

# MATH 829: Introduction to data mining and analysis

**Time & Place:** MWF 11:15AM – 12:05PM, Room: ALS 226 (Alison Hall)

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

Office hours details: Mon 3:00pm-4:00pm + Wed 3:00pm-4:00pm + by appointment.  
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 mining.
- Understand how to select a good model for data.
- Be able to analyse datasets using a modern programming language such as Python or R.

## 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:

Midterm (20%, **March 25th 2016**, in class), class project (20%), Weekly/biweekly homework (theoretical + programming problems), challenges.

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. Cheating receives a failing grade. For more details, please refer to the *Student Guide to University Policies* available at: <http://www.udel.edu/stuguide/15-16/code.html>.

**Faculty Statement on Disclosures of Instances of Sexual Misconduct** Please see the following link for information regarding the University's policy about sexual misconduct.