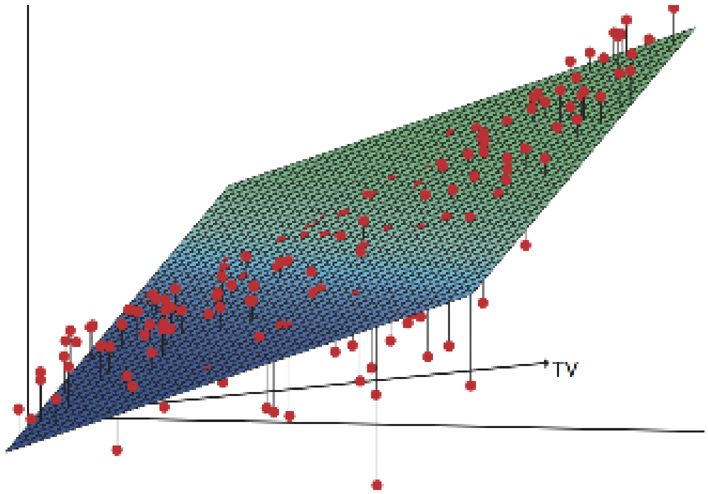


CS 189/289A Introduction to Machine Learning

[Jonathan Shewchuk](#)
jrs@cory.eecs.berkeley.edu
(Please send email only when Piazza is not feasible.)

Spring 2016
Mondays and Wednesday, 6:30–8:00 pm
2050 Valley Life Sciences Building

Instructions for the **CS 289A Project** are now available. Please find project partners and submit your proposal by **Monday, April 4**.

This class introduces algorithms for *learning*, which constitute an important part of Artificial Intelligence.

Topics include (not a complete list!)

- classification,
- regression,
- density estimation,
- clustering, and
- dimensionality reduction.

Best Links

- See the schedule of class and discussion section times and rooms.
- Access the CS 189/289A Piazza discussion group.
- If you want an instructional account, you can get one online. No more paper forms. Go to the same link if you forget your password or account name.

Prerequisites

- Math 53 (or another vector calculus course),
- Math 54 (or another linear algebra course),
- CS 70 (or other courses covering discrete math and probability), and
- CS 188 (Artificial Intelligence).

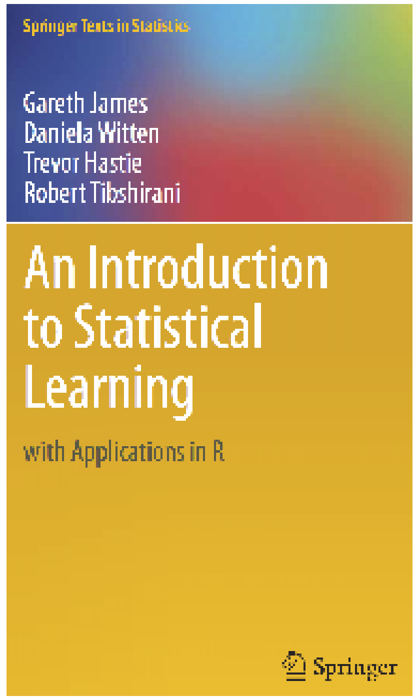
You should take the first three prerequisites quite seriously: if you don't have them, I strongly recommend not taking CS 189. CS 188 is not quite as essential, but still a very good idea.

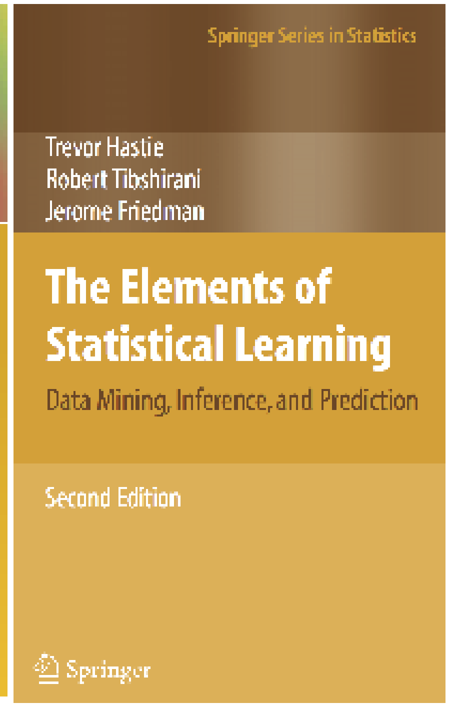
If you want to brush up on prerequisite material, Stanford's machine learning class provides nice reviews of linear algebra and probability theory. Other suggestions for review material appear in this Piazza post.

