Course Syllabus

Introduction to Data Science

INST 737
Spring 2019

Dr. Vanessa Frias-Martinez
vfrias@umd.edu

Class Meets
Online. Except for Wednesdays: January 30, February 27, March 27, and April 24 when we will meet from 14:00pm – 16:45pm at WDS #1114

Office Hours
Online. Please email me to set up a meeting.

Prerequisites
INST 627 and INFM 603

Course Communication
Announcements and sensitive information will be sent via Piazza. Please email me to discuss questions, absences, or accommodations. This is a link with helpful guidance on writing professional emails (ter.ps/email).

Learning Outcomes

Computers have made it possible, even easy, to collect vast amounts of data from a wide variety of sources. It is not always clear, however, how to use those data and how to extract useful information from data. This problem is faced in a tremendous range of scholarly, government, business, medical, and scientific applications. The purpose of this course is to teach some of the best and most general approaches to get the most out of data through clustering, classification, and regression techniques.

After successfully completing this course you will be able to:

- Understand state-of-the-art algorithms in the field data science
- Master the use of different R packages for applied machine learning tasks
- Envision, design and evaluate data science solutions for real-world problems

Recommended Resources

These books are not required but present an overview of the themes we will be discussing in class.

1. Machine Learning with R
Brett Lantz
PACKT Open Source
ISBN 9781782162148
Website: https://www.packtpub.com/big-data-and-business-intelligence/machine-learning-r-second-edition

2. R and Data Mining – Examples and Case Studies
Yanchang Zhao
Academic Press, Elsevier
ISBN: 9780123969637
Website: http://www.rdatamining.com/books/rdm

Course website: **https://piazza.com/umd/spring2019/inst737/home**

Activities, Learning Assessments, & Expectations for Students

This is a project-based class. Students will be required to carry out a data analysis project (with real data) including data collection, data cleaning, data analysis and visualization. For grading, the project will be divided into three milestones that students will need to present in class or recording a video. Online participation is highly encouraged through the piazza website. Students will be evaluated with these activities:

1. Semester-long Research Project

Students will be required to work on a semester-long, multi-student effort project. Students will be required to design and develop a full fledged data science project from data collection, to data cleaning, to applying different applied machine learning techniques and to discuss insights that can be extracted from the methods. This work will be divided into three graded deliverables (milestones):

*Milestone 1: dataset collection, cleaning, outlier detection, general statistics, plots
*Milestone 2: baseline computation, initial prediction/classification results
*Milestone 3: classifier comparison, visualizations

2. Project presentation and Discussion

Students will be required to present the results of each milestone either in class (milestones 1 and 2) or by recording a video (milestone 3).

3. Project Report and Code

Students are required to write a technical report with the main contributions of each of their milestones, and to submit the code that they have developed to achieve those results.

Campus Policies

It is our shared responsibility to know and abide by the University of Maryland's policies that relate to all courses, which include topics like:

- Academic integrity
- Student and instructor conduct
- Accessibility and accommodations
- Attendance and excused absences
- Grades and appeals
- Copyright and intellectual property

Please visit www.ugst.umd.edu/courserelatedpolicies.html for the Office of Undergraduate Studies' full list of campus-wide policies and follow up with me if you have questions.

Course-Specific Policies

No computers, phones or tablet devices are permitted during our class meeting unless used for class purposes.

I expect you to make the responsible and respectful decision to refrain from using your cellphone in class. If you have critical communication to attend to, please excuse yourself and return when you are ready. For more information about the science behind the policy watch: http://youtu.be/WwPaw3Fx5Hk

Late Work
Any assignment submitted up to three days after the deadline will get half credit. If you do this, please send me an email so that I know you have submitted it late. Assignments submitted after that will *not* be graded unless the student provides a formal letter (from the dean, doctor,...) to justify the special circumstances. If you envision not being able to meet a deadline, not being able to lead a paper presentation or discussion, not being able to participate in a discussion or not being able to attend your bi-weekly project presentation, please let me know one week in advance.

Get Some Help!

Taking personal responsibility for you own learning means acknowledging when your performance does not match your goals and doing something about it. I hope you will come talk to me so that I can help you find the right approach to success in this course, and I encourage you to visit http://tutoring.umd.edu to learn more about the wide range of campus resources available to you. In particular, everyone can use some help sharpen their communication skills (and improving their grade) by visiting ter.ps/writing and schedule an appointment with the campus Writing Center. Finally, if you just need someone to talk to, visit counseling.umd.edu.

Names/Pronouns and Self Identifications

The University of Maryland recognizes the importance of a diverse student body, and we are committed to fostering equitable classroom environments. I invite you, if you wish, to tell us how you want to be referred to both in terms of your name and your pronouns (he/him, she/her, they/them, etc.). The pronouns someone indicates are not necessarily indicative of their gender identity. Visit trans.umd.edu to learn more.

Additionally, how you identify in terms of your gender, race, class, sexuality, religion, and dis/ability, among all aspects of your identity, is your choice whether to disclose (e.g., should it come up in classroom conversation about our experiences and perspectives) and should be self-identified, not presumed or imposed. I will do my best to address and refer to all students accordingly, and I ask you to do the same for all of your fellow Terps.

Grades

Grades are not given, but earned. Your grade is determined by your performance on the learning assessments in the course and is assigned individually (not curved). If earning a particular grade is important to you, please speak with me at the beginning of the semester so that I can offer some helpful suggestions for achieving your goal.

All assessment scores will be posted on the course ELMS page. If you would like to review any of your grades (including the exams), or have questions about how something was scored, please email me to schedule a time for us to meet in my office.

Late work will not be accepted for course credit so please plan to have it submitted well before the scheduled deadline. I am happy to discuss any of your grades with you, and if I have made a mistake I will immediately correct it. Any formal grade disputes must be submitted in writing and within one week of receiving the grade.

Assessments	Weight
Milestone One	15%
Milestone Two	40%
Milestone Three	40%
Class Participation (online and in class)	5%

Final letter grades are assigned based on the percentage of total assessment points earned. To be fair to everyone I have to establish clear standards and apply them consistently, so please understand that being close to a cutoff is not the same this as making the cut (89.99 ≠ 90.00). It would be unethical to make exceptions for some and not others.

Final Grade Cutoffs									
+	97.00%	+	87.00%	+	77.00%	+	67.00%		
A	94.00%	B	84.00%	C	74.00%	D	64.00%	F	<60.0%
-	90.00%	-	80.00%	-	70.00%	-	60.00%		

Course Schedule

Class 1. Jan 30. Introduction to Data Science and R [in class]

Class 2. Feb 6. Managing and Understanding Data

Class 3. Feb 13. Probabilities and Data

Class 4. Feb 20. Regression Methods

Class 5. Feb 27. In-class Project Presentations (Milestone1) [HW due day before]

Class 6. March 6. Logistic Regressions and Naive Bayes Classifier

Class 7. March 13. Decision Trees and Random Forests

Spring Break

Class 8. March 27. In-class Project Presentations (Milestone2) [HW due day before]

Class 9. April 3. SVMs and Neural Networks

Class 10. April 10. Clustering

Class 11. April 17. Analyzing Model Performance

Class 12. April 24. In-class Project Clinic

Class 13. May 1. Project Video Presentations (Milestone3) [HW due day before]

Class 14. May 8. Project Video Presentations (Milestone3)