CSE 446: Machine Learning

Sergey Levine
svlevine@cs.washington.edu

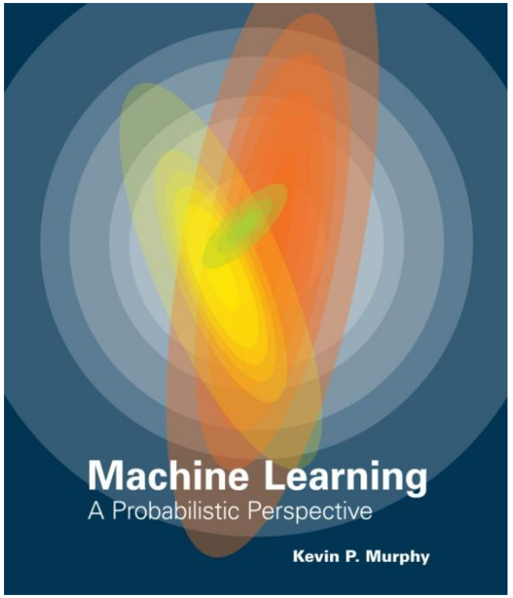
Slides adapted from Pedro Domingos, Carlos Guestrin, Luke Zettelmoyer, Ali Farhadi

Logistics

- Instructor: Sergey Levine
  - svlevine@cs.washington.edu
  - Office: CSE 528
  - Office hours: Monday 10:30 – 11:30
- TAs
  - Naozumi Hiranuma (CSE 220, Wed 1:30 – 2:30)
  - Akshay Srinivasan (CSE 218, Fri 11:00 – 12:00)
  - Isaac Tian (CSE 220, Tue 12:00 – 1:00)
- Website:

https://courses.cs.washington.edu/courses/cse446/16sp/

Textbook



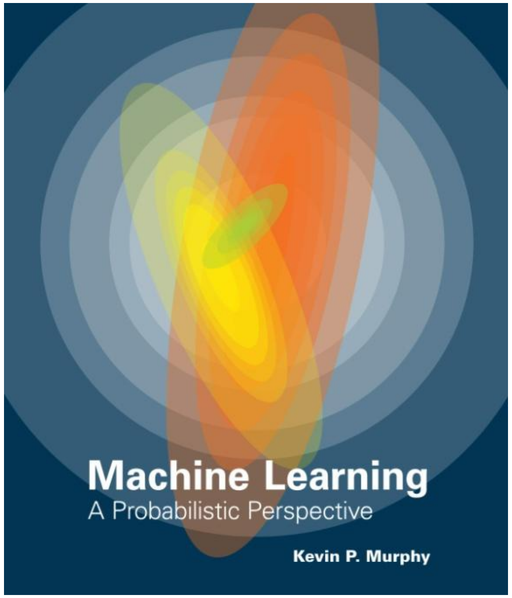
Machine Learning: a Probabilistic Perspective
Kevin Murphy,
MIT Press, 2013.

Optional:

- Pattern Recognition and Machine Learning, C. Bishop, Springer, 2007
- The Elements of Statistical Learning, Friedman, Tibshirani, Hastie, Springer, 2001
- Machine Learning, Mitchell, MacGraw Hill, 1997

Textbook

Machine Learning: a Probabilistic Perspective
Kevin Murphy,
MIT Press, 2013.



Readings (this week):

1.1 – 1.3 (1.4 optional): introduction to machine learning

16.2.1 – 16.2.4: decision trees (covered this week)

Chapter 2: background on probability – may be very useful for understanding 16.2.1 – 16.2.4

Assignments & Discussion

- Assignments graded via Gradscope: please let us know by Fri if you don't receive login information

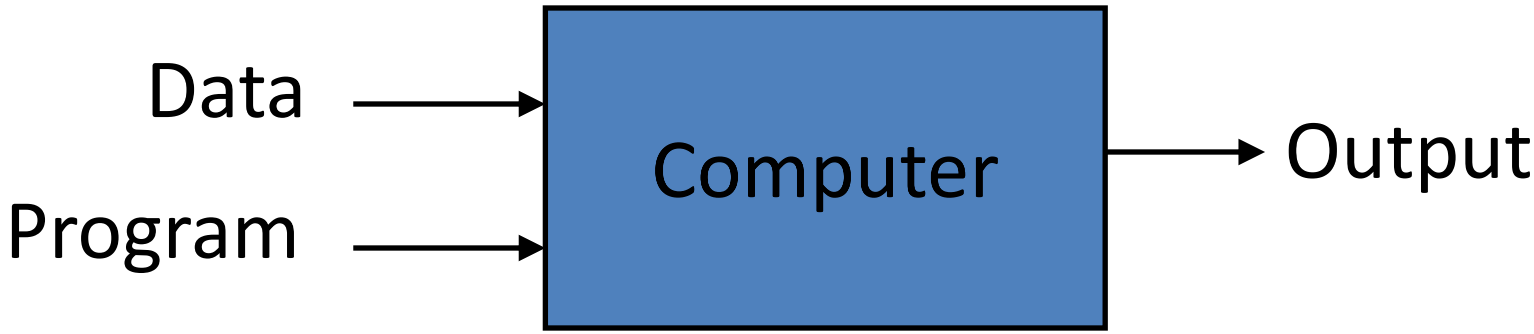
- Make an account with Piazza, so that you can post questions & discussion (see course website for instructions)

Grading

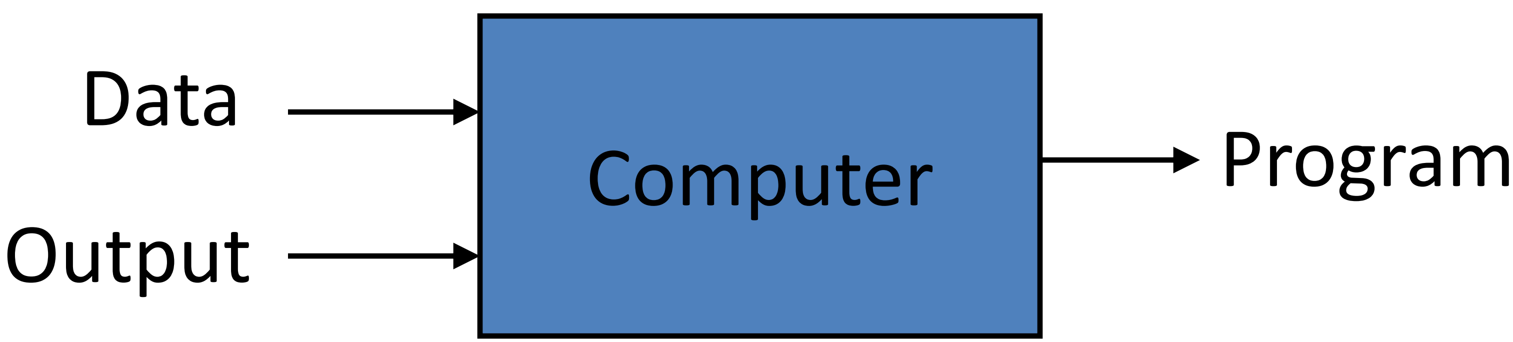
- Homeworks (65% of the grade):
  - 4 homeworks
  - First homework assigned week 2 and due week 4
  - Short answers
  - Programming (Python)
- Midterm exam (10% of the grade):
  - 6th week of class
  - Covers week 1 – 5
  - Will have review section
  - Open book, notes, etc., closed computer
- Final exam (20% of the grade):
  - Cumulative (covers everything)
  - Open book, notes, etc., closed computer
- Class participation (5% of the grade):
  - Speak up in class (it's fun)
  - Be sure to say your name

What is Machine Learning?

