HOUSECS 59.01: APPLIED MACHINE LEARNING
Fall 2018

INSTRUCTORS:

Shrey Gupta, Student Instructor, Duke University, shrey.gupta@duke.edu
Prof. Rebecca Steorts, Department of Statistical Science, rebecca.steorts@duke.edu

CLASS MEETINGS: Wednesdays (6–7:30pm), Keohane 4B, 402SEM

TEXT:
An Introduction to Statistical Learning by Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani

BULLETIN DESCRIPTION:
Introduction to topics in machine learning through an applied perspective. The course assumes basic fluency in programming and mathematics at the single-variable calculus level, and will include learning specific machine learning concepts (listed below), their historical origins, and existing and potential applications to modern society.

Machine learning concepts studied will include: classification (including naive Bayes, support vector machines, kernel methods, and neural networks), regression (including spline interpolation and linear and polynomial regression), mixture of Gaussians clustering, object detection (including convolutional neural networks, feature extraction, edge detection, and processing methods), principal component analysis, and evaluation of machine learning models.

DUKE COMMUNITY STANDARD:
Students are expected to adhere to the Duke Community Standard. On each assignment, students must reaffirm commitment to the Duke Community Standard. If a student is responsible for academic dishonesty on a graded item in this course, then the student will have an opportunity to admit the infraction and, if approved by the Office of Student Conduct, resolve it directly through a faculty-student resolution agreement; the terms of that agreement would then dictate the consequences. If the student is found responsible through the Office of Student Conduct and the infraction is not resolved by a faculty-student resolution agreement, then **the student will receive a failing (unsatisfactory) grade for the final grade in the course**.

COURSE ASSIGNMENTS:
Lectures will be focused on developing programming skills and techniques, and weekly readings will consist of reading from the textbook listed above and/or supplementary reading. A grade of satisfactory in this course requires satisfactory completion of **all** assignments in this course, including written and programming assignments as well as attendance.

PROGRAMMING ASSIGNMENTS:
There will be three (3) programming assignments throughout the semester. Programming assignments will be due on the assigned dates on the calendar below and should be completed before the start of class. The policy for turning in late assignments is detailed below. Each programming assignment will build up on topics covered in previous lectures, focusing on the concepts covered in more recent lectures. A specific rubric will be posted for each assignment.

- Students may work on programming assignments with a maximum of one (1) other individual in the class. However, both individuals should contribute *equally* to the assignment and understand *all* parts of the code written.
- Students are expected to write their adherence to the Duke Community Standard in a README for every assignment. Students are allowed to consult others outside of their group—limited to Duke students and faculty—about the assignment only in a general way, but not actually provide/receive code to/from other students. If assistance is received from other individuals (excluding the instructors), it should be cited in the README. **Students should be prepared to explain any program code they submit.**
- It is acceptable to use *small* pieces of outside code (found on the Internet or otherwise) due to the nature of this course—but not entire methods or programs. Using open source libraries and packages is allowed. If you are concerned whether using a piece of code is within the Duke Community Standard, please ask. *All code used should be properly cited.*
- **All submissions are subject to automated plagiarism detection.** Assignments will be randomly checked using the MOSS Plagiarism Detector.

FINAL PAPER:
House courses require one or more scholarly papers totaling approximately 1500 words in length or the equivalent of five (5) double spaced pages. A final paper of this length will be due at 12pm noon on Friday, December 7, 2018. The topic of this paper will be a write-up detailing the applied machine learning model created during the final programming assignment. It should include (1) a description of the problem being solved and data set(s) being used, (2) the general thinking process (including questions and solutions raised) when creating the model, and (3) an evaluation of the machine learning model using various methods mentioned during lecture.

LATE OR MISSED WORK:
Students are expected to arrive to class punctually and to submit all assignments on time. However, students will be allowed to submit up to one (1) programming assignment up to 48 hours (two days) late without penalty. No programming assignments will be accepted after 48 hours (two days) late. This extension does not apply to the final paper, whose due date is fixed.

Accommodation will be granted by the instructor via pre-arrangement, which requires that the situation qualify for one of four strictly defined types of university-sanctioned exceptions: personal emergencies or tragedies, an incapacitating illness, a religious holiday, or varsity athletic participation. In these instances, it is the student's responsibility to be aware of and follow all relevant university-wide policies, including appropriate notification of the instructor.

ATTENDANCE:
Students are required to attend at least 11 classes to receive a passing (satisfactory) grade in the course. That is, students may receive a maximum of three (3) unexcused absences.

GRADING:
House courses are graded only on a satisfactory/unsatisfactory basis. In order to receive a passing (satisfactory) grade, in addition to satisfying the attendance requirement, students must complete **all** assignments of this course with individual scores of 70% or greater.

This syllabus was adapted from Prof. Lillian B. Pierce's MATH 165FS syllabus and Prof. Owen L. Astrachan's COMPSCI 201 syllabus.

CLASS SCHEDULE:
Readings from An Introduction to Statistical Learning

Week 1 (August 29): Introduction to Machine Learning and Python
Readings: (see lecture notes)

Week 2 (September 5): History of Machine Learning
Readings: Chapters 1, 2.1

Week 3 (September 12): Splines and Linear and Polynomial Regression
Readings: Chapters 3, 7.1, 7.4-7.5

Week 4 (September 19): Evaluation of Machine Learning Models
Readings: Chapters 2.2, 5.1

Week 5 (September 26): *Review and Office Hours*

Week 6 (October 3): Naive Bayes Classifiers
Readings: (see lecture notes)
Due: Programming Assignment #1

Week 7 (October 10): Support Vector Machines and Kernel Methods
Readings: Chapter 9

Week 8 (October 17): Principal Component Analysis
Readings: Chapter 10.2

Week 9 (October 24): *Review and Office Hours*

Week 10 (October 31): Neural Networks
Readings: (see lecture notes)
Due: Programming Assignment #2

Week 11 (November 7): Convolutional Neural Networks
Readings: (see lecture notes)

Week 12 (November 14): Feature Extraction and Image Processing for Object Detection
Readings: (see lecture notes)

Week 13 (November 28): *Review and Office Hours*

Week 14 (December 5): Mixture of Gaussians Clustering
Readings: Chapters 10.1, 10.3
Due: Programming Assignment #3
Final Paper on Friday, 12/7 noon

*The Faculty Sponsor for the course, Prof. Rebecca Steorts, from the Department of Statistical Science, will attend three meetings.