CSE 446
Machine Learning

Instructor: Ali Farhadi

ali@cs.washington.edu

Slides adapted from Pedro Domingos, Carlos Guestrin, and Luke Zettelmoyer

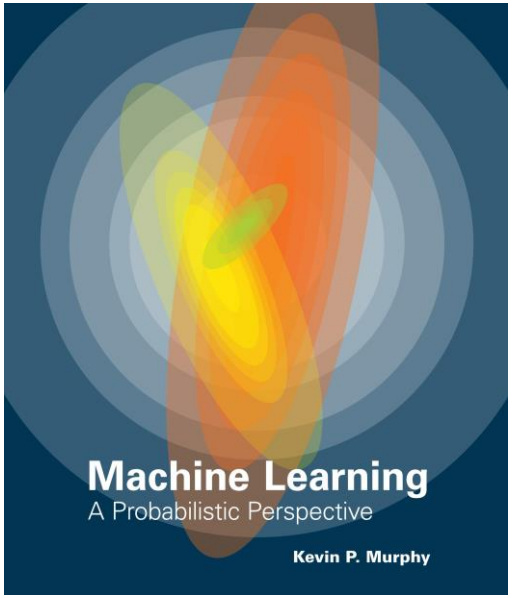
Logistics

- **Instructor:** Ali Farhadi
  - Email: ali@cs
  - Office: CSE 652
- **TAs:**
  - Dae Hyun Lee (dhlee4@u)
  - Jianyang Zhang (jianyz@uw)
  - Yue Zhang (yjzhang@uw)
  - Deric Pang (dericp@uw)
  - Qiang (Andrew) Yu (shift90@uw)

- **Web:** http://courses.cs.washington.edu/courses/cse446/17sp/
- Please read website carefully for academic integrity, late policy, etc.

Textbooks

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

Syllabus Overview:

- 3/28
  - Introduction
- 3/30
  - Decision Trees
- 4/4
  - Decision Trees
- 4/6
  - Point Estimation
  - Homework1 is available.
- 4/11
  - Linear Regression
- 4/13
  - Linear Regression
- 4/18
  - Naive Bayes
- 4/20
  - Homework1 is due before the class
  - Naive Bayes
  - Homework2 is available [.pdf][.tex][images]
- 4/25
  - logistic Regression
- 4/27
  - logistic Regression
- 5/2
  - logistic Regression
- 5/4
  - Perceptron
  - Homework2 is due before the class
  - Homework 3 is available
- 5/9
  - Perceptron
- 5/11
  - Support Vector Machines

- 5/16
  - Kernels
- 5/18
  - Boosting
  - Homework3 is due before the class. [Use Dropbox to submit your homework]
  - Homework 4 is available. [.pdf][.tex][country data]
- 5/23
  - Clustering
- 5/25
  - EM
- 5/30
  - Neural Networks
- 6/1
  - Neural Networks
  - Homework4 is due 11:59PM. [Use Dropbox to submit your homework]

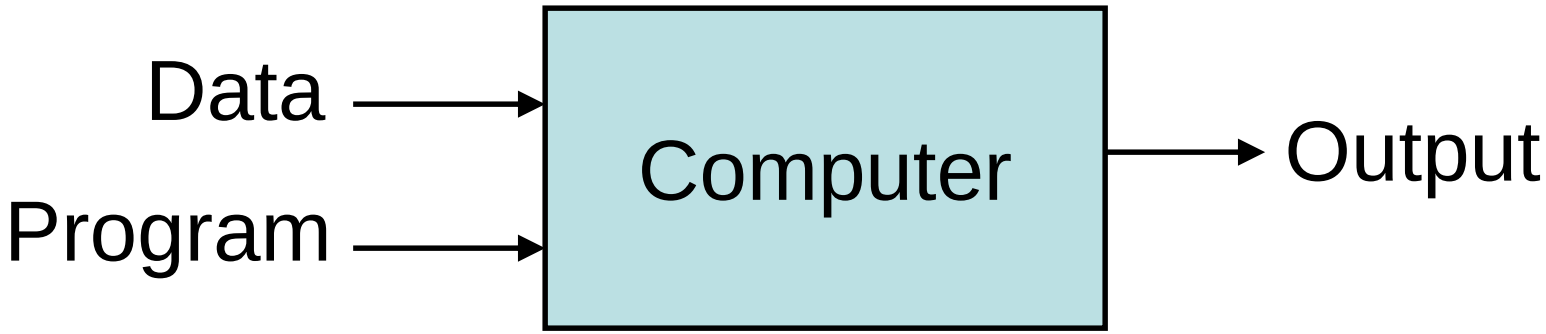
A Few Quotes

- “A breakthrough in machine learning would be worth ten Microsofts” (Bill Gates, Chairman, Microsoft)
- “Machine learning is the next Internet”
  (Tony Tether, Director, DARPA)
- Machine learning is the hot new thing”
  (John Hennessy, President, Stanford)
- “Web rankings today are mostly a matter of machine learning” (Prabhakar Raghavan, Dir. Research, Yahoo)
- “Machine learning is going to result in a real revolution”
  (Greg Papadopoulos, CTO, Sun)
- “Machine learning is today’s discontinuity”
  (Jerry Yang, CEO, Yahoo)

