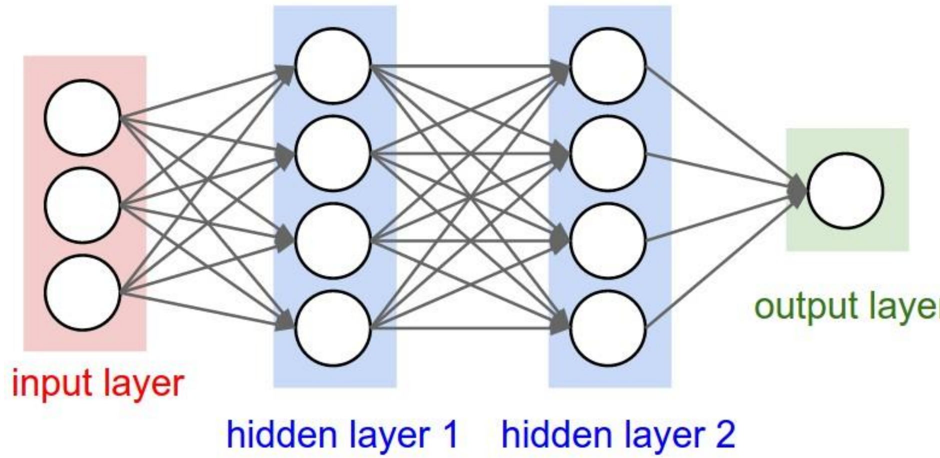
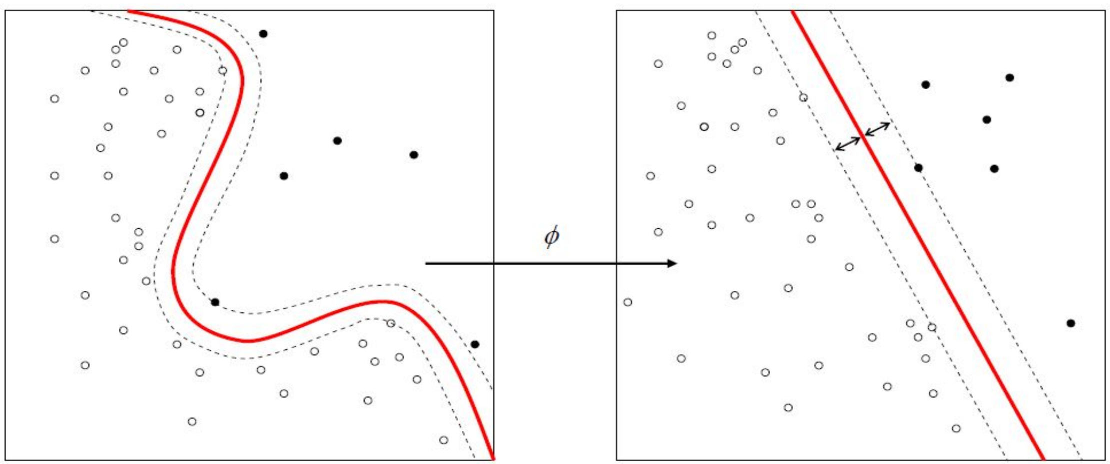
CS 5785/ORIE 5750/ECE 5414 Applied Machine Learning

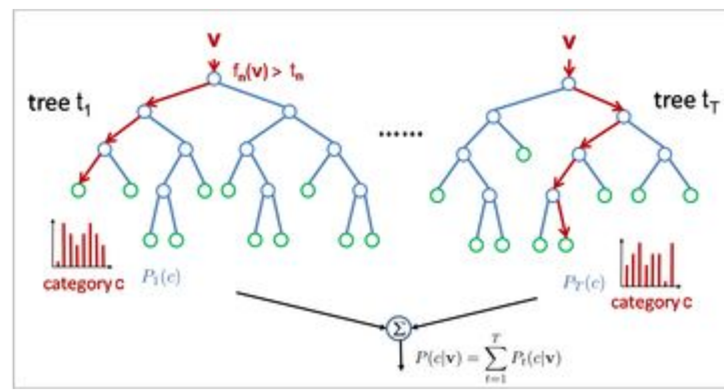
Fall 2019

Nathan Kallus

TAs: Yichun Hu, Xiaojie Mao







https://cs5785-2019.github.io/

Lecture 1

Course Logistics

https://cs5785-2019.github.io

Fa

Course Description

Learn and apply key concepts of modeling, analysis and validation from Machine Learning, Data Mining and Signal Processing to analyze and extract meaning from data. Implement algorithms and perform experiments on images, text, audio and mobile sensor measurements. Gain working knowledge of supervised and unsupervised techniques including classification, regression, clustering, feature selection, association rule mining and dimensionality reduction.

