

MATH 678: Stat Methods in Data Science Spring 2022 Course Syllabus

NJIT Academic Integrity Code: All Students should be aware that the Department of Mathematical Sciences takes the University Code on Academic Integrity at NJIT very seriously and enforces it strictly. This means that there must not be any forms of plagiarism, i.e., copying of homework, class projects, or lab assignments, or any form of cheating in quizzes and exams. Under the University Code on Academic Integrity, students are obligated to report any such activities to the Instructor.

Please be sure you read and fully understand our DMS Online Exam Policy.

COURSE INFORMATION

Course Description: This course introduces students to concepts in statistical methods used in data science, including data collection, data visualization and data analysis. Emphasis is on model building and statistical concepts related to data analysis methods. The course provides the basic foundational tools on which to pursue statistics, data analysis and data science in greater depth. Topics include sampling and experimental design, understanding the aims of a study, principles of data analysis, linear and logistic regression, resampling methods, and statistical learning methods. Students will use the R statistical software.

Number of Credits: 3

Prerequisites: MATH 661 or MATH 663, or permission by instructor.

Course-Section and Instructors:

Course-Section	Instructor
Math 678-102	Professor W. Guo

Office Hours for All Math Instructors: Spring 2022 Office Hours and Emails

Required Textbook:

Title	An Introduction to Statistical Learning: with Applications in R	
Author	Gareth James, et al	
Edition	2nd	
Publisher	Springer	
ISBN #	978-1071614174	

Reference Book	The Elements of Statistical Learning: Data Mining, Inference, and Prediction, by Hastie, Tibshirani,	
	and Friedman; Publisher: Springer, 2nd edition (2009); ISBN: 978-0387848570.	

University-wide Withdrawal Date: The last day to withdraw with a W is Monday, April 4, 2022. It will be strictly enforced.

POLICIES

DMS Course Policies: All DMS students must familiarize themselves with, and adhere to, the Department of Mathematical Sciences Course Policies, in addition to official university-wide policies. DMS takes these policies very seriously and enforces them strictly.

Grading Policy: The final grade in this course will be determined as follows:

Homework	25%
Project	20%
Midterm Exam	25%
Final Exam	30%

Your final letter grade will be based on the following tentative curve.

Α	90 - 100	C+	75 - 79
B+	85 - 89	с	70 - 74
В	80 - 84	F	0 - 69

Attendance Policy: Attendance at all classes will be recorded and is mandatory. Please make sure you read and fully understand the Math Department's Attendance Policy.

Exams: There will be one exam during the semester and a cumulative final exam during the final exam week:

Midterm Exam	March 23, 2022
Final Exam Period	May 6 - May 12, 2022

The final exam will test your knowledge of all the course material taught in the entire course. Make sure you read and fully understand the Math Department's Examination Policy. This policy will be strictly enforced.

Makeup Exam Policy: There will be NO MAKE-UP QUIZZES OR EXAMS during the semester. In the event an exam is not taken under rare circumstances where the student has a legitimate reason for missing the exam, the student should contact the Dean of Students office and present written verifiable proof of the reason for missing the exam, e.g., a doctor's note, police report, court notice, etc. clearly stating the date AND time of the mitigating problem. The student must also notify the Math Department Office/Instructor that the exam will be missed.

Cellular Phones: All cellular phones and other electronic devices must be switched off during all class times.

ADDITIONAL RESOURCES

Further Assistance: For further questions, students should contact their instructor. All instructors have regular office hours during the week. These office hours are listed on the Math Department's webpage for Instructor Office Hours and Emails.

Accommodation of Disabilities: The Office of Accessibility Resources and Services (OARS) offers long term and temporary accommodations for undergraduate, graduate and visiting students at NJIT.

