# Machine Learning PhD Handbook

Revised: September 6, 2019

# I. Introduction

The machine learning (ML) Ph.D. program is a collaborative venture between Georgia Tech's colleges of Computing, Engineering, and Sciences. 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 eight associated 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, Biomedical Engineering, Electrical and Computer Engineering, and Industrial and Systems Engineering in the College of Engineering; and the School of Mathematics in the College of Science. Faculty from these eight associated schools also supervise students in the program.

The program is administered by the Director, the Faculty Advisory Committee (FAC), the Academic Professional (AP), and the Academic Advisor. The roles of these entities are detailed below. The lifeblood of the program are the ML PhD students, and the Affiliated ML Faculty who advise, mentor, and conduct research with these students.

This handbook is meant to detail the inner workings of the ML PhD program for students, faculty, and administrators. The sections below describe the processes in the most important parts of the program, including admissions, coursework requirements, the qualifying exam, the PhD proposal, the dissertation and final defense, and the mechanisms for administration. A list of current personnel is included at the end.

# II. Admissions

External applications are only accepted for the Fall semester each year. The application deadline varies by home school.

The ML PhD admissions process works bottom-up through the home schools. The FAC representative coordinates the process in their respective units, working with the associate chair for graduate studies and other faculty associated with the ML PhD program. Admissions decisions are made by the home school, and then submitted to the ML FAC for final approval.

Support for incoming students (including guarantees of teaching assistantships and/or fellowships) is determined by the home schools.

After the admissions have been approved by the FAC, the home school will communicate the acceptance to the prospective student. The home school will also communicate all rejections.

## Transfer Admissions

Students can also transfer into the ML PhD if they are currently enrolled in another graduate program at Georgia Tech. Transfers are accepted at the ML PhD Director's discretion. A guideline for transfer admission is that the student have a commitment for a research assistantship for at least one year from a faculty member associated with the ML PhD program.

Transfer applications are submitted via email to the Program Director. The application should include a statement of purpose by the student, unofficial transcripts for all graduate work done at Georgia Tech and other institutions, and a letter from their advisor guaranteeing the research assistantship.

## Recruiting visits

Recruiting visits are organized by the home schools along with their other degree programs.

## New student orientation

There will be a new student orientation for all new ML PhD students the week before classes start before August. It will consist of a short presentation by the director on the degree requirements, an open discussion, and then a mixer with current PhD students and faculty. This will be in addition to the orientation the incoming students have with their home schools.

# III. Coursework Requirements

The coursework for the ML PhD consists of the following.

1. Core curriculum, 5 courses, 15 hours. The courses that can be used to fulfill this requirement are discussed in detail in Appendix A.
2. Area electives, 5 courses, 15 hours. The electives must be chosen from the list in Appendix B.
3. Doctoral minor, 3 courses, 9 hours. These courses must form a cohesive course of study, and must come from outside the area of machine learning; no ML courses (core or elective) can be used to fulfill this requirement. A minimum GPA of 3.0 must be achieved in courses used for the minor. The minor course of study is approved by the Academic Professional per the advice of the student's thesis advisor.
4. Responsible Conduct of Research, 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.
5. ML PhD Qualifying exam, 1 course, 3 hours (pass/fail). As described in the Qualifying Exam section above, the qualifying exam consists of a one semester literature study. To take the exam, students enroll in CS 7999 (Prep-Doctoral Qual Exams).

All coursework must be completed before the PhD proposal. An overall GPA of 3.0 is required for the PhD coursework.

The coursework requirements are subject to change. Students can satisfy any version of them that are in effect from the date of matriculation into the program to the date of the PhD Proposal.

More information about the Institute-level requirements for the PhD degree can be found here: http://www.catalog.gatech.edu/academics/graduate/doctoral-degree-info/

## Special problems

Special problem courses (e.g. ECE 890x) offer students the opportunity to be involved in research while earning course credit. Up to six special problem hours (two courses) can be used towards the ML PhD Area Elective Credits. Student performance for special problems courses must be certified by the faculty instructor. A short (5 page) report detailing the student's activities during the semester should be submitted to the supervising faculty during the last

week of class.  Faculty must also certify that the work done for a special problem course is not simultaneously being done for pay (as a GRA).

Only special problems courses taken in the first two years in the program will be counted as credit towards the ML PhD.