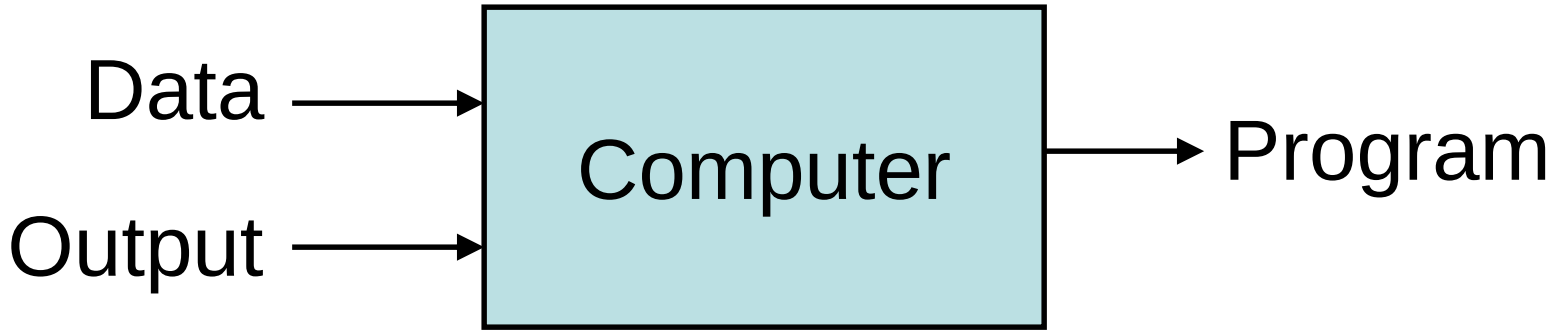
So What Is Machine Learning?

- Automating automation
- Getting computers to program themselves
- Writing software is the bottleneck
- Let the data do the work instead!
- The future of Computer Science!!!

Traditional Programming



Machine Learning



Magic?

No, more like gardening

- **Seeds** = Algorithms
- **Nutrients** = Data
- **Gardener** = You
- **Plants** = Programs

What is Machine Learning ?
(by examples)