If you are in need of accommodations due to a disability please contact Scott Janz, Associate Director of Disability Support Services at 973-596-5417 or via email at scott.p.janz@njit.edu. The office is located in Kupfrian Hall, Room 201. A Letter of Accommodation Eligibility from the Office of Accessibility Resources and Services office authorizing your accommodations will be required.

For further information regarding self identification, the submission of medical documentation and additional support services provided please visit the Office of Accessibility Resources and Services (OARS) website at:

https://www.njit.edu/studentsuccess/accessibility/

Important Dates (See: Spring 2022 Academic Calendar, Registrar)

Date	Day	Event	
January 18, 2022	Tuesday	First Day of Classes	
January 22, 2022	Saturday	Saturday Classes Begin	
January 24, 2022	Monday	Last Day to Add/Drop Classes	
March 14, 2022	Monday	Spring Recess Begins	
March 19, 2022	Saturday	Spring Recess Ends	
April 4, 2022	Monday	Last Day to Withdraw	
April 15, 2022	Friday	Good Friday - No Classes	
April 17, 2022	Sunday	Easter Sunday - No Classes	
May 3, 2022	Tuesday	Friday Classes Meet	
May 3, 2022	Tuesday	Last Day of Classes	
May 4 - May 5, 2022	Wednesday and Thursday	Reading Days	
May 6 - May 12, 2022	Friday to Thursday	Final Exam Period	

