

CSE 4/574 Introduction to Machine Learning

Mingchen Gao
mgao8@buffalo.edu
Davis 347

Slides adapted from Varun Chandola and Luke Zettlemoyer

Logistics

- Class webpage
 - <https://piazza.com/buffalo/fall2022/cse4574/info>
 - Lectures will be streamed and recorded.
- Office hours (on zoom)
 - <https://piazza.com/buffalo/fall2022/cse4574/staff>
 - Thursday 10 am – 12pm
- TAs
 - Graduate TAs
 - Wei Bo, weibo@buffalo.edu
 - Mingxi Lei, mingxile@buffalo.edu
 - Meng Ding, mengding@buffalo.edu

Prerequisites

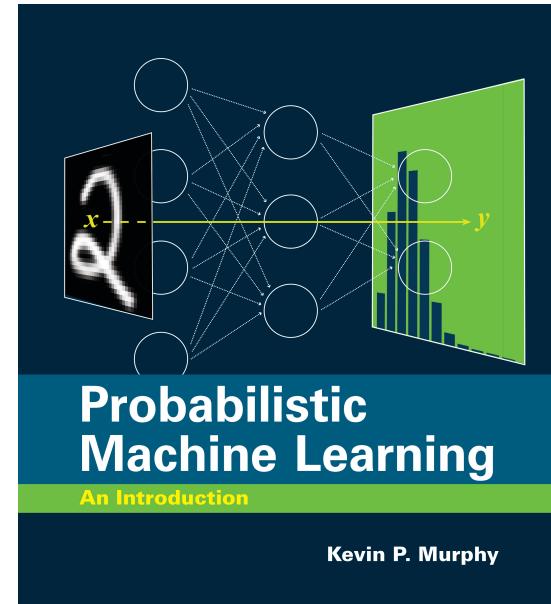
- Probability and Statistics
- Linear Algebra
- Python Programming

Topics Covered

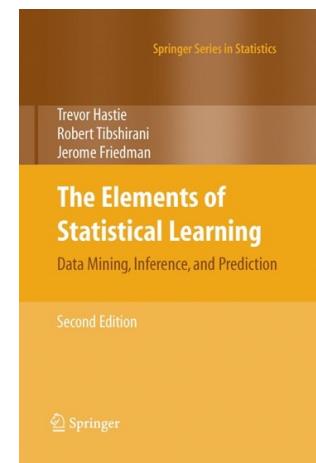
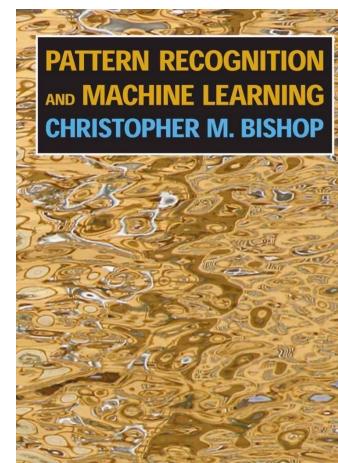
- Machine Learning Tools:
 - Bayesian Learning
 - Expectation Maximization
 - Optimization
- Machine Learning Algorithms:
 - Generative Models
 - Regression
 - Logistic Regression
 - Perceptron and Neural Networks
 - Convolutional Neural Networks, Recurrent Neural Networks
 - Graphical Models
 - Latent Linear Models
 - Support Vector Machines

Textbooks

- Kevin Murphy, Probabilistic Machine Learning, MIT Press, 2022.
<https://probml.github.io/pml-book/book1.html>



- Reference textbooks:
 - Christopher Bishop, Pattern Recognition and Machine Learning, Springer, 2007
 - Tom Mitchell, Machine Learning, McGraw-Hill, 1997
 - Trevor Hastie, Robert Tibshirani and Jerome Friedman, The Elements of Statistical Learning. Springer, 2009.



Evaluation

- Short weekly quizzes using Gradiance (12) – 20%
- Programming Assignments (3) – 30%
 - Group project of up to three students
 - Python, use UBLearn for all electronic submissions
- Mid-term Exam (in-class, open book) – 20%
 - Tentative 10/17/2022 Mon, in class
- Final Exam (open book) – 30%
 - 12/15/2022 Thu, 8am – 11am

Final Grade

Graduate

A [92.5; 100]
A- [87.5; 92.5)
B+ [82.5; 87.5)
B [77.5; 82.5)
B- [72.5; 77.5)
C+ [67.5; 72.5)
C [62.5; 67.5)
C- [57.5; 62.5)
D+ [52.5; 57.5)
D [0; 52.5)

Undergraduate

A [87.5; 100]
A- [82.5; 87.5)
B+ [77.5; 82.5)
B [72.5; 77.5)
B- [67.5; 72.5)
C+ [62.5; 67.5)
C [57.5; 62.5)
C- [52.5; 57.5)
D+ [47.5; 52.5)
D [0; 47.5)

Gradiance

- An online quiz system
- One quiz per week released on Monday by 8:59 am and due next Sunday by 11:59 pm
- 2-3 multiple choice problems about topics covered that week
- A warm up quiz (ungraded) is posted
- 10-minute delay between successive submissions
- Only 3 tries allowed, maximum score will be used
- Every wrong answer will result in 1 negative point per try

Class token: C0C2B237 use your UB email username
<http://www.newgradiance.com/services>.

Piazza

- Primary medium of communication
- All announcements, teaching slides, assignments, etc. will be made available through Piazza.
- Questions?
- General post to all (Name will be visible).
- Choose appropriate folder.
- Private post to instructor, TA.
- Interact.

