



PhD in Machine Learning

Course Requirements

The curriculum for the Machine Learning Ph.D. is built on a foundation of **five core** courses and **three electives** (plus the **Data Analysis Project** requirement). These five courses also comprise the required courses for the MS degree. Together with the Data Analysis Project requirement, these should be completed for MS degree within 3 years but many students do it in 2 or 2.5 years.

Note: If a student has taken some of the MLD core courses before joining the MLD PhD program, and has not counted these courses toward any other PhD-level degree, the student may count these courses toward the MLD PhD. In this situation the student will need to take fewer than 5 new core courses to graduate. A student must always take at least three elective courses while registered in the MLD PhD program, irrespective of any courses taken before joining the PhD program. Students who took 10-701 in Spring 2014 or earlier can use it as a core course, even if they weren't part of the MLD PhD program at the time they took 10-701.

A typical full-time, graduate course load during the first two years consists each term of two classes (at 12 graduate units per class) plus 24 units of advanced research. Thus, during the first two years, a student has the opportunity to take several elective classes in addition to the five required courses.

The ML curriculum joins courses with a Computer Science main theme and those with a Probability and Statistics main theme. These may be grouped, as follows:

In CS, relevant sub-fields include: Databases; Machine Learning, Data Mining and algorithms applications in areas such as Robotics, Information Retrieval and AI.

In Statistics (including Philosophy), the sub-fields include: Statistical modeling (e.g., hierarchical and times series); Bayes' Nets, Causation, and experimental design. The curriculum is based on core academic courses on Intermediate Statistics, Machine Learning, Statistical Machine Learning, Multimedia Databases, and Algorithms.

These core courses together provide a foundation in machine learning, statistics, probability, and algorithms.

- 10-715 Advanced Introduction to Machine Learning
- 10-702 Statistical Machine Learning
- 10-705 Intermediate Statistics

Plus any two of the following courses:

- 10-708 Probabilistic Graphical Models
- 10-725 Convex Optimization
- 15-826 Multimedia Databases and Data Mining
- 15-750 Algorithms or 15-853 Algorithms in the Real World

Possible electives

Data Analysis Project Requirement, in the second year, which serves in lieu of an MS thesis.

Here is a typical schedule for the first two years of study.

| FALL - 1st Year | SPRING - 1st Year |
|--|-------------------------------------|
| 10-715 Advanced Introduction to Machine Learning | 10-702 Statistical Machine Learning |
| 10-705 Intermediate Statistics | Core course or elective |
| 10-920 Research | 10-920 Research |
| FALL - 2nd Year | SPRING - 2nd Year |

| | |
|-------------------------|------------------------------------|
| Core course or Elective | Research for Data Analysis Project |
| Elective | Elective |
| 10-920 Research | 10-920 Research |
| | |

The Data Analysis Project requirement:

During the second year a Ph.D. student is required to demonstrate data analysis and machine learning skills in the context of a focused project. The [Data Analysis Project](#) may be carried out either at Carnegie Mellon or at a sponsoring corporate institution under the joint supervision of the sponsor and a ML faculty. It will be concluded by a written report (in lieu of a Masters Thesis) in which the student demonstrates an ability to approach data mining problems in a way that cuts across existing disciplinary boundaries. The requirement includes a presentation in the ML Journal Club and also the submission of a DAP Paper. Passing this requirement will be the judgment of the DAP committee.

Student must form an official "DAP committee" of three faculty to evaluate the document. The committee will consist of the advisor, the Journal club instructor(s), and one other faculty member selected by the student. The third member is typically someone with an interest in the analysis of the data set, and does not have to be an expert in ML or part of the student's thesis committee. The student should form the committee as early as possible during the DAP research process, and inform Diane of who the members are. Two faculty from the committee are required to attend the presentation.

The Third Year

During the third year, a Ph.D. student completes the elective course requirements. One of these three electives is taken from the offerings in Statistics. The other two advanced electives, chosen in consultation with the student's advisor, form a concentration in one of the allied disciplines with SCS, Biology, Philosophy, or GSIA. For those candidates seeking an academic position after completing the ML Ph.D. degree, the thoughtful selection of these three elective courses is particularly important. As in each of the first two years, coursework is supplemented by 24 units/term of research.

The Fourth Year and Beyond

A Ph.D. student typically presents a thesis proposal no later than the start of the fourth year, and then spends the fourth and sometimes fifth year working on their thesis research.

