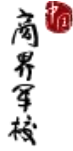




PHBS
北京大学汇丰商学院



Big Data Analysis

Module II, Academic Year 2025-26

Course Information

Instructor: Jian FENG, PhD, Assistant Professor of Finance, Peking University HSBC Business School

Office: PHBS Building, Room 649

Phone: 86-755-2603-4026

Email: jianfeng95@phbs.pku.edu.cn

Office Hour: TBA

Teaching Assistant: TBA

Phone: TBA

Email: TBA

Classes: Nov. 14,18,21,25,28, Dec. 2,5,9,12,16,19,23,26,30, Jan. 2,6,9,13

Lectures: Tue & Fri 10:30-12:20

Venue: PHBS Building, Room TBA

Course Website:

No course website.

1. Course Description

1.1 Context

Course overview: This course covers recent advancements in big data analysis, including machine learning (ML) methods, natural language processing (NLP), economic network analysis, and generative AI. Students will learn different algorithms, understand their applications in leading journal articles, and explore them in data work.

Prerequisites: A preliminary understanding of financial machine learning and Python is recommended but not required.

1.2 Textbooks and Reading Materials

Textbooks

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning: with applications in R*.

Sargent, T. J., Sargent, T. J., & Stachurski, J. (2024). *Economic networks: Theory and computation*.

Mitsotakis, D. (2023). *Computational mathematics: An introduction to numerical analysis and scientific computing with Python*.

Reading Materials and Research Papers will be provided in class.

2. Learning Outcomes

2.1 Intended Learning Outcomes

Learning Goals	Objectives	Assessment (YES with details or NO)
1. Our graduates will be effective communicators.	1.1. Our students will produce quality business and research-oriented documents.	NO
	1.2. Students are able to professionally present their ideas and also logically explain and defend their argument.	NO
2. Our graduates will be skilled in team work and leadership.	2.1. Students will be able to lead and participate in group for projects, discussion, and presentation.	YES. Performance of group presentation will be evaluated by other groups and the course instructor.
	2.2. Students will be able to apply leadership theories and related skills.	
3. Our graduates will be trained in ethics.	3.1. In a case setting, students will use appropriate techniques to analyze business problems and identify the ethical aspects, provide a solution and defend it.	NO
	3.2. Our students will practice ethics in the duration of the program.	NO
4. Our graduates will have a global perspective.	4.1. Students will have an international exposure.	NO

5. Our graduates will be skilled in problem-solving and critical thinking.	5.1. Our students will have a good understanding of fundamental theories in their fields.	YES. Understanding of big data algorithms and ability to use them in data work will be evaluated through coding assignments and in-class quizzes.
	5.2. Our students will be prepared to face problems in various business settings and find solutions.	
	5.3. Our students will demonstrate competency in critical thinking.	

2.2 Course specific objectives

2.3 Assessment/Grading Details

Coding Assignments: 10% + 15% * 2

In-class Quiz: 15% * 3

In-class Group Presentation: 15%

Course attendance issues (e. g., absence without justification) will be considered in grading. Final grades will be awarded in letter grades (e. g. A+, A-, ..., D+, D, F). Percentage of the distribution of each letter grade is limited:

(1) No more than 30% can receive A+, A and A minus;

(2) No more than 90% can receive B and above.

Grade	GPA	Note
A+	4.0	优秀/ Excellent
A		
A-	3.7	
B+	3.3	良好/ Good
B	3.0	
B-	2.7	
C+	2.3	一般/ Average
C	2.0	
C-	1.7	
D+	1.3	及格/ Fair
D	1.0	
F	0	不及格/ Fail

2.4 Academic Honesty and Plagiarism

It is important for a student's effort and credit to be recognized through class assessment. Credits earned for a student work due to efforts done by others are clearly unfair. Deliberate dishonesty is considered academic misconducts, which include plagiarism; cheating on assignments or examinations; engaging in unauthorized collaboration on academic work; taking, acquiring, or using test materials without faculty permission; submitting false or incomplete records of academic achievement; acting alone or in cooperation with another to falsify records or to obtain dishonestly grades, honors, awards, or professional endorsement; or altering, forging, or misusing a University academic record; or fabricating or falsifying of data, research procedures, or data analysis.

All assessments are subject to academic misconduct check. Misconduct check may include reproducing the assessment, providing a copy to another member of faculty, and/or communicate a copy of this assignment to the PHBS Discipline Committee. A suspected plagiarized document/assignment submitted to a plagiarism checking service may be kept in its database for future reference purpose.

Where violation is suspected, penalties will be implemented. The penalties for academic misconduct may include: deduction of honour points, a mark of zero on the assessment, a fail grade for the whole course, and reference of the matter to the Peking University Registrar.

AI tools requirements:

Using AI tools to complete assignments or assessments without the approval of the course instructor will be regarded as an act of academic dishonesty. Depending on the severity of the situation, penalties will be implemented in accordance with the provisions of the Peking University Graduate Student Handbook.

For more information of plagiarism, please refer to *PHBS Student Handbook*.

3. Topics, Teaching and Assessment Schedule

Course arrangements are tentative and depend on progress.

Week 1: Introduction to big data analysis, including Python language and basic data analysis.

Weeks 2-3: Basic machine learning methods, including linear algorithms with regularizations, dimension Reduction, clustering and classification methods, and basic numerical methods.

Weeks 4: Economic Network.

Weeks 5-7: Textual Analysis, Natural Language Processing.

Weeks 8-9: Big data and AI.

Assignments will include 2-3 take-home coding works. Arrangements are tentative and depend on progress.

Assignment I: Data preprocess, basic portfolio sorting, linear models.

Assignment II: Dimension reduction and clustering.

Assignment III: Textual analysis.