MA506: Probability & Statistical Inference

Fall 2021

Instructor:	Prashant Shekhar, PhD	**Email:**	shekharp@erau.edu
Class Time:	Tu, Thu: 5:15pm – 6:30pm	**Class Venue:**	Bldg IC Rm. 104
Student Hours (SH):	Tu, Thu: 4:00pm – 5:00pm	**SH Venue:**	Room 301.26, COAS.

Topics Included: This course will focus on using ideas from Probability and Statistics to solve problems in data science. The main topics included in the course are

1. Fundamentals of Statistical Learning

2. Linear Regression

3. Classification

4. Statistical Model Selection

5. Advanced topics in Regression/Classification

Why this course: The concepts that you learn in this course can be utilized to solve problems in the general area of machine learning and data science. The applications encompass multiple domains including healthcare, management, manufacturing, security and remote sensing etc.

Text Book: An Introduction to Statistical Learning with Applications in R by Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani, First Edition *Link*. I will also be following the excellent presentation slides prepared by the authors and available at *Link*.

Attendance: I will try to take attendance in every class. Although attendance is not mandatory, I encourage you to participate in class activities. This is because attendance is heavily correlated with the course grade. Also, your participation level in the class might inform some part of your final project grade.

Grading: Your grade will be determined as: Homework 55% + Tests 15% + Final (Project and Presentation) 30%. The grading is expected to follow the standard scale
A: 90% - 100%
B: 80% - 89.5%
C: 70% - 79.5%
D: 60% - 69.5%
F: <60%
However, based on the performance of the entire class, I might curve the grading scale later.

Homework: Your homework grade will be determined based on 5 programming oriented homeworks. You are required to use Python to solve homework problems. These exercises will test the ability of the students to apply the concepts in statistical learning on various categories of data sets.

Tests: You will have one main test (sometime in early November). Make-ups on the exam may be allowed only for valid extenuating circumstances when I am informed before the test takes place – please see me about conflicts as soon as they occur.

Final (Project and Presentation): During the semester you will be supervised to work on a project which combines classroom materials and real-world applications. The project together with the presentation is the final classroom assignment. It is supposed to be an individual project and I will work with each of you separately to identify a topic of your interest and find a relevant project in that domain. I will announce project topics, guidelines, and rubric in due course.

Academic Integrity: Embry-Riddle Aeronautical University maintains high standards of academic honesty and integrity in higher education. To preserve academic excellence and integrity, the University prohibits academic dishonesty in any form, including, but not limited to, cheating and plagiarism. More specific definitions of these violations and their consequences are described in the Dean of Students' Honor Codes and Student Policies.

Disability Services DSS Administration Office: Bldg 500; Contact: (386) 226-7916; email: dbdss@erau.edu
Testing Center: The Annex Building 2nd floor, room 217; Contact: (386) 226-2903; email: dbdss@erau.edu

- Student Disability Services: Students with disabilities who believe that they may need accommodations in this class are encouraged to contact the Office of Disability Services. Professors cannot make appropriate disability accommodations. Students are encouraged to register with DSS at the beginning of the term to better ensure that such accommodations are implemented in a timely fashion. Accommodations are not granted until official notice is received from DSS.

- DSS Testing Procedures: It is the responsibility of the student to notify DSS the date and time of test once s/he has been made aware of the scheduled test. DSS requires a 2 days minimum notification.

ERAU Coronavirus Updates: Information on testing, vaccinations, health services, procedures and frequently asked questions are available here.

- **Face Masks Strongly Encouraged**: Consistent with current recommendations of the Centers for Disease Control and Prevention, and Embry-Riddle's long-standing culture of safety, all students (vaccinated or unvaccinated) are strongly encouraged to wear face masks indoors especially during their in-person classes and in other group indoor settings, including faculty office hours.

- **Vaccinations Strongly Encouraged**: All students are strongly encouraged to receive a vaccination against Covid-19. Vaccinations are available at convenient campus locations.

MA 506 Probability and Statistical Inference

Instructor: Prashant Shekhar, PhD

Tentative Schedule for Fall 2021

Week Number	Section Number	Topic	Learning Outcome
Chapter 1: Introduction			
1 (T, Th)	-	Course Introduction Python Overview	1 2
Homework 1			
Chapter 2: Statistical Learning			
2 (T, Th)	2.1	Statistical Learning Python Examples	1 2
3 (T, Th)	2.2	Assessing Model Accuracy Python Examples	1, 3 2
Homework 2			
Chapter 3: Linear Regression			
4 (T, Th)	3.1	Simple Linear Regression Python Examples	4,5 2
5 (T, Th)	3.2	Multiple Linear Regression Python Examples	4,6 2
6 (T, Th)	3.3	More about Regression Python Examples	4,7 2
Homework 3			
Chapter 4: Classification			
7 (Th)	4.1–4.3	Classification Basics Python Examples	4,8,9 2
8 (T)	4.4	Linear Discriminant Analysis Python Examples	4,10 2
9 (T, Th)	4.5	Comparison of Classification Methods Python Examples	4,11 2
Homework 4			
Chapter 5: Resampling methods			
10 (T, Th)	5.1–5.2	Cross-Validation The Bootstrap Python Examples (Online)	4,12 4,12 2
11 (T)	–	Test	
Chapter 6: Linear Model Selection and Regularization			
12 (T, Th)	6.1	Subset Selection Python Examples	4,14 2
Homework 5			
13 (T)	6.2	Shrinkage Methods Python Examples	4,13 2
14 (T, Th)	6.3–6.4	Thinking in Higher Dimensions Python Examples	4,15,16 2
15 (T, Th)	-	Project Submission and Presentations	17

Learning outcome: After successful completion of this course, you will acquire knowledge to:

1. Understand the basics of statistical learning and its relation to machine learning.

2. Understand the basics of python and using it as a tool to solve problems in statistical learning.

3. Assess the quality of various statistical learning approaches based on various available metrics

4. Understand different problems in supervised learning.

5. Delve deeper into various aspects of linear regression.

6. Understand and implement multiple linear regression

7. Better understand various issues one might face while using linear regression as a tool to understand different properties of data.

8. Understand basics of classification and its relation to regression.

9. Use logistic regression as a tool to solve classification problems.

10. Understand and use discriminant analysis for classification.

11. Compare different classification models for your own problem.

12. Use various resampling approaches to make an intelligent choice of a model for your own data science related problem.

13. Use regularization as a way to produce better models.

14. Understand various ways of controlling the complexity of your statistical learning model.

15. Use dimensionality reduction as a tool to simplify your model to achieve better generalization.

16. Interpret data and models in higher dimensions.

17. Apply the concepts learnt in class to problems of practical importance.