CS 335 Machine Learning

Dan Sheldon

Spring 2019

What is Machine Learning?

What is Machine Learning?

- How do you program a computer to
  - Recognize faces?
  - Recommend movies?
  - Decide which web pages are relevant to a Google search query?

A Simple Task: Recognize Obama

A Simple Task: Recognize Obama

- Input: picture
- Output: yes/no

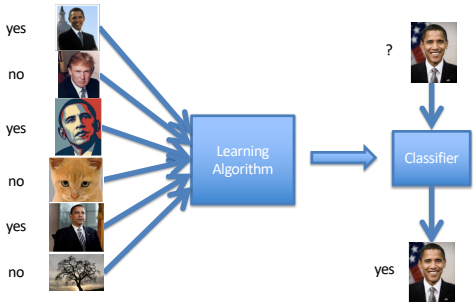
- Can you program this?
  - Probably not…
  - But you can *show* a computer how to solve this task

Examples

No

Yes

Learning from Examples



Labeled Examples

Discussion

- Is it easier to design a learning algorithm or an Obama recognizer?
- Is it more useful to have a learning algorithm or an Obama recognizer?
- What problems like this would you like to solve?

What is Machine Learning?

- *Machine learning is the practice of programming a computer to learn to solve a task through experience, rather than directly programming it to solve the task.*

Why is ML Important?

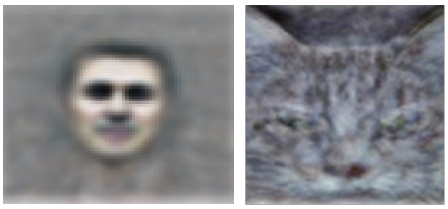
- What are some examples of ML in your day-to-day life?

ML makes the world go round.

ML Achievements

ML wins Jeopardy!

ML Achievements

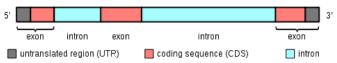


ML watches YouTube for three straight days!
(and learns to recognize cats)

http://www.npr.org/2012/06/26/155792609/a-massive-google-network-learns-to-identify
Building High-level Features Using Large Scale Unsupervised Learning
Quoc V. Le, Marc'Aurelio Ranzato, Rajat Monga, Matthieu Devin, Kai Chen, Greg S. Corrado, Jeffrey Dean, and Andrew Y. Ng

ML in Science

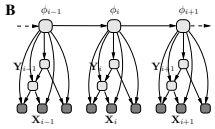
- Bioinformatics
  - Gene prediction



Computational Identification of Evolutionarily Conserved Exons

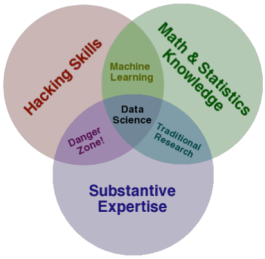
Adam Siepel
Center for Biomolecular Science and Engr.
University of California
Santa Cruz, CA 95064, USA
acs@soe.ucsc.edu

David Haussler
Howard Hughes Medical Institute and
Center for Biomolecular Science and Engr.
University of California
Santa Cruz, CA 95064, USA
haussler@soe.ucsc.edu



Hidden Markov model

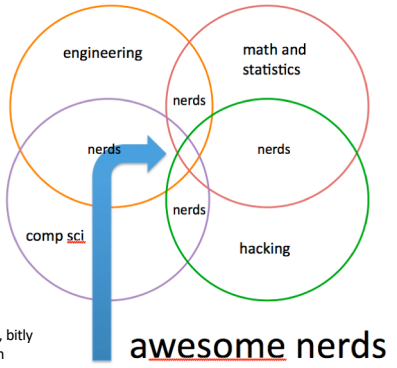
ML in Data Science



http://drewconway.com

Extracting insight and knowledge from data

Data scientists?



By Hilary Mason, bitly
From forbes.com

OK, but what are we actually going to do in this class?

Course Goals

- Learn how to design basic ML algorithms
- Learn about specific, widely used ML algorithms
- Learn tools to apply ML algorithms to real data and evaluate their performance

Course Structure

- Supervised learning
  - Learn from examples
- Unsupervised learning
  - Find patterns in data
- Probabilistic learning
  - Quantify uncertainty

"20% chance of rain"
"80% chance of survival"
"90% sure it is President Obama"

Course Structure

- There are many dimensions of ML methodology:
  - Supervised vs. unsupervised
  - Linear vs. nonlinear
  - Univariate vs. multivariate
  - Regression vs. classification
  - Binary vs. multiclass
  - Probabilistic or not
- We will touch on most of these

Logistics

Course Webpage

http://people.cs.umass.edu/~sheldon/teaching/cs335/index.html

- Entry point for all course information
  - Course policies (review)
  - Schedule
  - Slides
  - Homework
- Office hours:
  - Tue ~~4-5pm~~ 3:30-4:30pm, Clapp 200
  - Thu 1-2pm, Clapp 200

Math

- Warning: there is math in this course
  - Calculus
    - Derivatives
    - (Partial derivatives)
  - Probability
    - Sample space, events, conditional probability, discrete random variables, expected value
    - Review later in semester
  - Linear algebra
    - Matrices and vectors
    - (Vector norm, dot product, transpose, inverse)
    - (Manipulation of linear algebraic equations)
- Self-assessment in HW0

Python

- All programming in this course done in Python
  - **Required environment:** Anaconda 2018.12 for Python 3.7
  - Installed on CS lab computers
- Python session during class in ~1 week
- But you are largely responsible for learning on your own

What's Next?

- Homework 0 posted
  - By next Tuesday
    - Read course policies
    - Post on Piazza
  - By next Friday:
    - Get started with Python exercise
- Optional calculus review?
  - Derivatives / optima
  - TBD based on interest…

(If time) A First Supervised Learning Model

- Other slides and board work