http://www.cs.washington.edu/education/courses/cse446/13sp/

What's learning? Point Estimation

Machine Learning – CSE446

Carlos Guestrin
University of Washington

April 1, 2013

What is Machine Learning ?

Machine Learning

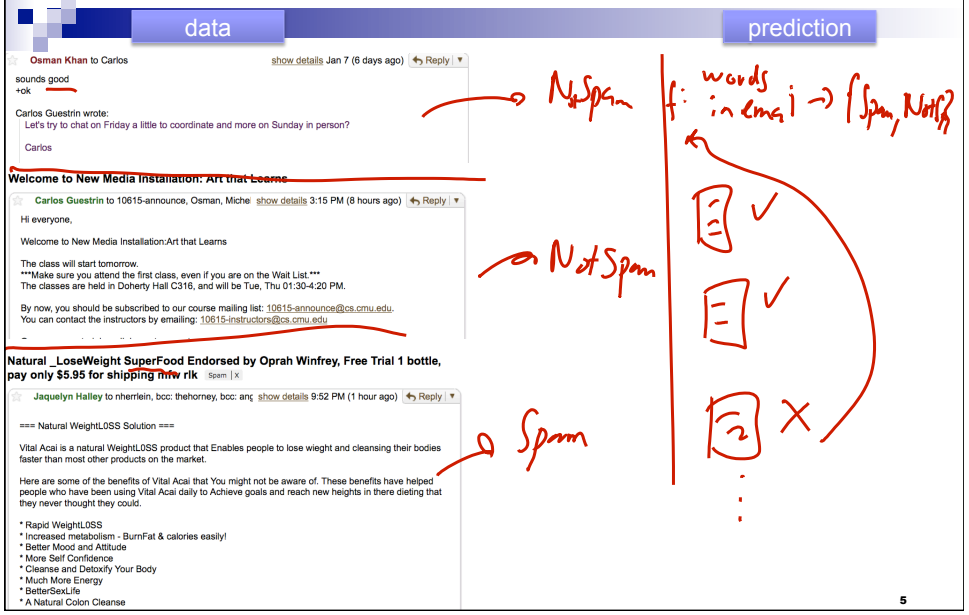
Study of algorithms that

- improve their performance
- at some task
- with experience

Classification

from data to discrete classes

Spam filtering



Text classification

Company home page

vs

Personal home page

vs

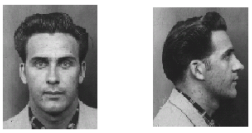
Univeristy home page

vs

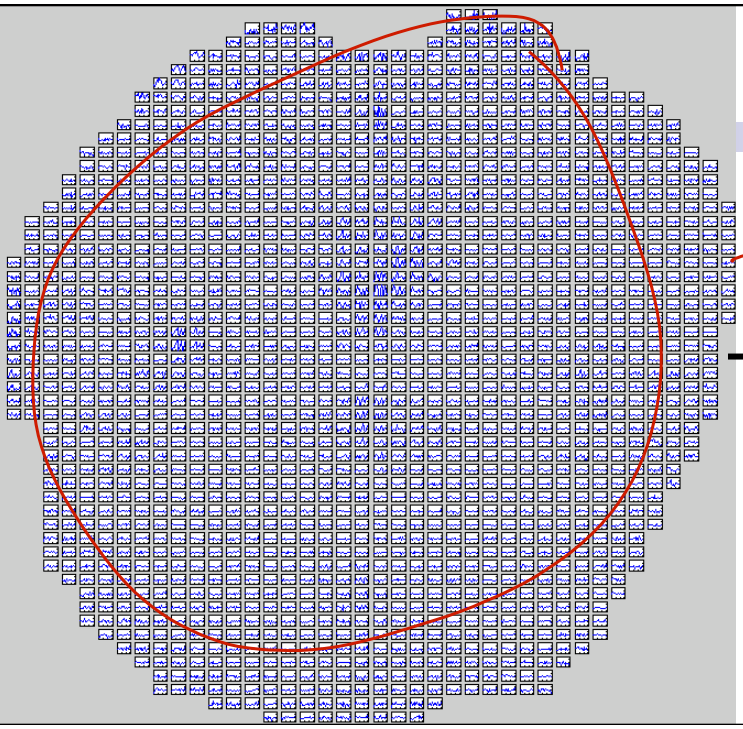
…

Object detection

(Prof. H. Schneiderman)