Office of Academic Integrity

<https://www.buffalo.edu/academic-integrity/instructors.html>

Academic Integrity Process



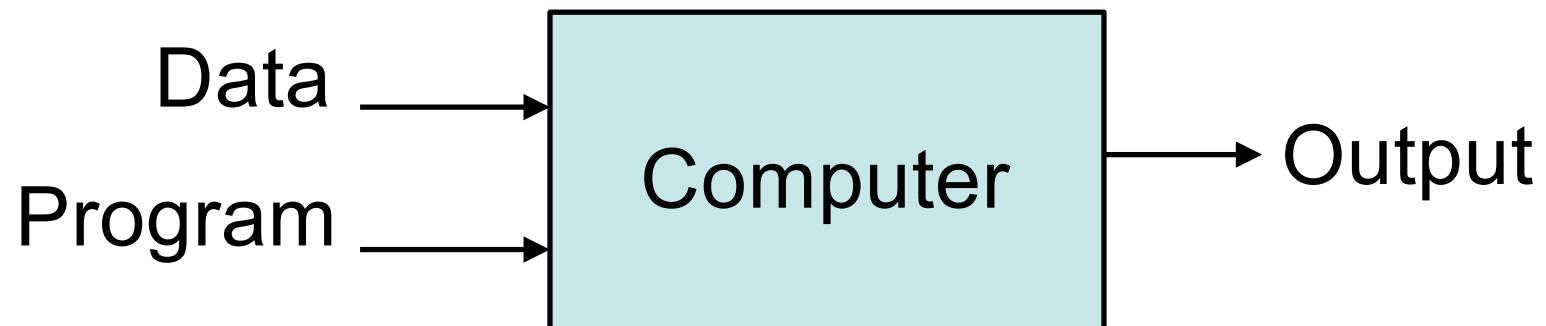
A Few Quotes

- “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)

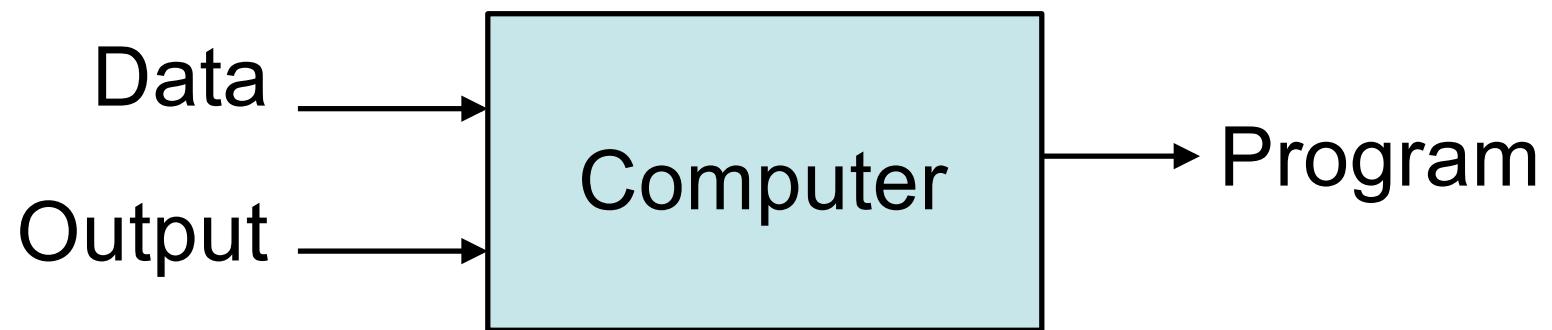
So What Is Machine Learning?

- A set of methods
 - Can automatically detect patterns in data
 - Use uncovered patterns to predict future data under uncertainty
- Let the data do the work instead!
- The future of Computer Science!!!

Traditional Programming



Machine Learning



What is Machine Learning ?

(by examples)

Classification

from data to discrete classes

Spam filtering

data

Osman Khan to Carlos
sounds good
+ok

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

Carlos

prediction



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 WeightLOSS Solution ===

Vital Acai is a natural WeightLOSS 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 WeightLOSS
- * 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



Spam
vs
Not Spam

Image classification

label = 5



label = 0



label = 4



label = 1



label = 9



label = 2



label = 1



label = 3



label = 1



label = 4



label = 3



label = 5



label = 3



label = 6



label = 1



label = 7



label = 2



label = 8



label = 6



label = 9



Object detection

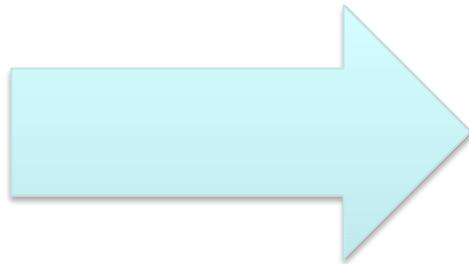
(Prof. H. Schneiderman)



