Advanced Machine Learning
Hanyang University, School of Intelligence

Course Information
Course #: ITG6013
Fall 2020
Thurs 9am-12pm
International Building (108), Room 307
Course Management System: https://learn.hanyang.ac.kr/

Instructor Information
Instructor: Casey Bennett
Office: FTC building (5th Floor)
Office Hours: MWF 2pm-3pm (other times by appointment)
Phone: TBD
Email: cabennet@hanyang.ac.kr
Home page: http://www.CaseyBennett.com

Course Description
This course focuses on the application of machine learning principles in Data Science and AI. The course will cover advanced modeling techniques, including ensemble learning (bagging/boosting), neural networks, bayesian networks, biologically-inspired computing, feature selection and feature engineering techniques, Markov models for temporal modeling to find patterns over time, extended linear models & kernel methods (support vector machines), and clustering techniques. We will also cover use of various packages like Scikit, Spark, XgBoost, and Keras. The course will focus on both the theoretical aspects of the techniques and connections to natural intelligence. State-of-the-art research will be explored through scientific paper reviews that highlight the use of advanced machine learning techniques to solve real world problems in medicine, finance, public policy, engineering, etc. Students will learn through hands-on problem-based learning and discussion.

Course Learning Goals
At the end of the course, students should be able to:
- understand the mechanics behind each machine learning method, as well as the respective pros and cons, for solving data science problems
- understand how to implement machine learning models on real world data using tools such as Python, Scikit-learn, Spark, XgBoost, and Keras
- understand methods to evaluate the performance of machine learning models
- understand how "information" in real world applications can be formulated as different data structures, and the effects of feature selection and other preprocessing techniques
- develop a sophisticated understanding of how to interpret and explain obtained ML results to diverse audiences
- understand the roots of various forms of machine learning from natural intelligence and human cognition, and scientific endeavors to create "artificial" intelligence

Recommended Books
None required, but good resources for background reading:
- *Data Mining: Practical Machine Learning Tools and Techniques* by Witten, Frank, and Hall, 4th

Edition, ISBN 978-0-12-374756-0
    - This book has a focus of practical applications and the use of the WEKA toolkit.
- *Probabilistic Graphical Models, Principles and Techniques* by Daphne Koller and Nir Friedman, ISBN 978-0-262-01319-2
    - This book has a focus on theoretical foundations of probabilistic graphical models http://mitpress.mit.edu/books/probabilistic-graphical-models
    - eBook available through the DePaul library at http://depaul.worldcat.org/title/data-mining-practical-machine-learning-toolsandtechniques/oclc/706802868&referer=brief_results
- *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* by Hastie, Tibshirani, Friedman
    - This book has a focus on theoretical foundations of data mining; PDF available at https://web.stanford.edu/~hastie/Papers/ESLII.pdf

Prerequisites
Some exposure to computer programming (Python) and basic statistics.

Grading
The grading scheme for the course listed below, along with descriptions of each component. The summary of the weights of each assignment for contributing to the final grade is as follows:

Assignment	Weight in final grade
Homework Assignments	25%
Paper Review	20%
Final Project	35%
Participation/Attendance	20%

Final Grades: A: 90% - 100%
B: 80% - 90%
C: 70% - 80%
D: 60% - 70%
F: less than 60

Assignments
Homework assignments:
There will be 5 homework assignments during the quarter. Students should plan to submit completed code, as well as a written summary answering the assignment questions. Make sure you completely answer all questions. Work to be submitted for the course is generally due one or two weeks after it was assigned; late submissions are allowed for 2 days (up to 48 hours after the due date/time), with a penalty of -10% per day. No late work will be accepted after two days since the assignment was due.

The assignments must be submitted online on the class Blackboard at https://learn.hanyang.ac.kr. Only legible, organized homework which shows your work will be graded. Include your name, section number, date, and homework number on the first page of your assignment. It is your responsibility to check that your files are uploaded correctly to Blackboard

Paper Review:

Throughout the quarter, the students will be also provided with a research paper each week related to the theory discussed in class. Each student will have to review these papers and participate in class discussion. Additionally, every few weeks during the course, the student will *pick any* of these papers read so far, and write a 2-3 page paper review to be submitted. These papers should be written as a *critique*, and cover three things: 1) problem was being addressed, 2) machine learning methods evaluated and results found, and 3) critique of what the authors did both good and bad.

Students should refer to the course schedule for due dates as to when each paper review is due.