Textbooks

Both textbooks for this class are available free online. Hardcover and Kindle/eTextbook versions are also available.

- Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani, *An Introduction to Statistical Learning with Applications in R*, Springer, New York, 2013. ISBN # 978-1-4614-7137-0. See Amazon for hardcover or eTextbook.
- Trevor Hastie, Robert Tibshirani, and Jerome Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, second edition, Springer, 2008. See Amazon for hardcover or Kindle.





Homework

You have a **total** of **5** slip days that you can apply to your semester's homework. We will simply not award points for any late homework you submit that would bring your total slip days over five.

Homework 1 is due February 10.

Homework 2 is due February 18.

Homework 3 is due March 3.

Homework 4 is due March 31.

The CS 289A Project has a proposal due **Monday, April 4**. The video and final report are due **Friday, May 6**.

Previous midterms (including this semester's) are available: Spring 2013, Spring 2014, Spring 2015, Fall 2015, Spring 2016.

Lectures

Lecture topics and readings will be added here as I figure them out.

Lecture 1 (January 20): Introduction. Read ESL, Chapter 1. My lecture notes (text). The screencast.

Lecture 2 (January 25): Linear classifiers. Perceptrons. Read parts of the Wikipedia Perceptron page. Optional: Read ESL, Section 4.5–4.5.1. My lecture notes (text). The screencast.

Lecture 3 (January 27): Gradient descent and the perceptron learning algorithm. The maximum margin classifier, aka hard-margin support vector machine (SVM). Read ISL, Section 9–9.1. My lecture notes (text). The screencast.

Lecture 4 (February 1): The support vector classifier, aka soft-margin support vector machine (SVM). Features and nonlinear decision boundaries. Read ESL, Section 12.2 up to and including the first paragraph of 12.2.1. My lecture notes (text). The screencast.

Lecture 5 (February 3): Machine learning abstractions (application/data, model, optimization problem, optimization algorithm). Common types of optimization problems: unconstrained, constrained (with equality constraints), linear programs, quadratic programs, convex programs. Optional: Read (selectively) the Wikipedia page on mathematical optimization. My lecture notes (text). The screencast.

Lecture 6 (February 8): Decision theory: the Bayes decision rule and optimal risk. Generative and discriminative models. Read ISL, Section 4.4.1. My lecture notes (text). The screencast.

Lecture 7 (February 10): Gaussian discriminant analysis, including quadratic discriminant analysis (QDA) and linear discriminant analysis (LDA). Maximum likelihood estimation (MLE) of the parameters of a statistical model. Read ISL, Section 4.4. Optional: Read (selectively) the Wikipedia page on maximum likelihood. My lecture notes (text). The screencast.

February 15 is Presidents' Day.

Lecture 8 (February 17): Eigenvectors, eigenvalues, and the eigendecomposition. The Spectral Theorem for symmetric real matrices. The quadratic form and ellipsoidal isosurfaces as an intuitive way of understanding symmetric matrices. Application to anisotropic normal distributions (aka Gaussians). Read Chuong Do's notes on the multivariate Gaussian distribution. My lecture notes (text). The screencast.

Lecture 9 (February 22): Anisotropic normal distributions (aka Gaussians). QDA and LDA revisited for anisotropic Gaussians. Read ISL, Sections 4.4 and 4.5. My lecture notes (text). The screencast.

Lecture 10 (February 24): Regression: fitting curves to data. The 3-choice menu of regression function + loss function + cost function. Least-squares linear regression as quadratic minimization and as orthogonal projection onto the column space. The design matrix, the normal equations, the pseudoinverse, and the hat matrix (projection matrix). Logistic regression; how to compute it with gradient ascent. Read ISL, Sections 4-4.3. My lecture notes (text). The screencast.