Example training images
for each orientation



Weather prediction



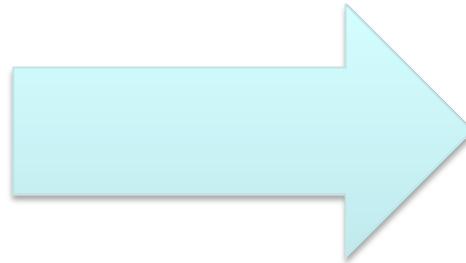
Regression

predicting a numeric value

Stock market



Weather prediction revisited

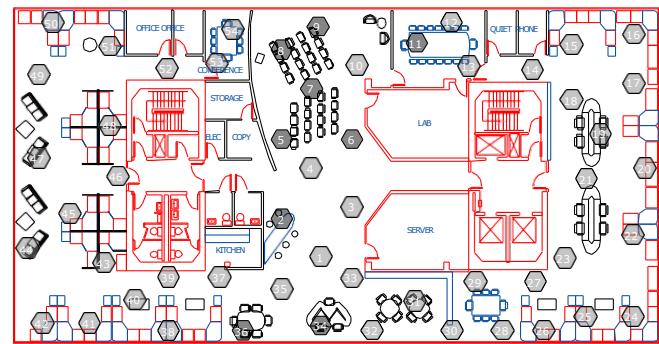


Temperature

72° F

Modeling sensor data

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

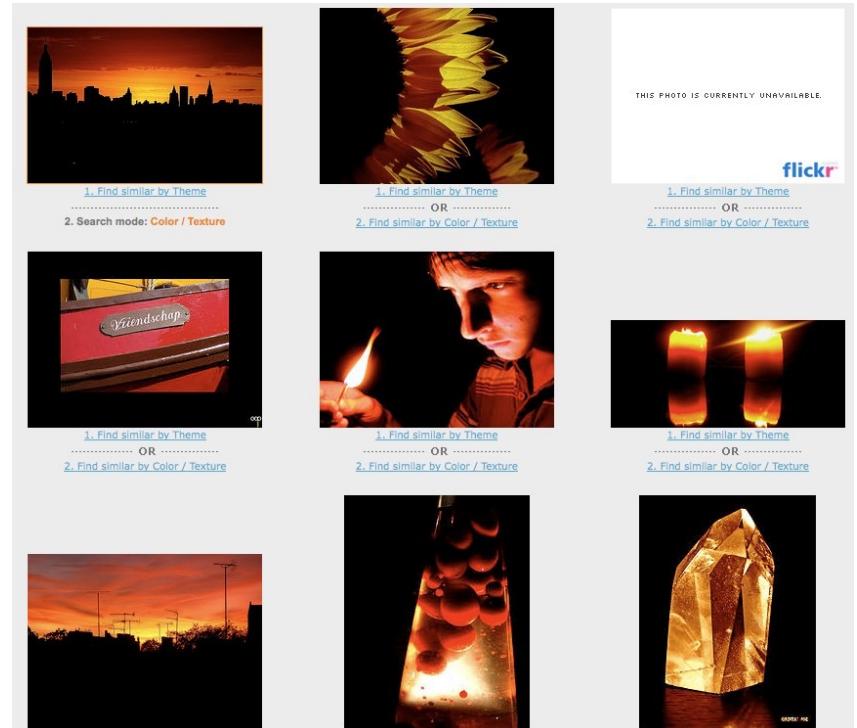
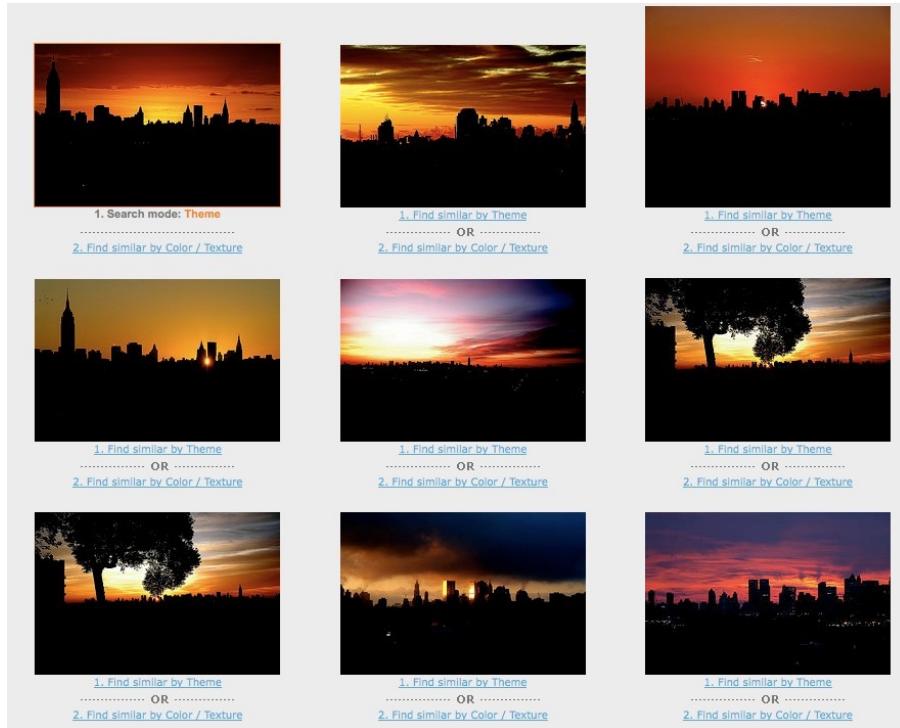


[Guestrin et al. '04]

Similarity

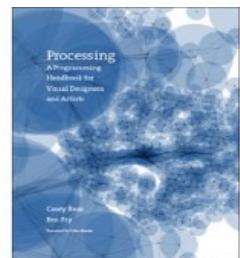
finding data

Given image, find similar images



<http://www.tiltomo.com/>

Recommender System: Collaborative Filtering



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.](#)

[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.](#)

Related Education & Training Services in Pittsburgh

[Learn HTML Coding](#)

[www.FullSail.edu](#) ▶ Earn Your Bachelor's Degree in Web Design and Development.