Prerequisites

CS 2800 or equivalent plus experience programming with Python or Matlab, or permission of the instructor.

Teaching staff

Instructor:

Prof. Nathan Kallus (www.nathankallus.com)

Office hours: Thursdays 12:20-13:20

TAs:

Yichun Hu, Xiaojie Mao

Office hours: Mondays 10–11AM, Tuesdays 1–2PM, Wednesdays 4–5PM

You?

Review Sessions

The first two TA office hours (Tuesday and Wednesday next week) will be (identical) review sessions, going over linear algebra, calculus, and probability.

Highly recommended!

Deep Learning Clinic

- Computer vision expert Jin Sun is holding a Deep Learning Clinic every Tuesday 9:30-10:45AM in 061
- https://cornelltech.github.io/deep-learning-clinic-2019-Fall/
- Very complementary to AML
- Covers neural net engineering
- Highly recommended!
  - Even if you do not have time/preparation, you can just attend the lectures and absorb the material

Communications

Everyone must join the Slack:

https://cs5785-2019fa-kallus.slack.com/

Course Requirements and Grading

- **Grade Breakdown:** Your grade will be determined by the assignments (40%), one prelim (20%), a final exam (30%), and participation including in-class quizzes (10%).

- **Homework:** There will be four assignments and an "assignment 0" for environment setup. Each assignment will have a due date for completion. Half of the points of the lowest-scoring assignment will count as extra credit, meaning the points received for homeworks 1, 2, 3, and 4 is calculated as (sum of scores) / 3.5.

- **Late Policy:** Each student has a total of **one** slip day that may be used without penalty.

- **External Code:** Unless otherwise specified, you are allowed to use well known libraries such as *scikit-learn, scikit-image, numpy, scipy,* etc. in the assignments. Any reference or copy of public code repositories should be properly cited in your submission (examples include *Github, Wikipedia, Blogs*). In some assignment cases, you are NOT allowed to use any of the libraries above, please refer to individual HW instructions for more details.

- **Collaboration:** You are encouraged (but not required) to work in groups of no more than 2 students on each assignment. Please indicate the name of your collaborator at the top of each assignment and cite any references you used (including articles, books, code, websites, and personal communications). If you're not sure whether to cite a source, err on the side of caution and cite it. You may submit just one writeup for the group. Remember not to plagiarize: all solutions must be written by members of the group.

- **Quizzes:** There will be surprise in-class quizzes to make sure you attend and pay attention to the class.

- **Prelim: October 22** in class. The exam is closed book but you are allowed to bring one sheet of written notes (Letter size, two-sided). You are allowed to use a calculator.

- **Final Exam: December 2 through December 9.** The final exam is take-home, open-internet, but must be done by your own group with thorough citations of all references used.


PLEASE BE ON TIME
memegenerator.net

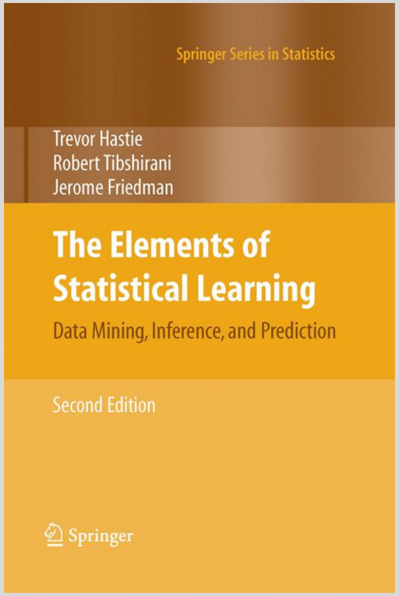
How to do well in class?

