

# Big Data Analytics

COMP4434

## Instructor Info —



HUANG Xiao



Office Hrs: Fri 14:30-16:30



PQ837



<https://web.comp.polyu.edu.hk/xiaohuang/>



xiaohuang@comp.polyu.edu.hk

## Course Info —



26 Hrs Lectures + 13 Hrs Lab



Friday



16:30 -18:20



GH201

## TA Info —



YANG Chang



chang.yang@connect.polyu.hk



Zhang Qinggang



qinggangg.zhang@connect.polyu.hk



Du Jiahe



jiahe.du@connect.polyu.hk



Yuan Zheng



yzheng.yuan@connect.polyu.hk



Dong Su



su.dong@connect.polyu.hk

## Overall Objectives

- 1) Introduce students the characteristics of big data (volume, velocity, variety, and veracity) and the challenges of big data analytics;
- 2) teach students in applying skills and tools to manage and analyze the big data.

## Grading Scheme

5%	Attendance/Participation
25%	Three Assignments
5%	Mini-project Presentation
25%	Mini-project Report
40%	Final Examination

Grades will roughly based on the following distribution: 20% of (A+, A, A-), 50% of (B+, B, B-), 20% of (C+, C, C-), 10% of (D+, D, F). Curving is at the discretion of the professor.

## Learning Objectives

Upon completion of the subject, students will be able to:

- Understand the characteristics of big data and the challenges of big data analytics and why existing technology is inadequate to analyze the big data
- Understand how to collect, manage, store, and query various form of big data
- Understand how to analyze big data using various quantitative methods
- Gain hands-on experience on large-scale analytics tools to solve some open big data problems
- Be able to conduct thorough analysis on the impact of big data for business decisions and strategy in real-world applications

## Academic Integrity

Academic Integrity refers to the honest and ethical manner in which academic work is done, whether it is an assignment, an examination, an oral presentation, a project, or a report. PolyU views Plagiarism as a serious disciplinary offence. It is a fundamental value that all students at PolyU are expected to uphold. Academic Integrity is central to the ideals of this course. Students are expected to be independently familiar with the Regulations of Academic Integrity and to recognize that their work in the course is to be their own original work that truthfully represents the time and effort applied. Violations of the Regulations are most serious and will be handled in a manner that fully represents the extent of the Regulations and that befits the seriousness of its violation.

## Grading Policy

Assignments have to be uploaded using Learn@PolyU. If homework is submitted after its due time, it will be considered a full day late. There will be a 10% deduction for homework that is up to two days late, and a 20% deduction for homework that is three days late. We will not accept any homework that is more than three days late. Plan your time carefully, and don't wait until the last minute to start an assignment so you have time to ask questions and get help. For assignments and exams, extensions and makeups will only be given in documented cases of serious illness or other emergencies. Leaving a phone message or sending an e-mail without confirmation is not acceptable.

## Class Schedule (tentative)

Week 1 (Jan 17)	Introduction to big data analytics	No lab
Week 2 (Jan 24)	Gradient descent algorithm	Lab 1
29 - 31 Jan	<a href="#">Lunar New Year Holidays</a>	No lab
Week 3 (Feb 7)	Logistic regression classifier, F1 score, & overfitting	Lab 2, HW1 Due
Week 4 (Feb 14)	Cross-validation, support vector machines & clustering	Lab 3
Week 5 (Feb 21)	Multilayer perceptron & backpropagation	Lab 4
Week 6 (Feb 28)	Convolutional Neural Networks	Lab 5, HW2 Due
Week 7 (Mar 7)	Recurrent Neural Networks	Lab 6
Week 8 (Mar 14)	MapReduce & Hadoop	Lab 7
Week 9 (Mar 21)	Recommender systems	Lab 8, HW3 Due
Week 10 (Mar 28)	Collaborative filtering & dimensionality reduction	Lab 9
4 Apr	<a href="#">Ching Ming Festival</a>	No lab
Week 11 (Apr 11)	PageRank	Lab 10
25 Apr - 13 May	Examination (covering all content in lectures and labs)	mini-project Due

