# MATH 4385/5385 Statistical Learning

## What will students learn from the course?

As part of the <u>Graduate Certificate in Data Analytics program</u>, we will teach <u>classical machine learning and</u> <u>statistical methods</u> with <u>hands-on R coding skills</u> for data analysis of real applications. The **main topics** include <u>bias-variance trade-off</u>, <u>multiple linear regression</u>, <u>classification</u> (logistic regression, linear discriminant analysis, KNN, Naïve Bayes), <u>resampling methods</u> (bootstrap and K-fold CV), <u>model selection</u>, <u>regularization</u> (lasso, ridge, elastic net), <u>principal component regression</u>, <u>non-linear models</u>, <u>tree methods</u> (random forest, boosting), <u>support vector machines</u>, and <u>unsupervised learning methods</u>.

Students will learn how these methods work, how to formulate questions from real problems, select appropriate methods, run analysis and modeling in **R**, interpret and present analysis results in a visualizable, professional, and insightful report.

### What is the prerequisite and what software and tools will be taught and used?

The **prerequisite** is **Math 2314** (Elementary Statistical Methods). The course will introduce and use the free open-source statistical software **R**, the integrated development environment **RStudio**, and the **R Markdown** tool. All lecture notes and projects will be written and provided in **R Markdown** documents. No prior experience of **R** is required.

#### What is the main textbook?

We will use the book with its corresponding **R** package **ISLR2**: James, G., Witten, D., Hastie, T. and Tibshirani, R. *An Introduction to Statistical Learning with Applications in* **R**, 2021, 2<sup>nd</sup> edition, Springer, NY. (ISBN 978-1071614174) (downloadable at <a href="https://www.statlearning.com/">https://www.statlearning.com/</a>)

#### How the course is taught?

It is a 100% **online** class **offered in each fall semester**. Weekly reading materials will be on **Canvas** including lecture **notes**, **videos**, and source files. Students manage time on their own. Four assignments (written homework + **R** projects), midterm exam, final exam, and final project will be the major submissions. **Online platform** on **Piazza.com** will be the forum for communications among students and the instructor.

#### **Current instructor**

Xueyan Sherry Liu, <u>xliu10@uno.edu</u>, Department of Mathematics at UNO.

to

	Khade		x - u x
	Ne Exit Code View Noto Session Bald Dubay Prolls Tools Halp	■ Hoh Stateware (Automatipat) × 記書 New 17日年の日回戸戸 Intel イトル	***
exts in Statistics	Moth Staticaming Aupdited Hind ×	Chapter 4. Classification	
	< >	Samey Lee, UNO Mana Department	East • 🕥 •
James Witten fastie Tibshirani	2 Little: "Chapter 4, classifieiton" 1 subtitle: "Statistical Learning_d" 4 author: "Merry Liu, UN Math Department" 5 state: "Sept 5, 2021" 6 autput: 7 pd[_documents]	<ol> <li>Logical engenesis</li> <li>Conditions of a set of encoding and threads a close of encoding and encoding and the set of encoding and t</li></ol>	4
Introduction	<ul> <li>Reader-includes;</li> <li>vuepackage(color)</li> <li>vuepackage(color)</li> <li></li></ul>	$\label{eq:rescaled} \begin{split} & r_{0} = + \sum_{i}\sum_{j} \sum_{i} v_{ij} < z_{ij} \\ & rescaled by the rescale state of the state of the$	
arning	15 knitr::opts_chunkiset(echo = IRUE) # change the in a document. echo=IRUE or FALSE to display not chunk. 16-****		sura e Coloradore
plications in R	15 19-30 1. Logistic regression 20 21 **Classification**: SyS is gualitative/categorid		
Edition	clatining an email is "spam" or "not spam" based patients in an emergency room bason on her sympt possible diagnoses: 22 "stroke", "drug overdose", "epileptic setzure". 23		e 'ISLR2'
Springer	24 **Bayes classifier**: Consider the probability observation ix US belong to class Sis. It assign 819 Chaster4 Classification :		r documentation.
	Controls Terminal × R Markalsum × John × John × v/ N >		۰