## Course transfer credit

If a student has completed graduate work at another institution, they can petition the FAC to receive transfer credit.  Students should request the class evaluation during their second term or later. The classes have to be at the graduate or senior undergraduate level. No classes that were used to satisfy an undergraduate degree can be used toward the ML Ph.D. degree. Up to 30 hours of credits earned toward a graduate degree with a grade of B and above at a different institution can be used towards the Ph.D. degree.  The approved classes will only count toward the Ph.D. degree course requirements and will not be transferred on the Georgia Tech transcript.

The course transfer credit process is coordinated by the programs Academic Professional.  The student should submit to the AP a short letter of petition describing which courses they believe they should receive credit for along with justification.  Along with this letter, an unofficial transcript from the institute where the classes were taken and syllabi for each of the courses in question should be included.  The program AP will work with academic professionals in the home units to determine course credit.  The FAC will have final approval.

## Academic Advising

For advice on which courses to take and when to take them, students should turn to the home school representative to the ML FAC (see the Program Personnel section), and/or their thesis advisor.

Advising on non-academic issues will done through the ML PhD Academic Professional (see Program Personnel section) and through the resources available in the student's home school.

# IV. Qualifying Exam

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

The ML Ph.D. qualifying exam consists of a focused literature review that will take place over the course of one semester. In the semester in which they want to take the qualifying exam, the student registers for CS 7999 (Prep-Doctoral Qual Exams). Before taking the exam, it is expected that the student has either taken or is currently enrolled in all five core required courses, and has completed their RCR course. As a guideline, students should plan to take the exam in their fourth semester in the ML PhD program.

At the beginning of the semester in which they are taking the qualifying exam, the student forms an exam committee consisting of three members of the ML faculty. The committee, in consultation with the student, will approve a course of study consisting of influential papers, books, or other intellectual artifacts relevant to the student's research field.

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

The oral exam should be scheduled on or before the last week of class in the semester; the written report should be submitted to the committee at least one week prior.

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. Each student will be allowed only two attempts at the exam.

After a successful exam, the committee fills out and signs the Qualifying Exam Report form and submits it to the ML Academic Professional.

**Student Guidelines:**

- If the student has a thesis advisor, the thesis advisor can be on the committee.
- At the beginning of the semester, the student should make one-on-one contact with each member of the committee to make the expectations clear.

- The written report should be presented to the committee at least one week before the oral exam.
- Ninety (90) minutes should be allocated for the exam. The talk should be no longer than 25 minutes; the rest of the time is meant for question and answer.

**Faculty committee guidelines:**

- Each faculty member should make their expectations about the content of both the written report and the oral exam clear at the beginning of the semester.
- Each committee member should come to the oral exam prepared with at least one technical question that tests the student's depth of knowledge about the artifacts on the syllabus.
- The evaluation forms should be filled out independently, and away from the site of the exam. Private discussion amongst committee members is fine, but the evaluation form should be completed and returned to the academic professional (by email or in person) individually. This allows each faculty member to provide an independent evaluation.
- If a student receives a 'fail' from any committee member, their case will be discussed at a year-end FAC meeting along with the faculty on the qualifying exam committee.

# V. PhD Proposal

The PhD proposal consists of short document describing the student's thesis, and a presentation to a proposal committee chosen by the student and approved by the ML FAC. All coursework must be completed before the proposal can be scheduled. The proposal must be at least six months before the PhD defense.

The proposal document should be solely authored by the student, and should be no longer than 30 pages. The proposal should describe the context of the research, including relevant references to the literature, the work that has been completed, and the research left to be done. It should be made clear throughout what the original research contributions of the thesis will be.

When the document is prepared, the student forms a Proposal Committee consisting of three faculty associated with the ML program, one of which is the thesis advisor. Once the committee approves the written document, the student submits the proposal, a letter describing the composition of the committee, and the Coursework Completion form to the Academic Professional.

After the committee and coursework are approved, the student schedules the proposal presentation. The proposal is announced through the academic office of the student's home school; this announcement must be made at least one week prior to the proposal presentation.

The proposal presentation consists of a short talk by the student (less than one hour), and questioning from the committee. Both the proposal document and the presentation are evaluated by the committee. 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.

After a successful proposal, the evaluation forms are submitted to the ML Academic Professional, who will sign Request for Admission to PhD Candidacy form, and communicate it to the Georgia Tech Graduate Studies office.