Traditional Programming



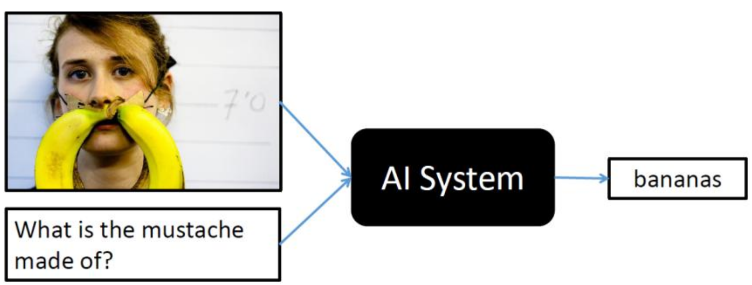
Machine Learning

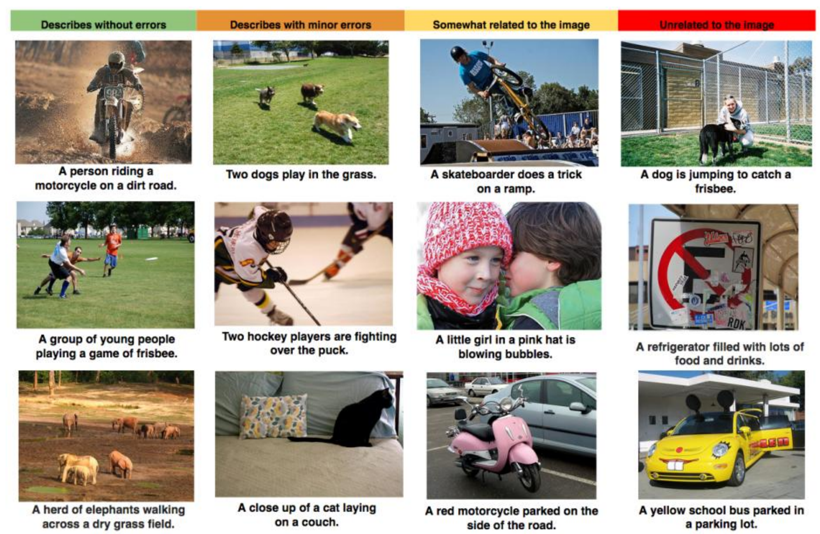


Why Machine Learning?

- Computers can simulate *anything*
- So why can't computers *do* everything?
  - Writing programs is difficult
  - Input data can be really complicated
  - That's why we have a software industry
- Is there a better way?
  - Let's have computers come up with their own programs

What Can it Do?





What Can't it Do?

- Great for detecting patterns
- Not so great at deeper understanding (yet)
- This slide will be revised next year (and the year after that...)

Winograd Schema

The city councilmen refused the demonstrators a permit because they feared violence.

The city councilmen refused the demonstrators a permit because they advocated violence.

Although they ran at about the same speed, Sue beat Sally because she had such a [good/bad] start. Who had a [good/bad] start?

The sculpture rolled off the shelf because it wasn't [anchored/level]. What wasn't [anchored/level]?

What is Machine Learning ?

(by examples)

Classification

from data to discrete classes

Spam filtering

data

Osman Khan to Carlos — show details Jan 7 (6 days ago) — Reply

sounds good
+ok

Carlos Guestrin wrote:

> Let's try to chat on Friday a little to coordinate and more on Sunday in person?
>
> Carlos

Welcome to New Media Installation: Art that Learns

Carlos Guestrin to 10615-announce, Osman, Michel — show details 3:15 PM (8 hours ago) — Reply

Hi everyone,

Welcome to New Media Installation:Art that Learns

The class will start tomorrow.
Make sure you attend the first class, even if you are on the Wait List.
The classes are held in Doherty Hall C316, and will be Tue, Thu 01:30-4:20 PM.

By now, you should be subscribed to our course mailing list: 10615-announce@cs.cmu.edu.
You can contact the instructors by emailing: 10615-instructors@cs.cmu.edu

Natural _LoseWeight SuperFood Endorsed by Oprah Winfrey, Free Trial 1 bottle, pay only $5.95 for shipping mfw rlk Spam | X

Jaquelyn Halley to nherrlein, bcc: thehorney, bcc: ang — show details 9:52 PM (1 hour ago) — Reply

=== Natural WeightL0SS Solution ===

Vital Acai is a natural WeightL0SS product that Enables people to lose wieght and cleansing their bodies faster than most other products on the market.

Here are some of the benefits of Vital Acai that You might not be aware of. These benefits have helped people who have been using Vital Acai daily to Achieve goals and reach new heights in there dieting that they never thought they could.

* Rapid WeightL0SS
* Increased metabolism - BurnFat & calories easily!
* Better Mood and Attitude
* More Self Confidence
* Cleanse and Detoxify Your Body
* Much More Energy
* BetterSexLife
* A Natural Colon Cleanse

