CS 341 Machine Learning

Instructor: Prof. Dan Sheldon

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...
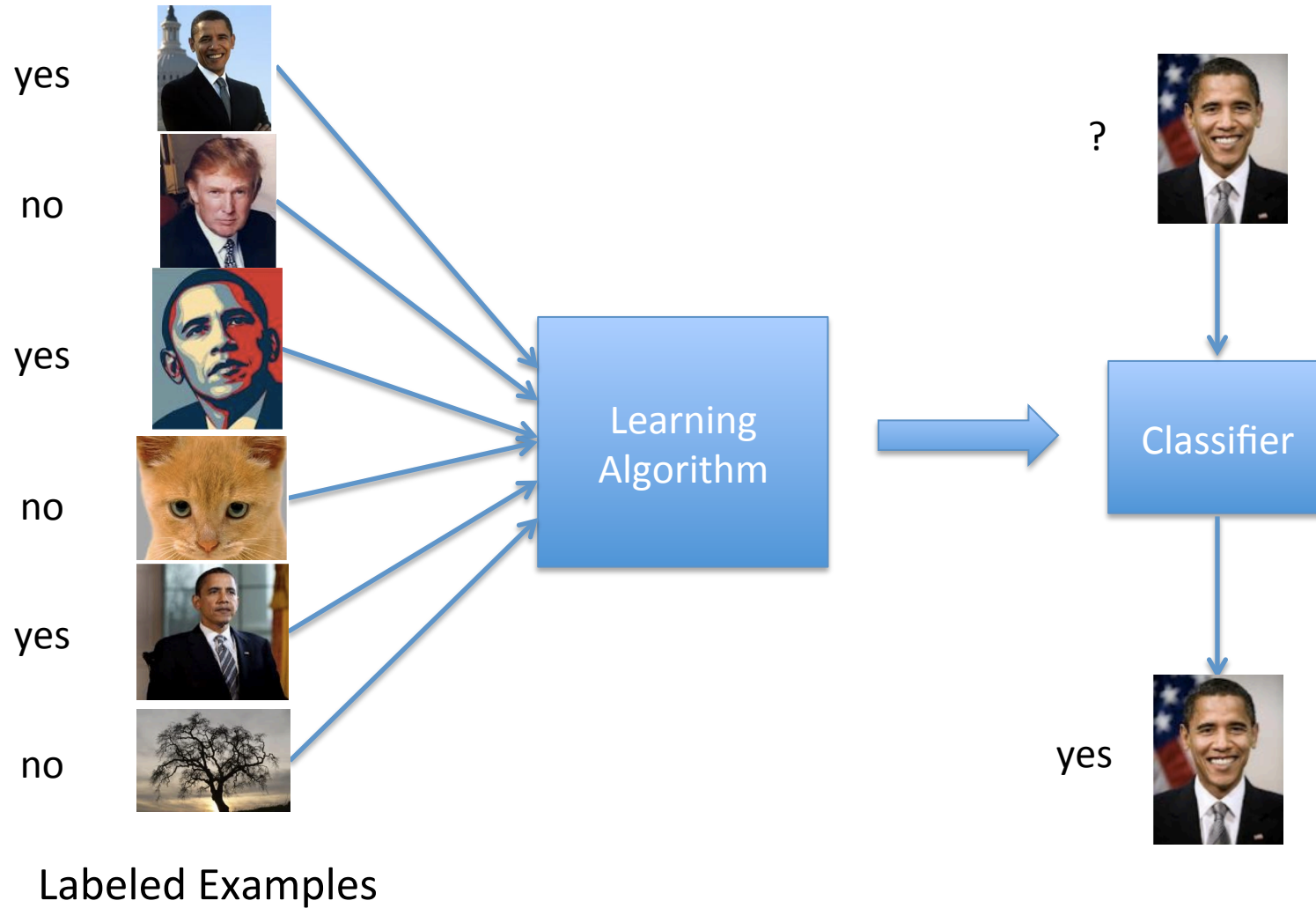 - But you can *show* a computer how to solve this task

Examples

No

Yes

Learning from Examples

yes

no

yes

no

yes

no

Labeled Examples

Learning Algorithm

Classifier

?

yes

Discussion

- Is it easier to devise a learning algorithm than it is to program an Obama recognizer?

