**Mathematics 6001-106 Syllabus**
**Nonlinear Optimization in Machine Learning**
Spring 2020

**Instructor:** Dr. Wenqing Hu
**Office:** Rolla 112
**Phone:** 573-341-4650
**Email:** huwen@mst.edu
**Classroom:** EECH 239
**Class time:** 12:00am-12:50am MWF
**Office Hours:** 08:00am-08:50am MWF
**Distant Section:** Class #73479

**Topics to be covered:** This course focuses on nonlinear optimization methods that are commonly used in modern statistical machine learning. The instructor plans to cover the following topic given permission of time:
1. background information in statistical machine learning;
2. fundamental knowledge about convex functions;
3. line search methods: steepest descent, line searches, back-propagation for neural networks;
4. gradient methods using momentum: Polyak's heavy-ball method, Nesterov's accelerated gradient method;
5. stochastic gradient methods and its variants: stochastic gradient descent (SGD), stochastic variance reduced gradient descent (SVRG), recent progresses on large-scale nonconvex optimization: perturbed gradient descent, escape from stationary points (saddle point, local minima), stochastic gradient descent Langevin dynamics (SGLD);
6. optimization using second order derivatives: Newton's method, quasi-Newton's method;
7. dual methods: dual coordinate descent/ascent, Stochastic Dual Coordinate Ascent;
8. Optimization with constraints: Subgradient projection method, Franke-Wolfe method, Alternating Direction Method of Multipliers;
9. Adaptive methods: SGD with momentum, AdaGrad, RMSProp, AdaDelta, Adam;
10. Optimization with discontinuous objective function: proximal operators and proximal algorithms.

**Textbook:** There will be no explicit textbook for the course, rather, the instructor will provide online lecture notes along with the progress of the course. However, the following monographs on this topic are recommended:
[1] J. Nocedal and S. J. Wright, *Numerical Optimization*, Second Edition, Springer, 2006.
[2] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press.
[3] Nesterov, Y., *Introductory Lectures on Convex Optimization*, Kluwer, 2004.
The course is also accompanied with some classical and very recent research papers related to its topic.

**Prerequisites:** Math 2222; Math 3108; one of Stat 3115, Stat 3117, Stat 5643. The course requires no prior knowledge in statistical machine learning. The only knowledge required are basic calculus (including multivariable calculus) and some elementary probability and statistics (such as expectation, variance, unbiased estimators, etc.).

**Coding:** Coding experience with either MATLAB or Python is not required but will be useful in understanding some of the algorithms that we will cover during the course. There will be several optional Programming Projects which are in the form of handouts that ask the students to run some given codes testing the performances of the algorithms discussed in class, but they do not count for credits. The codes for these projects will be drafted by the instructor and will be shared on the github account of the instructor.

**Canvas:** You should be enrolled in Math 6001-106 Canvas course. All class materials will be posted there.

**Attendance and Drop Policy:** You are expected to attend every class period. If you miss a class, it is your responsibility to find out what you missed, pick up any handouts, returned exams or homework, etc. If you incur three unexcused absences, you will receive an academic alert from me and you will be required to meet with me in my office to discuss your lack of attendance and its consequences. If you accumulate six unexcused absences you can expect to be dropped from the course.

**Homework:** You should make it a practice to do your homework in due time and you are expected to turn in homework regularly. A total of 4 homework sets will be assigned, graded, and returned during this course. Each homework set counts for 15 points, and in total 60 points. Solutions to the homework problems will be posted online in the Canvas MATH 6001-106 resource course.

**Final Project:** There will be a take-home final project assigned at the end of the semester. The final project is worth 40 points.

**Grading:** There will be 100 total points – 60 in homework and 40 in final project. You will need at least 90 points (90%) to receive an A, 80 points (80%) to receive a B, 70 points (70%) to receive a C, and 60 points (60%) to receive a D. If you earn less than 60 points you can expect to receive an F.

**Question/Concern Resolution:** If you ever have a question, problem, or concern about anything in this course, please come see me first. If your concern still is unresolved then consult our Interim Department Chair, Dr. Sam, in room 202 of the Rolla Building.

**Disability Support Services:** It is the policy and practice of Missouri University of Science and Technology to promote inclusive learning environments. If you have a documented disability you may be eligible for reasonable accommodations in compliance with university policy, the Americans with Disabilities Act of 1990, the Americans with Disabilities Amendment Act (ADAAA) of 2008, and Section 504 of the Rehabilitation Act of 1973. Please note, students are not encouraged to negotiate accommodations directly with professors.

To request accommodations or assistance, please self-identify with Disability Support Services (DSS), 203 Norwood Hall. For more information or to register for services, contact DSS at (573) 341-6655 or by email at dss@mst.edu.