Final Project:
The purpose of the final project is to demonstrate students' ability to apply the knowledge and the techniques learned during this course. The final project for this class is more extensive analysis task, chosen by students from among the topics we discuss. Final projects will include a presentation to the rest of the class at the end of the quarter, in place of a final exam. As part of your final project, you will also be asked to critique your classmates' projects during the final presentations

Whenever it is possible, it is recommended that the DL students attend the final presentations to participate in the live discussions of the final project. However, appropriate accommodations will be made for the DL students not being able to give the presentations in class; the DL students will still have to submit their critiques on the other projects.

Deliverables for the final project:
Proposal: One-page proposal describing the problem, data, proposed machine learning approach, and at least three references from the academic literature (e.g. peer-reviewed journal articles, use Google Scholar to search).

Presentation: Each project is to be presented using PowerPoint in a modified Pecha Kucha style, and the PPT file will have to be submitted to be published on course web site.

Report: The report will be written in the format of a paper (abstract, introduction, literature review, methodology, results, discussion, conclusions and future work). It should be approximately 6 pages long, single-spaced. The literature review for the final report consists of reading and summarizing about 5 to 6 published papers on the review topic. While the internet can serve as a good source of information, the Hanyang Library also has extensive holdings, most of them available electronically.

Participation
Students will also be graded based on their participation in in-class discussion and regular attendance. There will be a number of in-class group activities. The belief is that students learn better when they engage their own curiosity, rather than just engage in rote memorization. So bring your curiosity to class.

Software

The use of Python and the Scikit-Learn package will be taught in class. There will be also a lab tutorial scheduled during the first week of class to familiarize the students with the basics of Python and Scikit. Tutorials will also be conducted throughout the course on other packages like Spark, XgBoost, Keras, etc. Each week will have in-class coding lessons on each topic explaining the nuts and bolts of how to deploy machine learning models in applied settings. The Python/Scikit/Spark stack is a very common suite in real-world data science practice, and also a very in-demand skill set for employers at the moment.

While other software does exist to do machine learning, for this course use of Python is required, as code

templates in Python will be provided for homework coding assignments. For students, homework can be submitted as either basic python scripts (.py files) or as Jupyter notebooks (.ipynb files) … how to do so will be taught in class.

Course Schedule
The course schedule will be maintained on the course website on Blackboard.

Attendance
It is expected that you will attend every class and remain for the duration; it is the single most important action you can take in mastering the course objectives. Coming 15 minutes late or leaving 15 minutes constitutes an absence for the student. You are responsible for all material covered, assignments delivered or received, and announcements made in class sessions that you miss. For distance learning students, this means viewing the classes in a timely manner, participate in the discussion forum, and being sure to email or call in any questions that you have.

Email
Email is the primary means of communication between faculty and students enrolled in this course outside of class time. Students should be sure their email listed in Hanyang's system is correct.

Attitude
A professional and academic attitude is expected throughout this course. Measurable examples of nonacademic or unprofessional attitude include but are not limited to: talking to others when the instructor is speaking, mocking another's opinion, cell phones ringing, emailing, texting or using the internet whether on a phone or computer. If any issues arise a student may be asked to leave the classroom. The professor will work with the Office of Student Affairs to navigate such student issues.

Civil Discourse
Hanyang University is a community that thrives on open discourse that challenges students, both intellectually and personally. It is the expectation that all dialogue in this course is civil and respectful of the dignity of each student to become leaders. Any instances of disrespect or hostility can jeopardize a student's ability to be successful in the course

Cell Phones/On Call
If you bring a cell phone to class, it must be off or set to a silent mode. Should you need to answer a call during class, students must leave the room in an undisruptive manner. Out of respect to fellow students and the professor, texting is never allowable in class. If you are required to be on call as part of your job, please advise me at the start of the course.

Course Policies

Changes to Syllabus
This syllabus is subject to change as necessary during the quarter. If a change occurs, it will be thoroughly addressed during class, and an announcement will be posted on Blackboard and sent via email.

Academic Integrity and Plagiarism

This course will be subject to the university's academic integrity policy. More information can be found at the Office of Academic Affairs: https://academic.hanyang.ac.kr/home

Academic Policies

All students are required to manage their class schedules each term in accordance with the deadlines for enrolling and withdrawing as indicated in the University Academic Calendar. Information on enrollment, withdrawal, grading and incompletes can be found at More information can be found at the Office of Academic Affairs: https://academic.hanyang.ac.kr/home

Incomplete Grades

An incomplete grade is a special, temporary grade that may be assigned by an instructor when unforeseeable circumstances prevent a student from completing course requirements by the end of the term and when otherwise the student had a record of satisfactory progress in the course. All incomplete requests must be approved by the instructor of the course and the Associate Dean. Only exceptions cases will receive such approval. More information can be found at the Office of Academic Affairs: https://academic.hanyang.ac.kr/home