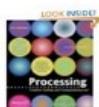
[Create Websites with HTML](#)

[http://www.unex.Berkeley.edu](#) ▶ Learn HTML Online, Start Anytime! with UC Berkeley Extension

[Intensive XSLT Training](#)

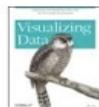
[www.objectdatalabs.com/course10.asp](#) ▶ OnSite or in NYC, LA, SFO, ORD, DC Will customize & train as few as 3

Customers Who Bought This Item Also Bought



[Processing: Creative Coding and Computational Art...](#) by Ira Greenberg

 (7) \$43.99



[Visualizing Data: Exploring and Explaining Data...](#) by Ben Fry

 (11) \$26.39



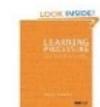
[Making Things Talk: Practical Methods for Conne...](#) by Tom Igoe

 (15) \$19.79



[Physical Computing: Sensing and Controlling the...](#) by Tom Igoe

 (20) \$19.00



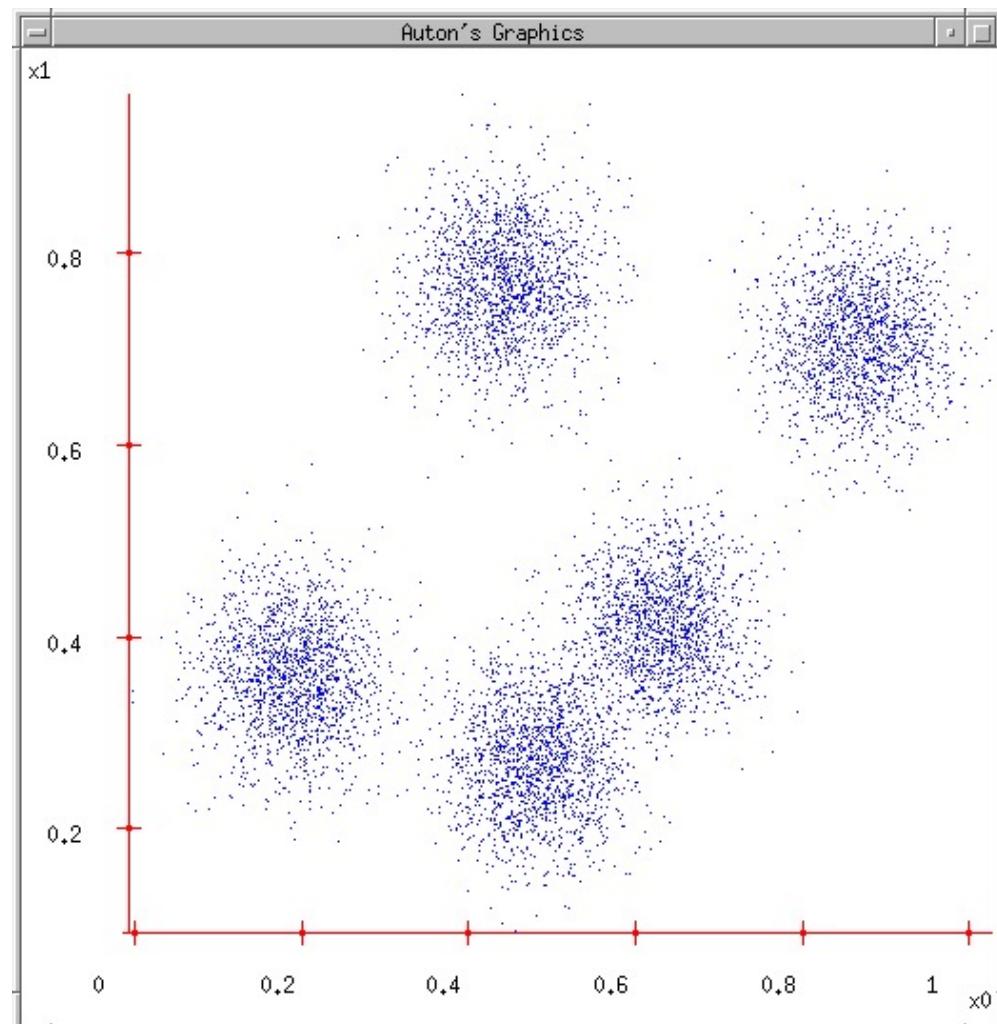
[Learning Processing: A Beginner's Guide to...](#) by Daniel Shiffman