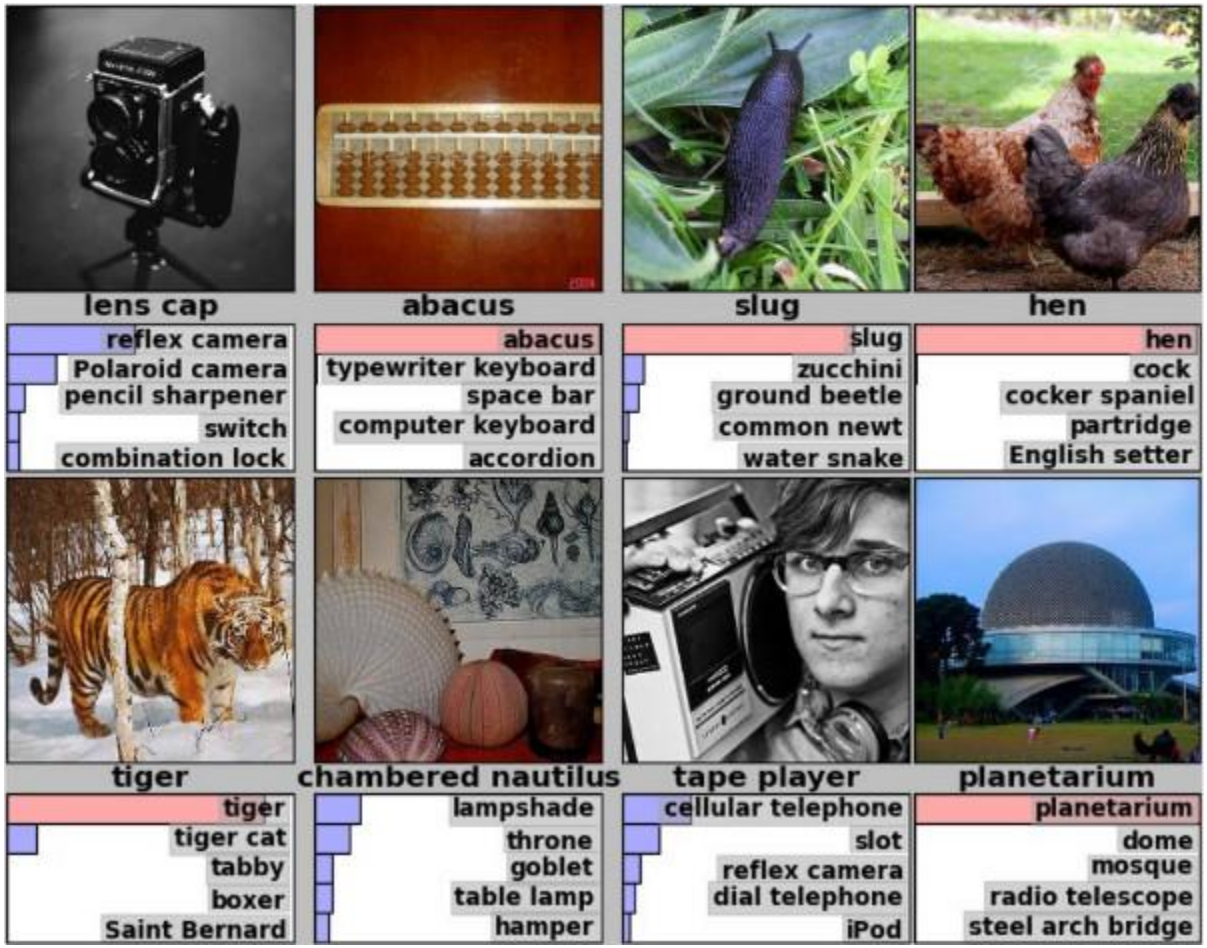
prediction

Spam

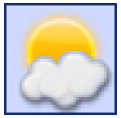
vs

Not Spam

Object classification



Weather prediction



Regression

predicting a numeric value

Stock market

Weather prediction revisted

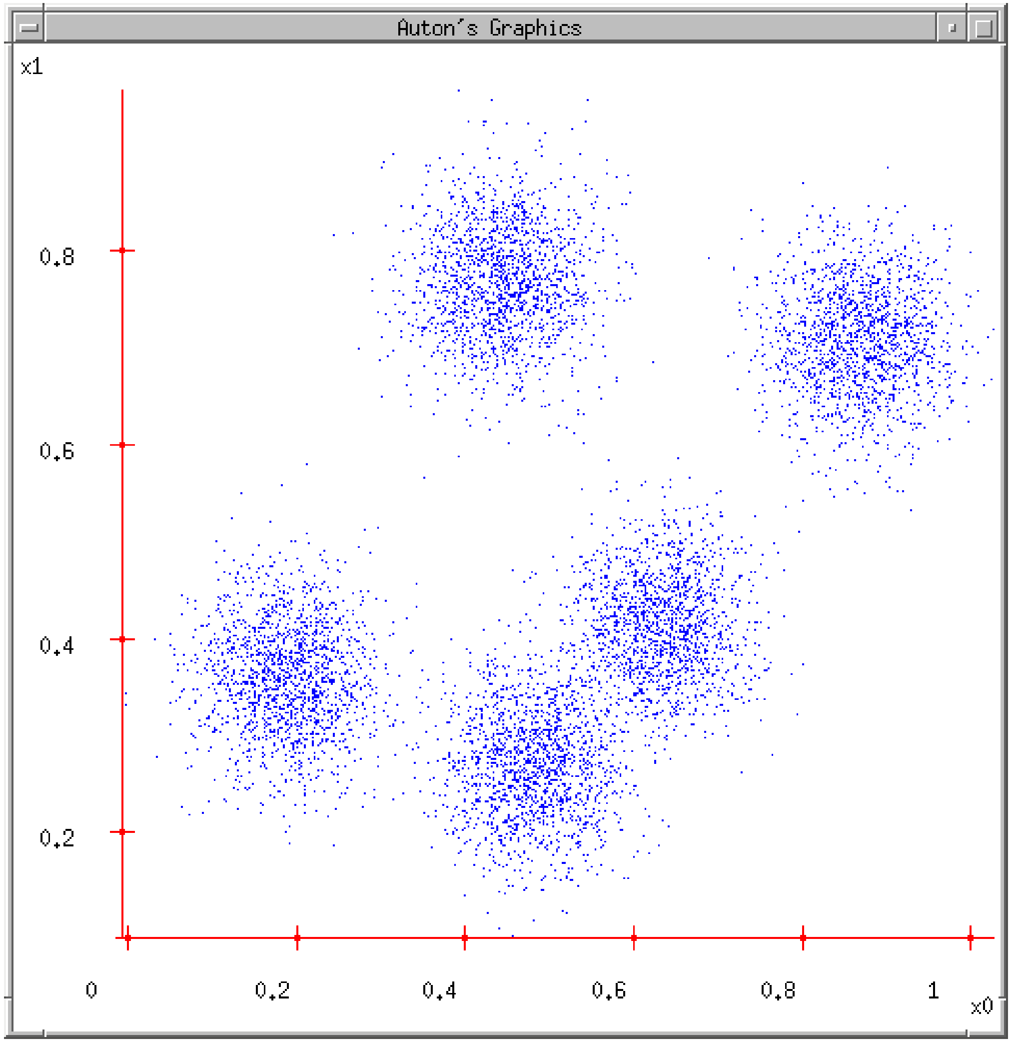
Temperature

72° F

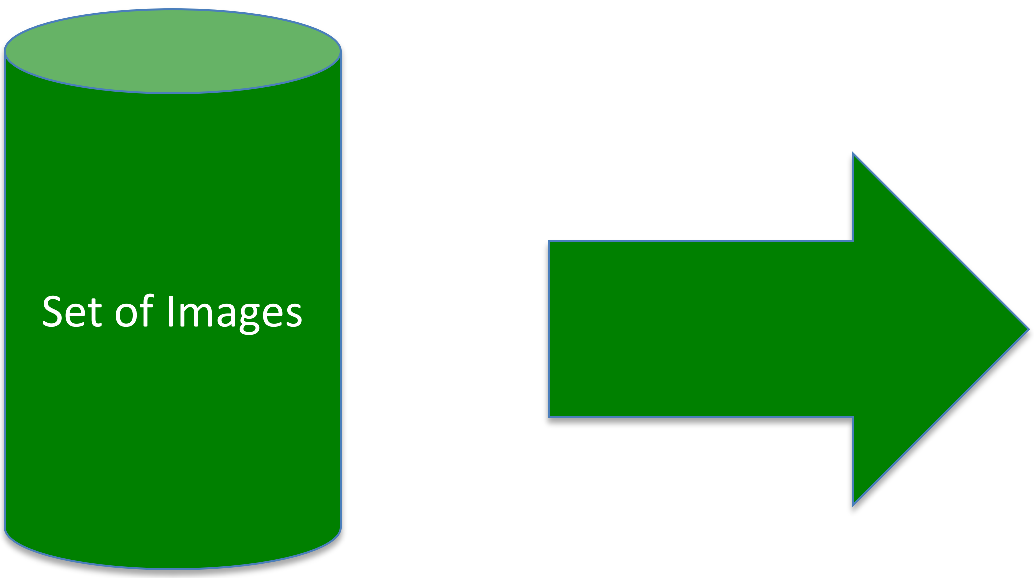
Clustering

discovering structure in data

Clustering Data: Group similar things

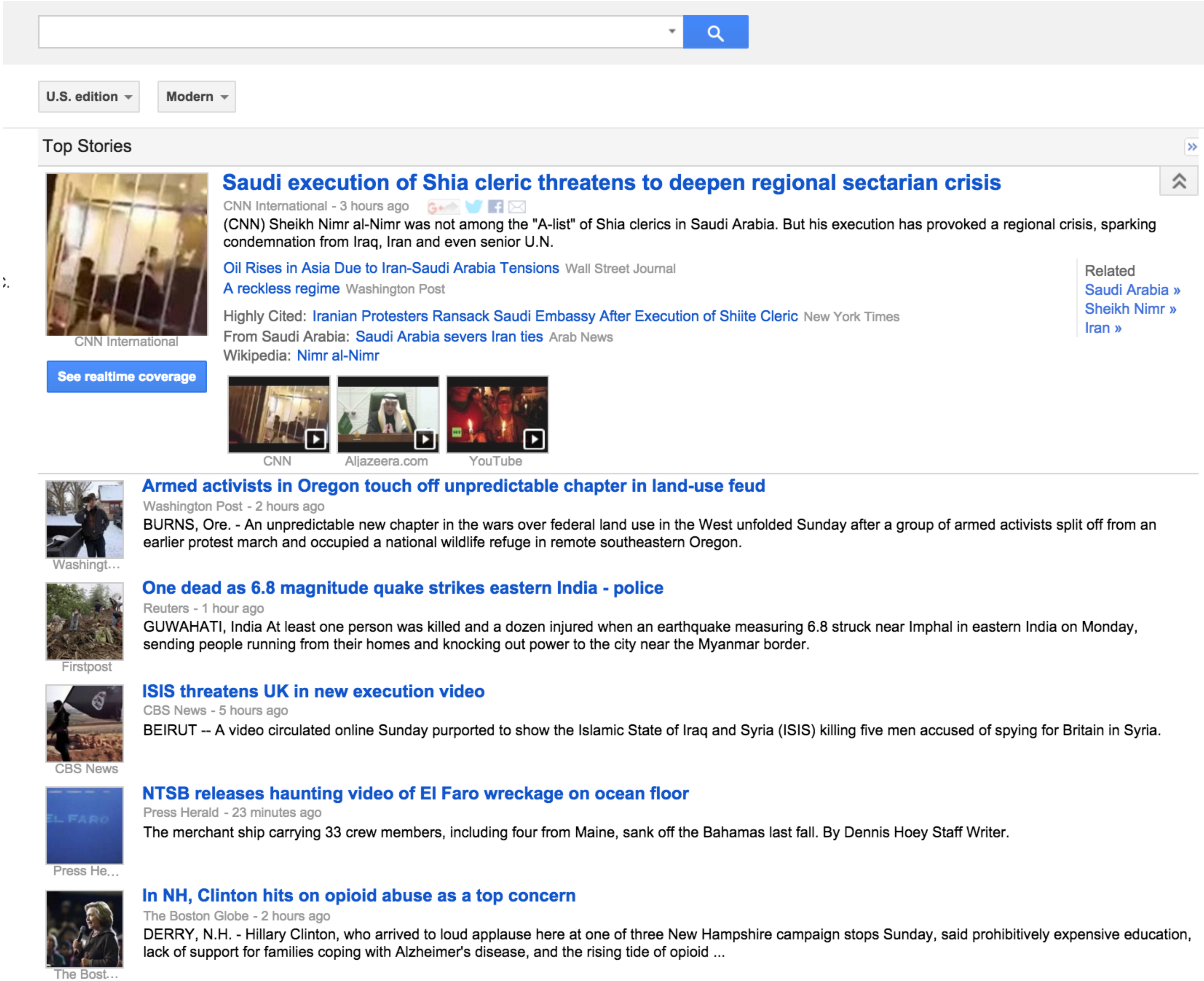


Clustering images



[Goldberger et al.]

Clustering News



U.S. edition | Modern

Top Stories

Saudi execution of Shia cleric threatens to deepen regional sectarian crisis
CNN International - 3 hours ago
(CNN) Sheikh Nimr al-Nimr was not among the "A-list" of Shia clerics in Saudi Arabia. But his execution has provoked a regional crisis, sparking condemnation from Iraq, Iran and even senior U.N.

Oil Rises in Asia Due to Iran-Saudi Arabia Tensions Wall Street Journal
A reckless regime Washington Post
Highly Cited: Iranian Protesters Ransack Saudi Embassy After Execution of Shiite Cleric New York Times
From Saudi Arabia: Saudi Arabia severs Iran ties Arab News
Wikipedia: Nimr al-Nimr

Related
Saudi Arabia »
Sheikh Nimr »
Iran »

CNN International
See realtime coverage

CNN | Aljazeera.com | YouTube

Armed activists in Oregon touch off unpredictable chapter in land-use feud
Washington Post - 2 hours ago
BURNS, Ore. - An unpredictable new chapter in the wars over federal land use in the West unfolded Sunday after a group of armed activists split off from an earlier protest march and occupied a national wildlife refuge in remote southeastern Oregon.
Washingt...

One dead as 6.8 magnitude quake strikes eastern India - police
Reuters - 1 hour ago
GUWAHATI, India At least one person was killed and a dozen injured when an earthquake measuring 6.8 struck near Imphal in eastern India on Monday, sending people running from their homes and knocking out power to the city near the Myanmar border.
Firstpost

ISIS threatens UK in new execution video
CBS News - 5 hours ago
BEIRUT -- A video circulated online Sunday purported to show the Islamic State of Iraq and Syria (ISIS) killing five men accused of spying for Britain in Syria.
CBS News

NTSB releases haunting video of El Faro wreckage on ocean floor
Press Herald - 23 minutes ago
The merchant ship carrying 33 crew members, including four from Maine, sank off the Bahamas last fall. By Dennis Hoey Staff Writer.
Press He...

