DEPARTMENT: Biostatistics and Bioinformatics

COURSE NUMBER: INFO534 **SECTION NUMBER:** 1

CREDIT HOURS: 2 **SEMESTER:** Spring 2021

COURSE TITLE: Applied Machine Learning in Public Health

CLASS HOURS AND LOCATION:
A single, two-hour lecture per week

INSTRUCTOR NAME: Max Lau

INSTRUCTOR CONTACT INFORMATION

EMAIL: msy.lau@emory.edu

PHONE:

SCHOOL ADDRESS OR MAILBOX LOCATION:

OFFICE HOURS

Teaching Assistant(s): TBD

COURSE DESCRIPTION

The elective course gives an introduction to machine learning techniques and theory, with a focus on its use in practical applications. The Applied Machine Learning course teaches you a wide-ranging set of techniques of supervised and unsupervised machine learning approaches using R as the programming language. During the course, a selection of topics will be covered in supervised learning, such as linear models for regression and classification, or nonlinear models such as neural networks, and in unsupervised learning such as clustering. The uses and limitations of these algorithms will be discussed, and their implementation will be investigated in programming assignments. The course also covers theoretical concepts such as inductive bias, the PAC and Mistake-bound learning frameworks, minimum description length principle, and Ockham's Razor. There will be a strong emphasis on the real-world context in which machine learning systems are used. The use of machine learning components in practical applications will be exemplified, and Public health realistic scenarios will be studied in application areas such as hospitalization metrics using electronic medical record data, clinical trials, natural language processing, image processing, and bioinformatics. The importance of the design and selection of features, and their reliability, will be discussed. In order to ground these methods the course includes

some programming and involvement in a semester-long research project. This is a programming course: **you will be required to write code**. Prerequisite: BIOS 500, BIOS 544 (or BIOS 545) or permission of instructor

INFO CONCENTRATION COMPETENCIES:

- Develop public health information systems to support public health efforts
- Identify software for the interface of data entry and statistical analysis
- Assess individual data elements and display results effectively and appropriately
- Apply standard statistical methods in the analysis of public health information

EVALUATION

The grade assignment will be based on:

4 Homework Assignments (20% each)
Final Exam (20%).

Grade scale*:

- A = 93 -- 100%
- A- = 90 -- 92%
- B+ = 87 – 89%
- B = 83 – 86%
- B- = 80 – 82%
- C = 65 – 79%
- F = <65%

*final grades are not rounded and the lower limit of each letter grade is inclusive, so e.g., 93.0 is an A, while 92.9 is an A-.

COURSE STRUCTURE

The course will be organized into weekly lectures consisting of a combination of electronic slides and demonstrations. Students are expected to ask and answer questions in class. Students are encouraged to bring a laptop to class to follow along with demonstrations.

Example assignments are shown below along with the associated core competency.

Assignment	**Competency**
Homework questions will be assigned that require students to develop code and justify use of machine learning algorithm	Identify software for the interface of data entry and statistical analysis
Homework assignments utilizing	Develop public health information systems to

electronic medical record data using a multitude of machine learning algorithms	support public health efforts
Homework assignments and exam will involve creation of functions and packages based on understanding specific infectious disease data sets	Apply standard statistical methods in the analysis of public health information
Homework assignments will require interpreting results of machine learning output, including assessing parameters and accuracy measures. This would include evaluating parameters in a neural network to best predict systolic blood pressure	Assess individual data elements and display results effectively and appropriately

COURSE POLICIES

Students are expected to attend lectures and ask questions during class. For computational assignments, students are encouraged, but not required, to bring a laptop to class to follow along with code demonstrations.

As the instructor of this course I endeavor to provide an inclusive learning environment. However, if you experience barriers to learning in this course, do not hesitate to discuss them with me and the Office for Equity and Inclusion, 404-727-9877.

RSPH POLICIES

Accessibility and Accommodations

Accessibility Services works with students who have disabilities to provide reasonable accommodations. In order to receive consideration for reasonable accommodations, you must contact the Office of Accessibility Services (OAS). It is the responsibility of the student to register with OAS. Please note that accommodations are not retroactive and that disability accommodations are not provided until an accommodation letter has been processed.

Students who registered with OAS and have a letter outlining their academic accommodations are strongly encouraged to coordinate a meeting time with me to discuss a protocol to implement the accommodations as needed throughout the semester. This meeting should occur as early in the semester as possible.

Contact Accessibility Services for more information at (404) 727-9877 or accessibility@emory.edu. Additional information is available at the OAS website at http://equityandinclusion.emory.edu/access/students/index.html

Honor Code

You are bound by Emory University's Student Honor and Conduct Code. RSPH requires that all material submitted by a student fulfilling his or her academic course of study must be the original work of the student. Violations of academic honor include any action by a student indicating dishonesty or a lack of integrity in academic ethics. *Academic dishonesty refers to cheating, plagiarizing, assisting other students without authorization, lying, tampering, or stealing in performing any academic work, and will not be tolerated under any circumstances.*

The RSPH Honor Code states: "Plagiarism is the act of presenting as one's own work the expression, words, or ideas of another person whether published or unpublished (including the work of another student). A writer's work should be regarded as his/her own property."
(http://www.sph.emory.edu/cms/current_students/enrollment_services/honor_code.html)

COURSE CALENDAR AND OUTLINE

Topics and dates are subject to change as the semester progresses.

Date	Topics
	Generalized Linear models
	Model selection and evaluation
	Supervised vs unsupervised
	SVM
	Decision trees, RF
	Neural networks, k-means
	Ethics, communication
	Measures of Accuracy
	Missing data
	Data generating process
	Feature generation