Date Lecture Sections Topic Assignment Week 1 Chapter 1 Introduction to Data Science Introduction to Data Science 1/19 Chapter 2 Statistical Learning; KNN Homework 1 2 Chapter 3 Linear Regression; R Lab Homework 2 4 Chapter 4 Logistic Regression Homework 2 4 Chapter 4 Logistic Regression Homework 2 4 Chapter 5 Chapter 4 Logistic Regression Homework 3 5 Chapter 5 Cross-Validation and Bootstrap Homework 3 6 Chapter 5 Cross-Validation and Bootstrap Homework 4 7 Chapter 6 Linear Model Selection; R Lab Homework 4 8 Chapter 6 Shrinkage Methods and Dimension Reduction Methods Homework 4 3/2 SPRING RECESS (NO CLASSES) 3/26 3/26 10 3/16 MIDTERM EXAM: Week Nethods; Bagging, Random Forests, Boosting Homework 5 12 Chapter 7 Nonlinear Modeling; R Lab Course Project 3/30 Chapter 8 Tree-Based Methods; Bagging, Random Fores		Course Outline			
1 IIIII/10Chapter 2Statistical Learning; KNNHomework 12 1/26Chapter 3Linear Regression; R LabHomework 13 2/2Chapter 3Linear Regression; R LabHomework 2Week 4 2/9Chapter 4Logistic RegressionHomework 24 2/9Chapter 4Logistic RegressionHomework 24 2/9Chapter 5Chapter 4LDA, QDA; R Lab5 2/16Chapter 5Cross-Validation and BootstrapHomework 36 2/23Chapter 6Linear Model Selection; R LabHomework 37 3/2Chapter 6Shrinkage Methods and Dimension Reduction MethodsHomework 48 3/9SPRING RECESS (NO CLASSES)Homework 49 3/16MIDTERM EXAM: Wednesday ~ March 23, 2022Course Project11 3/30Chapter 8Tree-Based Methods: Bagging, Random Forests, BoostingHomework 512 4/13Chapter 9Support Vector MachinesHomework 514 4/20Chapter 10Unsupervised Learning Homework 6Homework 6	Date	Lecture	Sections	Topic	Assignment
1/19	Week	1	Chapter 1	Introduction to Data Science	
Week 22 Chapter 2Chapter 2 AStatistical Learning; KNNHomework 11/263 3Chapter 3 ALinear Regression; R LabHomework 23/2/24Chapter 4 ALogistic RegressionHomework 24 4/2/94Chapter 4 ALogistic RegressionHomework 2Week 5 5/165Chapter 4 Chapter 5LDA, QDA; R LabHomework 36 6/2/236 Chapter 5Cross-Validation and Bootstrap Reduction Methods and Dimension Reduction MethodsHomework 47 3/26 Chapter 6Shrinkage Methods and Dimension Reduction MethodsHomework 48 3/9Chapter 6 ASPRING RECESS (NO CLASSES)Homework 49 3/166 AClapter 7Nonlinear Modeling; R LabCourse Project11 3/3077Chapter 8 ATree-Based Methods: Bagging, Random Forests, BoostingHomework 512 4/1310 AChapter 9 ASupport Vector MachinesHomework 513 4/1311 Chapter 10Chapter 10 Chapter 10Unsupervised Learning AHomework 614 4/2012 Chapter 10Chapter 10 Chapter 10Unsupervised Learning Chapter 10Homework 6					
2 112 Introduction of the second		2	Chantar 2	Statistical Learning, KNN	Homowork 1
1/26		Z	Chapter 2	Statistical Learning; KINN	HOILIEWOIK I
3 2/22AChapter 4Logistic RegressionHomework 2Week 4 2/95 2/16Chapter 4LDA, QDA; R LabHomework 3% 6 66 2/23Chapter 5 2/16Cross-Validation and Bootstrap 7 3/2Homework 3Week 7 7 3/26 7 7 7Chapter 6 7 7 3/2Linear Model Selection; R Lab 7 7 3/2Homework 4Week 8 9 3/167 7 7Chapter 6 7 7 7Sprinkage Methods and Dimension Reduction MethodsHomework 4Week 9 3/167 7 7Chapter 7 7 7Nonlinear Modeling; R Lab 7 7Course ProjectWeek 9 3/169 7 7Chapter 7 7 7Nonlinear Modeling; R Lab 7 7Course Project11 3/3010 7 7 7Chapter 8 7<					
2/2		3	Chapter 3	Linear Regression; R Lab	