- Go to the review sessions next week!
- Attend lectures
- Be on time
- Sit up front
- Ask questions
- Come to office hours
- Start on HW early
- Work with your partner
- Talk with your colleagues
- Read the additional reading

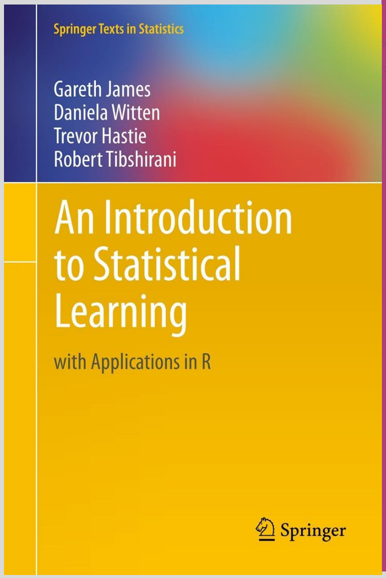
TA announcements about HW

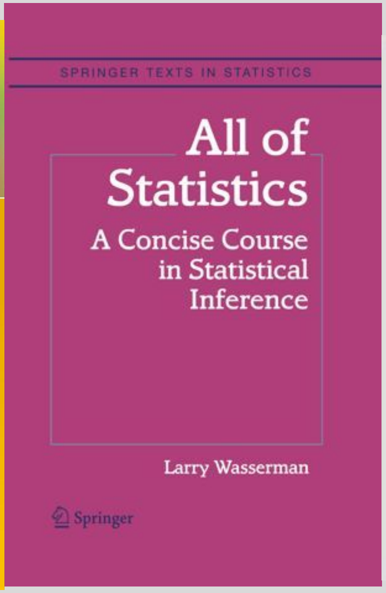
Required:

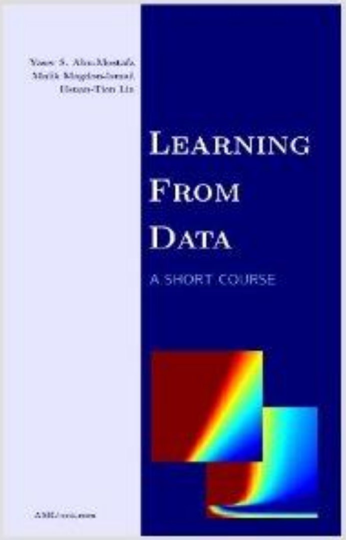
(available for free online)

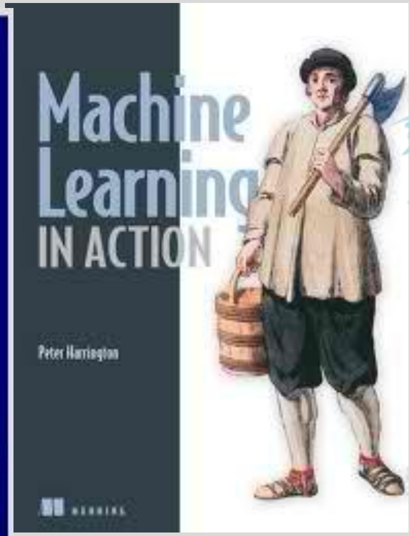


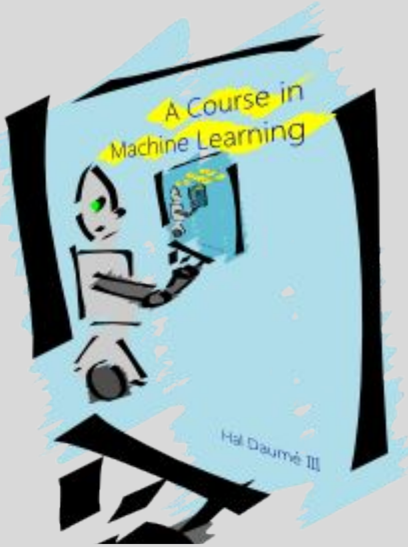
Recommended:











Machine Learning

ML = “Learning from data with computers”

Why?