Example training images
for each orientation



Reading a noun (vs verb)

[Rustandi et al., 2005]

Weather prediction

change snow cloudy

Show

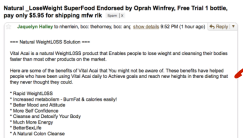
Seattle

$f : \ldots \rightarrow \{\text{Rain}\}$

The classification pipeline

Training

✓ not spam



✓ spam

N with data points

have label

functional class for $f \ldots$

f

Testing

using

don't have true label

X

predict

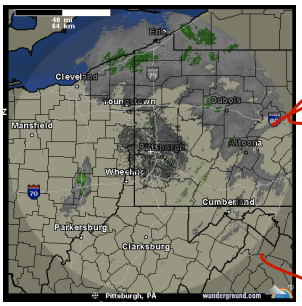
$\hat{f}(x) \rightarrow \{S, NS\}$

Regression

predicting a numeric value

Stock market

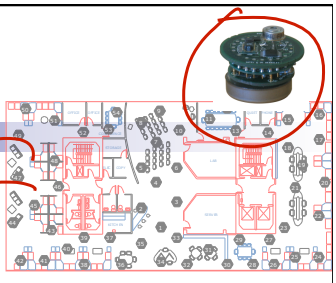
Weather prediction revisted

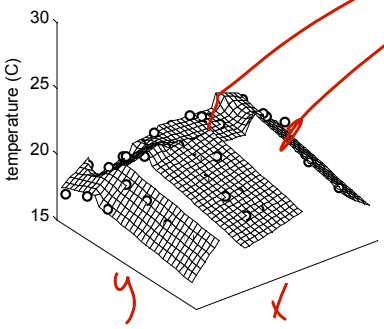


Temperature

63°F

Modeling sensor data





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

$\hat{f}(x,y) \rightarrow$ temp @ location

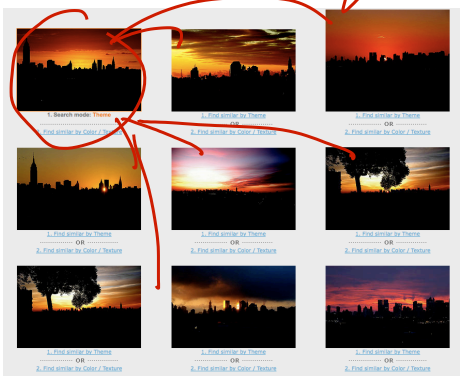
[Guestrin et al. '04]