- Is it more useful to have a learning algorithm, or an Obama recognizer?
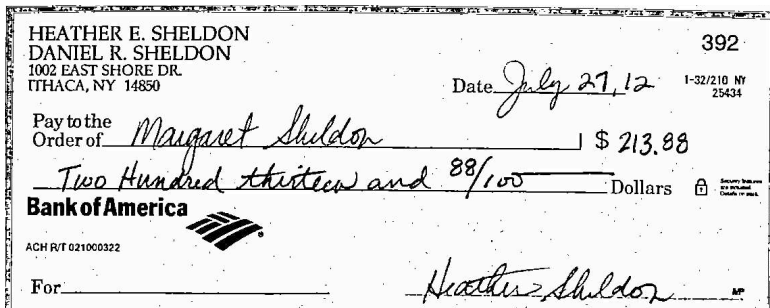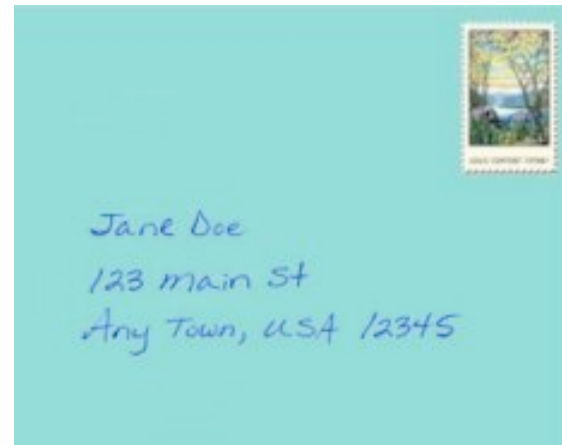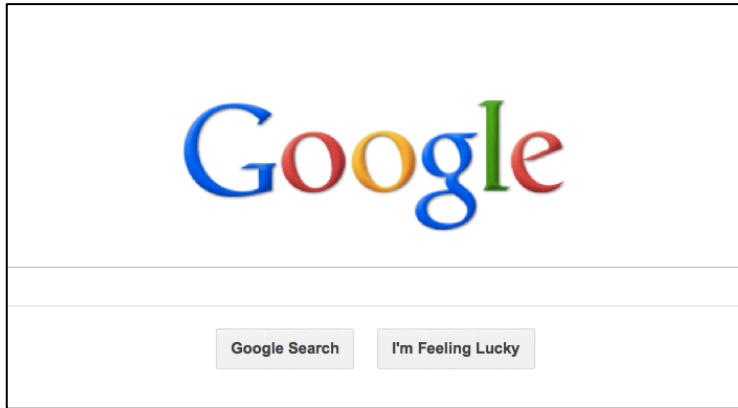
What is Machine Learning?

- *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 should I care about ML?

You tell me…

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

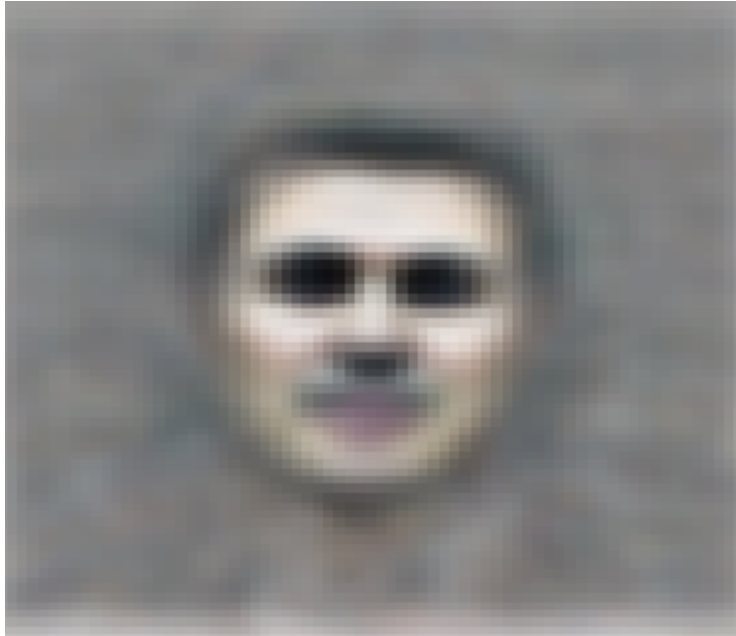# ML makes the world go round.