Research

[Responsible Conduct of Research Training](#)

Students must complete training for NSF and NIH grants. A copy of the certificate must be given to the MLD Business Manager, Colleen Everett.

It is expected that all Ph.D. students engage in active research from their first semester. Moreover, advisor selection occurs in the first month of entering the Ph.D. program, with the option to change at a later time. Roughly half of a student's time should be allocated to research and lab work, and half to courses until these are completed.

This [Research page](#) is a list of some of the projects for which ML faculty may be interested in recruiting students. Within each project there can be lines of research which range in size from a semester's work to an entire thesis (or beyond). So, this page is intended as a resource for students looking for a thesis advisor, for a Data Analysis project, or to collaborate for any other reason.



Relevant Courses offered for Spring 2015

| | |
|---|--|
| CORE COURSES: 10-701 Machine Learning 10-702 Statistical Machine Learning 10-708 Probabilistic Graphical Models 10-725 Convex Optimization 15-750 Graduate Algorithms | ML JOURNAL CLUB 10-915, ML PhD students must register if you plan to satisfy either the Speaking Skills or Data Analysis Project (DAP) requirements. Incoming PhD students should register in Spring. |
| Research Courses: 10-920 Grad Reading & Research 10-930 Dissertation Research | Students should register for 10-920 R & R until they propose. After you propose, register for Dissertation Research. |

Suggested Research Depth Electives:

For ML PhD students, two advanced electives, chosen in consultation with the student's advisor, form a research depth concentration. Approved Research Depth electives are listed below.
 Full list of [Approved Electives](#)

| | |
|---|---|
| Research Depth in AI: 10-708 Probabilistic Graphical Models 10-725 Convex Optimization 15-780 Graduate Artificial Intelligence | Research Depth in CNBC Track: 03-762 Systems Neuroscience 36-759 Statistical Models of the Brain Applicable Courses from the University of Pittsburgh http://www.cmu.edu/hub/registration/undergraduates/cross/outgoing.html |
| Research Depth in Algorithms & Theory 10-725 Convex Optimization 15-859 (E) Special Topics in Theory: Advanced Algorithms | Research Depth in NLP or Text Analysis: 10-708 Probabilistic Graphical Models 11-741 Machine Learning for Text Mining 11-745 Adv. Statistical Learning Seminar (6) 11-761 Language and Statistics |
| Research Depth in Computational Biology: 02-710 Computational Genomics 10-708 Probabilistic Graphical Models | |
| Research Depth in Computer Vision: 10-725 Convex Optimization 16-720 Computer Vision 16-822 Geometry-Based Methods in Vision 16-824 Learning Based Methods in Vision | |

Suggested Electives from Statistics

(For ML PhD Students, one elective or courses combined for a total of 12 units must be chosen from Statistics)
 36-728 Time Series
 36-752 Adv. Probability Overview

Other electives from SCS approved but don't have a category:

10-704 Information Processing & Learning
 11-755 Machine Learning for Signal Processing
 18-755 Networks in the Real World



Electives

ML PhD students:

One elective or courses combined for a total of 12 units must be chosen from Statistics. You must also have a research depth of 24 units.

Both Statistics and Tepper offer "mini" half-term courses. Two such "mini" courses are equivalent to one (12 unit) graduate course.

Suggested Electives from Statistics

36-703 Intermediate Probability
36-707 Regression Analysis
36-708 Experimental Design, 6 units, A4 mini
36-709 Linear Models, 6 units, A3 mini
36-720 Discrete Multivariate Analysis, 6 units, A2 mini
36-722 Applied Continuous Multivariate Analysis, 6 units, A4 mini
36-724 Applied Bayesian Methods, 6 units, A3 mini
36-728 Time Series
36-737 Applied Multilevel & Hierarchical Models, 6 units, A2 mini
36-752 Adv. Probability Overview
36-754 Adv. Probability
36-755 Advanced Statistical Theory I
36-759 Statistical Models of the Brain
36-781 Advanced Statistical Methods I mini, 6 units, A1 mini
36-782 Advanced Statistical Methods II mini, 6 units, A2 mini
36-786 Bayesian Theoretical Statistics 1, 6 units, A1 mini
36-787 Bayesian Theoretical Statistics 2, 6 units, A2 mini
36-825 Statistics Journal Club
36-835 Foundations of Statistics Seminar
36-900 Selected Topics of the Contemporary Frontiers of High Dimensional Inference
36-905 Seminar on Latent Variable Models, 6 units