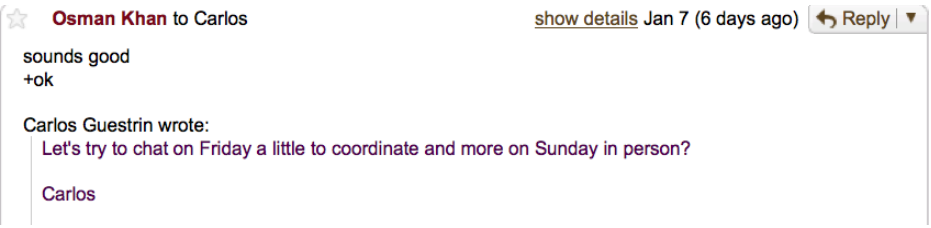
Classification

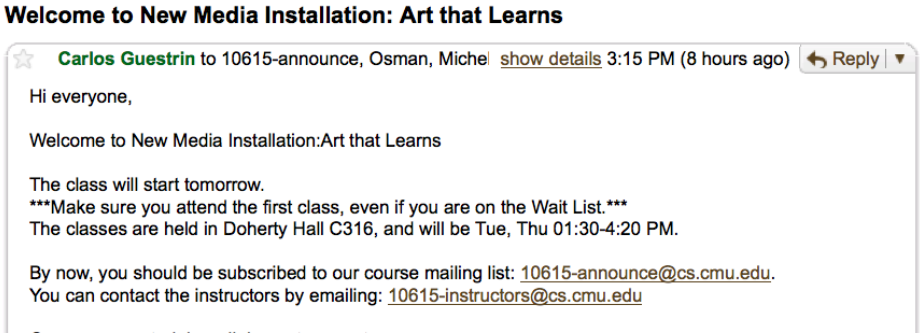
from data to discrete classes

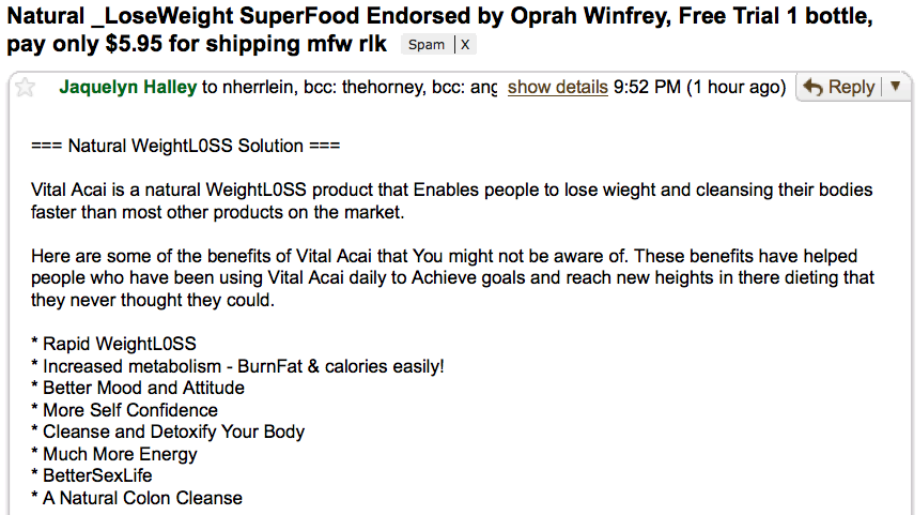
Spam filtering

data

prediction





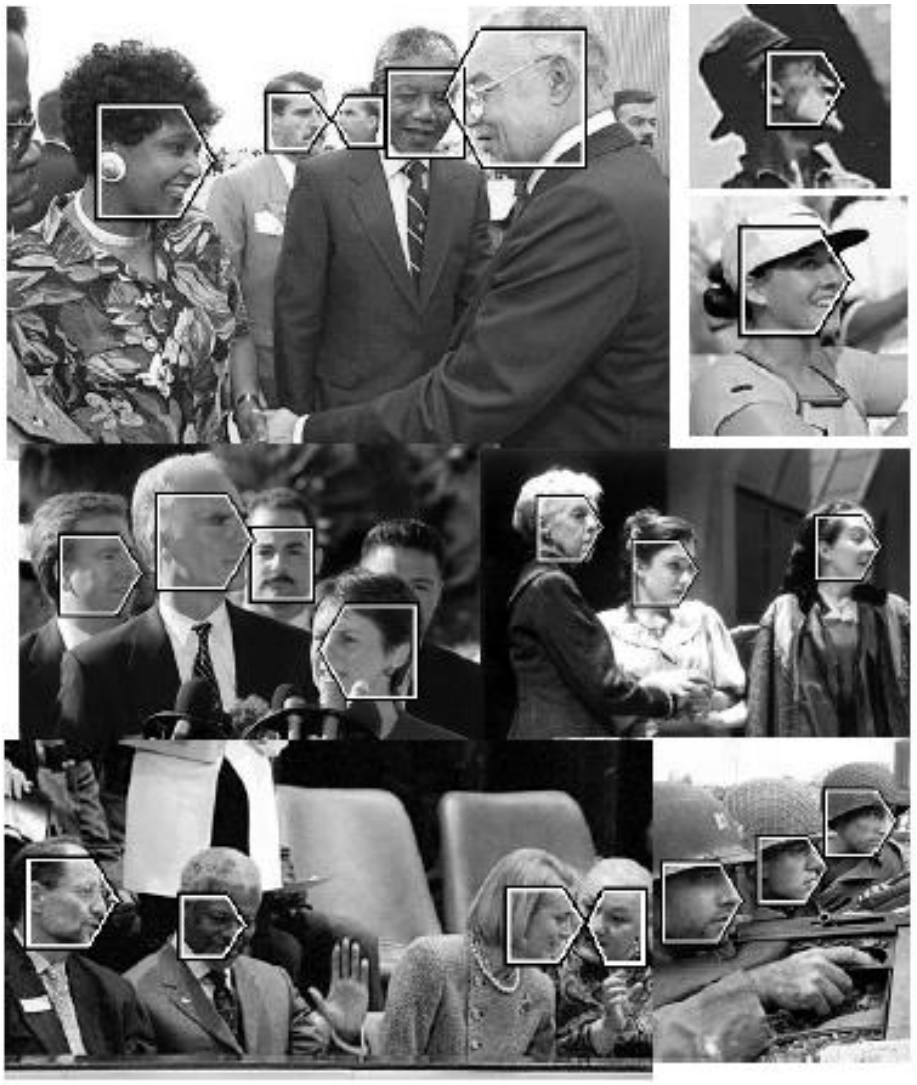


Spam
vs
Not Spam

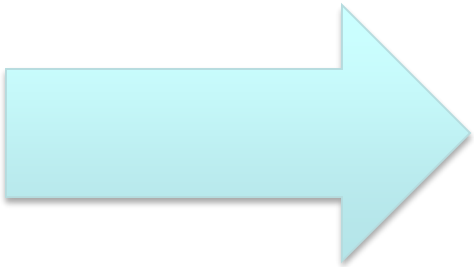
Object detection

(Prof. H. Schneiderman)

Example training images
for each orientation



Weather prediction

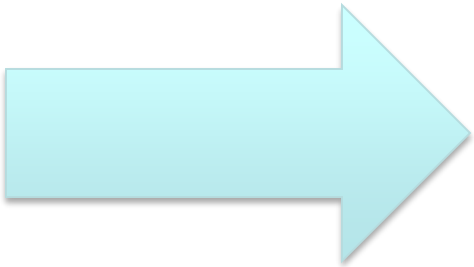


Regression

predicting a numeric value

Stock market

Weather prediction revisited

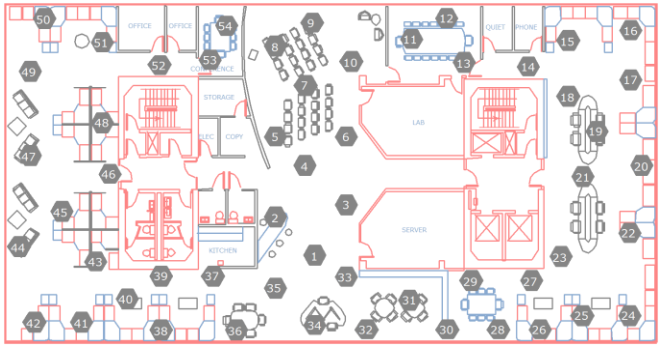


Temperature

72° F

Modeling sensor data

- Measure temperatures at some locations
- Predict temperatures throughout the environment