Lecture 11 (February 29): My Mom's 18th birthday (not kidding). Newton's method and its application to logistic regression. LDA vs. logistic regression: advantages and disadvantages. ROC curves. Weighted least-squares regression. Least-squares polynomial regression. Read ISL, Sections 4.4.3, 7.1, 9.3.3; ESL, Section 4.4.1. My lecture notes (text). The screencast. Happy birthday, Mom!

Lecture 12 (March 2): Statistical justifications for regression. The empirical distribution and empirical risk. How the principle of maximum likelihood motivates the cost functions for least-squares linear regression and logistic regression. The bias-variance decomposition; its relationship to underfitting and overfitting; its application to least-squares linear regression. Read ESL, Sections 2.5 and 2.9. Optional: Read the Wikipedia page on the bias-variance trade-off. My lecture notes (text). The screencast.

Lecture 13 (March 7): Ridge regression: penalized least-squares regression for reduced overfitting. How the principle of maximum *a posteriori* (MAP) motivates the penalty term (aka Tikhonov regularization). Kernels. Kernel ridge regression. The polynomial kernel. Read ISL, Sections 6.2-6.2.1 and ESL, Sections 12.3-12.3.2. Optional: This CrossValidated page on ridge regression is pretty interesting. My lecture notes (text). The screencast.

Lecture 14 (March 9): Kernel perceptrons. Kernel logistic regression. The Gaussian kernel. Subset selection. Lasso: penalized least-squares regression for reduced overfitting and subset selection. Read ISL, Sections 6-6.1.2 and the last part of 6.1.3 on validation; and ESL, Sections 3.4-3.4.3. Optional: Read ISL, Section 9.3.2 if you're curious about kernel SVM. My lecture notes (text). The screencast.

Lecture 15 (March 14): Decision trees; algorithms for building them. Entropy and information gain. Read ISL, Sections 8-8.1. My lecture notes (text). The screencast.

The Midterm takes place in class on **Wednesday, March 16**. You are permitted one "cheat sheet" of letter-sized (8½" × 11") paper.

March 21–25 is Spring Recess.

The Final Exam takes place on **Friday, May 13, 3–6 PM**. (CS 189 is in exam group 19.)

Discussion Sections and Teaching Assistants

Sections begin to meet on January 28. Sections 102 and 103 are cancelled.

Thursday, 10 am	101	Tuomas	102 Latimer
Thursday, 1 pm	104 105	Rohan Brian	228 Dwinelle B56 Hildebrand
Thursday, 2 pm	106 107	Rohan Tuomas	105 Dwinelle 255 Dwinelle
Thursday, 3 pm	108 115	Marvin Daylen	179 Dwinelle 405 Soda
Thursday, 4 pm	110 111	Marvin Aldo	255 Dwinelle 254 Dwinelle
Thursday, 5 pm	116	Vashisht	83 Dwinelle

Friday, 9 am	112	Aldo	81 Evans
Friday, 10 am	113	Shaun	71 Evans
Friday, 11 am	114	Brian	85 Evans
Friday, 2 pm	109	Shaun	110 Barker

Your teaching assistants:
Rohan Chitnis, ronuchit@berkeley.edu
Tuomas Haarnoja, haarnoja@berkeley.edu
Brian Hou, brian.hou@berkeley.edu
Vashisht Madhavan, vashisht.madhavan@berkeley.edu
Aldo Pacchiano, pacchiano@berkeley.edu
Shaun Singh, sdsingh@berkeley.edu
Ajay Solanky, ajaysolanky@berkeley.edu
Daylen Yang, daylen@berkeley.edu

Marvin Zhang, zhangmarvin@berkeley.edu

Grading

- **40%** for homeworks.
- **20%** for the Midterm (Wednesday, March 16, in class).
- CS 189: **40%** for the Final Exam (Friday, May 13, 3–6 PM; CS 189 is in exam group 19.)
- CS 289A: **20%** for the Final Exam.
- CS 289A: **20%** for a project.

Supported in part by the National Science Foundation under Awards ACI-9875170, CMS-9980063, CCR-0204377, CCF-0430065, CCF-0635381, IIS-0915462, and CCF-1423560, in part by a gift from the Okawa Foundation, and in part by an Alfred P. Sloan Research Fellowship.