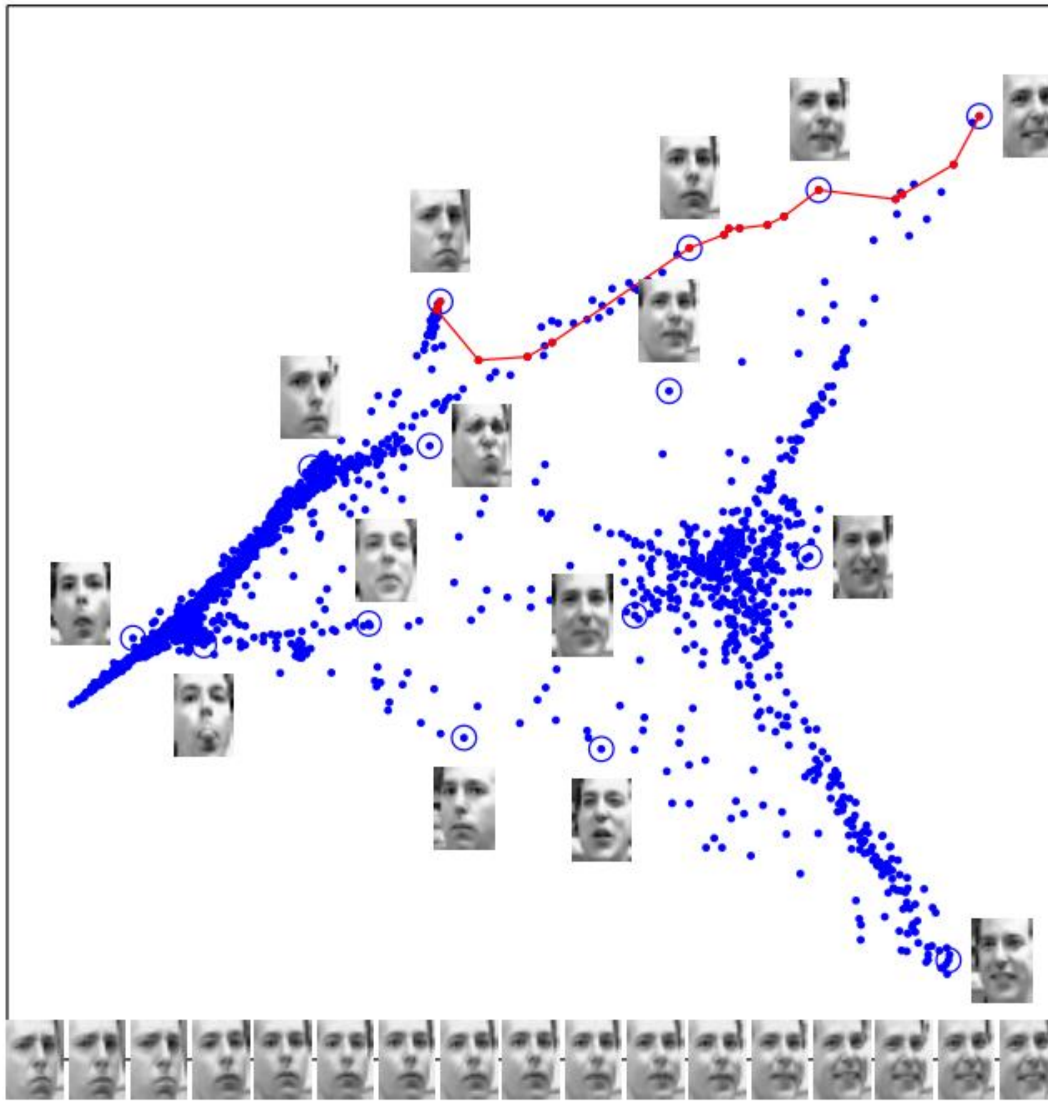
In NH, Clinton hits on opioid abuse as a top concern
The Boston Globe - 2 hours ago
DERRY, N.H. - Hillary Clinton, who arrived to loud applause here at one of three New Hampshire campaign stops Sunday, said prohibitively expensive education, lack of support for families coping with Alzheimer's disease, and the rising tide of opioid ...
The Bost...

Embedding

visualizing data

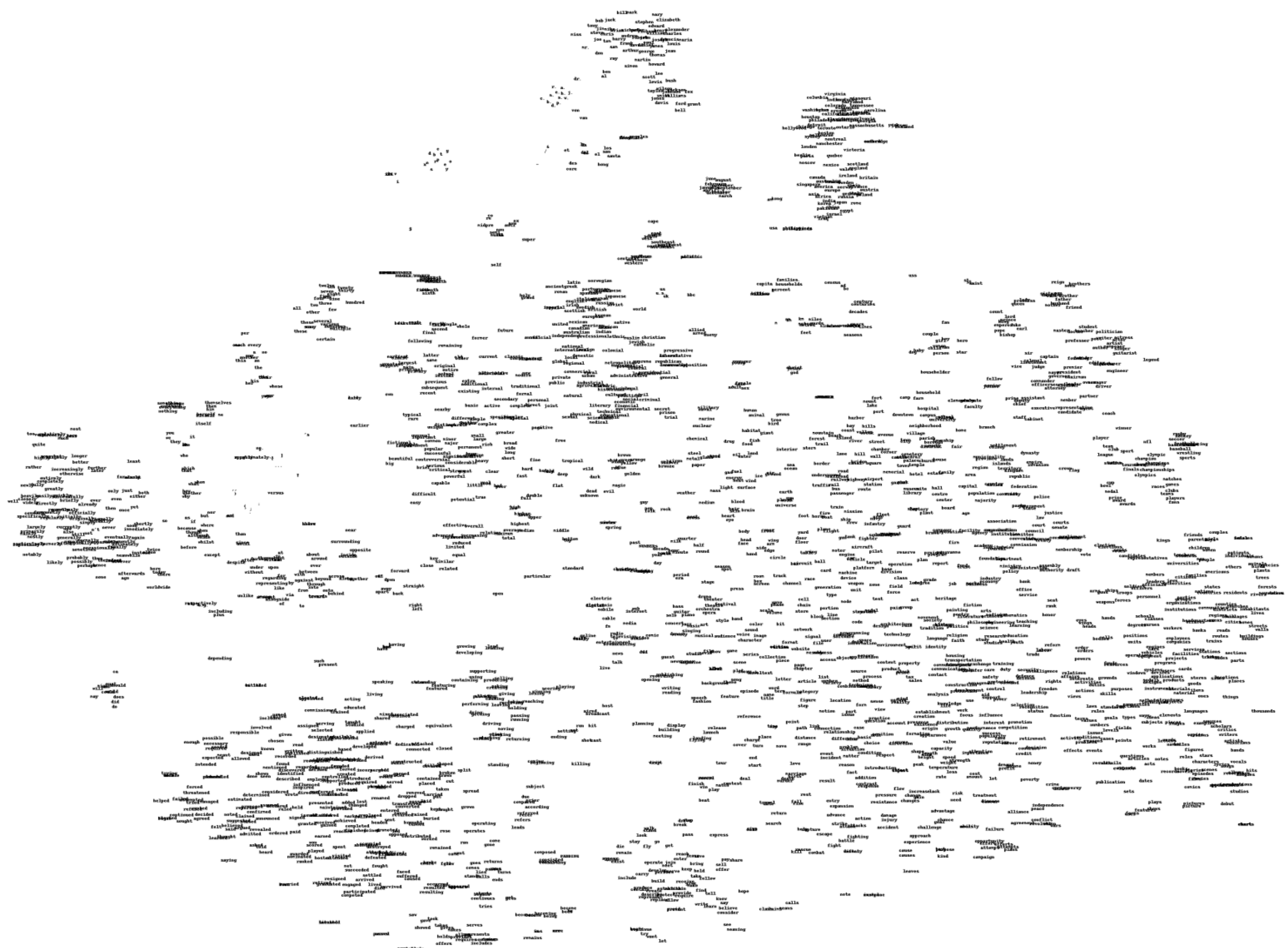
Embedding images

- Images have thousands or millions of pixels.
- Can we give each image a coordinate, such that similar images are near each other?