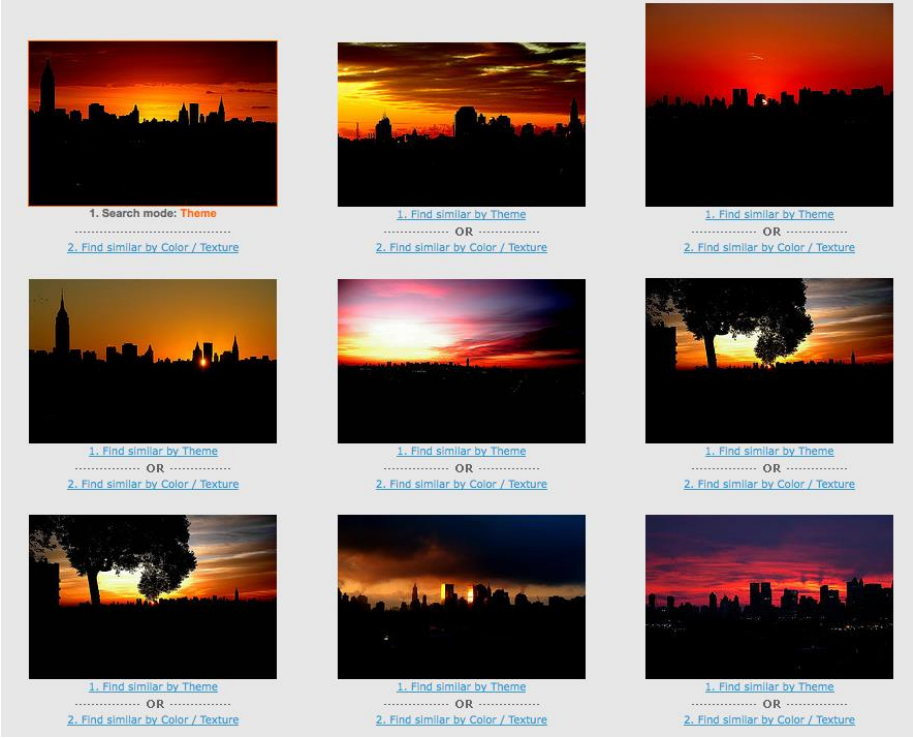


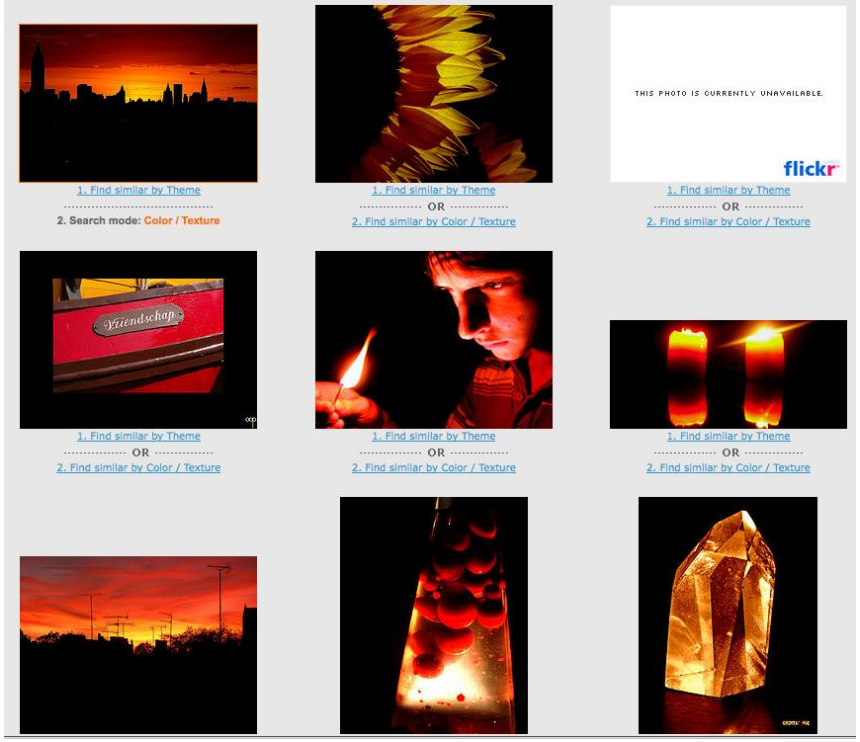
[Guestrin et al. ’04]

Similarity

finding data

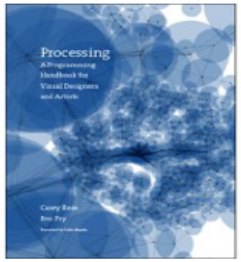
Given image, find similar images





http://www.tiltomo.com/

Collaborative Filtering



See larger image

Share your own customer images

Publisher: learn how customers can search inside this book.

Please tell the publisher:
I'd like to read this book on Kindle

Don't have a Kindle? Get yours here.

Processing: A Programming Handbook for Visual Designers and Artists (Hardcover)
by Casey Reas (Author), Ben Fry (Author), John Maeda (Foreword)
★★★★★ (13 customer reviews)

Available from these sellers.

31 new from $47.95 **8 used** from $43.56

Get Free Two-Day Shipping
Get Free Two-Day Shipping for three months with a special extended free trial of Amazon Prime™. Add this eligible textbook to your cart to qualify. Sign up at checkout. See details.

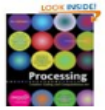
Related Education & Training Services in Pittsburgh (What's this?) | Change location

Learn HTML Coding
www.FullSail.edu - Earn Your Bachelor's Degree in Web Design and Development.

Create Websites with HTML
http://www.unex.Berkeley.edu - Learn HTML Online, Start Anytime! with UC Berkeley Extension

Intensive XSLT Training
www.objectdatalabs.com/course10.asp - OnSite or in NYC, LA, SFO, ORD, DC Will customize & train as few as 3

Customers Who Bought This Item Also Bought



Processing: Creative Coding and Computational A... by Ira Greenberg
★★★★½ (7) $43.99

