

Spring 2-1-2018

# M 462.01: Theoretical Basics of Big Data Analytics and Real Time Computation Algorithms

Brian M. Steele

University of Montana - Missoula, [brian.steele@umontana.edu](mailto:brian.steele@umontana.edu)

Let us know how access to this document benefits you.

Follow this and additional works at: <https://scholarworks.umt.edu/syllabi>

---

## Recommended Citation

Steele, Brian M., "M 462.01: Theoretical Basics of Big Data Analytics and Real Time Computation Algorithms" (2018). *Syllabi*. 7785.  
<https://scholarworks.umt.edu/syllabi/7785>

This Syllabus is brought to you for free and open access by the Course Syllabi at ScholarWorks at University of Montana. It has been accepted for inclusion in Syllabi by an authorized administrator of ScholarWorks at University of Montana. For more information, please contact [scholarworks@mso.umt.edu](mailto:scholarworks@mso.umt.edu).

# Syllabus: Math 462 Data Analytics Theory

**Instructor:** Brian Steele. Office: Math 314. 243-5396. [brian.steele@umontana.edu](mailto:brian.steele@umontana.edu)

## Course description

Co-convenes with M 562. Offered spring. Prereq., M 221 and two other Mathematics / Statistics classes at the 200-level or above, or consent of instr. The main goal of this course is to provide students with the opportunity to acquire conceptual knowledge and theoretical understanding of mathematical methods applicable to data analytics and real-time computations.

In this projects-oriented course, a complex data analytics problem, stock market forecasting, will motivate most of the course topics. As a group, we will attempt to develop methods and supporting code to forecast stock prices.

## Course Format

Three meetings per week MWF 2:00-2:50 a.m. Friday will usually, but not always, be reserved for the discussion of mathematically-advanced aspects of the subject material. After a methodological topic is introduced, students will apply the method to the problem by coding the method, or using existing software.

## Topics

### Predictive Analytics

1. Objective functions
2. Optimization methods: least squares, direct search, steepest descent, Newton's method
3. Accuracy assessment and randomization techniques
4. Cluster analysis
5. Artificial neural networks

### Data reduction methods

1. Dimension reduction: eigenvector techniques, projection onto lower-dimensional subspaces
2. Discrete wavelet transform

## Textbook

We will use several textbooks posted online on the Moodle website. These include but are not limited to:

1. *Algorithms for Data Science* Steele, B., Chandler, J., and Reddy, S.
2. *An Introduction to Statistical Learning* James, G., Witten, D., Hastie, T., and Tibshirani, R.
3. *Principal Components Analysis* Jolliffe, I.
4. *Deep Learning* Goodfellow, I., Bengio, Y., and Courville, A.

## Grading

Course grade is based on homework. Homework assignments often consist of programming a mathematical method (usually in Python) though some mathematical calculations will be assigned. I acknowledge that students differ widely respect to mathematical preparation and domain knowledge and will tailor the assignments accordingly (particularly those enrolled in M 562). Approximately 10 homeworks will be assigned and the point value will vary between 15 and 30. Homework will be collected approximately every 10 days (depending on the difficult of the task). Students are encouraged to work together. The final is take-home, and is due on the scheduled day of the final, Monday May 7.

Students enrolled in the graduate-level course, M 562, will give expository talks to the class related to the methods or the subject-matter aspects of the project. Total time engaged in expository talks usually will not exceed 50 minutes.

There will be time after class on Mondays and Wednesdays to get assistance from the instructor (in addition to Tuesday and Thursday office hours) and to work with other students.