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
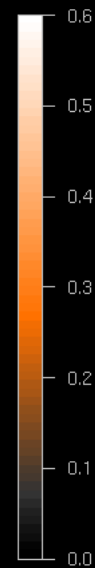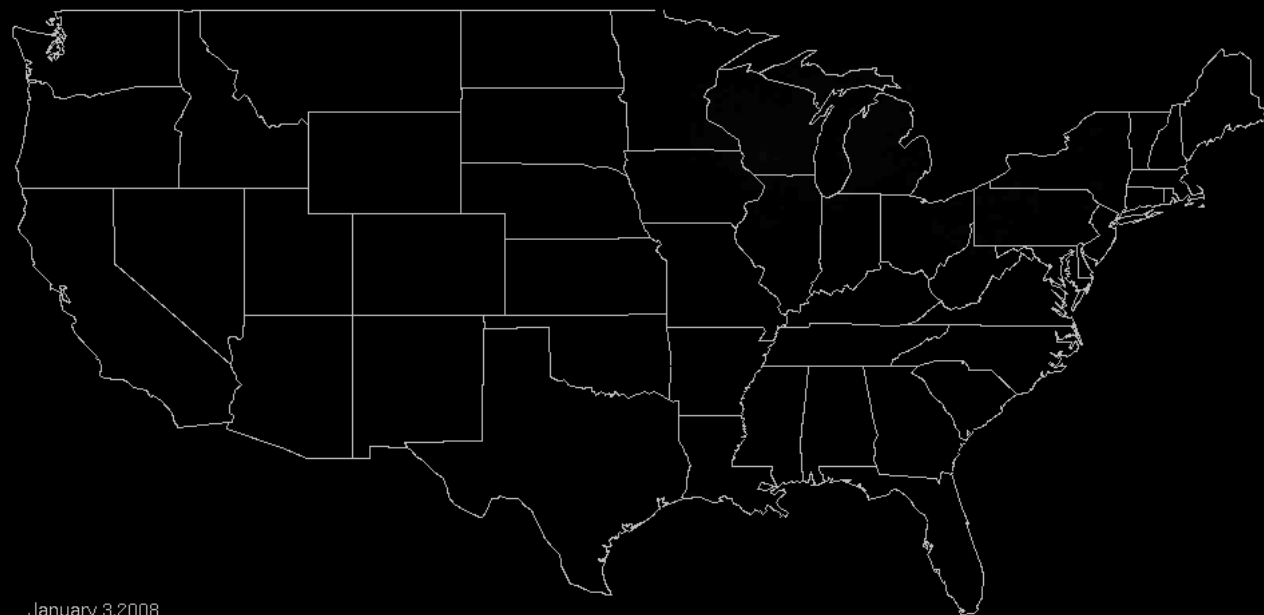Quoc V. Le, Marc'Aurelio Ranzato, Rajat Monga, Matthieu Devin, Kai Chen, Greg S. Corrado, Jeffrey Dean, and Andrew Y. Ng

ML in Science

- Bioinformatics
  - Gene expression
  - Does this gene (section of DNA) do anything?

- Species distribution modeling…

Species Distribution Modeling

January 3,2008

Jan · Apr · Jun · Sep · Dec
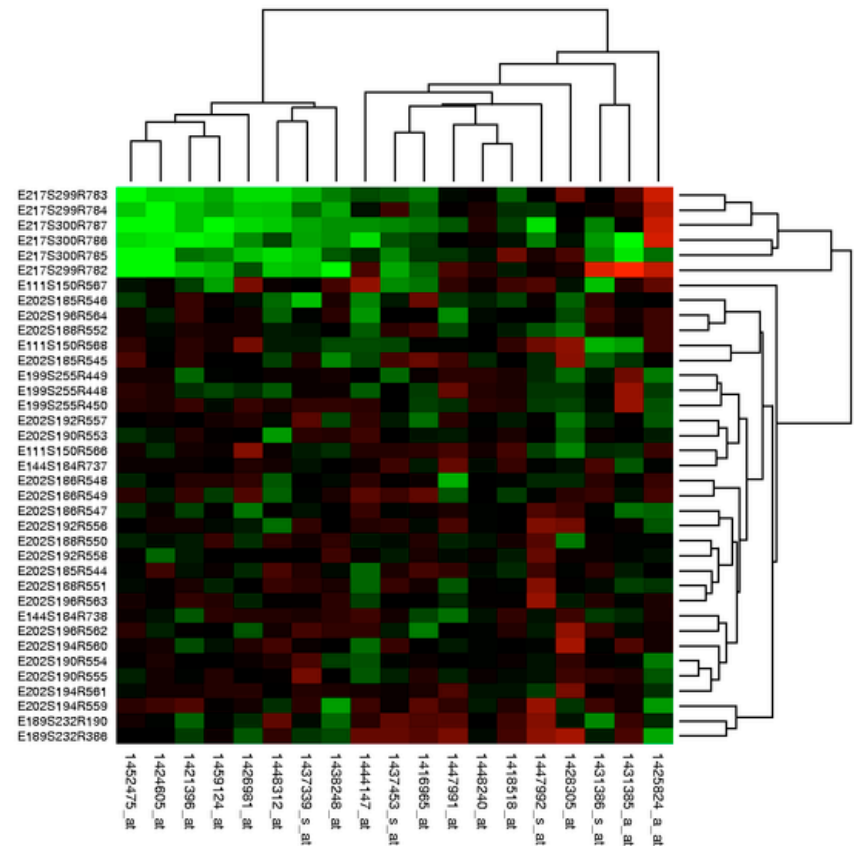
Big Data → Big $$$$

- "Analyzing large data sets—so called big data—will become a key basis of competition, underpinning new waves of productivity growth, innovation, and consumer surplus…"

- "The United States alone faces a shortage of 140,000 to 190,000 people with analytical expertise and 1.5 million managers and analysts with the skills to understand and make decisions based on the analysis of big data."