- Automation of human tasks
  - AKA “AI”
  - For speed: handwriting & face recognition
  - For ease: self-driving cars
- Find complex patterns in big and complex data beyond human capacities
  - AKA “data science”
  - Can learn from very many examples
  - Can learn from very rich data

Machine Learning

Problems/methods we cover in this class will fall into two broad categories

- Supervised learning
  - Mostly predictive
  - Learning from examples
  - Regression, classification

- Unsupervised learning
  - Mostly descriptive
  - Understanding data without a target
  - Clustering, dimensionality reduction

Coding ML

- We’ll use python+matplotlib+pandas+sklearn
  - Best compromise of powerful scripting language and most common data wrangling and ML tools
- Other languages/packages of note:
  - TensorFlow, PyTorch, Keras -- graph computation for easily coding neural networks
  - R -- the scripting language of *true* data scientists
  - StatsModels -- core R functionality for python
  - Julia -- ease of python with speed of C but less ML
  - Stan -- scripting language for Bayesian inference
  - Tons of other stuff that you should *never* touch: Weka, anything matlab, SPSS, Stata

How is this section different?

- For the most part it really isn’t
  - Most of the material is the same
  - Both sections are an intro to ML with a complete practitioner’s toolkit
- In terms of methodology and examples, we’ll have a bit more emphasis on ML for data science
  - E.g., inference, regularized regression, causality, ...

Date		No.	Topic	Readings			Assignments
8/29	Thu	1	Class introduction	Ch. 1-2			#0: Setup Environment;
9/3	Tue	2	Bayes rate, Confusion matrix, ROC curve	Ch. 2			#1: kNN, Linear Regression, Logistic Regression
9/5	Thu	3	Linear regression	Sec. 3.1-3.2			
9/10	Tue	4	Logistic regression	Sec. 4.4			
9/12	Thu	6	Shrinkage, subset selection, cross validation	Sec. 3.3-3.4			Due 9/19 12:01AM
9/17	Tue	8	Histograms, kernel smoothing	Sec. 6.6.1-6.6.2			#2: Shrinkage, Naive Bayes, SVD, PCA
9/19	Thu	9	Naive Bayes, bag of words	Sec. 6.6.3			
9/24	Tue	10	SVD, PCA	Sec. 14.5.1, 14.6			
9/26	Thu	11	Similarities, multidimensional scaling	Sec. 14.8			Due 10/3 12:01AM
10/1	Tue	12	k-means, Gaussian mixture model, EM Algorithm	Sec. 13.2, 14.3			#3: Dimensionality reduction, Clustering
10/3	Thu	13	k-medoids, hierarchical clustering	Sec. 8.5.0, 8.5.1			
10/8	Tue	14	Guest lecture, Xiaojie Mao: Causality and Machine Learning	Spirtes	AoS Ch 16		
10/10	Thu	15	Guest lecture: Angela Zhou, Ethics and Fairness in Machine Learning	Kleinberg et al	ProPublica	WashPost	Due 10/17 12:01AM
10/15	Tue	**Fall break**					
10/17	Thu	**Prelim review**					Practice prelim released 10/10
10/22	Tue	**Prelim (in class) covers up to 10/3/2018**					
10/24	Thu	16	Classification and Regression Trees	Sec. 9.2			
10/29	Tue	17	Bagging, Random Forests, AdaBoost	Ch. 15			#4: Ensembles, SVM, Nerual Nets
10/31	Thu	18	Support Vector Machines	Sec. 4.5, Ch. 12			
11/5	Tue	19	Kernel Machines	Sec. 12.3.3, Sec. 14.5.4			
11/7	Thu	21	Neural Networks	Sec. 11.3-11.6			
11/12	Tue	22	Optimization, SGD, back-propagation				
11/14	Thu	22	Convolutional Networks	Sec. 11.7			
11/19	Tue	23	Advanced neural net topics I				Due 12/26 12:01AM
11/21	Thu	23	Advanced neural net topics II				
11/26	Tue	20	*Guest lecture: Andrew Bennett, Reinforcement Learning*				
11/28	Thu	**Thanksgiving Break**					
12/3	Tue	**Final Exam (take home) released 12/2; report due 12/10 12:01AM; reviews due 12/12 12:01AM**					
12/5	Thu	**Course Review Session**					
12/10	Tue	**NeurIPS conference: no classes**					

