Fall 9-1-2018

M 561.01: Advanced Data Science Analytics

Javier P. Alvaro
University of Montana, Missoula

M461 - Data Science Analytics
M561 - Advanced Data Science Analytics

Instructor information:

Instructor: Javier Pérez Álvaro
Office: M205A
Email: javier.perez-alvaro@mso.umt.edu
Office hours: See http://www.umt.edu/people/perezalvaro for up-to-date OH.

Course Format:
Meetings: Monday, Wednesday, Friday 9:00-9:50 a.m., Math 306. Class time is split (50/50) between lecture and individual and small group work on programming and algorithm development.

Course content:
Algorithms are the machinery behind the data analytics (the subject matter and focal point of the course). To be good at data analytics, one must be a competent programmer and have experience with the data and the algorithms of data science. To gain an understanding of algorithm design and good programming practices, students will work with a set of prototypical algorithms that are representative of data analytics. To learn how to function as a data scientist in a relatively short time, the student will be actively engaged in turning the algorithms into code and using them with real data. We shall cover most, but not all of the material in chapters 1 through 8 of the text book:

1. Data mappings and the concepts of data reduction. Similarity measures and distance metrics. List, set, and dictionary comprehension.

2. Scalable algorithms and associative statistics. Computing univariate and multivariate statistics from massively large data sets.

3. Distributed computing using MapReduce algorithms and the Hadoop environment. Basics of the command line. Utilizing Elastic MatReduce.

4. Data visualization.

5. Linear regression. Hypothesis testing, regression with factors, interaction, and residual analysis.

6. Healthcare analytics.

7. Cluster analysis. Hierarchical and k-means methods. Optimality.

Learning outcomes:

1. Understand the purpose of data reduction and information extraction (e.g., associative statistics and data mapping).

2. Develop understanding and practical experience regarding reduction of massive data sets and data streams.

3. Understand the mechanics of distributed computing.

4. Ability to implement algorithms for processing massively large data sets. Ability to compute histograms, correlation matrices, and linear regression estimators using massively large data sets.

5. Understand the objectives of multiple regression and examining model assumptions. Ability to carry out and interpret hypothesis tests.

6. Competency using Python and R.

Textbook: Algorithms for Data Science (ISBN-10: 3319457950)

Homework: Homework exercises emphasizing applications of the algorithms and/or theory will be assigned weekly. Students are responsible for completing 4 to 6 tutorials per month (due the first meeting of each month except September). Tutorials are oriented toward gaining proficiency in programming and algorithm design.

Final Project: In lieu of a final, students will be graded on their final project. Students must work on projects in groups of one, two or three individuals. The final meeting time (for presentation of projects) is Friday, December 14. Students are responsible for a written paper (75% of the project grade) and oral presentation (25% of the project grade).

Grading policy: Your course grade will be based on homework, tutorials, and the final project

Item	Percentage
Homework	40%
Tutorials	40%
Final project	20%

Student Conduct: All students need to be familiar with the Student Conduct Code. You can find in the the "A to Z Index" on the UM home page. All students must practice academic honesty. Academic misconduct is subject to an academic penalty by the course instructor and/or a disciplinary sanction by the University.

Accommodation: The University of Montana assures equal access to instruction through collaboration between students with disabilities, instructors and Disability Services for Students (DSS). If you think that you may have a disability adversely affecting your academic performance, and you have not already registered with DSS, please contact DSS in Lommasson Center 154 or call 406.243.2243. I will work with you and DSS to provide an appropriate accommodation.

Important note: Announcements made in class are considered addenda to this syllabus. Make sure you stay informed as the progress of the class.