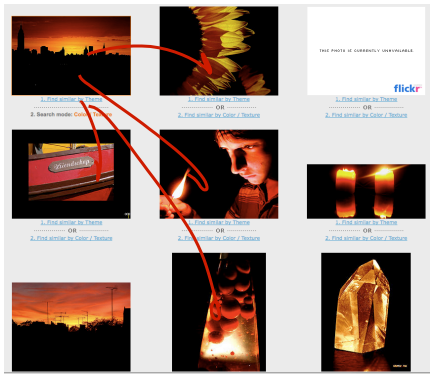
Similarity

finding data

Given image, find similar images

learning the similarity model for a particular task



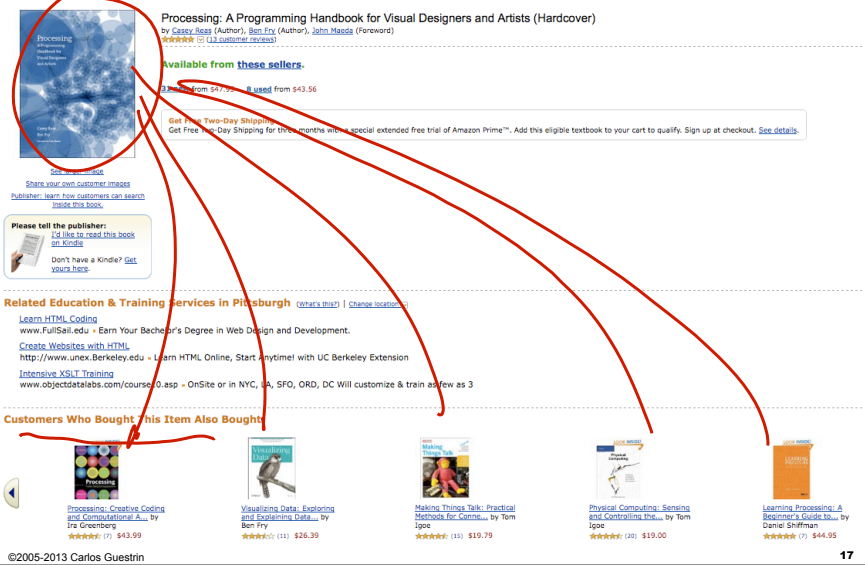


Similarity = sun sets image about same "thing"

