

Foundation of Data Science

Department: Fudan International Summer Session 2026

Course Code	GEIS30010		
Course Title	Foundation of Data Science		
Credit	2	Credit Hours	36+3 tutorial hours (one credit hour is 45 minutes)
Course Nature	<input type="checkbox"/> Specific General Education Courses <input type="checkbox"/> Core Courses <input checked="" type="checkbox"/> General Education Elective Courses <input type="checkbox"/> Basic Courses in General Discipline <input type="checkbox"/> Professional Compulsory Courses <input type="checkbox"/> Professional Elective Courses <input type="checkbox"/> Others		
Course Objectives	The course, Foundations of Data Science, provides a comprehensive introduction to modern data science with a strong emphasis on the mathematical and algorithmic foundations of machine learning. The course content is primarily based on the undergraduate course, from University of Chicago and UC-Berkeley.		
Course Description	Students will develop practical Python programming skills for data analysis while gaining rigorous understanding of the mathematical principles that underpin modern machine learning. The course adopts a "learn by doing" approach: every mathematical concept is immediately applied to real-world datasets and problems. By the end of this course, students will be able to: <ol style="list-style-type: none"> 1. Write Python programs for data analysis, visualization, and modeling. 2. Understand key linear algebra concepts (vectors, matrices, projections, SVD) and their role in machine learning. 3. Apply fundamental statistical reasoning including sampling, hypothesis testing, and confidence intervals. 4. Formulate and solve regression and classification problems using least squares, regularization, and kernel methods. 5. Understand dimensionality reduction via PCA and its applications. 6. Implement gradient-based optimization and understand the basics of neural networks and backpropagation. 7. Apply unsupervised learning methods including k-means clustering and mixture models. 8. Critically evaluate model performance using cross-validation, bias-variance tradeoffs, and appropriate metrics. 		

Course Requirements:

The pre-requisite of this class is basic high school algebra and an inquisitive mind. There is no requirement on prior programming experience. Each student is expected to have a computer. Either Windows or Mac is fine.

Teaching Methods:

Lectures

Instructor's Academic Background:

Dr. Roxie Jiang received her Ph.D. in Computer Science from University of Chicago. Focusing on the intersection of AI for Science and computer vision, her work centers on theoretical innovation and applications in generative models and representation learning. She published top-tier applications in ML conferences. Looking ahead, she plans to advance a unified framework that brings together generative and representation learning, incorporate physical priors to enhance neural simulators, and explore cross-modal unified foundation models for applications in Scientific AI.

Dr. Hao Xiong received his Ph.D. from Shanghai Jiao Tong University. His research focuses on graph representation learning, network mining, quantum machine learning, large model architecture and combinatorial optimization.

Email:

Roxie_jiang@fudan.edu.cn

Course Schedule:

	Topic	Reading	Lab
1	Introduction	Chapters 1 - 2	Lab1: Arrays and Tables
	Expressions and data types	Chapters 3 - 4	
	Lab time		
2	correlation and regression	Chapter 5	Lab2: Least squares
	Least squares	Chapter 6	
	Lab time		
3	Functions and apply	Chapter 7	Lab3: Kernel Methods
		Chapter 8 intro, 8.1	
	Lab time		
4	Dimensionality reduction	8.2-8.5	Lab4: PCA

	SVD, PCA	9.1-9.2	
	Lab time		
5	Unsupervised learning	9.3 - 9.5	Lab5: K-means
	K-means	Chapter 10	
	Lab time		
6	Basics of neural network	11.1	Lab6: Neural Network
	Clustering	11.2-11.4	
	Lab time		
7	Bootstrap	Chapter 12	Lab 7: Bootstrap
	Statistical Inference	Chapter 13	
	Lab time		
8	Model Evaluation	13.4, 14.1-14.2	No Lab
	Review	14.3-14.4	
	Free time		
9	Advanced regression	14.5, 14.6	Lab8: Model validation
	Validation	15.1 - 15.2	
	Lab time		
10	Parameter regularization	15.3, 15.5, Chapter 16	Lab9: Regularization
	Prediction	Chapter 17	
	Lab time		
11	Comprehensive review	Chapter 18	Lab10: Final review & practice
	Wrap up and finish lab 10	No reading	
12	Final Exam		

The design of class discussion or exercise, practice, experience and so on:

Exercise and practice

Grading & Evaluation

Your final grade will be determined via the following percentages:

Lecture participation points: 10%

Labs: 50%

Final: 40%

Important grading policies:

- You must score at least 55% on the final exam to pass the course. If you score lower than 55% on the final, you will receive an F for the course, regardless of your overall average.
- Every student should follow the policy on pair programming.
- According to Fudan University's policy, there is a threshold on the percentage of students who may receive A or A- in a class. Please keep this policy in mind.
- There will be no make - up exam.

Teaching Materials & References:

- Textbook for our class will be the freely available awesome textbook, "Computational and Inferential Thinking - The Foundations of Data Science", by Ani Adhikari and John DeNero.
- A reference textbook very useful for AI is "Artificial Intelligence: A Modern Approach" by S. Russell and P. Norvig.
- There will be a reading assignment for most of the days. It is expected that you complete the reading assignment before the start of the lecture.
- You should score at least 55% in the final exam to get a passing grade for this class, regardless of your overall percentage.
- There is an optional final project that will be treated as extra credit.