# VI. PhD Dissertation and Final Defense

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.

The following steps are taken to fulfill the PhD thesis requirement:

1. The student writes a complete draft of their thesis. It is expected that the thesis advisor will read and provide feedback during this process. Institute guidelines and requirements for formatting the document can be found in the *Georgia Institute of Technology Graduate Thesis/Dissertation Guidelines and Procedures*, available at: http://www.gradadmiss.gatech.edu/theses-dissertations

2. The student forms their PhD thesis committee. The PhD thesis committee consists of five faculty members comprising
    a. Three ML PhD faculty, including the student's advisor. These faculty are the reading committee.
    b. One faculty member external to the ML PhD program, and one additional member of the Georgia Tech faculty. The external member can be any academic faculty not affiliated with the ML PhD committee, including faculty at other universities. External non-academic faculty that are research active can also be used as external members; in this case, a CV must be provided for approval by the ML FAC.

c. One additional Georgia Tech faculty member. This additional member can come from inside or outside the ML PhD program.

3. The student submits the thesis to the reading committee, giving them a review period of at least 14 days.

4. After the review period, the reading committee can either recommend the thesis for defense, or send it back to the student for revisions. To recommend for defense, the reading committee signs the Reading Committee Report along with the thesis advisor.

5. The student, in conjunction with the thesis committee, schedules a time and location for the oral defense. This is communicated officially through the Final Defense Memo. The thesis advisor submits the Reading Committee Report and the Final Defense Memo to both the ML FAC and the academic office of the student's home school. The FAC will then approve the thesis committee, or ask the student to add/remove members.

   The Reading Committee Report and the Final Defense Memo must be signed and delivered to the FAC and the home school academic office at least 14 days in advance of the defense.

6. The home school academic office will announce the oral defense through the usual official and unofficial channels.

7. The student defends their dissertation at the oral defense. The defense is open to the public, and consists of four segments: a presentation by the student, questions from the audience at large, questions from the thesis committee, and a private deliberation and discussion period by the committee. After a successful defense, the each committee member fills out a Dissertation Evaluation Form and signs the Institute Certificate of Thesis Approval.

8. The student submits the Dissertation Evaluation Forms to the ML FAC and the Certificate of Thesis Approval to the academic office of their home school. The home school academic office signs the Certificate of Thesis Approval and submits it to the Georgia Tech Graduate Studies - VPR office.

9. The student submits their thesis electronically to the Georgia Tech Graduate Studies - VPR office. Information on how this is is accomplished can be found in the *Georgia Institute of Technology Graduate Thesis/Dissertation Guidelines and Procedures*, available at:
   http://www.gradadmiss.gatech.edu/theses-dissertations

# VII. Administration and Governance

## Program Director

The director is in charge of the overall administration and policy directions for the program. They also chair the Faculty Advisory Committee, and approve transfers into the ML PhD program from students enrolled in other PhD programs at Georgia Tech.

## Faculty Advisory Committee

The FAC consists of one faculty member from each school involved with the program, and is chaired by the Program Director. Duties of the FAC include

- Admissions. Each FAC member will coordinate with their school on admissions offers into the ML PhD program. The FAC member coordinates the review of applications within the school with the relevant faculty and corresponding associate chair within the school, then supplies a list of candidates for the FAC to approve.
- Committee approval. The FAC will have ultimate approval over ML PhD student thesis proposal and thesis defense committees.
- Course transfer credit. The FAC will have ultimate approval over course credit transfer from other universities. (The process is coordinated by the Academic Professional, as detailed above.)
- Curriculum revisions. The FAC will periodically review the curriculum. Revisions to the requirements will be submitted to the program Academic Professional, who will shepherd them through approval by the Institute Graduate Curriculum Committee.

## Affiliated Faculty Members

Faculty members associated with the program can serve as the thesis advisor for any student in the program, no matter their academic unit. Associated faculty must have an appointment in one of the participating home schools at Georgia Tech.

Associated faculty members are also expected to serve on ML PhD qualifying exam, proposal, and PhD defense committees.

A list of faculty members currently associated with the program will be kept at [http://ml.gatech.edu/phd/affiliated-faculty](http://ml.gatech.edu/phd/affiliated-faculty)

To become affiliated with the ML PhD program, email the Program Director.  Affiliates will be approved by the  FAC.

