CMSE 820: Mathematical Foundations of Data Science

Instructor Information:

• **Instructor:** Yang Yang yangy5@msu.edu
• **Zoom:** https://msu.zoom.us/j/8393634403
• **Lectures:** M W F 10:20-11:10am
• **Office hours (via Zoom):** T Th 10-11am

Course Description:

This course serves as the mathematical theory foundation for data science, with a concentration on basic mathematical principles and algorithms in data sciences. The purpose of this course is to build the fundamental mathematical knowledge in the context of data sciences. Topics covered will include the mathematical theory of linear regression, ridge regression/l2-regularization, LASSO regression/l1-regularization, principal component analysis, multi-dimensional scaling, perceptron, support vector machine, spines, kernel methods, and concentration of measure. Other selected topics will be added if time permits. The content relies heavily on reasoning. Most of the problems in the homework assignments and exams are mathematical proofs.

A strong background in mathematics is required for participants, see the prerequisites below. Students must be proficient with the prerequisite requirements. Historically, students lacking the prerequisites have found it challenging to keep up with the course delivery speed.

Caution: Here is what this course **will** teach regarding data sciences:
• Mathematical theory
• Algorithm analysis
• Algorithm design

Caution: Here is what this course **will NOT** teach regarding data sciences:
• Data processing
• Algorithm implementation
• Coding
• Use of software packages

Prerequisites:

A strong background in calculus, linear algebra, and probability/statistical theory is essential in order to succeed in this course. In addition, knowledge of optimization theory and functional analysis are required for a few advanced topics. The minimum prerequisites include:
• CMSE 802 or equivalent experience
• Differential equation at the level of MTH 235/255H/340 + 442/347H + 442
• Linear algebra at the level of MTH 390/317H
• Probability and statistics at the level of STT 231

Textbook and References:

There is no textbook for the course. Handwritten lecture notes will be posted on the D2L course page. The following references are recommended:

• *Learning with Kernels*, by Schlkopf and Smola. Available online through the MSU library.
• *Foundations of Machine Learning*, by Mohri, Rostamizadeh and Talwalkar. Available online through the MSU library.
• *Topics in Mathematics of Data Science*, MIT open course.
• *A Mathematical Introduction to Data Science* by Yuan Yao, available online.
• *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* by Trevor Hastie, Robert Tibshirani, and Jerome Friedman

Course Grading:

Homework   40%
Midterm   20% (Date TBD)
Final   30% (Date TBD)
Participation   10%

The Midterm covers everything up to that point. The Final will be cumulative and covers all the topics of the semester. These are "closed book" exams; no notes, computers, phones, or other aides are permitted during the exam. Homework assignments will be uploaded to D2L weekly.

Other Information:

Academic Honesty: Intellectual integrity is the foundation of the scientific enterprise. In all instances, you must do your own work and give proper credit to all sources that you use in your homework, reports and oral presentations – any instance of submitting another person's work, ideas, or wording as your own counts as plagiarism. This includes failing to cite any direct quotations in your essays, research paper, class debate, or written presentation. The MSU College of Natural Science adheres to the policies of academic honesty as specified in the General Student Regulations 1.0, Protection of Scholarship and Grades, and in the all-University statement on Integrity of Scholarship and Grades, which are included in Spartan Life: Student Handbook and Resource Guide. Students who plagiarize will receive a 0.0 in the course. In addition, University policy requires that any cheating offense, regardless of the magnitude of the infraction or punishment decided upon by the professor, be reported immediately to the dean of the student's college.

It is important to note that **plagiarism in the context of this course includes, but is not limited to,** directly copying another student's solutions to homework problems; copying materials from online sources, textbooks, or other reference materials *without citing those references in your source code or documentation*, or having somebody else do your homework on your behalf. Any work that is done in collaboration with other students should state this explicitly, and have their names as well as yours listed clearly.

More broadly, we ask that students adhere to the Spartan Code of Honor academic pledge, as written by the Associated Students of Michigan State University (ASMSU): "As a Spartan, I will strive to uphold values of the highest ethical standard. I will practice honesty in my work, foster honesty in my peers, and take pride in knowing that honor is worth more than grades. I will carry these values beyond my time as a student at Michigan State University, continuing the endeavor to build personal integrity in all that I do."

Accommodations: If you have a university-documented learning difficulty or require other accommodations, please provide me with your VISA as soon as possible and speak with me about how I can assist you in your learning. If you do not have a VISA but have been documented with a learning difficulty or other problems for which you may still require accommodation, please contact MSU's Resource Center for People with Disabilities (355-9642) in order to acquire current documentation.