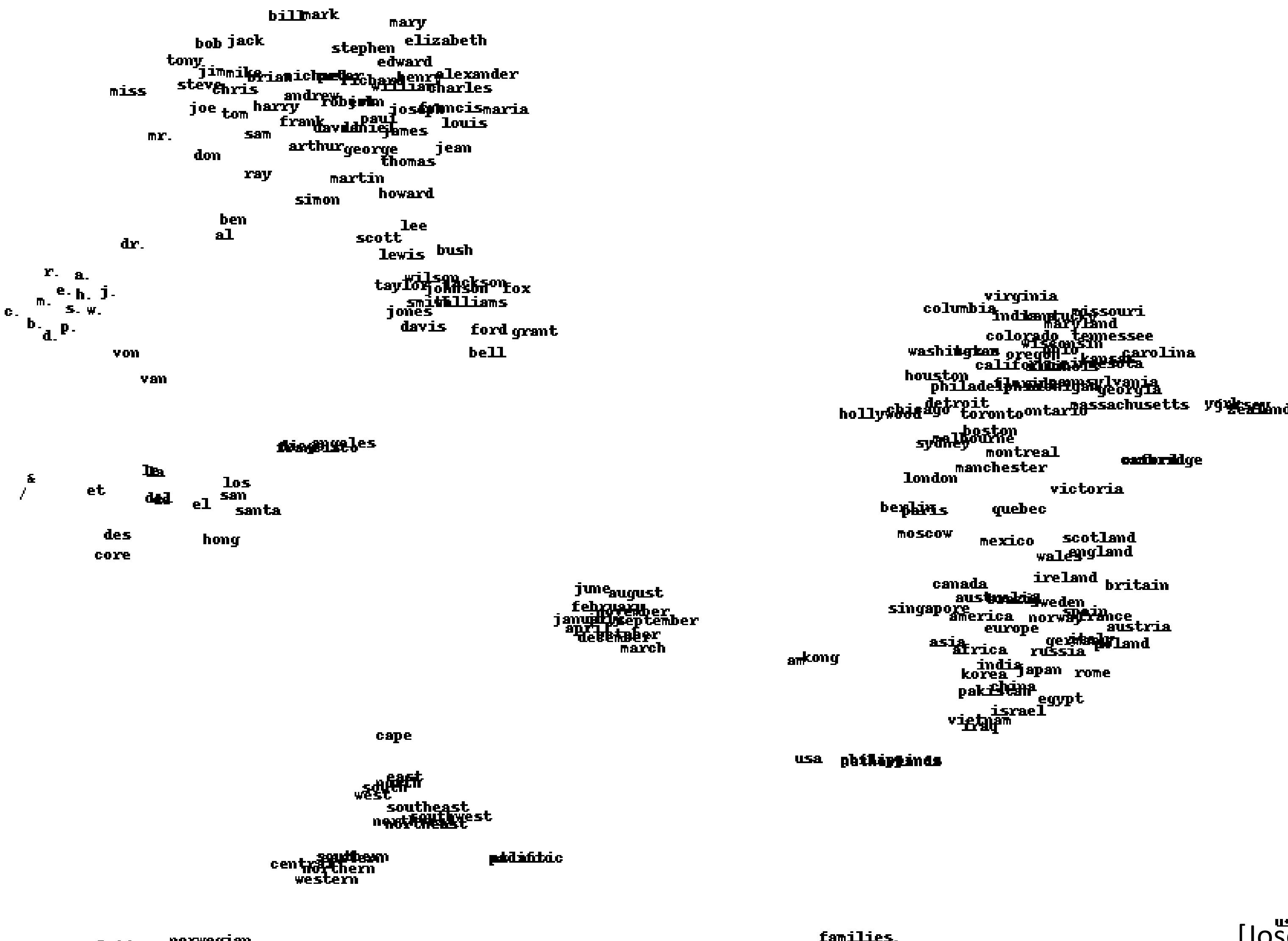
[Saul & Roweis '03]

Embedding words



[Joseph Turian]

Embedding words (zoom in)

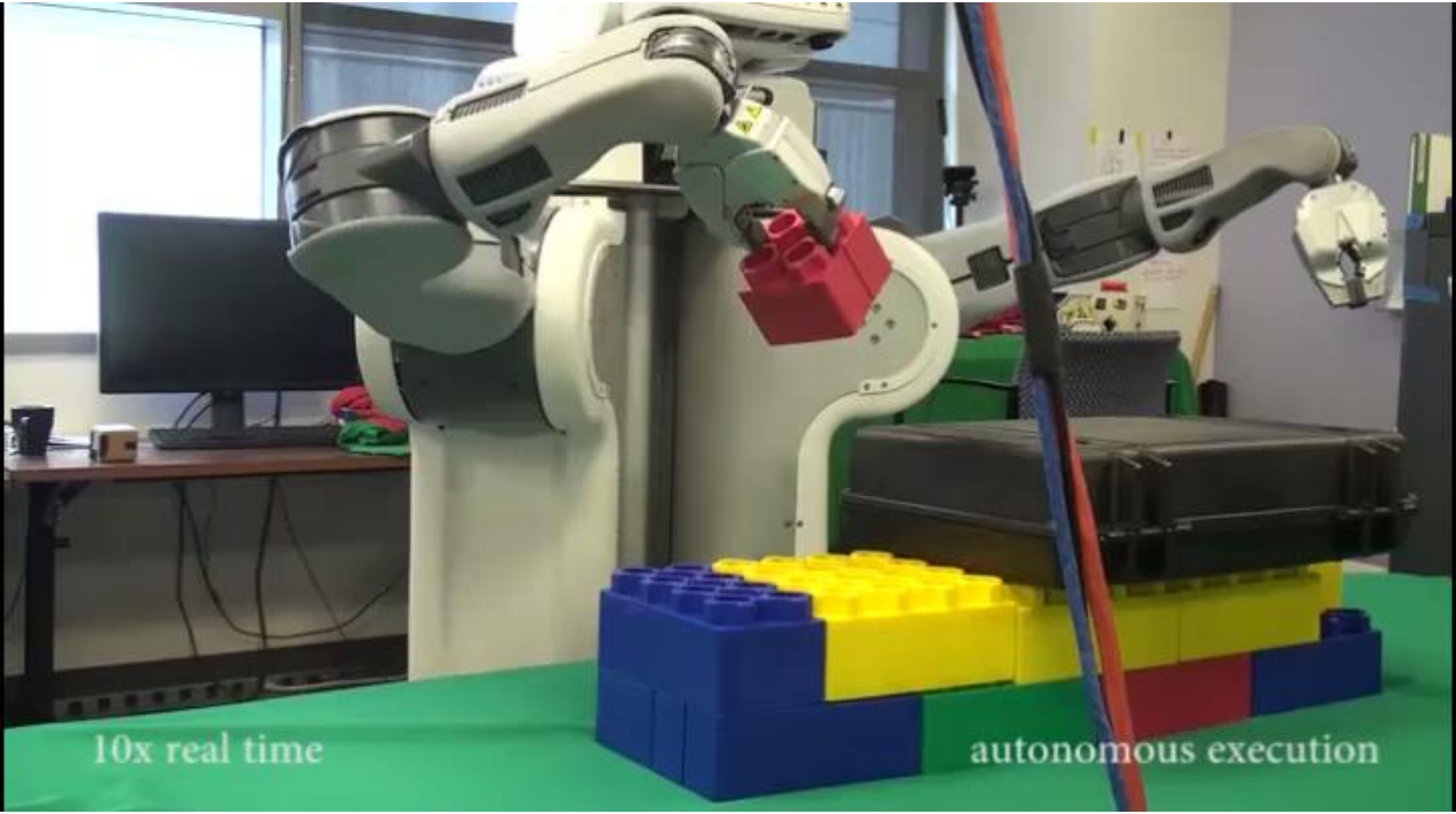


[Joseph Turian]

Reinforcement Learning

training by feedback

Learning to act



Taxonomy

- Three basic problem settings
  - supervised learning
    - predict y from x
    - classification, regression
  - unsupervised learning
    - model p(x), evaluate how likely x is, understand x
    - clustering, embedding
  - reinforcement learning
    - learn to make decisions
    - really just a generalization of supervised learning with weak supervision

Supervised Learning: find f

- Given: Training set $\{(x_i, y_i) \mid i = 1 \dots n\}$
- Find: A good approximation to $f : X \rightarrow Y$

Examples: what are X and Y ?