# VIII. Current Personnel

## Director

Justin Romberg, jrom@ece.gatech.edu

## Academic Professional

Rebecca Wilson, rebecca.wilson@cc.gatech.edu

## Academic Advisor

Stephanie Niebuhr, stephanie.niebuhr@cc.gatech.edu

## Operations and Support

Kyla Hanson, khanson@cc.gatech.edu

## Faculty Advisory Committee

Each school affiliated with the ML PhD program has a member on the FAC.

Aerospace Engineering (AE): Evangelos Theodorou, evangelos.theodorou@gatech.edu
Biomedical Engineering (BME): May Wang, maywang@gatech.edu
Computational Science and Engineering (CSE): Polo Chau, polo@gatech.edu
Computer Science (CS): Jacob Abernethy, prof@gatech.edu
Electrical and Computer Engineering (ECE): David Anderson, dva@ece.gatech.edu
Industrial Systems Engineering (ISyE): Yao Xie, yao.xie@isye.gatech.edu
Interactive Computing (IC): Zsolt Kira, zkira@gatech.edu
Mathematics (MATH): Vladimir Koltchinskii, vlad@math.gatech.edu
Chair (Program Director): Justin Romberg, jrom@ece.gatech.edu

# Appendix A: Core Courses

The Machine Learning PhD core curriculum consists of courses from five different areas: Mathematical Foundations, Intermediate Statistics, ML Theory and Methods, Data Models, and Optimization. With the exception of the Foundations and Data Models course, the requirements can be met with different courses in different schools.

**Mathematical Foundations of Machine Learning.** This required course is the gateway into the program, and will cover 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.

ECE/ISYE/CS/BMED 8843/8843/8803/8813, Mathematical Foundations of Machine Learning

**Intermediate Statistics**. The purpose of this requirement is to expose students to the main concepts in mathematical statistics. It can be met through any one of the three courses listed below. While these courses emphasize different material, they are all centered on mathematical analysis of fundamental problems in statistics.

ISYE 6412, Theoretical Statistics

ECE 7251, Signal Detection and Estimation

MATH 6262, Statistical Estimation

**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 7616, Pattern Recognition

CSE/ISYE 6740, Computational Data Analysis

ECE 6254, Statistical Machine Learning

ECE 6273, Methods of Pattern Recognition with Applications to Voice

**Data Models: Probabilistic Graphical Models and ML in High Dimensions**. This course is cross-listed between the College of Computing (CS) and schools of CSE, ECE, and ISYE. The course will provide students with an introduction to the theory and practice of graphical models, one of the most dominant frameworks in machine learning and artificial intelligence.

ECE/ISYE/CS 8803, Probabilistic Graphical Models in Machine Learning

**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 6669, Deterministic Optimization

ISYE 7683, Advanced Nonlinear Programming

# Appendix B: Elective Courses

In addition to meeting the five 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.

Up to two elective courses can be Special Topics courses as described in the Coursework Requirements section above.

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

AE 6505, Kalman Filtering

BMED 6700, Biostatistics

ECE 6558, Stochastic Systems

ECE 6601, Random Processes

ECE 6605, Information Theory

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

MATH 6263, Testing Statistical Hypotheses

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

**Advanced Theory**:

To build a deeper understanding of foundations of ML.

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 7545, Machine Learning Theory

ECE 6283, Harmonic Analysis and Signal Processing

ECE 6555, Linear Estimation

ISYE 7682, Convexity

MATH 6112, Advanced Linear Algebra

MATH 6221, Advanced Classical Probability Theory

MATH 6580, Introduction to Hilbert Space

MATH 7338, Functional Analysis

MATH 7586, Tensor Analysis

MATH 88XX, Special Topics: Mathematical Foundations of Learning Theory

MATH 88XX, Special Topics: High Dimensional Probability and Statistics

**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 8813BHI, Biomedical and Health Informatics

BMED 8813MHI, mHealth Informatics

BMED 8813MLB, Machine Learning in Biomedicine

BMED 8823ALG, OMICS Data and Bioinformatics Algorithms

CS 6440, Introduction to Health Informatics

CS 6465, Computational Journalism

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 7636, Computational Perception

CS 7646, Machine Learning for Trading

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 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

HS 6000, Healthcare Delivery

MATH 6759, Stochastic Processes in Finance

MATH 6783, Financial Data Analysis

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

CS 6505, Computability and 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

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

**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