Similarity → color

http://www.tiltomo.com/

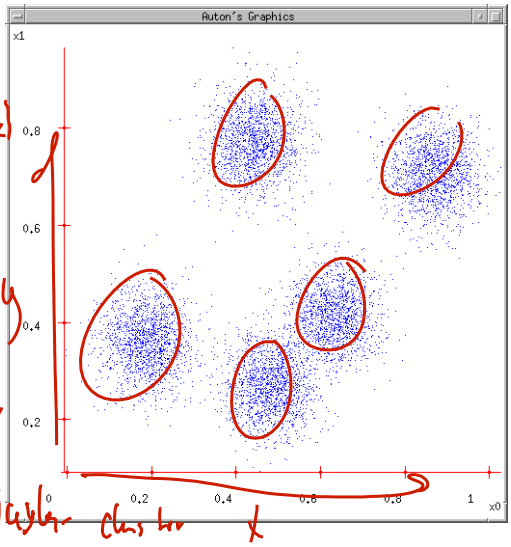
Similar products



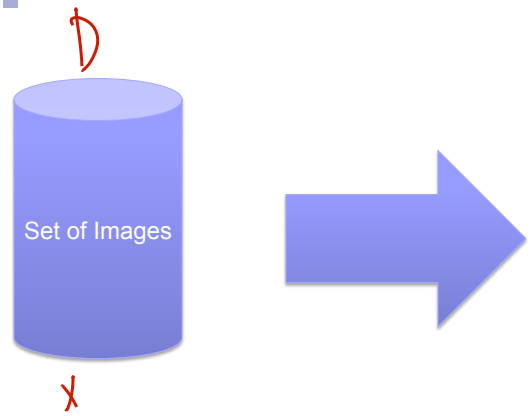
Clustering

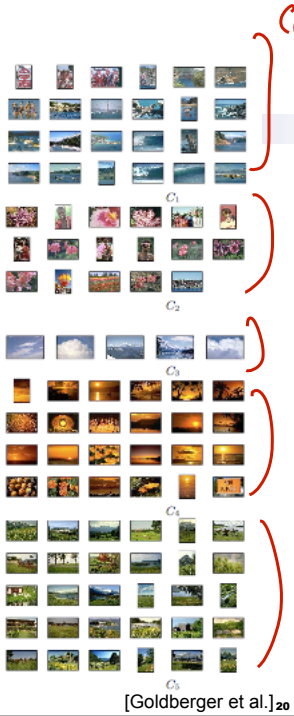
discovering structure in data

Clustering Data: Group similar things



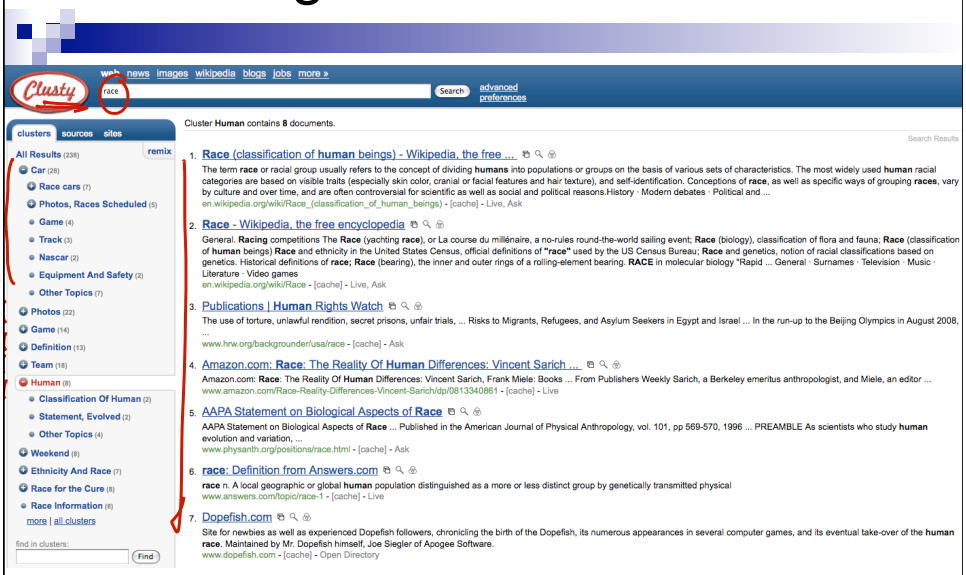
Clustering images





[Goldberger et al.]

Clustering web search results



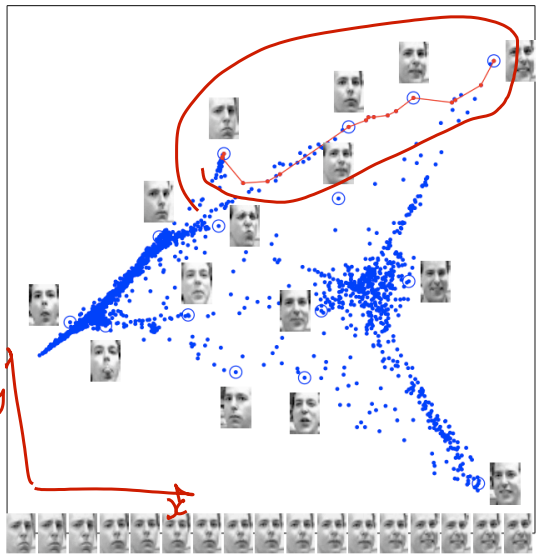
Embedding

visualizing data

Embedding images

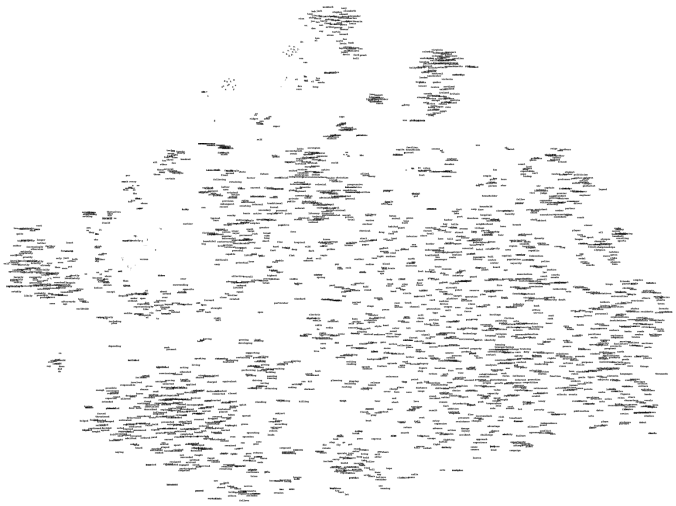
Images have thousands or millions of pixels.