Fundamentals of supervised learning
Topics: classification, regression, regularization
Methods: kNN, linear regression, logistic regression, LASSO, Ridge regression

Probabilistic ML
Topics: inference, smoothing, Bayes law
Methods: OLS inference, bootstrap, kernel density estimation, kernel regression, Naive Bayes

Unsupervised learning
Topics: dimensionality reduction, clustering
Methods: SVD, PCA, NMF, MDS, k-means, EM, k-medoids, Ward's clustering

Cautionary tales in ML

Advanced supervised learning
Topics: trees, ensembles, kernels, neural nets
Methods: CART, RF, AdaBoost, SVM, kernel SVM, kernel PCA, neural networks, CNNs, autoencoders, GANs, RNNS

Date		No.	Topic
8/29	Thu	1	Class introduction
9/3	Tue	2	Bayes rate, Confusion matrix, ROC curve
9/5	Thu	3	Linear regression
9/10	Tue	4	Logistic regression
9/12	Thu	6	Shrinkage, subset selection, cross validation
9/17	Tue	8	Histograms, kernel smoothing
9/19	Thu	9	Naive Bayes, bag of words
9/24	Tue	10	SVD, PCA
9/26	Thu	11	Similarities, multidimensional scaling
10/1	Tue	12	k-means, Gaussian mixture model, EM Algorithm
10/3	Thu	13	k-medoids, hierarchical clustering
10/8	Tue	14	Guest lecture, Xiaojie Mao: Causality and Machine Learning
10/10	Thu	15	Guest lecture: Angela Zhou, Ethics and Fairness in Machine Learning
10/15	Tue		**Fall break**
10/17	Thu		**Prelim review**
10/22	Tue		**Prelim (in class) covers up to 10/**
10/24	Thu	16	Classification and Regression Trees
10/29	Tue	17	Bagging, Random Forests, AdaBoost
10/31	Thu	18	Support Vector Machines
11/5	Tue	19	Kernel Machines
11/7	Thu	21	Neural Networks
11/12	Tue	22	Optimization, SGD, back-propagation
11/14	Thu	22	Convolutional Networks
11/19	Tue	23	Advanced neural net topics I
11/21	Thu	23	Advanced neural net topics II
11/26	Tue	20	*Guest lecture: Andrew Bennett, Reinforcement Learning*
11/28	Thu		**Thanksgiving Break**
12/3	Tue		**Final Exam (take home) released 12/2; report du**
12/5	Thu		**Course Revie**
12/10	Tue		**NeurIPS conferer**

Basics of probability

- $P(A \mid B) = P(A,B) / P(B)$
- RV X is assignment of probabilities to values
- $E[X] = \mathrm{Sum}(P(X=x)\, x)$ for discrete RV X
- $E[X] = \mathrm{Int}(p(x)\, x)$ for continuous RV X with density p
- $f(x) = E[Y \mid X=x] = \mathrm{Sum}(P(Y=y \mid X=x)\, y)$
- $E[Y \mid X] = f(X)$ (a transform of RV X)
- $E[Y] = E[E[Y \mid X]]$ (iterated expectations)
- $\mathrm{Var}(X) = E[X^2] - E[X]^2$
- $\mathrm{Var}(Y) = E[\mathrm{Var}(Y|X)] + \mathrm{Var}(E[Y|X])$ (total var)

Basics of linear algebra

- Matrix A: m by n real values
- Vector v: n by 1 real values
- Av: m-dim vector with (Av)_j=Sum_i(A_ji v_i)
- Similarly, if B is n by k then AB is m by k
- λ and v are eigenvalue/eigenvector of n by n square matrix A if Av = λv

Go to review sessions!!!

Supervised Learning

MNIST Handwritten Digits

Y = 10 classes (0-9)
X = 28x28 pixels in grayscale (vector of 784 positive numbers denoting brightness)

60k training examples
10k testing examples

http://yann.lecun.com/exdb/mnist

Sign in

kaggle

Sign Up About Hosting Center All Competitions Users Forums Wiki Blog Data Science Jobs

Digit Recognizer

9 months to go

Wednesday, July 25, 2012 Knowledge • 478 teams Friday, July 26, 2013

Dashboard

Public Leaderboard

This leaderboard is calculated on approximately 25% of the test data.
The final results will be based on the other 75%, so the final standings may be different.

