Mathematical Foundations of Machine Learning

Instructor

Justin Romberg, School of Electrical and Computer Engineering
jrom@ece.gatech.edu

Goal

The purpose of this course is to provide first-year ML PhD students with a solid mathematical background for two of the pillars of modern data science: linear algebra and applied probability.

Recommended Textbook

Extensive course notes will be provided. Almost all of the material is covered in the book Theodoridis S. (2015) Machine Learning: A Bayesian and Optimization Perspective (Academic Press).

Grading

30% Homework, 10 assignment
20% Midterm Exam 1
20% Midterm Exam 2
25% Final Exam
5% Participation

Prerequisites

There are no formal course prerequisites. Students should have had exposure to the basics of linear algebra, probability, and statistics. Student should also have basic programming skills.

Learning Objectives

As part of this course, students …

1. Formulate machine learning algorithms using the language of linear algebra and probability.

2. Analyze and compute the solutions to least-squares problems in the context of regression.
3. Become familiar with basic computational methods from optimization.
4. Quantify the influence of one random variable on another using tools from conditional probability.
5. Gain exposure to the fundamental techniques in parameter estimation and hypothesis testing.

Course Educational Outcomes

Upon successful completion of this course, students should be able to …

1. Analyze data in different domains defined by different basis expansions or kernels.
2. Solve regression problems on both small and large scales.
3. Describe and analyze the solutions to least-squares problems using the singular value decomposition.
4. Solve unconstrained optimization programs using gradient descent and Newton's method
5. Compute estimates of parameters from data using maximum likelihood.
6. Compute estimates of parameters from data using Bayes rule.
7. Describe the role of prior knowledge in solving parameter estimation problems.
8. Describe the bias-variance trade-off in estimation problems.
9. Describe the role of regularization in controlling bias and variance.
10. Design optimal procedures for hypothesis testing.
11. Apply algorithms for optimization to statistical estimation problems.
12. Model the dependencies in the joint distribution of data using a graphical model.
13. Use principal components analysis in service of dimensionality reduction.
14. Describe the role that low-rank approximations of matrices play in machine learning algorithms.
15. Describe the role that sparsity plays in modeling and processing high dimensional data.

Academic Integrity

Academic dishonesty will not be tolerated. This includes cheating, lying about course matters, plagiarism, or helping others commit a violation of the Honor Code. Plagiarism includes reproducing the words of others without both the use of quotation marks and citation. Students are reminded of the obligations and expectations associated with the Georgia Tech Academic Honor Code and Student Code of Conduct, available online at www.honor.gatech.edu.

Learning Accommodations

If needed, we will make classroom accommodations for students with documented disabilities. These accommodations must be arranged in advance and in accordance with the Office of Disability Services (http://disabilityservices.gatech.edu).

Outline of Topics

1. Vector space basics
    a. linear vector spaces, linear independence
    b. norms and inner products
    c. bases and orthobases
    d. examples: B-splines, cosines/Fourier, radial basis functions, etc.
    e. linear approximation (closest point in a subspace, least-squares I)

2. Linear Estimation
    a. Examples: classical regression/recovering a function from point samples, imaging, etc.
    b. The Singular Value Decomposition (SVD)
    c. least-squares solutions and the pseudo-inverse
    d. stable inversion and regularization
    e. kernels
        i. Mercer's theorem
        ii. RKHS, representer theorem
    f. Computing least-squares solutions
        i. LU and Cholesky factorizations
        ii. steepest descent and conjugate gradients
        iii. low rank updates for online least-squares

3. Probability and random vectors
    a. Review: joint pdfs, random vectors, conditional probability, Bayes rule
    b. Multivariate Gaussian
    c. random processes

4. Statistical estimation
    a. Best linear unbiased estimator and weighted least-squares
    b. Maximum likelihood
        i. unconstrained optimization
        ii. Stochastic gradient descent
    c. Bayesian estimation
        i. Bayesian interpretation of least squares, regularization
    d. Hypothesis testing

5. Modeling
    a. Probabilistic graphical models
        i. Hidden markov models
        ii. Gaussian graphical models
    b. Subspace models
        i. principal components analysis
        ii. low rank approximation (Eckart-Young theorem)
        iii. structured matrix factorization (e.g. NNMF, dictionary learning)
    c. Sparsity