Can we give each image a coordinate, such that similar images are near each other?



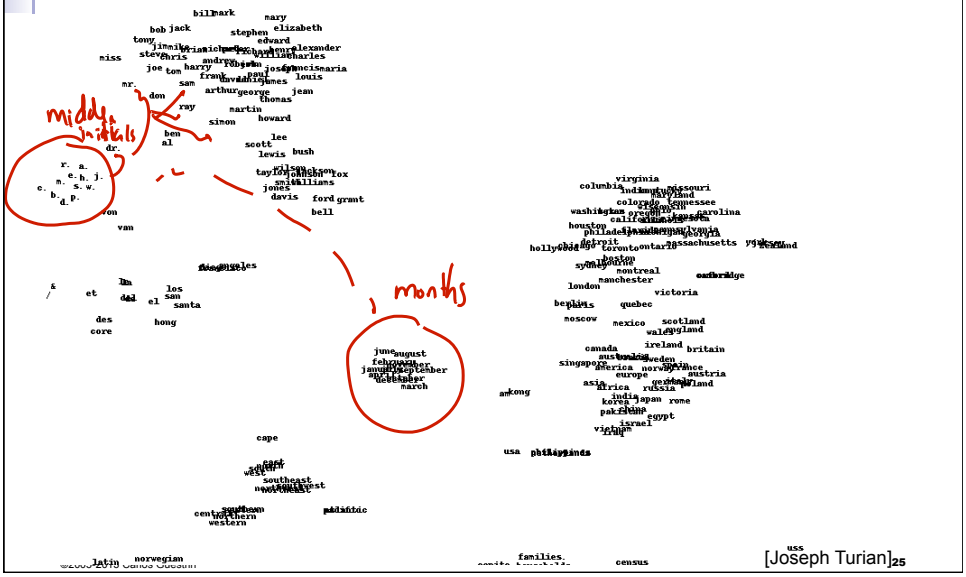
[Saul & Roweis '03]

Embedding words



[Joseph Turian]

Embedding words (zoom in)



[Joseph Turian]25

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

[Ng et al. ’05]

Bringing it all together…

Combining video, text and audio

HURLEY: Uh ... the Chinese people have water.
(Sayid and Kate go to check it out.)

[EXT. BEACH - CRASH SITE]

(Sayid holds the empty bottle in his hand and questions Sun.)

SAYID: (quietly) Where did you get this?
(He looks at her.)

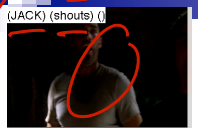
[EXT. JUNGLE]

(Sawyer is walking through the jungle. He reaches a spot. He kneels down and looks back to check that no one's followed him.

Taskar et al.

Automatically Discovered and Labeled Actions

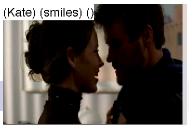
shout



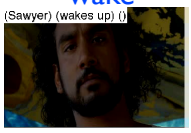
sit down

follow

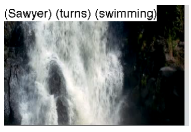
smile



wake



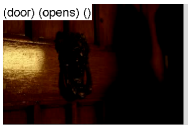
swim



grab

kiss

open door



point

Growth of Machine Learning

One of the most sought for specialties in industry today!!!!

- 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, especially with **Big Data**
  - 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

Syllabus

- Covers a wide range of Machine Learning techniques – from basic to state-of-the-art
- You will learn about the methods you heard about:
  - Point estimation, regression, naïve Bayes, logistic regression, nearest-neighbor, decision trees, boosting, perceptron, overfitting, regularization, dimensionality reduction, PCA, error bounds, VC dimension, SVMs, kernels, margin bounds, K-means, EM, mixture models, semi-supervised learning, HMMs, graphical models, active learning, reinforcement learning…
- Covers algorithms, theory and applications
- **It's going to be fun and hard work ☺**

