

Big Data Ecosystems

EEL 6935: Big Data Ecosystems (Spring 2018)

Announcements:

We will use all kinds of cloud resources for the course projects, including GENI, Amazon Web Services, Google Cloud, NSFCLOUD, and GatorCloud. We appreciate the support from the National Science Foundation grants, Amazon AWS in Education Faculty/Research grants, GENI and GENI Rack grants.

We are using the e-learning course management system, <https://lss.at.ufl.edu/> (<https://lss.at.ufl.edu/>) (<http://elearning.ufl.edu/>) (<http://elearning.ufl.edu/>). Go to the Learning Support Services (<http://lss.at.ufl.edu/>) homepage, click "e-Learning Login" button, and enter your GatorLink username and password.

This course involves intensive programming and extensive software systems. We use many professional tools for coding, project management, and documentation, e.g., Asana, Trello, GitHub, Google Drive etc.

Syllabus

Instructor:

Dr. Xiaolin (Andy) Li
Office: 433 NEB
Office Hours: W 2pm-2:50pm
Email: [andyli-at-ece \(suffix .ufl.edu\)](mailto:andyli-at-ece (suffix .ufl.edu))
Web: <http://www.andyli.ece.ufl.edu/> (<http://www.andyli.ece.ufl.edu/>)

Teaching Assistants:

Mr. Yanjun Li
Mr. Xiyao Ma
Mr. Guohao Yu
Mr. Xiaoyong Yuan

Office: 406 NEB
Office Hours: MWF 1pm-2pm
Email: yanjun.li, maxiy, guohaoyu, chbrian (suffix .ufl.edu)

Class Meeting Time and Place:

Time: MWF 8 (3pm-3:50pm)
Place: NEB 202

Course Objective and Description:

Big data features high volume, high velocity, and high variety. The tremendous big data generated from natural systems, engineered systems, and human activities require new capabilities in algorithms and systems to explore insights and make decisions. To address the challenges of big data, this course covers the full spectrum of big data ecosystems: algorithms, systems, and big data analytics at scale. It consists of an overview of representative data mining, statistics, and machine learning algorithms (particularly deep learning), a thorough coverage of big data analytics software stack and underlying large-scale systems, and a holistic methodology on the design of big data ecosystems. Real-world case studies will be explored in science, engineering, business, and health.

Prerequisite and Co-requisites: machine learning, data science, statistics, or instructor approval.

Textbook:

Deep Learning, Ian Goodfellow, Yoshua Bengio, and Aaron Courville, MIT Press, 2016. [Link (<http://www.deeplearningbook.org/>)]

Other References:

Recent conference papers and online resources/documents
Machine Learning: A Probabilistic Perspective, Kevin P. Murphy, MIT Press, 2012.
Artificial Intelligence: A Modern Approach, Stuart Russell and Peter Norvig, Prentice Hall, 3rd Edition, 2009.
Pattern Recognition and Machine Learning, Christopher M. Bishop, Springer, 2007.
Reinforcement Learning: An Introduction, Richard S. Sutton and Andrew G. Barto, 1st Edition, 1998; 2nd Edition, 2017. [Link (<http://www.incompleteideas.net/book/the-book.html>)]

Machine Learning, Tom M. Mitchell, McGraw-Hill, 1997.

Mining of Massive Datasets, Jure Leskovec, Anand Rajaraman, and Jeffrey David Ullman, Cambridge University Press, 2nd Edition, 2014. [Link (<http://mmds.org/>)]

Data Mining: Concepts and Techniques, Jiawei Han, Micheline Kamber, Morgan Kaufmann, 3rd Edition, 2011.

The Fourth Paradigm: Data-Intensive Scientific Discovery, Tony Hey, Stewart Tansley, and Kristine Tolle, Microsoft Research, 2009.

Cloud Computing for Data Intensive Applications, X. Li and J. Qiu (Eds.), Springer, 2014.

Hadoop YARN, Arun Murthy, Vinod Kumar Vavilapalli, Doug Eadline, Jeffrey Markham, Joseph Niemiec, Addison Wesley, 2014.

Hadoop: The Definitive Guide, Tom White, O'Reilly Media, 4th Edition, 2015.

Data-Intensive Text Processing with MapReduce, Jimmy Lin and Chris Dyer, 2010.

The Go Programming Language, Alan A.A. Donovan and Brian W. Kernighan, Addison-Wesley, 2015.