See someone using multiple accounts?
Let us know.

#	Δ1w	Team Name * in the money	Score	Entries	Last Submission UTC (Best Submission – Last)
1	-	Mikalai Drabovich *	0.99614	9	Fri, 28 Sep 2012 17:28:46
2	-	Mathew Monfort	0.99471	3	Mon, 03 Sep 2012 03:00:18
3	-	Xing Xu	0.99271	1	Fri, 10 Aug 2012 18:17:47
4	-	Analyst	0.99086	11	Sun, 30 Sep 2012 15:11:58
5	-	waronzevon	0.98857	21	Sat, 06 Oct 2012 00:46:22 (-3.5d)
6	-	Josef Slavicek	0.98700	1	Thu, 20 Sep 2012 17:03:22

276	↓20	BK	0.96614	3	Wed, 01 Aug 2012 07:17:04 (-44.3h)
276	↓20	Amit	0.96614	15	Thu, 11 Oct 2012 09:46:06
278	↓20	rippy	0.96586	1	Sun, 07 Oct 2012 17:00:34
278	new	Chris Kennedy	0.96586	2	Mon, 22 Oct 2012 03:58:07 (-1h)
280	↓21	Katalyst	0.96571	7	Sat, 06 Oct 2012 15:50:35 (-45.1d)
280	↓21	Jeremy Miller	0.96571	1	Mon, 01 Oct 2012 10:41:16
280	↓21	ross mckinlay	0.96571	2	Tue, 16 Oct 2012 18:09:39 (-3.8h)
	↓21	KNN, K=10	0.96557		
283	↓21	Tomato	0.96557	1	Fri, 27 Jul 2012 09:15:45
283	↓21	Michael Schwab	0.96557	1	Fri, 27 Jul 2012 18:52:37
283	↓21	Sergey Dolgopolov	0.96557	1	Sat, 28 Jul 2012 06:43:48
283	↓21	kudzai	0.96557	2	Sat, 28 Jul 2012 08:46:49 (-0h)
283	↓21	garcimore	0.96557	2	Sun, 29 Jul 2012 23:32:57 (-5.5h)
283	↓21	geekmarcus	0.96557	2	Wed, 05 Sep 2012 12:14:03 (-36.7d)
283	↓21	tcamolesi	0.96557	1	Tue, 31 Jul 2012 16:51:25

276	↓20	BK	0.96614	3	Wed, 01 Aug 2012 07:17:04 (-44.3h)
276	↓20	Amit	0.96614	15	Thu, 11 Oct 2012 09:46:06
278	↓20	rippy	0.96586	1	Sun, 07 Oct 2012 17:00:34
278	new	Chris Kennedy	0.96586	2	Mon, 22 Oct 2012 03:58:07 (-1h)
280	↓21	Katalyst	0.96571	7	Sat, 06 Oct 2012 15:50:35 (-45.1d)
280	↓21	Jeremy Miller	0.96571	1	Mon, 01 Oct 2012 10:41:16
280	↓21	ross mckinlay	0.96571	2	Tue, 16 Oct 2012 18:09:39 (-3.8h)
	↓21	KNN, K=10	0.96557		
283	↓21	[illegible]	0.96557	1	Fri, 27 Jul 2012 09:15:45
283	↓21	[illegible]	0.96557	1	Fri, 27 Jul 2012 18:52:37
283	↓21	[illegible]	0.96557	1	Sat, 28 Jul 2012 06:43:48
283	↓21	[illegible]	0.96557	2	Sat, 28 Jul 2012 08:46:49 (-0h)
283	↓21	garcimore	0.96557	2	Sun, 29 Jul 2012 23:32:57 (-5.5h)
283	↓21	geekmarcus	0.96557	2	Wed, 05 Sep 2012 12:14:03 (-36.7d)
283	↓21	tcamolesi	0.96557	1	Tue, 31 Jul 2012 16:51:25

BENCHMARK INFO

Treating the images as 784-dimensional vectors, for each test image we find the 10 nearest training images (in Euclidean distance). Then we have these 10 "nearest neighbors" vote on what digit the test image is.