## No Textbook. Here are Several Reference Books.

1. Mining of Massive Datasets. Jure Leskovec, Anand Rajaraman, Jeff Ullman.

<http://www.mmds.org>

2. Introduction to Data Mining. Pang-Ning Tan, Michael Steinbach, Anuj Karpatne, Vipin Kumar.

<https://www-users.cse.umn.edu/~kumar001/dmbook/index.php>

3. Probabilistic Machine Learning: An Introduction. Kevin Patrick Murphy. MIT Press, 2022.

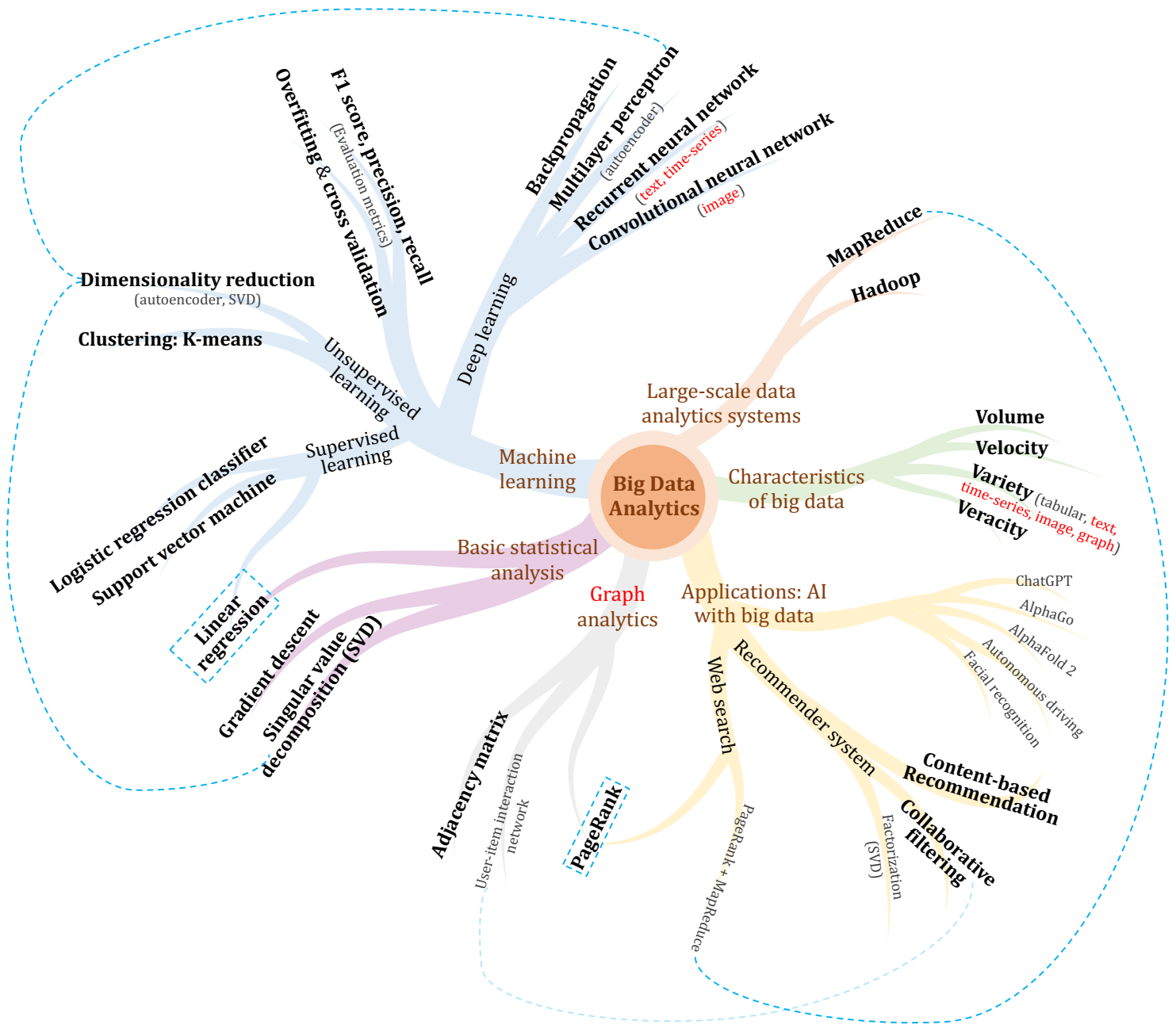
<https://probml.github.io/pml-book/book1.html>

4. Deep learning with PyTorch. Eli Stevens, Luca Antiga, Thomas Viehmann, Soumith Chintala. 2020.

Resource List in PolyU Library:

[https://julac-hkpu.alma.exlibrisgroup.com/leganto/public/852JULAC\\_HKPU/lists?courseCode=COMP4434&auth=SAML](https://julac-hkpu.alma.exlibrisgroup.com/leganto/public/852JULAC_HKPU/lists?courseCode=COMP4434&auth=SAML)

## Mind Map



You will gain hands-on experience in using Python, especially scikit-learn and PyTorch, during lab tutorials.

Take one of the five sessions	Mon	16:30-17:20	PQ604A
	Thu	12:30-13:20	PQ604B
	Thu	15:30-16:20	PQ604A
	Thu	16:30-17:20	PQ604A
	Fri	14:30-15:20	PQ604A

## Diversity and Inclusivity Statement

I consider this classroom to be a place where you will be treated with respect, and I welcome individuals of all ages, backgrounds, beliefs, ethnicities, genders, gender identities, gender expressions, national origins, religious affiliations, sexual orientations, ability - and other visible and non-visible differences. All members of this class are expected to contribute to a respectful, welcoming and inclusive environment for every other member of the class.

## Accommodations for Students with Disabilities

If you are a student with learning needs that require special accommodation, contact the Resources and Support Section at 2766 6800 or [srss.info@polyu.edu.hk](mailto:srss.info@polyu.edu.hk), as soon as possible, to make an appointment to discuss your special needs and to obtain an accommodations letter. Please e-mail me as soon as possible in order to set up a time to discuss your learning needs.