Visualizing Data: Exploring and Explaining Data... by Ben Fry
★★★½☆ (11) $26.39

Making Things Talk: Practical Methods for Conne... by Tom Igoe
★★★★½ (15) $19.79

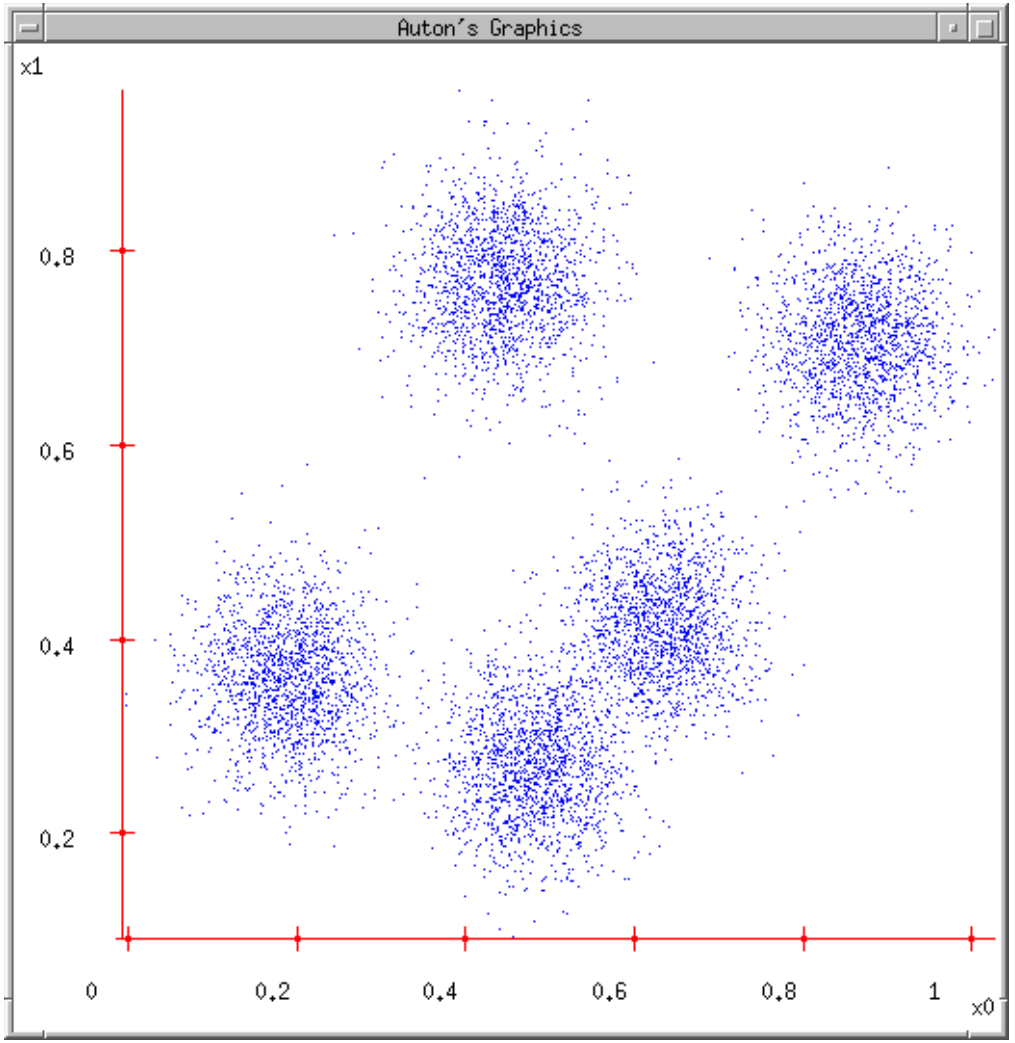
Physical Computing: Sensing and Controlling the... by Tom Igoe
★★★★½ (20) $19.00

Learning Processing: A Beginner's Guide to... by Daniel Shiffman
★★★★★ (7) $44.95

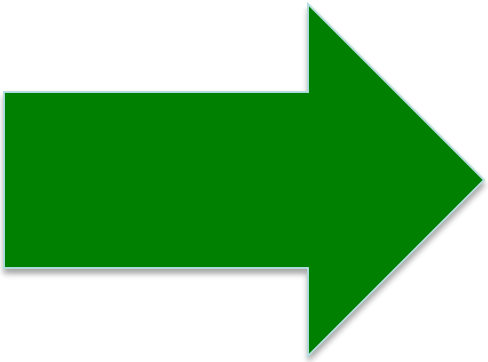
Clustering

discovering structure in data

Clustering Data: Group similar things



Clustering images



[Goldberger et al.]

Clustering News

U.S. edition | Modern

Top Stories

CNN International

See realtime coverage

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

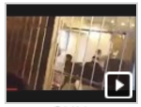
A reckless regime Washington Post

Highly Cited: Iranian Protesters Ransack Saudi Embassy After Execution of Shiite Cleric New York Times

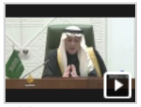
From Saudi Arabia: Saudi Arabia severs Iran ties Arab News

Wikipedia: Nimr al-Nimr

Related
Saudi Arabia »
Sheikh Nimr »
Iran »



CNN



Aljazeera.com

YouTube

Washingt…

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.

Firstpost

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.

CBS News

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.

Press He…

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.