 (7) \$44.95

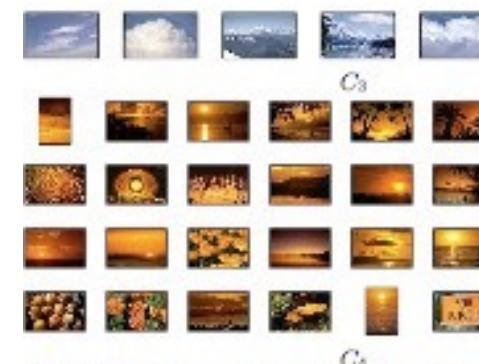
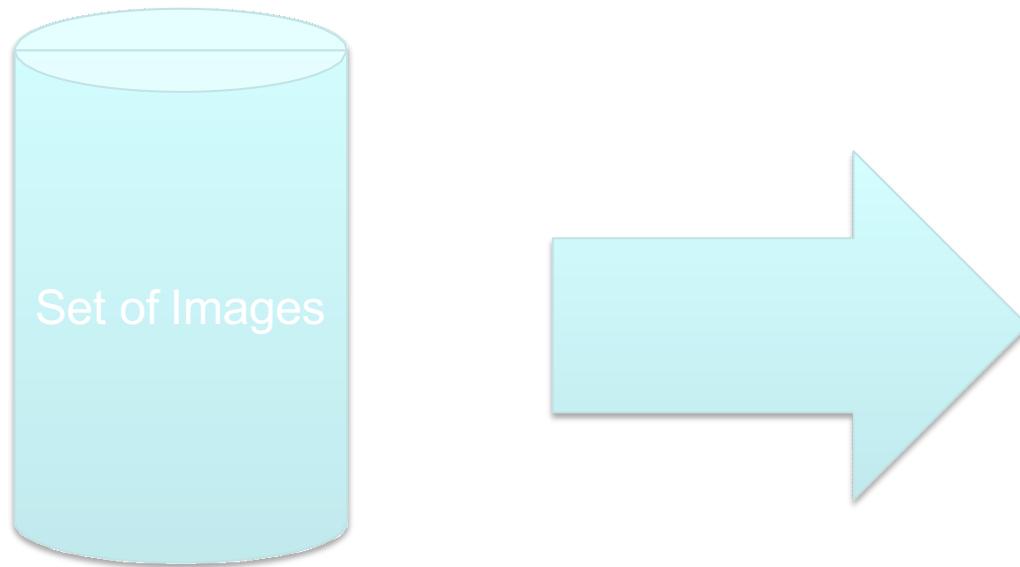
Clustering

discovering structure in data

Clustering Data: Group similar things



Clustering images



[Goldberger et al.]

Clustering web search results

web news images wikipedia blogs jobs more »
race Search advanced preferences

clusters sources sites Search Results

Cluster Human contains 8 documents.

1. [Race \(classification of human beings\) - Wikipedia, the free ...](#)
The term **race** or racial group usually refers to the concept of dividing **humans** into populations or groups on the basis of various sets of characteristics. The most widely used **human** racial categories are based on visible traits (especially skin color, cranial or facial features and hair texture), and self-identification. Conceptions of **race**, as well as specific ways of grouping **races**, vary by culture and over time, and are often controversial for scientific as well as social and political reasons. History · Modern debates · Political and ...
[en.wikipedia.org/wiki/Race_\(classification_of_human_beings\)](http://en.wikipedia.org/wiki/Race_(classification_of_human_beings)) - [cache] - Live, Ask

2. [Race - Wikipedia, the free encyclopedia](#)
General. **Racing** competitions The **Race** (yachting **race**), or La course du millénaire, a no-rules round-the-world sailing event; **Race** (biology), classification of flora and fauna; **Race** (classification of human beings) **Race** and ethnicity in the United States Census, official definitions of "race" used by the US Census Bureau; **Race** and genetics, notion of racial classifications based on genetics. Historical definitions of **race**; **Race** (bearing), the inner and outer rings of a rolling-element bearing. **RACE** in molecular biology "Rapid ... General · Surnames · Television · Music · Literature · Video games
en.wikipedia.org/wiki/Race - [cache] - Live, Ask

3. [Publications | Human Rights Watch](#)
The use of torture, unlawful rendition, secret prisons, unfair trials, ... Risks to Migrants, Refugees, and Asylum Seekers in Egypt and Israel ... In the run-up to the Beijing Olympics in August 2008, ...
www.hrw.org/backgrounder/usa/race - [cache] - Ask

4. [Amazon.com: Race: The Reality Of Human Differences: Vincent Sarich ...](#)
Amazon.com: **Race**: The Reality Of Human Differences: Vincent Sarich, Frank Miele: Books ... From Publishers Weekly Sarich, a Berkeley emeritus anthropologist, and Miele, an editor ...
www.amazon.com/Race-Reality-Differences-Vincent-Sarich/dp/0813340861 - [cache] - Live

5. [AAPA Statement on Biological Aspects of Race](#)
AAPA Statement on Biological Aspects of **Race** ... Published in the American Journal of Physical Anthropology, vol. 101, pp 569-570, 1996 ... PREAMBLE As scientists who study **human** evolution and variation, ...
www.physanth.org/positions/race.html - [cache] - Ask

6. [race: Definition from Answers.com](#)
race n. A local geographic or global **human** population distinguished as a more or less distinct group by genetically transmitted physical
www.answers.com/topic/race-1 - [cache] - Live

7. [Dopefish.com](#)
Site for newbies as well as experienced Dopefish followers, chronicling the birth of the Dopefish, its numerous appearances in several computer games, and its eventual take-over of the **human race**. Maintained by Mr. Dopefish himself, Joe Siegler of Apogee Software.
www.dopefish.com - [cache] - Open Directory

