THE UNIVERSITY OF HONG KONG
FACULTY OF BUSINESS AND ECONOMICS

ECON 3225: Big Data Economics

GENERAL INFORMATION

Instructor: Dr. Jasmine Y. HAO and Dr. Ping YU
Email: haoyu@hku.hk; pingyu@hku.hk
Office: KKL 927; KKL 1108
Phone: 3917-7759, 2857-8358
Consultation times: TBA

Teaching time: 13:30-16:20, Thursday
Teaching location: KK102

Tutor: TBA

Pre-requisite(s): Introductory econometrics (ECON2280) or causal inference (ECON3284)
Co-requisite(s): None
Mutually exclusive: None

Course website: Available through HKU Portal e-learning
Other important details: None

COURSE DESCRIPTION

This course introduces students to fundamental ideas, important methods and popular techniques in big data analysis and machine learning. Combining statistical theory, computational tools, and hands-on experience with real data, this course will provide students with a solid basis for handling big data in economics, finance, and management.

The primary focus of this course is on application instead of methodological rigor. Hence, the use of mathematics will be limited to an elementary level. However, students are expected to have a strong background in statistics and/or econometrics. In particular, the course will assume that participants have an understanding of statistical inference using t-tests and have prior experience of interpreting the results of multiple linear regression. We will review these topics briefly during the course.

Because of the emphasis on hands-on experience, students are expected not to be scared by data and coding. Previous experience with statistical software and knowledge about computer programming is an advantage but not required. Homework assignments are designed to familiarize students with the necessary programing language. For programing, we will use the statistical package R via a front-end called RStudio. Both R and RStudio are free and open source.

COURSE OBJECTIVES

1. Provide students a broad overview of the most popular data science methods in economics and business studies
2. Enhance students' analytical ability to apply appropriate methods in different contexts
3. Equip students with a basic toolkit that can be directly used for their own research

FACULTY LEARNING GOALS (FLGs)

FLG1: Acquisition and internalization of knowledge of the programme discipline
FLG2: Application and integration of knowledge
FLG3: Inculcating professionalism
FLG4: Developing global outlook
FLG5: Mastering communication skills
FLG6: Cultivating leadership

COURSE LEARNING OUTCOMES (CLOs)

Course Learning Outcomes	Aligned Faculty Learning Goals (FLGs)
CLO1. Gain a solid understanding of the principles of applying data science to social	FLG 1, FLG 2

sciences CLO2. Demonstrate a solid grounding in recent developments in big data methods, including state-of-the-art machine learning techniques and their suitability to solve important economic, finance, and business problems. CLO3. Demonstrate ability to address questions of interest by using applied data science and econometric techniques. CLO4. Demonstrate facility with implementing the techniques covered in the course using statistical software on real-world datasets	 FLG 1, FLG 2, FLG 5 FLG 1, FLG 2, GLG 4, FLG 5 FLG 1, FLG 2, FLG 3

COURSE TEACHING AND LEARNING ACTIVITIES

Course Teaching and Learning Activities	Expected Study Hours	Study Load (% of study)
T&L1. Lectures	36	30
T&L2. Weekly problem sets	36	30
T&L3. Computer programming	24	20
T&L4. Tutorial/self-learning sessions	24	20
Total	120	100

Assessment Methods	Brief Description (Optional)	Weight	Aligned Course Learning Outcomes
A1. Four Problem sets		80	CLO1, CLO2, CLO3, CLO4
A2. One Final Project		20	CLO1, CLO2, CLO3, CLO4
	Total	100%	

STANDARDS FOR ASSESSMENT

Course Grade Descriptors

A	Strong evidence of superb ability to fulfill the intended learning outcomes of the course at all levels of learning: describe, apply, implement, evaluate and synthesis.
B	Strong evidence of ability to fulfill the intended learning outcomes of the course at all levels of learning: describe, apply, implement, evaluate and synthesis.
C	Evidence of adequate ability to fulfill the intended learning outcomes of the course at low levels of learning; such as describe and apply, but not at high levels of learning such as evaluate and synthesis.
D	Evidence of basic familiarity with the subject.
F	Little evidence of basic familiarity with the subject.

Assessment Rubrics for Each Assessment (Please provide us the details in a separate file if the space here is not enough)

Problem sets are mainly extracted from the textbook, including both analytical and empirical exercises. A few other

exercises are not from the textbook, but are helpful to understand the course materials. The project asks the students to predict a quantitative or qualitative outcome using the techniques learned in this course or even out of this course. The best prediction will get the full score, and the worst prediction will get a pass score. Students can form a team up to five members to do the problem sets and the project. The team members for each of the four assignments and the project need not be the same, i.e., each student can join at most five teams.

COURSE CONTENT AND TENTATIVE TEACHING SCHEDULE

Lecture 1. Introduction to Statistical Learning
Lecture 2. Classification
Lecture 3. Clustering
Lecture 4: Model Selection and Regularization
Lecture 5: Principal Components Analysis
Lecture 6*: Text as Data
Lecture 7: Moving Beyond Linearity
Lecture 8: Tree-Based Methods
Lecture 9: Support Vector Machines
Lecture 10: Deep Learning

REQUIRED/RECOMMENDED READINGS & ONLINE MATERIALS (e.g. journals, textbooks, website addresses etc.)

Required: (**ISLR2**) *An introduction to Statistical* Learning, 2nd edition. James, Witten, Hastie, and Tibshirani. Springer. 2021. (The digital version of this book is freely available on https://hastie.su.domains/ISLR2/ISLRv2_website.pdf)

Useful references:

Data Science for Business: *What you need to know about data mining and data-analytic thinking.* Provost and Fawcett. O'Reilly. 2013.

Business Data Science: *combining machine learning and economics to optimize, automate, and accelerate business decisions*. Taddy. McGraw Hill. 2019.

MEANS/PROCESSES FOR STUDENT FEEDBACK ON COURSE

☐ conducting mid-term survey in additional to SETL around the end of the semester

☒ Online response via Moodle site

☒ Others: ____SETL_____ (please specify)

COURSE POLICY (e.g. plagiarism, academic honesty, attendance, etc.)

1. This is an active learning course, and attendance and participation are extremely important. Please observe appropriate classroom etiquette and be considerate to others. In particular, laptop use should be limited to course-related activities, and cell phones are not allowed in class.

2. All the turn-ins must be typewritten.

3. Plagiarism and cheating are serious academic offenses, so copying other teams' answers is not permitted even with consent.

ADDITIONAL COURSE INFORMATION (e.g. e-learning platforms & materials, penalty for late assignments, etc.)

All course materials can be downloaded from Moodle. Late assignments and project are not acceptable for whatever reasons. To avoid any risk, start your assignments and project early (the assignments indicate clearly which problems can be solved after each lecture).