

MATH 637: Mathematical Techniques in Data Science

Time & Place: MWF 12:20PM – 1:10PM, Room: EWG 204 (Ewing Hall)

Instructor: Dominique Guillot, dguillot@udel.edu, Office: Ewing Hall 534.

Office hours details: TBA.

Venue: EWG 534.

Course description: The course provides an introduction to the fundamental techniques used in data mining. The main objective of the course is to develop a good mathematical understanding of the common tools that are used to analyse modern datasets. The course will also provide hands-on experience in data analysis through practical homework and class projects.

Goals of the course:

- Become familiar with the basic methods used to analyse modern datasets.
- Understand the mathematical theory and the standard models used in data science.
- Understand how to select a good model for data.
- Be able to analyse datasets using a modern programming language such as Python.

List of topics:

Linear methods for regression (subset selection, ridge, lasso), Logistic regression. Analysis of the convergence and complexity of common algorithms. Linear discriminant analysis, Principal component analysis, Additive Models, Kernel Smoothing. Cross-validation, Bootstrap, Support Vector Machines, Cluster analysis (K-means, spectral clustering). Undirected graphical models, Network models, Neural Networks.

Textbook:

T. Hastie, R. Tibshirani, J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer, 2009.

Pdf available at: <http://statweb.stanford.edu/~tibs/ElemStatLearn/>

Assessment plan:

60% homework (theoretical + programming problems), 40% class project.

Grading system: $A > 94\%$, $A- \geq 90\%$, $B- \geq 80\%$, $C- \geq 70\%$, $D- \geq 60\%$, $F < 60\%$. Subject to change (to your advantage only).

Prerequisites

Probability theory and basic statistics (e.g. MATH 350), Multivariable calculus (e.g. MATH 222), Linear Algebra (e.g. MATH 349), Optimization background (e.g. MATH 529) desirable but not necessary, basic computing skills.

Classroom Conduct: As a common courtesy to the instructor and your fellow classmates, please refrain from using your cell phone and computer in the classroom (unless instructed otherwise).

Academic Integrity Statement: All University of Delaware policies regarding ethics and honorable behavior apply to this course. Any student caught cheating will *automatically receive a failing course grade (F)*. Cheating includes copying during exams/quizzes, using an electronic device to communicate or obtain information during an exam, copying homework, taking material from an outside source without references, or any other form of plagiarism. Any material taken from an outside source must be properly cited. Other serious disciplinary measures (e.g. expulsion from the University) may also be applied to cheating students. For more details, please refer to the *Student Guide to University Policies* available at: <http://www1.udel.edu/stuguide/19-20/code.html>.

Miscellaneous: Please note that it is difficult to discuss mathematics questions by email due to the lack of an appropriate interface. You are encouraged to ask your technical questions during office hours or before/after the class.

Faculty Statement on Disclosures of Instances of Sexual Misconduct: If, at any time during this course, I happen to be made aware that a student may have been the victim of sexual misconduct (including sexual harassment, sexual violence, domestic/dating violence, or stalking), I am obligated to inform the university's Title IX Coordinator. Please visit <https://sites.udel.edu/sexualmisconduct/policies/> for more details about the University's policy about sexual misconduct.