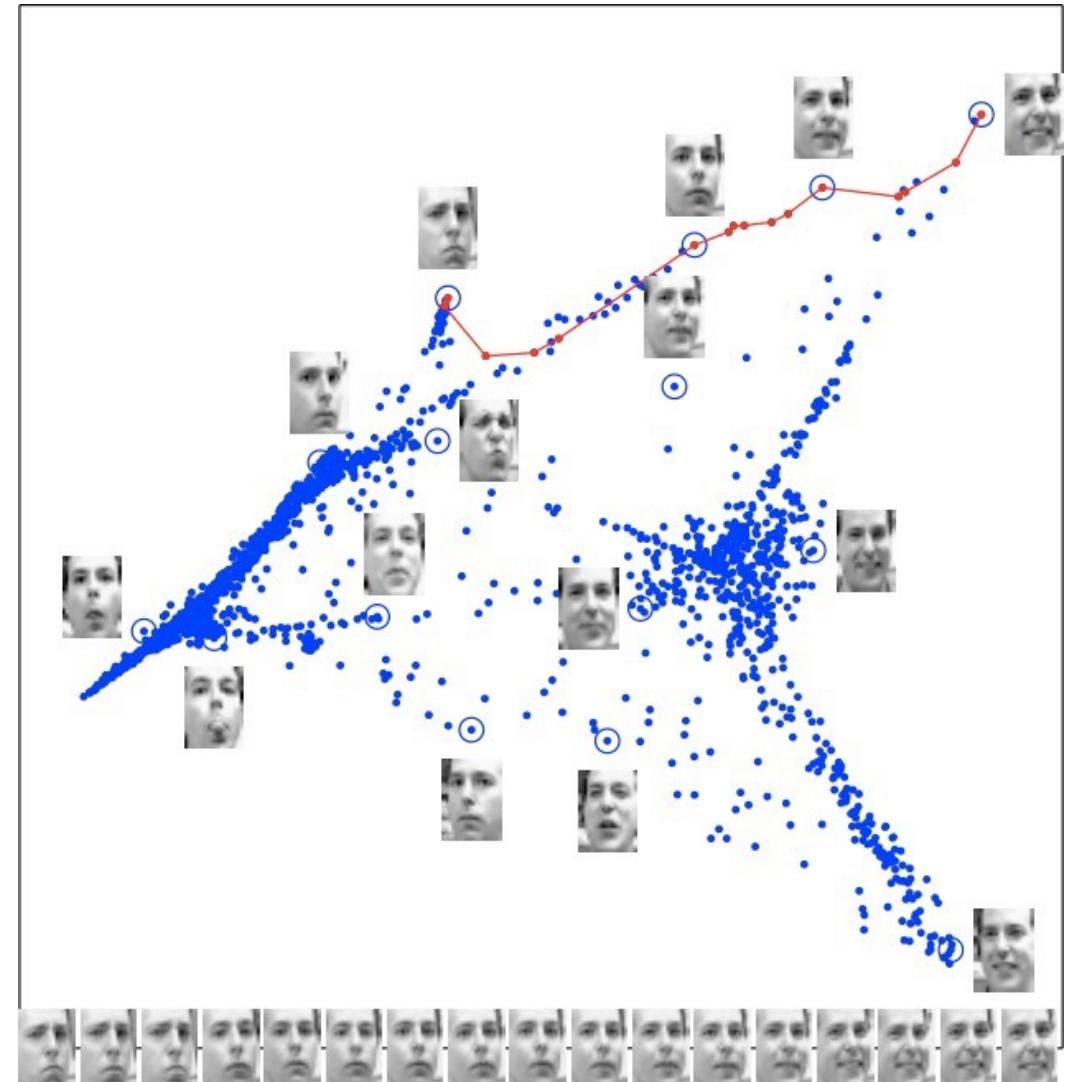
find in clusters:

Embedding

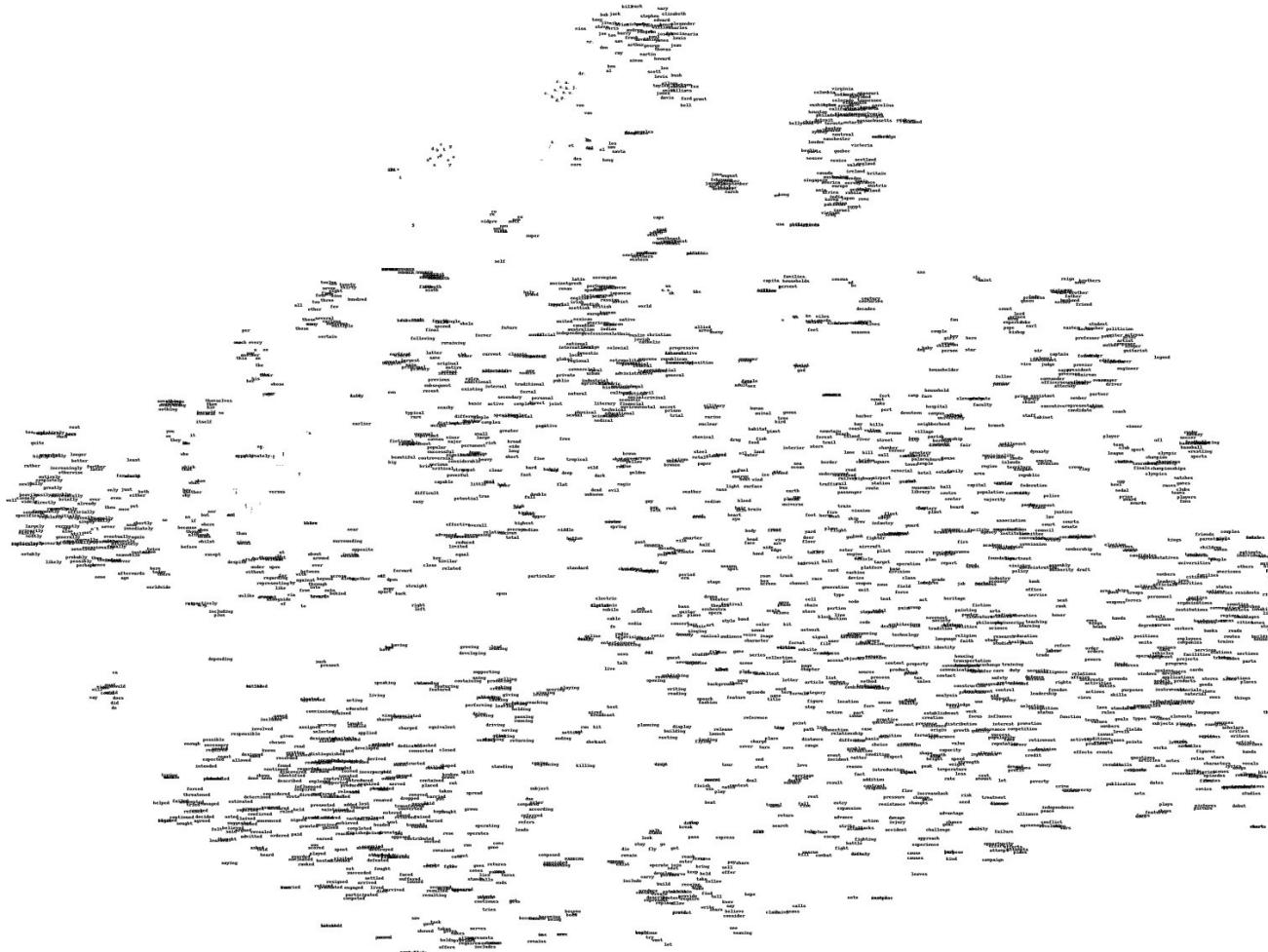
**discovering latent factor
visualizing data**

