# Machine Learning

Computer Sciences 760
Spring 2018

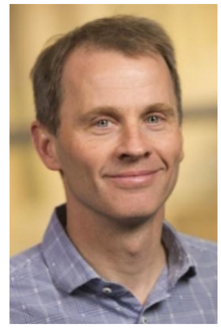
www.biostat.wisc.edu/~craven/cs760/

## Class enrollment

- typically the class is limited to 30
- we've allowed 90 to register
- ~ 70 are on the waiting list

- unfortunately, many on the waiting list will not be able to enroll
- but 760 is now offered every semester

# Instructors



## Mark Craven

email: craven@biostat.wisc.edu

office hours: 3-4:30 Wednesday, or by appointment

office: 4775A Medical Sciences Center

## David Page

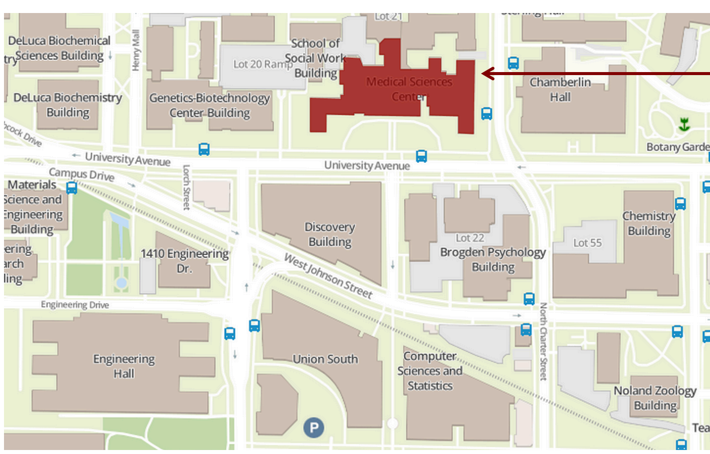
email: page@biostat.wisc.edu

office hours: 2:30-4 Friday, or by appointment

office hours room: 1153/4 Discovery Building

# Finding my office

- 4775A Medical Sciences Center
- easiest to enter from Charter St. and take elevator immediately to your right




enter here

# TAs

Daniel Griffin

email: dgriffin@cs.wisc.edu
office hours: 11:00-noon Monday and Wednesday
office: 4384 Computer Sciences

Viswesh Periyasamy

email: viswesh@cs.wisc.edu
office hours: 4:00-5:00 Tuesday and Thursday
office: 4710 Medical Sciences Center

# Monday, Wednesday *and* Friday?

- we'll have 28 lectures in all, just like a standard TR class
- most weeks we won't meet on Fridays
- but we will meet for the first three Fridays
- ***see the schedule on the course page***

## Expected background

- CS 540 (Intro to Artificial Intelligence) or equivalent
  - search
  - first-order logic
  - unification
  - deduction
- good programming skills
- basics of probability
- calculus, including partial derivatives

## Learning objectives

1. Students will understand what a learning system should do.
2. Students will distinguish among a variety of learning settings: supervised learning, unsupervised learning, reinforcement learning, active learning.
3. Students will employ a broad toolbox of machine-learning methods: decision trees, nearest neighbor, linear and logistic regression, neural nets, Bayesian networks, SVMs, ensemble methods.
4. Students will understand fundamental underlying theory: bias-variance tradeoff, PAC learning, mistake-bound theory.
5. Students will know how to characterize how well learning systems work, and they will employ sound experimental methodology for evaluating learning systems: cross validation, ROC and PR curves, hypothesis testing.

# Course requirements

- daily quizzes: ~14%
- 4 homework assignments: ~36%
  - programming
  - computational experiments (e.g. measure the effect of varying parameter $x$ in algorithm $y$)
  - some written exercises
- final exam: ~30%
- group project (4-5 students per group): ~20%

# TopHat for quizzes

- we will use TopHat for in-class quizzes
- each student will have to set up an account and purchase a subscription ($16 for the semester, $20 for the year)
- see https://kb.wisc.edu/luwmad/page.php?id=59937

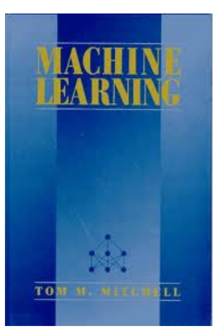
# Programming assignments

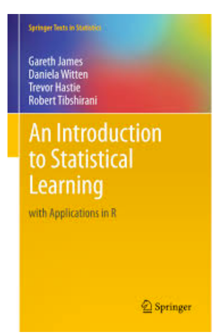
- for the programming assignments, you can use
  - C
  - C++
  - Java
  - Perl
  - Python
  - R
- programs must be callable from the command line
- programs must run on the CS department Linux servers

# Course readings

Buy one of two recommended books

- *Machine Learning*. T. Mitchell. McGraw Hill, 1997.
- *An Introduction to Statistical Learning.* G. James, D. Witten, T. Hastie, R. Tibshirani, Springer 2017.