The Bost…

In NH, Clinton hits on opioid abuse as a top concern

The Boston Globe - 2 hours ago

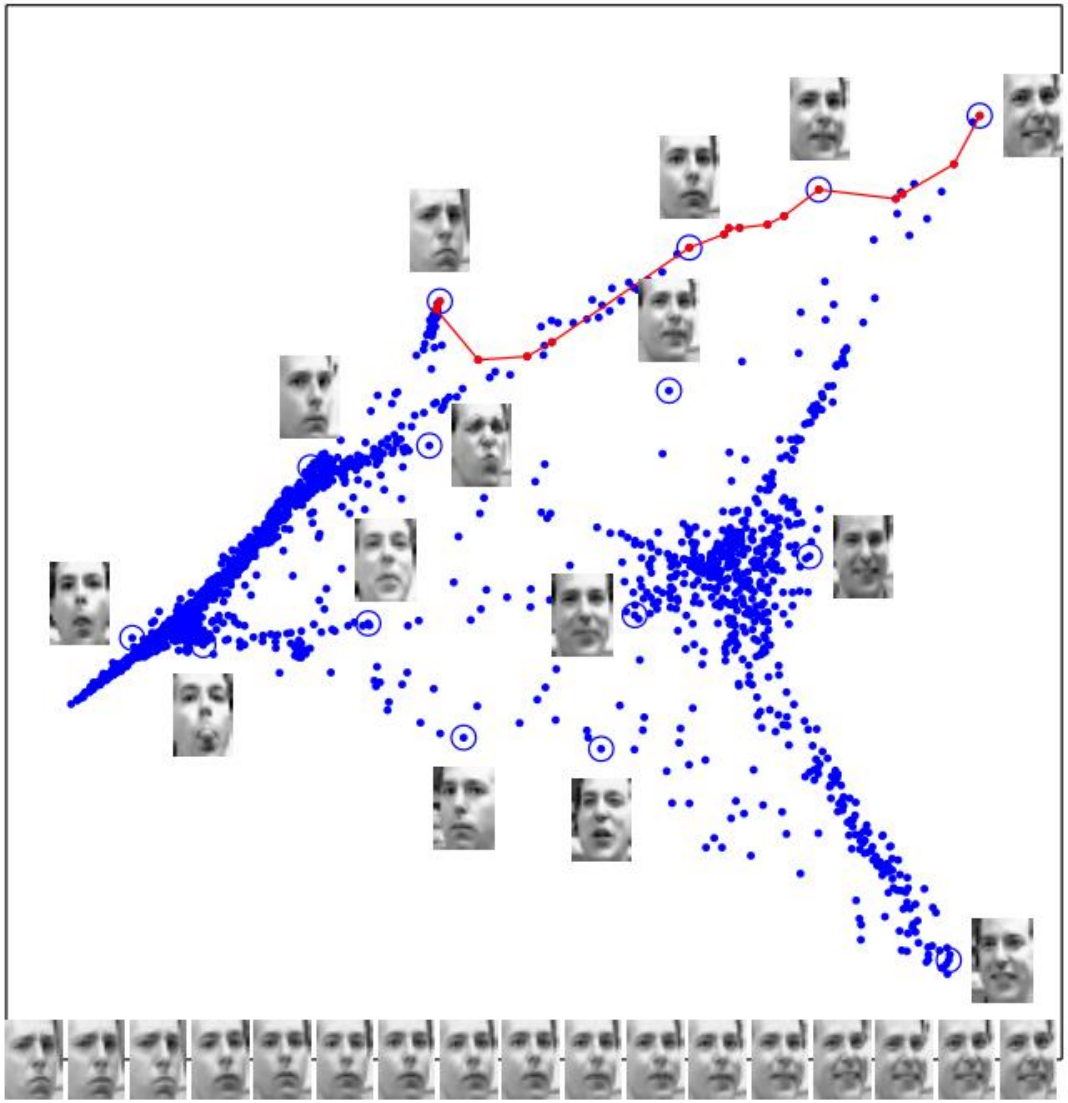
DERRY, N.H. - Hillary Clinton, who arrived to loud applause here at one of three New Hampshire campaign stops Sunday, said prohibitively expensive education, lack of support for families coping with Alzheimer's disease, and the rising tide of opioid ...