Source: Big data: The next frontier for innovation, competition, and productivity, McKinsey & Company, 2011

Source: Big data: The next frontier for innovation, competition, and productivity, McKinsey & Company, 2011

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

Course Overview

- **Supervised learning**
  - Learn from examples

- **Unsupervised learning**
  - Find patterns in data

- **Probabilistic learning**
  - Quantify uncertainty
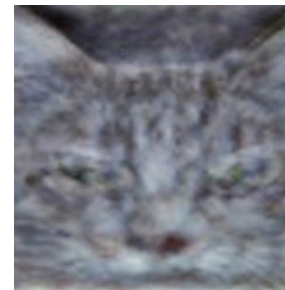
"20% chance of rain"
"80% chance of survival"
"90% sure it is President Obama"

Course Organization

- Course is organized around *problems* and *algorithms*
  - See a list of topic on the webpage

- But the goal is to understand ML *building blocks* and *general principles*
  - Mathematical tools
    - Probability, statistics, basic optimization
  - ML principles
    - Problem formulations, notation, patterns
  - Methodology
    - How to evaluate, diagnose problems, and troubleshoot
  - Applications
    - Understand ML in the world and get your hands dirty

Logistics

- Instructor
  - Prof. Dan Sheldon
  - http://people.cs.umass.edu/~sheldon

- Office hours (Clapp 200; come see me!)
  - Mon 4-5pm
  - Wed 9-10am
  - By appointment. Mon/Wed only

- Course webpage
  - http://people.cs.umass.edu/~sheldon/teaching/2012fa/ml/index.html
  - Please check frequently
  - Ella: assignments
    - https://ella.mtholyoke.edu/portal/site/COMSC-341-01-FA12

Logistics

- Email me!
  - Response time: 24 hours / 48 hours on weekend
  - Best way to ask administrative questions
  - HW questions: please come to office hours

- Accommodation letters

- I'm new here
  - But I'm here because of you!
  - Please come talk to me
 - We will work out any kinks

Prerequisites

- Prerequisites
  - Data structures (know how to program)
  - Discrete math (not be afraid of math)

- Warning: there is math in this course
  - Calculus
  - Probability
  - Linear algebra

- I will teach necessary material
  - Math
  - MATLAB programming

Textbook and Other Resources

- Required textbook:
  - Russell and Norvig, *Artificial Intelligence: A Modern Approach*
  - **Third Edition** required

- Coursera ML course
  - https://class.coursera.org/ml/lecture/preview
  - Andrew Ng

- Additional books on reserve at library
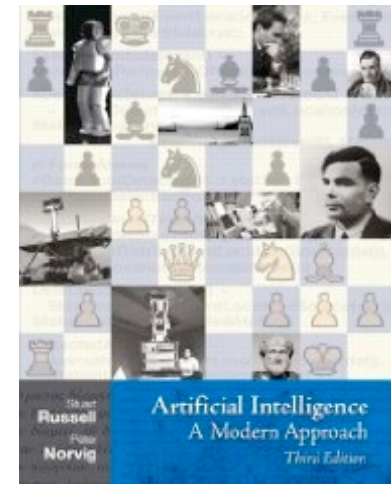
Coursework

- Daily assignments: 20%
- Problem sets: 40%
- Project: 30%
- Class participation: 10%

Coursework

- Daily exercises
  - Due at beginning of each class
  - < 1 hour, 1–3 problems
  - Solidify material covered in lecture
  - **Must be done independently**

- Problem sets
  - Every two weeks, until project
  - Written problems, programming, experiments
  - **I encourage you to work in groups, but you must**
    - *Write your own code & solutions*
    - *Run your own experiments*
    - *Cite all group members and sources*

Coursework

- Final project
  - *Creative* application/exploration of ML
    - Apply ML to scientific data
    - Gather dataset from your hobby/sport/passion
    - Explore an ML topic not covered in this class
    - Prove something about ML
  - Presentation on last day of class
  - Paper in format of short scientific paper
  - More details and suggested topics later in course
  - **Work in groups of up to 3 students**

Late Policy

- Problem sets and project milestones
  - Five free late days (24 hours)
  - *For project milestones, charged to entire group*
  - Must submit by ella

- Daily Exercises
  - No credit for late submission, but lowest 4 dropped

- Final project report
  - No credit for late submission

- (For emergencies, talk to me)