ROLLINS SCHOOL OF PUBLIC HEALTH
EMORY

DEPARTMENT: Biostatistics

COURSE NUMBER: 731 **SECTION NUMBER:** **SEMESTER:** Fall

CREDIT HOURS: 2

COURSE TITLE: Advanced Statistical Computing

INSTRUCTOR NAME: Hao Wu. PhD

INSTRUCTOR CONTACT INFORMATION:

EMAIL: hao.wu@emory.edu

PHONE:

SCHOOL ADDRESS OR MAILBOX LOCATION: The Rollins School of Public Health

OFFICE HOURS

COURSE DESCRIPTION (3-4 Sentences)

This course covers the theories and applications of some common statistical computing methods. Topics include Markov chain Monte Carlo (MCMC), hidden Markov model (HMM), Expectation-Maximization (EM) and Minorization-Maximization (MM), and optimization algorithms such as linear and quadratic programming. The class has two main goals for students: (1) learn the general theory and algorithmic procedures of some widely used statistical models; (2) develop fluency in statistical programming skills. The class puts more emphasis on implementation instead of statistical theories. Students will gain computational skills and practical experiences on simulations and statistical modeling.

This course requires significant amount of programming. Each set of homework involves the implementation of certain algorithms using high-level programming language (such as Matlab or R).

Prerequisite: BIOS 510, 511 and prior programming experience, or permission from one of the instructors.

EVALUATION

Homework (90%). Three sets of homework, each worth 30% of the final grade.
Class participation (10%).

LEARNING OBJECTIVES OR COMPETENCIES OF THE COURSE

Upon successfully completing this course, students will be able to:

1. Understand the mathematical properties and computational procedures of the algorithms covered in the class.
2. Implement the algorithms using high-level programming languages such as Matlab or R.
3. Apply the methods to model real world data.

LEARNING OBJECTIVES OR COMPETENCIES FOR THE DEPARTMENT OR PROGRAM TO WHICH THE COURSE CONTRIBUTES

Upon completion of the PhD degree in Biostatistics the graduate will be able to:

- Use a variety of statistical computer packages
- Conduct appropriate statistical analyses
- Conduct complex statistical analyses for a broad range of applications

TENTATIVE SCHEDULE

- **Lecture 1**: MCMC I. Course information. Monte Carlo methods, random variate generation.
- **Lecture 2**: MCMC II. Metropolis-Hasting algorithm. Gibbs sampler.
- **Lecture 3**: MCMC III. Check for convergence, techniques to accelerate Markov chain mixing.
- **Lecture 4**: MCMC IV. Implementation of MCMC, examples of MCMC applications.
- **Lecture 5**: Optimization. Optimization Algorithms, Newton-Raphson, Quasi-Newton-Raphson, Iteratively Re-weighted Least Squares for Generalized Linear Regression.
- **Lecture 6**: EM I. EM algorithm and applications.
- **Lecture 7**: EM II. EM algorithm extensions, SEM algorithm, EM gradient algorithm, ECM algorithm.
- **Lecture 8**: MM. MM algorithm and applications.
- **Lecture 9**: HMM I. Introduction to hidden Markov model. Viterbi algorithm.
- **Lecture 10**: HMM II. Forward-backward algorith. Applications of HMM.
- **Lecture 11**: Linear programming I. Introduction. Simplex algorithm.
- **Lecture 12**: Linear programming II. Duality. Interior point method.
- **Lecture 13**: Linear programming III. Applications of LP: quantile regression, LASSO.
- **Lecture 14**: Quadratic programming. Support vector machine.