Embedding

visualizing data

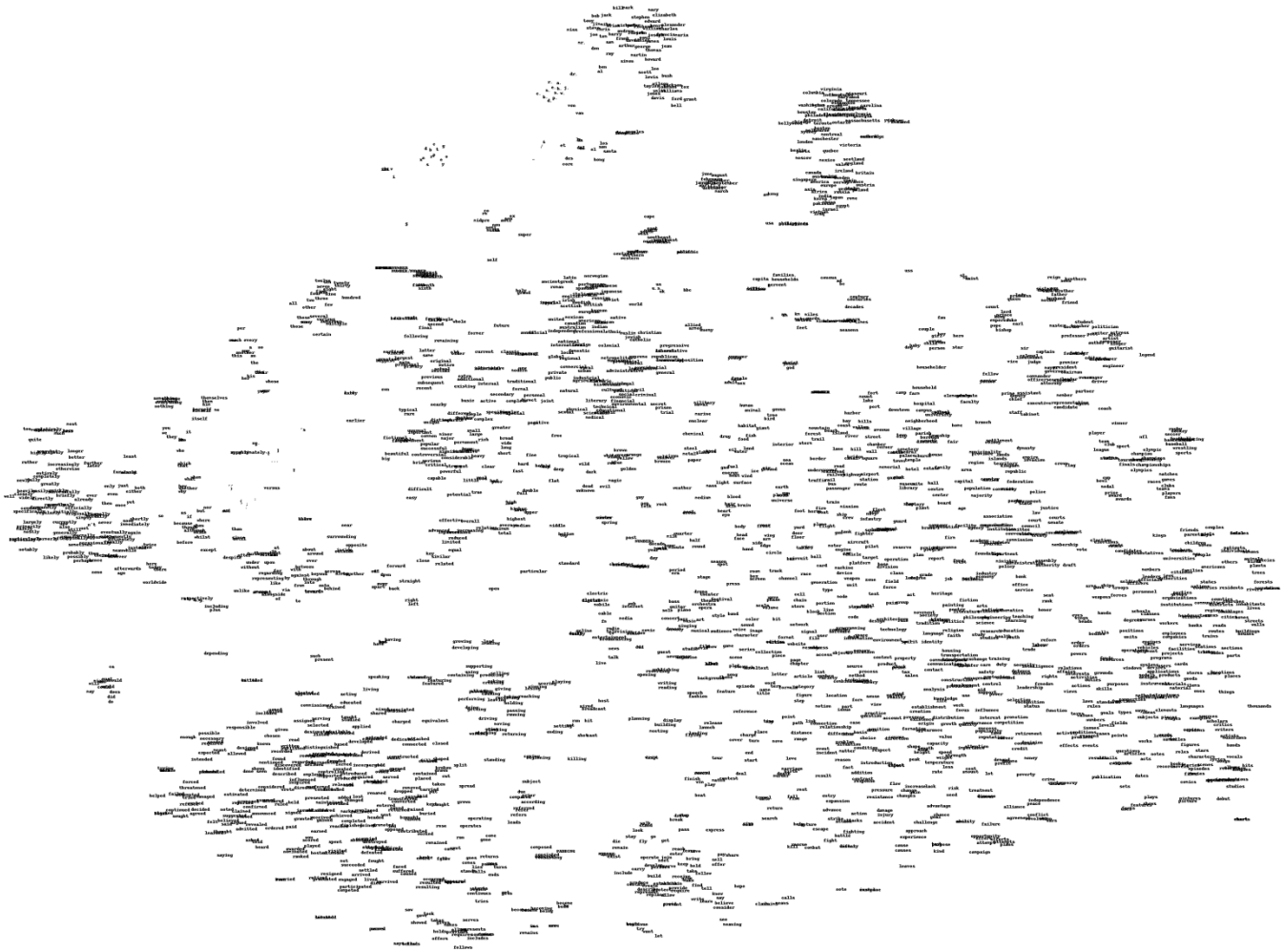
Embedding images

- 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

- Reinforcement learning
- An agent
  - Makes sensor observations
  - Must select action
  - Receives rewards
    - positive for “good” states
    - negative for “bad” states

Growth of Machine Learning

- Machine learning is preferred approach to
  - Speech recognition, Natural language processing
  - Computer vision
  - Medical outcomes analysis
  - Robot control
  - Computational biology
  - Sensor networks
  - …
- This trend is accelerating
  - Improved machine learning algorithms
  - Improved data capture, networking, faster computers
  - Software too complex to write by hand
  - New sensors / IO devices
  - Demand for self-customization to user, environment

Supervised Learning: find f

- Given: Training set $\{(x_i, y_i) \mid i = 1 \ldots n\}$
- 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: $dd, 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, future digit images
- Features: The attributes used to make the digit decision
  - Pixels: (6,8)=ON
  - Shape Patterns: NumComponents, AspectRatio, NumLoops
  - …

0

1

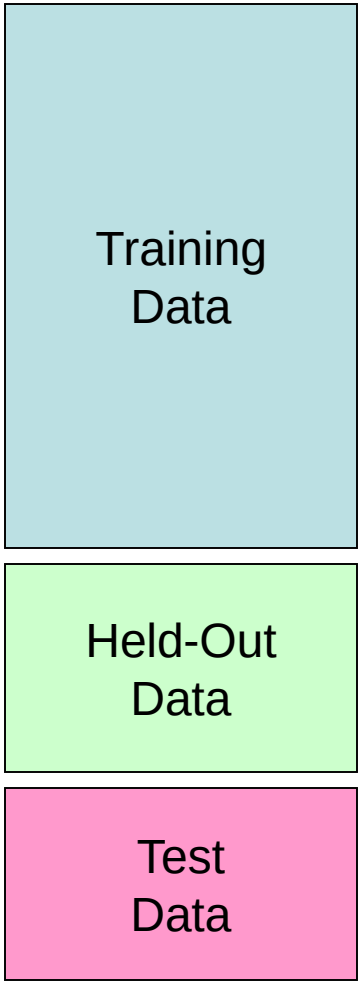
2

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

Most General Hypothesis Space

Consider all possible boolean functions over four input features!

- 2^{16} possible hypotheses
- 2^{9} are consistent with our dataset
- How do we choose the best one?

x_1	x_2	x_3	x_4	y
0	0	0	0	?
0	0	0	1	?
0	0	1	0	0
0	0	1	1	1
0	1	0	0	0
0	1	0	1	0
0	1	1	0	0
0	1	1	1	?
1	0	0	0	?
1	0	0	1	1
1	0	1	0	?
1	0	1	1	?
1	1	0	0	0
1	1	0	1	?
1	1	1	0	?
1	1	1	1	?

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:

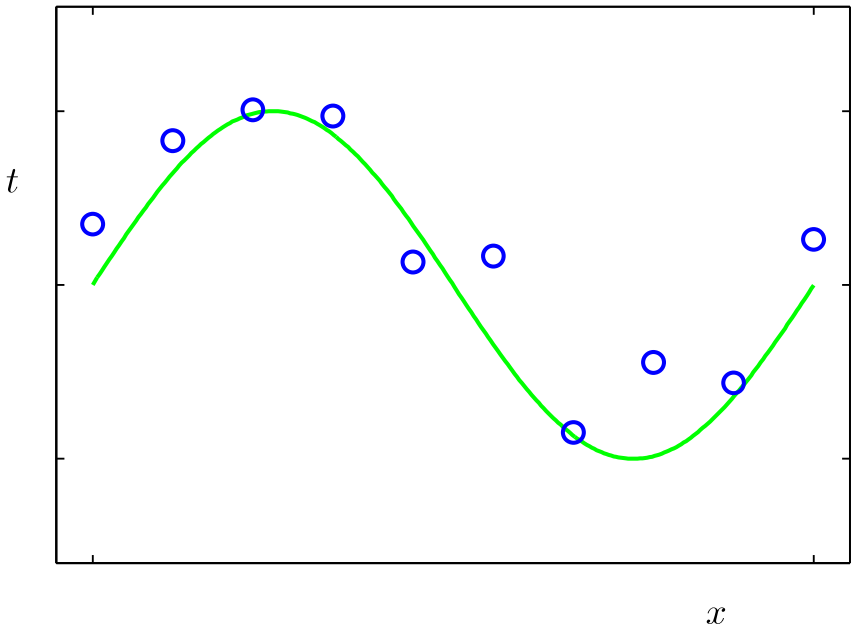
Example	x_1	x_2	x_3	x_4	y
1	0	0	1	0	0
2	0	1	0	0	0
3	0	0	1	1	1
4	1	0	0	1	1
5	0	1	1	0	0
6	1	1	0	0	0
7	0	1	0	1	0

Another Sup. Learning Problem

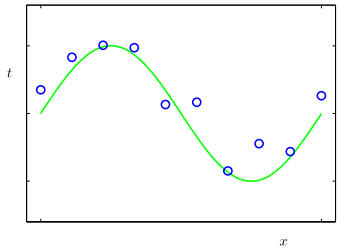
- Consider a simple, regression dataset:
  - $f : X \rightarrow Y$
  - $X = \hat{A}$
  - $Y = \hat{A}$

- 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?

