DEPARTMENT: BIOS

COURSE NUMBER: 534 **SECTION NUMBER:**

CREDIT HOURS: 3 **SEMESTER:** Spring 2019

COURSE TITLE: Machine Learning

CLASS HOURS AND LOCATION: Tuesday and Thursday, 4:00-5:20pm

INSTRUCTOR NAME: Tianwei Yu

INSTRUCTOR CONTACT INFORMATION

EMAIL: tianwei.yu@emory.edu

PHONE: (404)727-7671

SCHOOL ADDRESS OR MAILBOX LOCATION: 1518-002-3AA

OFFICE HOURS

Teaching Assistant(s): Yunchuan Kong, Teng Fei

COURSE DESCRIPTION

This course covers fundamental machine learning theory and techniques. The topics include basic theory, classification methods, model generalization, clustering, and dimension reduction. The material will be conveyed by a series of lectures and projects.

MPH/MSPH FOUNDATIONAL COMPETENCIES:

- Analyze quantitative and qualitative data using biostatistics, informatics, computer-based programming and software, as appropriate
- Select communication strategies for different audiences and sectors

CONCENTRATION COMPETENCIES:

- Use statistical software for data management and exploratory data analysis.
- Communicate the results of statistical analyses to a broad audience.

EVALUATION

The grade assignment will be based on:

Participation in class and discussions (10%);

Three projects (20% each);

Final project (30%). The instructor will assign a dataset. Each student should research the background of the data, conduct dimension reduction/data visualization to summarize the data, conduct variable selection and predictive modeling, and write a thorough data analysis report in research article style.

COURSE STRUCTURE

The course will be organized into weekly lectures consisting of a combination of electronic slides, whiteboard problem solving, and computational demonstrations. Students are expected to ask and answer questions in class.

MPH/MSPH Foundational Competency assessed	Representative Assignment
Use statistical software for data management and exploratory data analysis.	Project assignments will involve the exploratory analysis of real data sets.
Select communication strategies for different audiences and sectors	Project assignments require students to interpret machine learning techniques to applied researchers with no statistics/machine learning background.
BIOS Concentration Competencies assessed	**Representative Assignment**
Use statistical software for data management and exploratory data analysis.	Project assignments will require programming in Python and exploratory analysis of real data sets.
Communicate the results of statistical analyses to a broad audience.	Project assignments will require interpreting the results.

COURSE POLICIES

Students are expected to attend lectures and ask questions during class. After taking the

course, the students are expected to have working knowledge the following areas:
(1) Recognize the scope of machine learning methods and where to apply them. (2) Have working knowledge in the areas of classification, clustering and dimension reduction. (3) Understand and be able to judge the performance of machine learning algorithms. (4) Understand model generalization, including variance, bias, and their trade off, and the issue of over-fitting.

Textbook:

The elements of statistical learning. Hastie, Tibshirani & Friedman.
Python Machine Learning. Raschka & Mirjalili

Other references:

Pattern classification. Duda, Hart & Stork.
Data clustering: theory, algorithms and application. Gan, Ma & Wu.
An introduction to Statistical Learning: with Applications in R. James, Witten, Hastie, Tibshirani.

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.

Lecture 1	Introduction	
Lecture 2	Statistical background	
Lecture 3	Stat decision theory 1	
Lecture 4	Stat decision theory 2	
Lecture 5	Density estimation	
Lecture 6	Basis expansion 1	
Lecture 7	Basis expansion 2	
Lecture 8	Linear Machine	First project assigned
Lecture 9	Support Vector Machine 1	
Lecture 10	Support Vector Machine 2	
Lecture 11	Boosting 1	
Lecture 12	Boosting 2	
Lecture 13	Tree and Forest 1	First project due
Lecture 14	Tree and Forest 2	Second project assigned
Lecture 15	Bump hunting	
Lecture 16	Hidden Markov Model 1	
Lecture 17	Hidden Markov Model 2	
Lecture 18	Neural Networks 1	
Lecture 19	Neural Networks 2	Second project due
Lecture 20	Neural Networks 3	Third project assigned
Lecture 21	Clustering 1	
Lecture 22	Clustering 2	
Lecture 23	Clustering 3	
Lecture 24	Dimension reduction 1	
Lecture 25	Dimension Reduction 2	
Lecture 26	Dimension Reduction 3	
Lecture 27	Model generalization 1	Third project due;

		Final project assigned
Lecture 28	Model generalization 2	
Tuesday – 5/9		Final project due