data Preprocessing and Univariate Linear Regression
4	Multivariate Linear Regression, Correlation and Perceptron
5	Artificial Neural Network, Support Vector Machine and k Nearest Neighbors (k-NN) algorithm
6	Clustering algorithms
7	Data Management 1
8	Lab 1: Statistics and Data Mining*
9	Quiz +Data Management 2
10	Visualization
11	Lab 2: SQL and Tableau*
12	Security
13	Revision

Course Aims

- This course introduces principal concepts of data management and analysis.
- It covers various topics including database management, data analytics, data mining, data visualization, and data privacy.
- It is expected that students can grasp practical skills about how to collect, store, analyze, and visualize data.

Expected Learning Outcomes

- Knowledge
  - Describe fundamentals of database management
  - Explain concepts of data analysis techniques and data mining algorithms
  - Describe and explain concepts of data visualization
  - Describe concepts and legal foundations of data security and privacy
- Professional Skill
  - Formulate SQL queries on the database
  - Conduct statistical analysis and design visualization to present analysis results

Assessment methods

- Continuous Assessment (40%)
  - Lab1 (12%),
  - Lab2 (12%),
  - Quiz (16%)
- Examination (60%)
  - Final examination

- Import Notices
  - Plagiarism: Students who plagiarized and who were plagiarized will be given zero mark.
  - Final Exam: In order to pass this course, students should attain at least 30% of the final examination mark.
  - Cheating in exam: **Students who cheated in the exam/quizzes may receive a failure grade of the course and may defer their study for one year**.
  - A cumulative GPA of at least 2.50 for graduation

Support

Post	**Post your question on Piazza: https://piazza.com/class/m051mzibx1s7l0/**
Email	Email your instructors or TA
Appointment	Make appointment for individual consultation with instructor and TA

Lab arrangement

- We have two labs for COMP7990 on Week 8 (Lab 1) and Week 11 (Lab 2).
- Please be mindful that you may be assigned to different labs due to the limited capacity of each lab.
- You are not required to bring your own device to the lab. However, you may also do that if you wish.