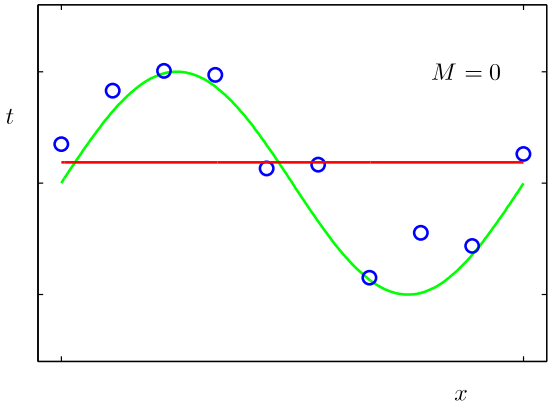
Dataset: 10 points generated from a sin function, with noise

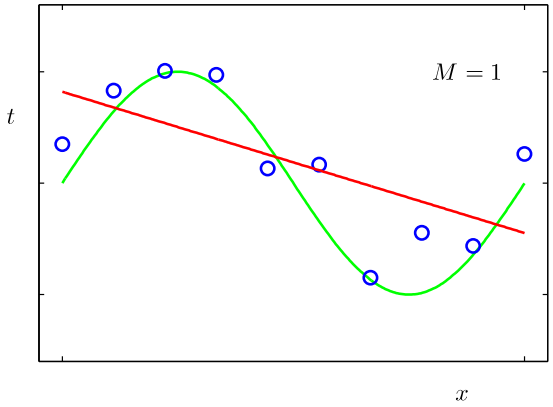


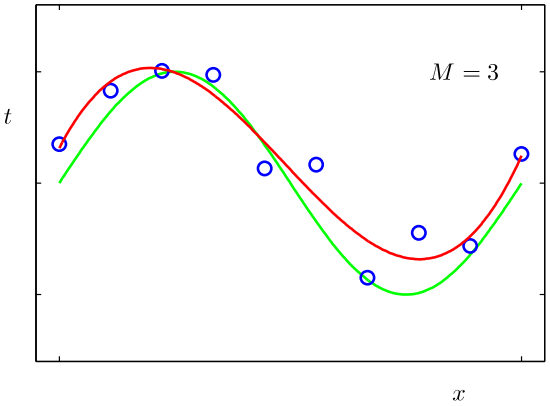
Hypo. Space: Degree-N Polynomials

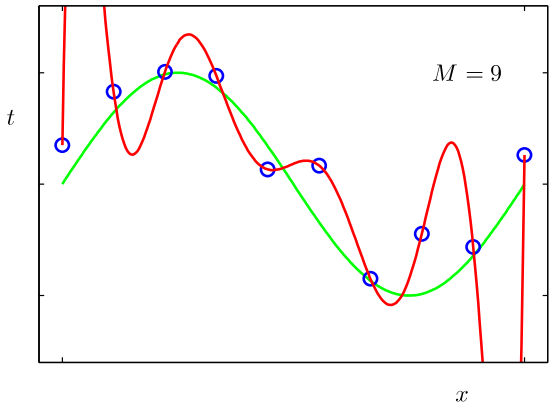


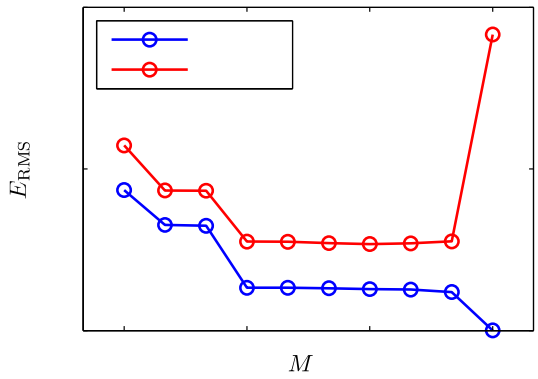
- Infinitely many hypotheses
- None / Infinitely many are consistent with our dataset
- How do we choose the best one?











Key Issues in Machine Learning

- What are good hypothesis spaces?
- How to find the best hypothesis? (algorithms / complexity)
- 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)

Logistics: Evaluation

- 4 homeworks (70% total)
  - Assigned in weeks 2,4,6,8
  - Due two weeks later
  - **Can take time: start early!!!!**
- Final example (25%)
- Course participation (5%)
  - includes in class, message board, etc.

Homeworks

- HW1: Decision Trees
  - Release: 4/6 , Due: 4/20
- HW2: Classifiers
  - Release: 4/20, Due: 5/4
- HW3: SVMs and Ensembles
  - Release: 5/4, Due: 5/18
- HW4: Clustering and dimensionality Reduction
  - Release: 5/18, Due: 6/1

Calibration

- Linear Algebra
- Eigenvectors
- Covariance
- Entropy
- Conditional Entropy
- Least Squares
- Gradient
- Gradient descent

CS
EE
Math
Stat
Others
Year