STAT/DSBA 6115 Statistical Learning with Big Data
Syllabus

Instructor Information:

Name: Dr. Jiancheng Jiang

Office: Fretwell 350A

Office Number: 704-687-0631

Email: jjiang1@uncc.edu

Office Hours: 1:30 pm – 3:00 pm, TR or by appointment

Scheduled Meeting Times					
Type	Time	Days	Where	Date Range	Schedule Type
Class	5:30 pm - 8:15 pm	T	Denny 106	Aug 22, 2022 - Dec 15, 2022	Lecture

Course Information:

Course Name:	Statistical Learning with Big Data
Prerequisite:	STAT 5110 or STAT 5123, or permission of the department
Textbook:	There is no required textbook for this course. The course material will be a selective combination of a number of excellent books in statistical learning, data mining, and machine learning, such as

An Introduction to Statistical Learning with Applications in R, by G. James, D. Witten, T. Hastie and R. Tibshirani, Springer, 2013.

(The university has full access to the eBook of the textbook.)

Course Description: This course provides students a survey of major statistical learning methods and concepts for both supervised and unsupervised learning including resampling methods, support vector machines, model selection and regularization, tree-based methods and ensembles, statistical graphics. Students learn how and when to apply statistical learning techniques, their comparative strengths and weaknesses, and how to critically evaluate the performance of learning algorithms in case studies in financial investment, gene identification, and feature selection in high-dimensional spaces. Software R will be used for simulations and for real data analysis.

Course Objectives:

By providing students a survey of major statistical learning methods, the course is designed for students to learn how and when to apply statistical learning techniques, how to critically evaluate the performance of learning algorithms, and how to build predictive models. Students completing this course should be able to apply basic statistical learning methods to build predictive models, to properly tune and select statistical learning models, to correctly assess model fit and error, and to build an ensemble of learning algorithms for data.

Instructional Method:

The primary format for instruction will be lectures in classroom.

Means of Student Evaluation:

Attendance (5%): Attendance is required. For each un-excused absence, you will lose 10% of the attendance points. If you miss 10 or more than 10 classes, you will receive a U for the final grade without exception. If you miss a class, it will be your responsibility to obtain the assignment and find out what material is covered.

Pop quizzes (5%): Several pop quizzes will be given either online or at any time in lectures. All pop quizzes will be open book and open notes. No make up for missed quizzes if there is no legitimate excuse.

Homework assignments (20%): Reading relevant sections in reference books after classes is always expected. Homework will be assigned in class or through Canvas. Homework will be due at the beginning of class on the due date. No late homework will be accepted.

Course project (15%): Students will make groups to work on a course project at the end of the semester.

Midterm exam (25%): You will be informed in advance about the date and the material coverage of the midterm exam. The midterm exam may be open or closed book. If it is closed book, you may choose to prepare a formula sheet for reference for the exam.

Final exam (30%): The final exam may be open or closed book. A formula sheet and some selected tables will be provided for the final exam if appropriate. The final exam will contain 20% multiple choice, true/false, or blank-filling questions and 80% free response questions. If a student misses the final exam, his/her final grade for the course would be either an I or a U, depending on his/her performance on previous work.

Course letter grades will follow the scale listed below (not curved):

90-100%	A
80-89%	B
70-79%	C
0-69%	U

Software and Computing:

Students are encouraged to bring a laptop to each class. The course will make extensive use of the free R software environment for statistical computing and graphics. Introductory tutorials/labs will be provided for students unfamiliar with R. Students are encouraged to install R on their laptops before the semester begins. Copies of R may be downloaded from http://www.rproject.org.

Class Policies:

Exam make-up policy: No make-up exam will be given without prior notification and documented acceptable excuse. If you miss an exam due to a legitimate excuse, you need to call or email me either before the exam or within 24 hours after the exam in order to be allowed a make-up exam.

Cheating: Cheating in any form will result in a "U" for the course grade and may be reported to the University.

Classroom policy: Coming to class late or leaving class early is NOT allowed. Any distraction behavior such as TALKING or LAUGHING or YELLING in class is prohibited. Electronic devices other than a laptop also cause distraction during classes and are thus not permitted either. More classroom policies can be found at

> http://legal.uncc.edu/legal-topics/classroom-policies-and-practices

Academic integrity: The UNC Charlotte is committed to the highest standards of academic integrity and honesty. Students are expected to be familiar with these standards regarding academic honesty and to uphold the policies of the University in this respect. Students are particularly urged to familiarize themselves with the provisions of the Code of Student Behavior at http://legal.uncc.edu/policies/up-407

Disability services: UNC Charlotte is committed to access to education. If you have a disability and need academic accommodations, please provide a letter of accommodation from Disability Services early in the semester. For more information on accommodations, contact the Office of Disability Services at 704-687-0040 or visit their office in Fretwell 230.

Grief absence policy for students: UNC Charlotte recognizes that a time of bereavement is very difficult for a student. The University therefore provides the following rights to students facing the loss of a family member through the Grief Absence Policy for Students (GAPS): Students will be excused for funeral leave and given the opportunity to earn equivalent credit and to demonstrate evidence of meeting the learning outcomes for missed assignments or assessments in the event of the death of a member of the student's family.

Statement on Diversity: UNC Charlotte strives to create an inclusive academic climate in which the dignity of all individuals is respected and maintained. Therefore, we create diversity that includes, but is not limited to ability/disability, age, culture, ethnicity, gender, language, race, religion, sexual orientation, and socio-economic status.

Others: The instructor reserves the right to revise this syllabus if it is necessary.

<u>Tentative Schedule:</u>

- Week 1. Introduction to statistical learning and examples
- Week 2. Overview of supervised learning
- Week 3. Statistical learning with linear models – part I
- Week 4. Statistical learning with linear models – part II
- Week 5. Statistical learning with classification: logistic regression, LDA and QDA
- Week 6. Resampling methods and variable selection: CV, bootstrap
- Week 7.
    1. Model selection: subset selection and shrinkage
    2. Midterm exam
- Week 8. Regularization
- Week 9. Nonparametric smoothing: spline methods
- Week 10. Kernel smoothing
- Week 11. Decision trees
- Week 12. Regression trees: bagging and boosting
- Week 13. Support vector machines
- Week 14. Unsupervised learning: PCA
- Week 15.
    1. Unsupervised learning: Clustering
    2. Final exam