DOCTOR OF PHILOSOPHY WITH A MAJOR IN MACHINE LEARNING

The Doctor of Philosophy with a major in Machine Learning program has the following principal objectives, each of which supports an aspect of the Institute's mission:

1. Create students that are able to advance the state of knowledge and practice in machine learning through innovative research contributions.
2. Create students who are able to integrate and apply principles from computing, statistics, optimization, engineering, mathematics and science to innovate, and create machine learning models and apply them to solve important real-world data intensive problems.
3. Create students who are able to participate in multidisciplinary teams that include individuals whose primary background is in statistics, optimization, engineering, mathematics and science.
4. Provide a high quality education that prepares individuals for careers in industry, government (e.g., national laboratories), and academia, both in terms of knowledge, computational (e.g., software development) skills, and mathematical modeling skills.
5. Foster multidisciplinary collaboration among researchers and educators in areas such as computer science, statistics, optimization, engineering, social science, and computational biology.
6. Foster economic development in the state of Georgia.
7. Advance Georgia Tech's position of academic leadership by attracting high quality students who would not otherwise apply to Tech for graduate study.

All PhD programs must incorporate a standard set of Requirements for the Doctoral Degree.

The central goal of the PhD program is to train students to perform original, independent research. The most important part of the curriculum is the successful defense of a PhD Dissertation, which demonstrates this research ability. The academic requirements are designed in service of this goal.

The curriculum for the PhD in Machine Learning is truly multidisciplinary, containing courses taught in nine schools across three colleges at Georgia Tech: the Schools of Computational Science and Engineering, Computer Science, and Interactive Computing in the College of Computing; the Schools of Aerospace Engineering, Chemical and Biomolecular Engineering, Industrial and Systems Engineering, Electrical and Computer Engineering, and Biomedical Engineering in the College of Engineering; and the School of Mathematics in the College of Science.

Summary of General Requirements for a PhD in Machine Learning

- Core curriculum (4 courses, 12 hours). Machine Learning PhD students will be required to complete courses in four different areas: Mathematical Foundations, Probabilistic and Statistical Methods in Machine Learning, ML Theory and Methods, and Optimization.
- Area electives (5 courses, 15 hours).
- Responsible Conduct of Research (RCR) (1 course, 1 hour, pass/fail). Georgia Tech requires that all PhD students complete an RCR requirement that consists of an online component and in-person training. The online component is completed during the student's first semester enrolled at Georgia Tech. The in-person training is satisfied by taking PHIL 6000 or their associated academic program's in-house RCR course.
- Qualifying examination (1 course, 3 hours). This consists of a one-semester independent literature review followed by an oral examination.
- Doctoral minor (2 courses, 6 hours).
- Research Proposal. The purpose of the proposal is to give the faculty an opportunity to give feedback on the student's research direction, and to make sure they are developing into able communicators.
- PhD Dissertation.

Almost all of the courses in both the core and elective categories are already taught regularly at Georgia Tech. However, two core courses (designated in the next section) are being developed specifically for this program. The proposed outlines for these courses can be found in the Appendix. Students who complete these required courses as part of a master's program will not need to repeat the courses if they are admitted to the ML PhD program.

Core Courses

Machine Learning PhD students will be required to complete courses in four different areas. With the exception of the Foundations course, each of these area requirements can be satisfied using existing courses from the College of Computing or Schools of ECE, ISyE, and Mathematics.

Machine Learning core:

Mathematical Foundations of Machine Learning. This required course is the gateway into the program, and covers the key subjects from applied mathematics needed for a rigorous graduate program in ML. Particular emphasis will be put on advanced concepts in linear algebra and probabilistic modeling. This course is cross-listed between CS, CSE, ECE, and ISyE.

ECE 7750/ISYE 7750/CS 7750/CSE 7750 Mathematical Foundations of Machine Learning

Probabilistic and Statistical Methods in Machine Learning

- ISYE 6412, Theoretical Statistics
- ECE 7751/ISYE 7751/CS 7751/CSE 7751 Probabilistic Graphical Models
- MATH 7251 High Dimension Probability
- MATH 7252 High Dimension Statistics

Machine Learning: Theory and Methods. This course serves as an introduction to the foundational problems, algorithms, and modeling techniques in machine learning. Each of the courses listed below treats roughly the same material using a mix of applied mathematics and computer science, and each has a different balance between the two.

- CS 7545 Machine Learning Theory and Methods
- CS 7616, Pattern Recognition
- CSE 6740/ISYE 6740, Computational Data Analysis
- ECE 6254, Statistical Machine Learning
- ECE 6273, Methods of Pattern Recognition with Applications to Voice

Optimization. Optimization plays a crucial role in both developing new machine learning algorithms and analyzing their performance. The three courses below all provide a rigorous introduction to this topic;

each emphasizes different material and provides a unique balance of mathematics and algorithms.