Functional Programming in Scala, Paul Chiusano and Runar Bjarnason, Manning, 2014.

Python Essential Reference, David M. Beazley, Addison-Wesley, 2009.

Course Homepage:

<http://www.andyli.ece.ufl.edu/index.php/teaching/bigdata/> (<http://www.andyli.ece.ufl.edu/index.php/teaching/bigdata/>)

Course Outline (tentative):

Introduction

Basic Algorithms

Tools: **Python**, R

Big Data Stack

Tools: **Spark**, Hadoop, Cassandra, ZooKeeper

Deep Learning

Regression, Classification, Dimension Reduction, Recommender Systems, Mining Data Streams

DNN, CNN, RNN, DRL, Neural Computer

Applications in vision, language, game, self-driving, health, business, IoT, security, science, engineering

Tools: Caffe2, CNTK, MXNet, **TensorFlow**, **PyTorch**

Case Studies: Vision, NLP, Genomics, Security; Business, Health, IoT, Surveillance, Science, Engineering

Grading Policies:

Class participation and contribution (bonus): 5%

Homework assignments, reading summary, and paper presentation: 40%

Programming assignments (30%)

Reading Summaries (5%)

Paper Presentation and Demo (5%)

Course Project: 60%

Proposal (5%)

Midterm Presentation (10%)

Final Presentation, Demo/Code, Poster, Report (45%)

Note: Homework and programming assignments are due by 11:59 pm of the due date (unless announced in class otherwise). Late homework (non-programming) will NOT be accepted. Late program penalty is 10% per day, according to the timestamp of your online submission. Only when verifiable extenuating circumstances can be demonstrated will extended assignment due dates be considered. Verifiable extenuating circumstances must be reasons beyond the control of the students, such as illness or accidental injury. Poor performance in class is not an extenuating circumstance. Inform your instructor of the verifiable extenuating circumstances in advance or as soon as possible. In such situations, the date and nature of the extended due dates for the assignments will be decided by the instructor.

Attendance Policy:

Attendance is required. Students are responsible for any material covered in class. Lots of the materials covered in class will not be in the textbook. Announcements about homework, projects, programming assignments, etc. may be made in class or online or by emails. Students are encouraged to check the course Canvas and Asana systems regularly.

Collaboration Policy:

Discussion of techniques and ideas covered in class is encouraged. However, every line of all assignments must be your own (or your team's). In programming assignments, discussion of techniques in a natural language (such as English) is allowed, but a discussion in a computer or algorithmic language is not allowed. (Computer language discussions and questions are to be limited to the language and should not concern the assignment.) **Stealing, giving or receiving any code, drawings, diagrams, texts or designs (from others or Internet) is not allowed.** Project reports should be written in your own words; apparent copy (**ONE sentence or more**) is assumed as **plagiarism, if not quoted**. Students who do not comply with the above-described collaboration policy will receive a grade of **F** in the course. Furthermore, the case will be **reported** to the University Officials.

Honesty Policy

All students admitted to the University of Florida have signed a statement of academic honesty committing themselves to be honest in all academic work and understanding that failure to comply with this commitment will result in disciplinary action. This statement is a reminder to uphold your obligation as a UF student and to be honest in all work submitted and exams taken in this course and all others.

Accommodation for Students with Disabilities

Students Requesting classroom accommodation must first register with the Dean of Students Office. That office will provide the student with documentation that he/she must provide to the course instructor when requesting an accommodation.

UF Counseling Services

Resources are available on-campus for students having personal problems or lacking clear career and academic goals. The resources include:

- * University Counseling Center, 301 Peabody Hall, 392-1575, Personal and Career Counseling.
- * SHCC mental Health, Student Health Care Center, 392-1171, Personal and Counseling.
- * Center for Sexual Assault/Abuse Recovery and Education (CARE), Student Health Care Center, 392-1161, sexual assault counseling.
- * Career Resource Center, Reitz Union, 392-1601, career development assistance and counseling.

Software Use

All faculty, staff, and students of the University are required and expected to obey the laws and legal agreements governing software use. Failure to do so can lead to monetary damages and/or criminal penalties for the individual violator. Because such violations are also against University policies and rules, disciplinary action will be taken as appropriate. We, the members of the University of Florida community, pledge to uphold ourselves and our peers to the highest standards of honesty and integrity.