Week 4 2/94Chapter 4 2 2Logistic RegressionHomework 2 4 2 4Week 6 2/235Chapter 4 2LDA, QDA; R LabHomework 3Week 6 6 2/236Chapter 5Cross-Validation and Bootstrap 4Homework 3Week 7 3/27Chapter 6Linear Model Selection; R LabHomework 4Week 8 3/98Chapter 6Shrinkage Methods and Dimension Reduction MethodsHomework 4Week 9 3/168Chapter 6Shrinkage Methods and Dimension Reduction MethodsHomework 4Week 9 3/169MIDTERM EXAM: Wednesday ~ March 23, 2022Course Project10 3/3010Chapter 7Nonlinear Modeling; R LabCourse Project11 3/3010Chapter 8Tree-Based Methods: Bagging, Random Forests, BoostingHomework 513 4/1311Chapter 9Support Vector MachinesHomework 614 4/2012Chapter 10Unsupervised Learning Homework 6Homework 6					
4 2/9Week 5 2/165Chapter 4LDA, QDA; R LabWeek 6 2/236Chapter 5Cross-Validation and Bootstrap 12/23Homework 36 2/2361Week 7 3/27Chapter 6Linear Model Selection; R Lab7 3/27Chapter 6Shrinkage Methods and Dimension Reduction MethodsHomework 48 3/98Chapter 6Shrinkage Methods and Dimension Reduction MethodsHomework 49 3/16999Week 10 3/2311Nonlinear Modeling; R LabCourse Project11 3/301010Chapter 7Nonlinear Modeling; R LabCourse Project12 4/611Chapter 8Tree-Based Methods: Bagging, Random Forests, BoostingHomework 512 4/1311Chapter 9Support Vector Machines113 4/1312Chapter 10Unsupervised LearningHomework 614 4/2012Chapter 10Unsupervised LearningHomework 6			Chapter 4	Logistic Dogrossion	Homowork 2
2/9		4	Chapter 4	Logistic Regression	Homework 2
511112/16Week6Chapter 5Cross-Validation and BootstrapHomework 362/23Week7Chapter 6Linear Model Selection; R Lab-77Chapter 6Shrinkage Methods and Dimension Reduction MethodsHomework 488Chapter 6Shrinkage Methods and Dimension Reduction MethodsHomework 4993/16Week10103/30Week9Chapter 7Nonlinear Modeling; R LabCourse Project113/30Week10Chapter 8Tree-Based Methods: Bagging, Random Forests, BoostingHomework 513144/13Week12Chapter 10Unsupervised LearningHomework 6144/20Week12Chapter 10Unsupervised LearningHomework 614141516					
2/16		5	Chapter 4	LDA, QDA; R Lab	
Week 6 2/236 Chapter 5Cross-Validation and BootstrapHomework 3Week 7 3/27Chapter 6Linear Model Selection; R Lab1Week 8 3/98Chapter 6Shrinkage Methods and Dimension Reduction MethodsHomework 48 3/9991Homework 29 3/169SPRING RECESS (NO CLASSES)1Week 9 3/2391Chapter 7Nonlinear Modeling; R LabWeek 10 3/239Chapter 7Nonlinear Modeling; R LabCourse Project11 3/3010Chapter 8Tree-Based Methods: Bagging, Random Forests, BoostingHomework 512 4/611 Chapter 9Chapter 9Support Vector MachinesHomework 613 4/1312 Chapter 10Chapter 10 Unsupervised LearningHomework 6					
6 1 1 1 2/23 - - - - Week 7 Chapter 6 Linear Model Selection; R Lab - 7 - - - - - Week 8 Chapter 6 Shrinkage Methods and Dimension Reduction Methods Homework 4 3/9 - - - - - Week 8 Chapter 6 SPRING RECESS (NO CLASSES) - - 9 3/16 - - - - - Week 10 -		6	Chapter 5	Cross-Validation and Bootstran	Homework 3
Week 7 3/27 Chapter 6Chapter 6 		0	Chapter 5	Cross-vanuation and Dootstrap	Home work 5
7 3/2Week 8 3/98Chapter 6Shrinkage Methods and Dimension Reduction MethodsHomework 48 3/99SPRING RECESS (NO CLASSES)Homework 49 3/16SPRING RECESS (NO CLASSES)9 3/1610 3/23Week 10 3/30Week 11 3/30Week 4/6Tree-Based Methods: Bagging, Random Forests, BoostingHomework 512 4/6Week 4/13Homework 614 4/20Week 4/20Week 4/20Week 4/20Week 4/20Week 4/20Week 4/20Week 4/20Week 4/20Week 4/20Week 4/20Week 4/20Week 4/20	2/23				
3/2Image: symbol sy		7	Chapter 6	Linear Model Selection; R Lab	
Week 8 3/98Chapter 6Shrinkage Methods and Dimension Reduction MethodsHomework 43/999 </td <td></td> <td></td> <td></td> <td></td> <td></td>					