- ECE 8823, Convex Optimization: Theory, Algorithms, and Applications
- ISYE 6661, Linear Optimization
- ISYE 6663, Nonlinear Optimization
- ISYE 7683, Advanced Nonlinear Programming

Electives

After core requirements are satisfied, all courses listed in the core not already taken can be used as (appropriately classified) electives.

In addition to meeting the core area requirements, each student is required to complete five elective courses. These courses are required for getting a complete breadth in ML. These courses must be chosen from at least two of the five subject areas listed below. In addition, students can use up to six special problems research hours to satisfy this requirement.

i. **Statistics and Applied Probability**: To build breadth and depth in the areas of statistics and probability as applied to ML.

- AE 6505, Kalman Filtering
- AE 8803 Gaussian Processes
- BMED 6700, Biostatistics
- ECE 6558, Stochastic Systems
- ECE 6601, Random Processes
- ECE 6605, Information Theory
- ISYE 6402, Time Series Analysis
- ISYE 6404, Nonparametric Data Analysis
- ISYE 6413, Design and Analysis of Experiments
- ISYE 6414, Regression Analysis
- ISYE 6416, Computational Statistics
- ISYE 6420, Bayesian Statistics
- ISYE 6761, Stochastic Processes I
- ISYE 6762, Stochastic Processes II
- ISYE 7400, Adv Design-Experiments
- ISYE 7401, Adv Statistical Modeling
- ISYE 7405, Multivariate Data Analysis
- ISYE 8803, Statistical and Probabilistic Methods for Data Science
- ISYE 8813, Special Topics in Data Science
- MATH 6221, Probability Theory for Scientists and Engineers
- MATH 6266, Statistical Linear Modeling
- MATH 6267, Multivariate Statistical Analysis
- MATH 7244, Stochastic Processes and Stochastic Calculus I
- MATH 7245, Stochastic Processes and Stochastic Calculus II

ii. Advanced Theory: To build a deeper understanding of foundations of ML.

- AE 8803, Optimal Transport Theory and Applications
- CS 7280, Network Science
- CS 7510, Graph Algorithms
- CS 7520, Approximation Algorithms
- CS 7530, Randomized Algorithms
- CS 7535, Markov Chain Monte Carlo Algorithms
- CS 7540, Spectral Algorithms
- CS 8803, Continuous Algorithms
- ECE 6283, Harmonic Analysis and Signal Processing
- ECE 6555, Linear Estimation
- ISYE 7682, Convexity
- MATH 6112, Advanced Linear Algebra
- MATH 6241, Probability I
- MATH 6262, Advanced Statistical Inference
- MATH 6263, Testing Statistical Hypotheses
- MATH 6580, Introduction to Hilbert Space
- MATH 7338, Functional Analysis
- MATH 7586, Tensor Analysis
- MATH 88XX, Special Topics: High Dimensional Probability and Statistics

iii. **Applications:** To develop a breadth and depth in variety of applications domains impacted by/with ML.

- AE 6373, Advanced Design Methods
- AE 8803, Machine Learning for Control Systems
- AE 8803, Nonlinear Stochastic Optimal Control
- BMED 6780, Medical Image Processing
- BMED 6790/ECE 6790, Information Processing Models in Neural Systems
- BMED 7610, Quantitative Neuroscience
- BMED 8813BHI, Biomedical and Health Informatics
- BMED 8813MHI, mHealth Informatics
- BMED 8813MLB, Machine Learning in Biomedicine
- BMED 8823ALG, OMICS Data and Bioinformatics Algorithms
- CHBE 6745, Data Analytics for Chemical Engineers
- CHBE 6746, Data-Driven Process Engineering
- CS 6440, Introduction to Health Informatics
- CS 6465, Computational Journalism
- CS 6471, Computational Social Science
- CS 6474, Social Computing
- CS 6475, Computational Photography
- CS 6476, Computer Vision
- CS 6601, Artificial Intelligence
- CS 7450, Information Visualization
- CS 7476, Advanced Computer Vision
- CS 7630, Autonomous Robots
- CS 7632, Game AI
- CS 7636, Computational Perception
- CS 7643, Deep Learning
- CS 7646, Machine Learning for Trading
- CS 7647, Machine Learning with Limited Supervision
- CS 7650, Natural Language Processing
- CSE 6141, Massive Graph Analysis
- CSE 6240, Web Search and Text Mining
- CSE 6242, Data and Visual Analytics
- CSE 6301, Algorithms in Bioinformatics and Computational Biology
- ECE 4580, Computational Computer Vision
- ECE 6255, Digital Processing of Speech Signals
- ECE 6258, Digital Image Processing
- ECE 6260, Data Compression and Modeling
- ECE 6273, Methods of Pattern Recognition with Application to Voice