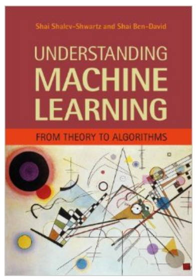


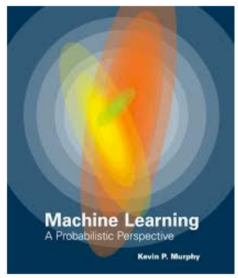



# Course readings

Also readings from two on-line books

- *Machine Learning: A Probabilistic Perspective*. K. Murphy. MIT Press, 2012
- *Understanding Machine Learning: From Theory to Algorithms*. S. Shalev-Shwartz and S. Ben-David. Cambridge University Press, 2014.



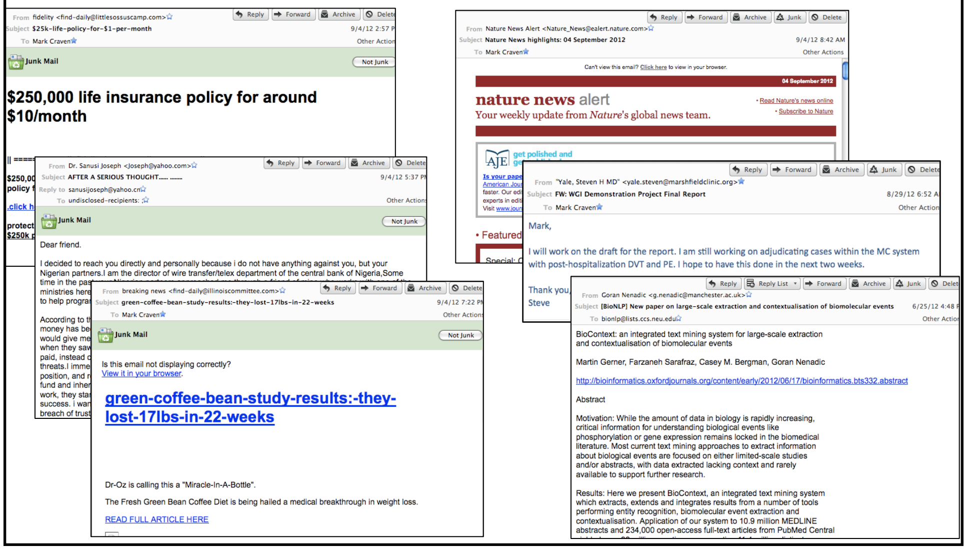




- additional on-line articles, surveys, and chapters

# What is machine learning?

- the study of algorithms that
  improve their performance $P$
  at some task $T$
  with experience $E$
- to have a well defined learning task,
  we must specify: $< P, T, E >$

# ML example: spam filtering

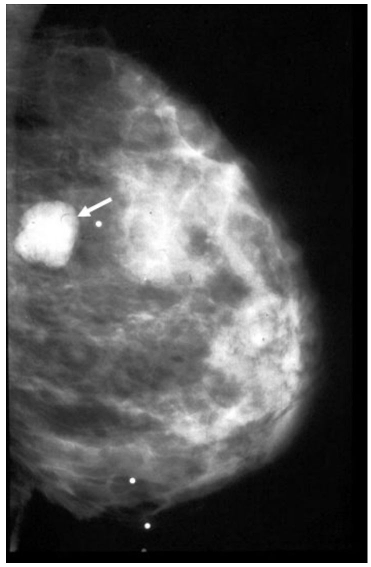



# ML example: spam filtering

- $T$ : given new mail message, classify as spam vs. other
- $P$ : minimize misclassification costs
- $E$ : previously classified (filed) messages

# ML example: mammography

[Burnside et al., *Radiology* 2009]



# ML example: mammography

- $T$ : given new mammogram, classify each abnormality as benign vs. malignant
- $P$ : minimize misclassification costs
- $E$ : previously encountered patient histories (mammograms + subsequent outcomes)

# ML example: predictive text input

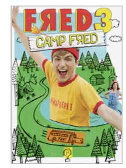

# ML example: predictive text input

- $T$ : given (partially) typed word, predict the word the user intended to type
- $P$ : minimize misclassifications
- $E$ : words previously typed by the user
  (+ lexicon of common words + knowledge of keyboard layout)

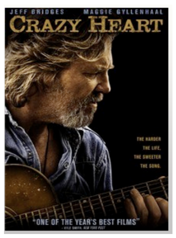
domain knowledge

# ML example: Netflix Prize






Our best guess for Mark:






Our best guess for Mark:

# ML example: Netflix

- $T$ : given a user/movie pair, predict the user's rating (1-5 stars) of the movie
- $P$ : minimize difference between predicted and actual rating
- $E$ : histories of previously rated movies (user/movie/rating triples)

## ML example: reinforcement learning to control an autonomous helicopter

video of Stanford University autonomous helicopter from http://heli.stanford.edu/

## ML example: autonomous helicopter

- $T$ : given a measurement of the helicopter’s current state (orientation sensor, GPS, cameras), select an adjustment of the controls
- $P$ : maximize reward (intended trajectory + penalty function)
- $E$ : state, action and reward triples from previous demonstration flights

# Assignment

- for Friday, read
  - Chapter 1 of Mitchell **or**
  - Chapter 1 of Murphy **or**
  - Chapter 1 and Section 2.1 of James *et al*.

  - article by Dietterich on web site

- set up TopHat account
- check out www.biostat.wisc.edu/~craven/cs760/