- Spam Detection
  - Map email to {Spam,Ham}
- Digit recognition
  - Map pixels to {0,1,2,3,4,5,6,7,8,9}
- Stock Prediction
  - Map new, historic prices, etc. to (the real numbers)

Example: Spam Filter

- Input: email
- Output: spam/ham
- Setup:
  - Get a large collection of example emails, each labeled "spam" or "ham"
  - Note: someone has to hand label all this data!
  - Want to learn to predict labels of new, future emails

- Features: The attributes used to make the ham / spam decision
  - Words: FREE!
  - Text Patterns: CAPS
  - Non-text: SenderInContacts
  - …

Dear Sir.

First, I must solicit your confidence in this transaction, this is by virture of its nature as being utterly confidencial and top secret. …

TO BE REMOVED FROM FUTURE MAILINGS, SIMPLY REPLY TO THIS MESSAGE AND PUT "REMOVE" IN THE SUBJECT.

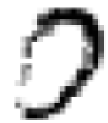
99 MILLION EMAIL ADDRESSES FOR ONLY $99

Ok, Iknow this is blatantly OT but I'm beginning to go insane. Had an old Dell Dimension XPS sitting in the corner and decided to put it to use, I know it was working pre being stuck in the corner, but when I plugged it in, hit the power nothing happened.

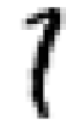
Example: Digit Recognition

- Input: images / pixel grids
- Output: a digit 0-9
- Setup:
  - Get a large collection of example images, each labeled with a digit
  - Note: someone has to hand label all this data!
  - Want to learn to predict labels of new digit images

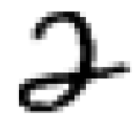
- Features: The attributes used to make the digit decision
  - Pixels: (6,8)=ON
  - Shape Patterns: NumComponents, AspectRatio, NumLoops
  - …



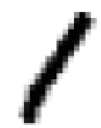
0



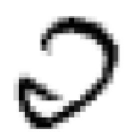
1



2



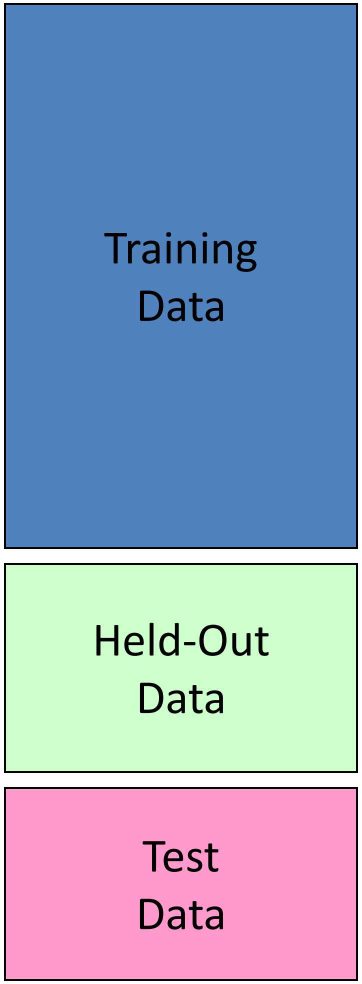
1



??

Important Concepts

- Data: labeled instances, e.g. emails marked spam/ham
  - Training set
  - Held out set (sometimes call Validation set)
  - Test set
- Features: attribute-value pairs which characterize each x
- Experimentation cycle
  - Select a hypothesis f to best match training set
  - (Tune hyperparameters on held-out set)
  - Compute accuracy of test set
  - Very important: never "peek" at the test set!
- Evaluation
  - Accuracy: fraction of instances predicted correctly
- Overfitting and generalization
  - Want a classifier which does well on *test* data
  - **Overfitting**: fitting the training data very closely, but not generalizing well
  - We'll investigate overfitting and generalization formally in a few lectures



A Supervised Learning Problem

- Consider a simple, Boolean dataset:
  - $f : X \rightarrow Y$
  - $X = \{0,1\}^4$
  - $Y = \{0,1\}$
- Question 1: How should we pick the *hypothesis space*, the set of possible functions f ?
- Question 2: How do we find the best f in the hypothesis space?

Dataset:

Example	x_1	x_2	x_3	x_4	y
1	0	0	1	0	0
2	0	1	0	0	0
3	0	0	1	1	1
4	1	0	0	1	1
5	0	1	1	0	0
6	1	1	0	0	0
7	0	1	0	1	0

A Restricted Hypothesis Space

Consider all conjunctive boolean functions.

- 16 possible hypotheses
- None are consistent with our dataset
- How do we choose the best one?

Rule	Counterexample
$\Rightarrow y$	1
$x_1 \Rightarrow y$	3
$x_2 \Rightarrow y$	2
$x_3 \Rightarrow y$	1
$x_4 \Rightarrow y$	7
$x_1 \wedge x_2 \Rightarrow y$	3
$x_1 \wedge x_3 \Rightarrow y$	3
$x_1 \wedge x_4 \Rightarrow y$	3
$x_2 \wedge x_3 \Rightarrow y$	3
$x_2 \wedge x_4 \Rightarrow y$	3
$x_3 \wedge x_4 \Rightarrow y$	4
$x_1 \wedge x_2 \wedge x_3 \Rightarrow y$	3
$x_1 \wedge x_2 \wedge x_4 \Rightarrow y$	3
$x_1 \wedge x_3 \wedge x_4 \Rightarrow y$	3
$x_2 \wedge x_3 \wedge x_4 \Rightarrow y$	3
$x_1 \wedge x_2 \wedge x_3 \wedge x_4 \Rightarrow y$	3