- ECE 6550, Linear Systems and Controls
- ECE 8813, Network Security
- ISYE 6421, Biostatistics
- ISYE 6810, Systems Monitoring and Prognosis
- ISYE 7201, Production Systems
- ISYE 7204, Info Prod & Ser Sys
- ISYE 7203, Logistics Systems
- ISYE 8813, Supply Chain Inventory Theory
- HS 6000, Healthcare Delivery
- MATH 6759, Stochastic Processes in Finance
- MATH 6783, Financial Data Analysis

iv. Computing and Optimization: To provide more breadth and foundation in areas of math, optimization and computation for ML.

- AE 6513, Mathematical Planning and Decision-Making for Autonomy
- AE 8803, Optimization-Based Learning Control and Games
- CS 6515, Introduction to Graduate Algorithms
- CS 6550, Design and Analysis of Algorithms
- CSE 6140, Computational Science and Engineering Algorithms
- CSE 6643, Numerical Linear Algebra
- CSE 6644, Iterative Methods for Systems of Equations
- CSE 6710, Numerical Methods I
- CSE 6711, Numerical Methods II
- ECE 6553, Optimal Control and Optimization
- ISYE 6644, Simulation
- ISYE 6645, Monte Carlo Methods
- ISYE 6662, Discrete Optimization
- ISYE 6664, Stochastic Optimization
- ISYE 6679, Computational methods for optimization
- ISYE 7686, Advanced Combinatorial Optimization
- ISYE 7687, Advanced Integer Programming

v. **Platforms**: To provide breadth and depth in computing platforms that support ML and Computation.

- CS 6421, Temporal, Spatial, and Active Databases
- CS 6430, Parallel and Distributed Databases
- CS 6290, High-Performance Computer Architecture
- CSE 6220, High Performance Computing
- CSE 6230, High Performance Parallel Computing

Qualifying Examination

The purpose of the Qualifying Examination is to judge the candidate's potential as an independent researcher.

The Ph.D. qualifying exam consists of a focused literature review that will take place over the course of one semester. At the beginning of the second semester of their second year, a qualifying committee consisting of three members of the ML faculty will assign, in consultation with the student and the student's advisor, a course of study consisting of influential papers, books, or other intellectual artifacts relevant to the student's research interests. The student's focus area and current research efforts (and related portfolio) will be considered in defining the course of study.

At the end of the semester, the student will submit a written summary of each artifact which highlights their understanding of the importance (and weaknesses) of the work in question and the relationship of this work to their current research. Subsequently, the student will have a closed oral exam with the three members of the committee. The exam will be interactive, with the student and the committee discussing and criticizing each work and posing questions related the students current research to determine the breadth of student's knowledge in that specific area.

The success of the examination will be determined by the committee's qualitative assessment of the student's understanding of the theory, methods, and ultimate impact of the assigned syllabus.

The student will be given a passing grade for meeting the requirements of the committee in both the written and the oral part. Unsatisfactory performance on either part will require the student to redo the entire qualifying exam in the following semester year. Each student will be allowed only two attempts at the exam.

Students are expected to perform the review by the end of their second year in the program.

Doctoral Dissertation

The primary requirement of the PhD student is to do original and substantial research. This research is reported for review in the PhD dissertation, and presented at the final defense.
As the first step towards completing a dissertation, the student must prepare and defend a Research Proposal. The proposal is a document of no more than 20 pages in length that carefully describes the topic of the dissertation, including references to prior work, and any preliminary results to date. The written proposal is submitted to a committee of three faculty members from the ML PhD program, and is presented in a public seminar shortly thereafter. The committee members provide feedback on the proposed research directions, comments on the strength of writing and oral presentation skills, and might suggest further courses to solidify the student's background. Approval of the Research Proposal by the committee is required at least six months prior to the scheduling of the PhD defense. It is expected that the student complete this proposal requirement no later than their fourth year in the program.
The PhD thesis committee consists of five faculty members: the student's advisor, three additional members from the ML PhD program, and one faculty member external to the ML program. The committee is charged with approving the written dissertation and administering the final defense. The defense consists of a public seminar followed by oral examination from the thesis committee.

Doctoral minor (2 courses, 6 hours):

The minor follows the standard Georgia Tech requirement: 6 hours, preferably outside the student's home unit, with a GPA in those graduate-level courses of at least 3.0. The courses for the minor should form a cohesive program of study outside the area of Machine Learning; no ML core or elective courses may be used to fulfill this requirement and must be approved by your thesis advisor and ML Academic Advisor. Typical programs will consist of three courses two courses from the same school (any school at the Institute) or two courses from the same area of study.