Embedding images

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



Embedding words



Embedding words (zoom in)

billmark mary
 bob jack stephen elizabeth
 tony edward
 miss jimmie brianchrisrichardhenryalexander
 steve andrew williamcharles
 joe harry roberson
 tom frank paul josephfrancismaria
 mr. sam davida james louis
 don arthurgeorgejean
 ray thomas
 simon martin
 howard
 dr. ben lee
 al scott lewis bush
 r. a. wilsonjacksonfox
 m. e. h. j. smithwilliams
 c. b. s. w. jones davis fordgrant
 d. p. von bell
 van

 et da los san santa
 des core hong

 june august
 february september
 january october
 april november
 march

 cape
 east
 west
 southeast
 southwest
 northeast

 center southern
 northern
 western

 pacific

 latin norwegian
 uss

mary
 elizabeth
 edward
 alexander
 charles
 francis
 maria
 louis
 jean
 thomas
 bush
 fox
 williams
 grant
 bell

 virginia
 columbia missouri
 indiana maryland
 colorado tennessee
 wisconsin
 washington oregon northcarolina
 houston philadelphia atlanta
 detroit toronto ontario
 hollywood boston
 sydney montreal cambridge
 london manchester victoria
 beijing quebec
 moscow mexico scotland
 wales england
 ireland britain
 australia sweden
 singapore spain
 america norway austria
 europe
 asia germany
 africa russia
 india japan rome
 korea pakistan egypt
 vietnam israel
 iraq

amkong
 usa philippines

families
 census

[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

**Robot Motor Skill
Coordination with EM-based
Reinforcement Learning**

Petar Kormushev, Sylvain Calinon,
and Darwin G. Caldwell

Italian Institute of Technology

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 $\{(x_i, y_i) \mid i = 1 \dots n\}$
- 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 \mathbb{R} (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 virtue of its nature as being utterly confidential and top secret.



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

99 MILLION EMAILADDRESSES FOR ONLY \$99



Ok, I know 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



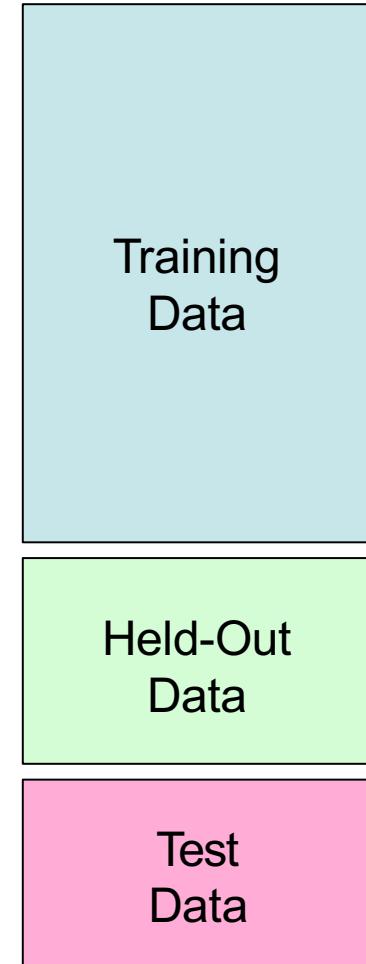
??

Machine Learning Categories

- Supervised Learning
 - Classification
 - Regression
- Unsupervised Learning
 - Clustering
 - Latent Factors
 - Matrix Completion
- Reinforcement Learning