Suggested Depth Requirement Electives from SCS

AI:

10-725 Convex Optimization
10-708 Probabilistic Graphical Models
15-780 Graduate Artificial Intelligence
15-857 Analytical Performance Modeling & Design of Computer Systems
15-859 (M) Randomized Algorithms
15-887 Planning, Execution, and Learning

Algorithms & Theory:

10-725 Convex Optimization
15-855 Computational Complexity Theory
15-857 Analytical Performance Modeling & Design of Computer Systems
15-859 Special Topics in Theory - check for appropriate topics
15-859 (B) Machine Learning Theory
15-859 (E) Special Topics in Theory: Advanced Algorithms
16-811 Mathematical Fundamentals for Robotics
21-801 Adv. Topics Discrete Math (Random Graphs)

Computational Biology:

02-750 Automation of Biological Research: Robotics & Machine Learning
02-710 Computational Genomics
10-708 Probabilistic Graphical Models

Computer Vision:

10-725 Convex Optimization
16-720 Computer Vision
16-822 Geometry-Based Methods in Vision

Databases:

15-823 Advanced Database Topics

NLP or Text Analysis:

10-708 Probabilistic Graphical Models

10-710/11-763 Structured Prediction for Language & Other Discrete Data

10-802/11-772 Analysis of Social Media

11-711 Algorithms for NLP

11-741 Machine Learning for Text Mining

11-744 Experimental Information Retrieval

11-745 Advanced Statistical Learning Seminar (6 units)

11-761 Language and Statistics

11-762 Language and Statistics II

11-773 Text-Driven Forecasting

Robotics:

02-750 Automation of Biological Research: Robotics & Machine Learning

15-887 Planning, Execution, and Learning

16-811 Mathematical Fundamentals for Robotics

16-831 Statistical Techniques in Robotics

16-899C Adaptive Control and Reinforcement Learning

Other electives from SCS approved but don't have a Depth Requirement category:

10-704 Information Processing & Learning

11-745 Adv. Statistical Learning Seminar (6 units)

11-755 Machine Learning for Signal Processing

15-830 Computational Methods in Sustainable Energy

18-755 Networks in the Real World

Suggested Depth Requirement Electives for CNBC Track

03-762 Advanced Cellular Neuroscience

03-763 Systems Neuroscience

15-883 Computational Models of Neural Systems

36-759 Statistical Models of the Brain

85-719 Introduction to Parallel Distributed Processing

85-765 Cognitive Neuroscience (12 units)

Applicable Courses from the University of Pittsburgh (Please see

<http://www.cmu.edu/hub/registration/undergraduates/cross/outgoing.html>)

NROSCI 2100 Cellular & Molecular Neurobiology

NROSCI 2102/2103 Systems Neurobiology

MATH 3375 Computational Neuroscience

Suggested Concentration Electives from School of Public Policy & Management:

10-830/90-904 Research Seminar in Machine Learning & Policy, 6 units, A3 mini

10-831/90-921 Special topics in Machine Learning & Policy, 6 units, A4 mini

Suggested Concentration Electives from Tepper (Must follow Tepper special registration rules)

Finance Track:

45-814 Options

46-926 Linear Models/Equity Portfolio Management

46-929 Financial Time Series Analysis

46-944 Stochastic Calc Fin 1

Marketing Track:

15-892 Foundations of Electronic Marketplaces (CS course)

47-800 Intermediate Microeconomic Analysis

47-741 Seminar in Marketing I

47-742 Seminar in Marketing II

47-743 Seminar in Marketing III

47-744 Analytical and Structural Marketing Models

45-821 Marketing with Electronic & Social Media

45-824 Database Marketing

Information Systems Track:

47-800 Intermediate Microeconomic Analysis

45-870 Management of Information Systems

45-871 Information Strategy, Systems and Economics

47-951 Seminar in Information Systems I

47-952 Seminar in Information Systems II

47-953 Seminar in Information Systems III

47-954 Seminar in Information Systems IV

NOTE: Tepper courses are on the mini-system. 45-* and 46-* are Master level courses and the 47-

*** are PhD level courses Suggested**

Concentration Electives from Philosophy

80-605 Rational Choice

80-614 Logic in Artificial Intelligence

80-616 Probability and Artificial Intelligence

80-621 Causality in the Social Sciences