Artificial Intelligence and Big Data Computing Programming

COMP5113

Instructor Info

HUANG Xiao

Office Hrs: Thu 15:30 - 17:30

PQ837

https://www4.comp.polyu.edu.hk/~xiaohuang

xiaohuang@comp.polyu.edu.hk

Course Info

Thursday

18:30 -21:20

V312

TA Info

DU Jiahe

jiahe.du@connect.polyu.hk

DONG Junnan

junnan-hanson.dong@connect.polyu.hk

Overview

The objectives of this subject are: 1) provide students with the programming and real-world problem-solving skills in AI and big data computing; 2) equip students with the knowledge and skills in utilizing big data tools and techniques under various application scenarios; 3) offer hands-on experience to students under big data platforms (e.g., Hadoop, Python, NoSQL), and on the implementation of deep learning models.

Grading Scheme

25%	Three Assignments
15%	Midterm
20%	Mini-project
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 concept of artificial intelligence and big data computing, including terminologies, challenges, and applications;
- understand the major techniques, programming languages, and tools for big data computing;
- be familiar with big data computing tools;
- understand the typical machine learning algorithms, especially deep learning models;
- design and implement deep learning models to solve real-world problems.

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 (Sept 7)	Introduction to AI and big data	
Week 2 (Sept 14)	Gradient descent & overfitting & evaluation metrics & cross validation	
Week 3 (Sept 21)	Logistic regression classifier & support vector machines	
Week 4 (Sept 28)	Clustering & multilayer perceptron	HW1 due
Week 5 (Oct 5)	Backpropagation	
Week 6 (Oct 12)	Convolutional neural networks	
Week 7 (Oct 19)	Midterm (covers content learned in Lectures 1-5)	
Week 8 (Oct 26)	Recurrent neural networks	HW2 due
Week 9 (Nov 2)	MapReduce & Hadoop & CNN Implementation	
Week 10 (Nov 9)	Application I. Recommender systems: Collaborative filtering	
Week 11 (Nov 16)	SVD & Dimensionality reduction	HW3 due
Week 12 (Nov 23)	Application II. Web search: PageRank	
Week 13 (Nov 30)	Presentations	
Dec 20	Final Exam 19:00-21:00 at room SH1	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.

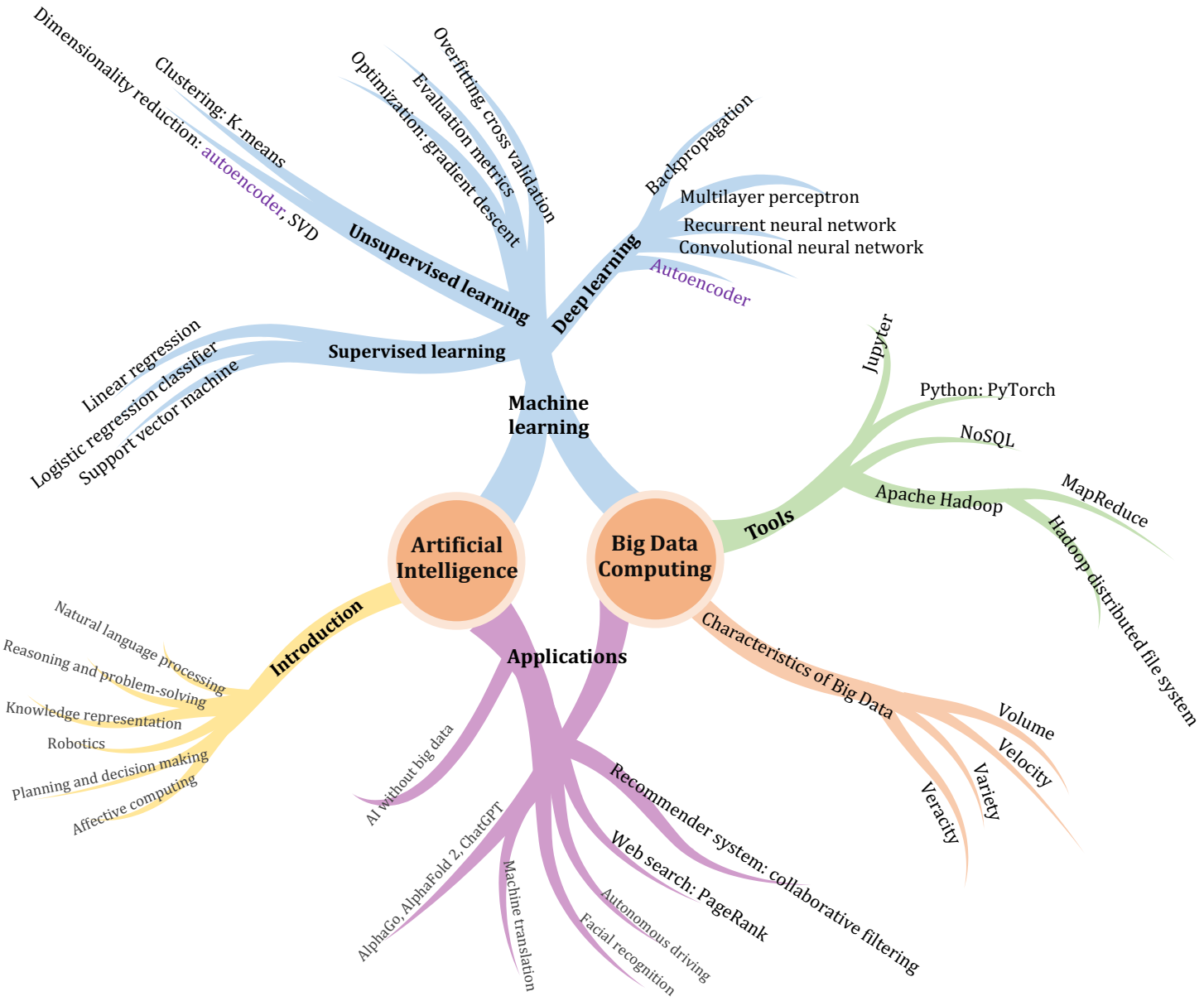
`https://probml.github.io/pml-book/book1.html`

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`

Mind Map



Diversity and Inclusivity Statement

I consider this online 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, 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.