Important Concepts

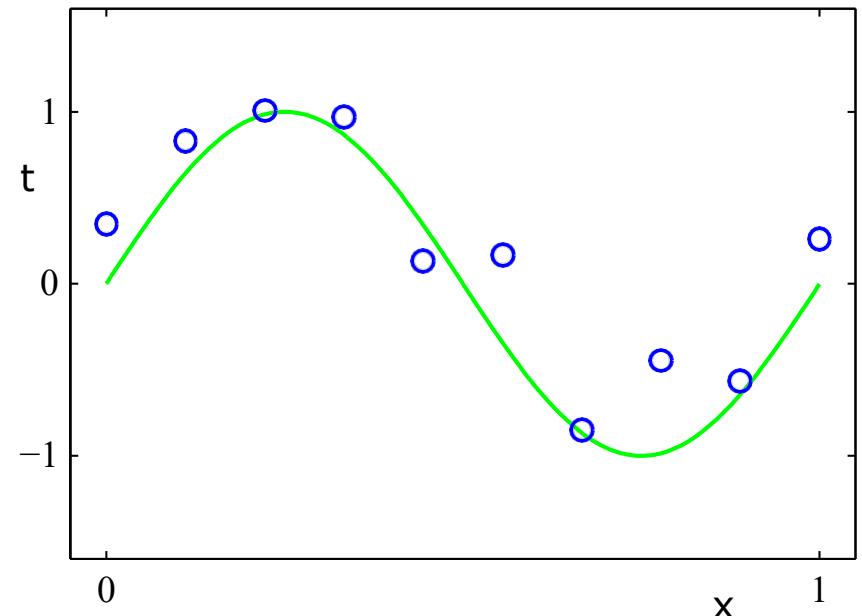
- **Data:** labeled instances, e.g. emails marked spam/ham
 - Training set
 - Held out set (sometimes call Validation set)
 - Test set
- **Features:** attribute-value pairs which characterize each x
- **Experimentation cycle**
 - Select a hypothesis f to best match training set
 - (Tune hyperparameters 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 Example

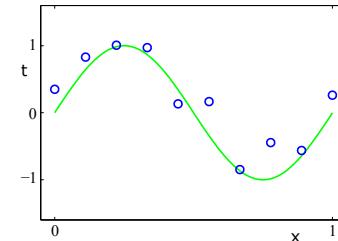
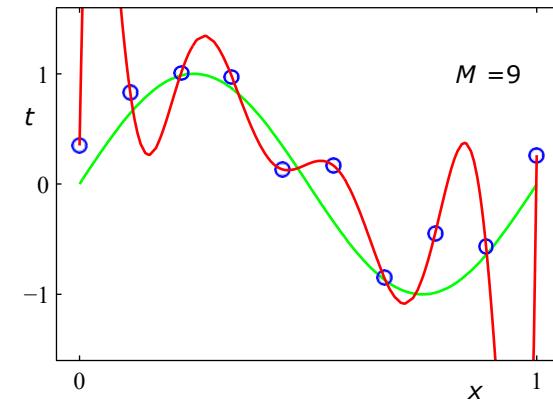
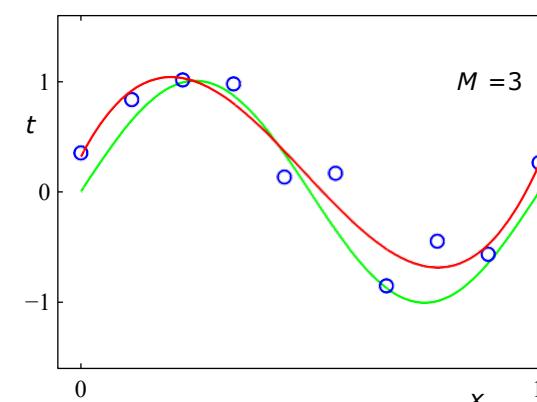
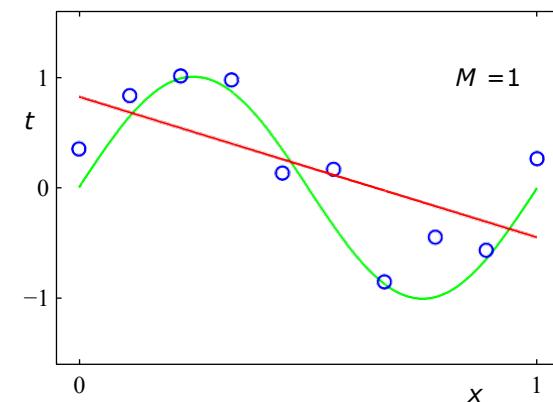
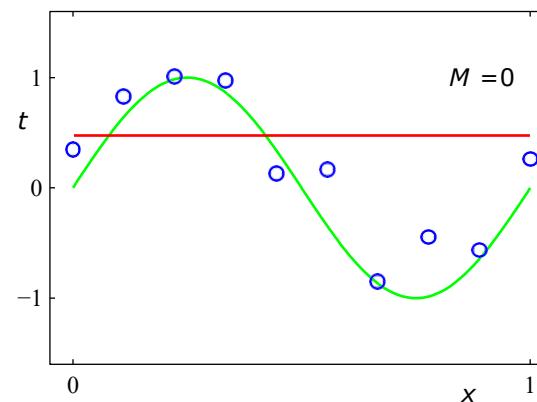
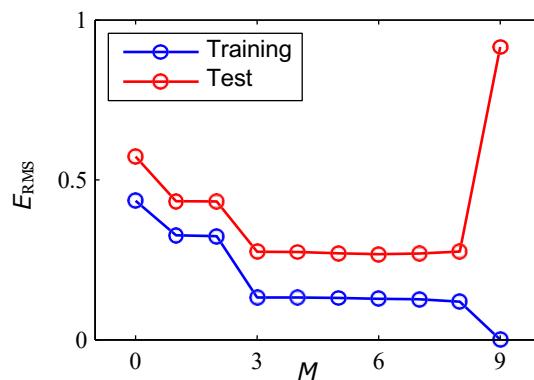
- Consider a simple, regression dataset:
 - $f: X \rightarrow Y$
 - $X = \mathbb{R}$
 - $Y = \mathbb{R}$
- **Question 1:** How should we pick the *hypothesis space*, the set of possible functions f ?
- **Question 2:** How do we find the best f in the hypothesis space?

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)

Checklist

1. Sign-up for Piazza
2. Sign-up for Gradiance, try warm-up quiz
3. Read the department's academic integrity policy
4. Reading materials: Murphy book Chapter 1