81Reduction Methods3/9WeekSPRING RECESS (NO CLASSES)9-3/16-Week-10-3/23-Week9Chapter 7Nonlinear Modeling; R LabCourse Project11-3/30-Week1012-4/6-13-4/13-Week1212Chapter 913-4/13-Week1212Chapter 10Week1214-4/20-Week1213-4/20-14-4/20-Week1214-4/20-15-16-17-18-19-19-19-10-11-12-13-14-14-4/20-19-19-19-19-19-19-19-19-19-19-19-19-19-19- <t< td=""><td></td><td>8</td><td>Chapter 6</td><td>Shrinkage Methods and Dimension</td><td>Homework 4</td></t<>		8	Chapter 6	Shrinkage Methods and Dimension	Homework 4
Week 9 3/16SPRING RECESS (NO CLASSES)3/16MIDTERM EXAM: Week 3/2310 3/23MIDTERM EXAM: Wednesday ~ March 23, 20223/23Chapter 7Week 11 3/309Chapter 7Nonlinear Modeling; R LabWeek 4/610Chapter 8Tree-Based Methods: Bagging, Random Forests, Boosting12 4/6Chapter 9Week 13 4/1311Chapter 9Support Vector Machines13 4/13Chapter 10Week 14 4/2012Week12Meek1214 4/20Chapter 10Week12Meek14Meek <t< td=""><td></td><td>0</td><td>Chapter 0</td><td>8</td><td>Homework 4</td></t<>		0	Chapter 0	8	Homework 4
93/16Image: constraint of the symbol o	3/9				
3/16Image: constraint of the symbol of the symb				SPRING RECESS (NO CLASSES)	
Week 10 3/239Chapter 7MIDTERM EXAM: Wednesday ~ March 23, 2022Week 11 3/309Chapter 7Nonlinear Modeling; R LabCourse Project11 3/3000Chapter 8Tree-Based Methods: Bagging, Random Forests, BoostingHomework 512 4/610Chapter 8Tree-Based Methods: Bagging, Random Forests, BoostingHomework 512 4/611Chapter 9Support Vector MachinesHomework 513 4/1312Chapter 10Unsupervised LearningHomework 614 4/2012Chapter 10Students Project PresentationDeadline of the					
10				ΜΙΟΤΕΡΜ ΕΧΑΜ·	
3/23Image: symbol display bound in the symbol display bound i					
11 3/3010Chapter 8Tree-Based Methods: Bagging, Random Forests, BoostingHomework 512 4/610Chapter 8Tree-Based Methods: Bagging, Random Forests, BoostingHomework 512 4/611Chapter 9Support Vector MachinesHomework 513 4/1311Chapter 9Support Vector MachinesHomework 614 4/2012Chapter 10Unsupervised LearningHomework 614 4/2011Chapter 10Deadline of the	3/23				
3/30Image: symbol s		9	Chapter 7	Nonlinear Modeling; R Lab	Course Project
Week 12 4/610Chapter 8 aTree-Based Methods: Bagging, Random Forests, BoostingHomework 512 4/611Chapter 9Support Vector MachinesHomework 5Week 4/1311Chapter 9Support Vector MachinesHomework 613 4/1312Chapter 10Unsupervised LearningHomework 6Week 4/2012Chapter 10Students Project PresentationDeadline of the					
12Image: Constant of the state o		10	Chanter 8	Tree-Based Methods: Bagging, Random	Homework 5
4/6Image: constraint of the symbolImage: constraint of the symbolImage: constraint of the symbolWeek11Chapter 9Support Vector MachinesImage: constraint of the symbol13Image: constraint of the symbolImage: constraint of the symbolImage: constraint of the symbol4/13Image: constraint of the symbolChapter 10Image: constraint of the symbolImage: constraint of the symbolWeek12Chapter 10Image: constraint of the symbolImage: constraint of the symbolImage: constraint of the symbolWeekImage: constraint of the symbolStudents Project PresentationImage: constraint of the symbol		10			
13 4/13Image: Sector of the s	4/6				
4/13		11	Chapter 9	Support Vector Machines	
Week 12 Chapter 10 Unsupervised Learning Homework 6 14 4/20 -					
14 1 1 1 4/20 1 1 1 Week 1 1 1 Students Project Presentation Deadline of the		12	Chapter 10	Unsupervised Learning	Homework 6
Week Students Project Presentation Deadline of the		± -	Chapter 10		
5 nroject report			1	Students Project Presentation	
4/27	15 4/27				project report

Week		Reading Day 1	
16			
5/4			
Week		FINAL EXAM:	
17		Wednesday ~ May 11, 2022	
5/11			

Updated by Professor W. Guo - 1/13/2022 Department of Mathematical Sciences Course Syllabus, Spring 2022