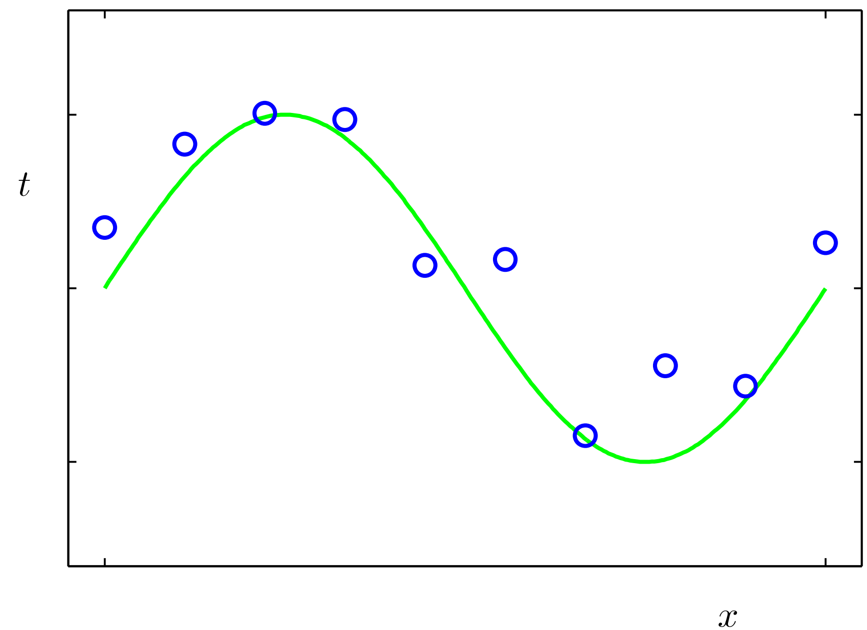
Dataset:

Example	x_1	x_2	x_3	x_4	y
1	0	0	1	0	0
2	0	1	0	0	0
3	0	0	1	1	1
4	1	0	0	1	1
5	0	1	1	0	0
6	1	1	0	0	0
7	0	1	0	1	0

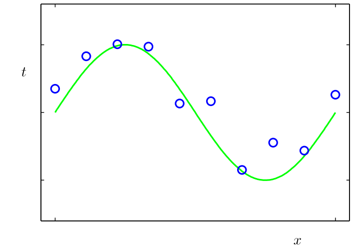
Another Sup. Learning Problem

- Consider a simple, regression dataset:
  - $f : X \rightarrow Y$
  - $X =$ Â
  - $Y =$ Â
- Question 1: How should we pick the *hypothesis space*, the set of possible functions f ?
- Question 2: How do we find the best f in the hypothesis space?

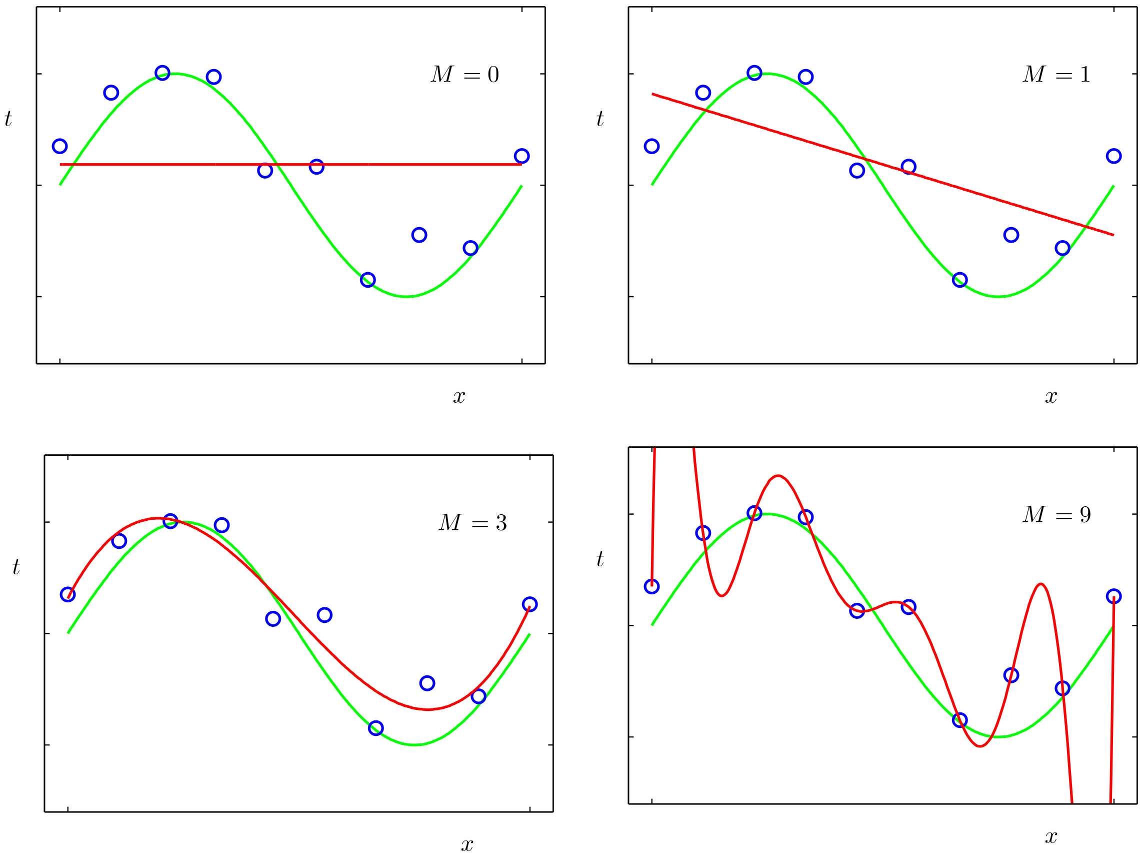
Dataset: 10 points generated from a sin function, with noise

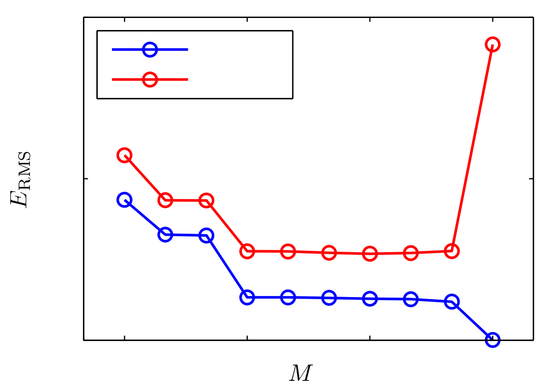


Hypo. Space: Degree-N Polynomials



- Infinitely many hypotheses
- None / Infinitely many are consistent with our dataset
- How do we choose the best one?





Key Issues in Machine Learning

- What are good hypothesis spaces?
- How to find the best hypothesis?
- How to optimize for accuracy of unseen testing data? (avoid overfitting, etc.)
- Can we have confidence in results? How much data is needed?
- How to model applications as machine learning problems? (engineering challenge)

Summary & Takeaways

- Parts of a machine learning algorithm
  - Data (input x and output y)
  - Hypothesis space (e.g. all boolean functions)
  - Objective (what makes one “incorrect” answer better/worse than another “incorrect” answer)
  - Algorithm (how do we get the least “incorrect” answer)
- Important concepts
  - Want high accuracy on unseen test data, but only get to see training data (this is why ML is almost but not quite the same thing as optimization)
  - Central challenge in machine learning: generalization

Course Summary

- Week 1: introduction, decision trees
- Week 2: finish decision trees, point estimation
- Week 3: linear regression
- Week 4: naïve Bayes
- Week 5: neural networks
- Week 6: learning theory
- Week 7: model ensambles
- Week 8: clustering & dimensionality reduction
- Week 9: support vector machines (SVMs)
- Week 10: reinforcement learning, last-minute gift ideas