Prerequisites

- Formally:
  - Either CSE 326 or CSE 332; either STAT 390, STAT 391, or CSE 312
- Probabilities
  - Distributions, densities, marginalization…
- Basic statistics
  - Moments, typical distributions, regression…
- Algorithms
  - Dynamic programming, basic data structures, complexity…
- Programming
  - R will be very useful, but we'll help you get started
- We provide some background, but the class will be fast paced
- Ability to deal with "abstract mathematical concepts"

Optional R tutorial

- There are many resources to get started with R online
- We'll run an ***optional*** tutorial:
  - Thursday 4/4 @6pm
  - Location TBD

Staff

- Three Great TAs: Great resource for learning, interact with them!
  - **Danielle Bragg**
    Office hours: Wednesdays 3:30-5:30pm
  - **Daryl Hansen**
    Office hours: Thursdays 1:30-3:30pm
  - **James McQueen**
    Office hours: Tuesdays 9:30-11:30am
  - Prof: **Carlos Guestrin**
    Office hours: Fridays 1:30-2:30pm



Communication Channels

- Only channel for announcements, questions, etc. – Google Group:
  - https://groups.google.com/forum/?fromgroups#!forum/cse446-spr13
  - Subscribe!
  - All non-personal questions should go here
  - Answering your question will help others
  - Feel free to chime in
- For e-mailing instructors about personal issues, use:
  - cse446-staff@cs.washington.edu

Text Books

- Required Textbook:
  - Machine Learning: a Probabilistic Perspective; Kevin Murphy
- Optional Books:
  - Pattern Recognition and Machine Learning; Chris Bishop
  - The Elements of Statistical Learning: Data Mining, Inference, and Prediction; Trevor Hastie, Robert Tibshirani, Jerome Friedman
  - Machine Learning; Tom Mitchell
  - Information Theory, Inference, and Learning Algorithms; David MacKay

Grading

- 4 homeworks (40%)
  - First one goes out 4/4
    - Start early, Start early
- Final project (20%)
  - Details out around April 29th
  - Projects done individually, or groups of two students
- Midterm (15%)
  - Wed., May 8th in class
- Final (25%)
  - TBD by registrar, probably 6/12/2013, 8:30-10:20am

Homeworks

- Homeworks are hard, start early ☺
- Due in the beginning of class
- 33% subtracted per late day
- All homeworks **must be handed in**, even for zero credit
- Use Catalyst to submit homeworks

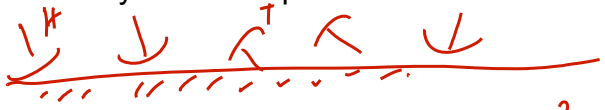
- Collaboration
  - You may **discuss** the questions
  - Each student writes their own answers
  - Write on your homework anyone with whom you collaborate
  - Each student must write their own code for the programming part
  - **Please don't search for answers on the web, Google, previous years' homeworks, etc.**
    - please ask us if you are not sure if you can use a particular reference

Enjoy!

- ML is becoming ubiquitous in science, engineering and beyond
- It's one of the hottest topics in industry today
- This class should give you the basic foundation for applying ML and developing new methods
- The fun begins…

Your first consulting job

- A billionaire from the suburbs of Seattle asks you a question:
  - He says: I have thumbtack, if I flip it, what's the probability it will fall with the nail up?
  - You say: Please flip it a few times:



$P(H) = 3/5$

  - You say: The probability is:
  - **He says: Why???**
  - You say: Because…

Thumbtack – Binomial Distribution

model

- P(Heads) = θ, P(Tails) = 1-θ

observed HHTTH

$P(D|\theta) = P(HHTTH) = \theta\,\theta\,(1-\theta)\,(1-\theta)\,\theta = \theta^3(1-\theta)^2$

learning task: choose $\hat{\theta}$ pick θ that maximizes the likelihood of seeing these observations according to my model

MLE

- Flips are i.i.d.:
  - Independent events
  - Identically distributed according to Binomial distribution

iid

- Sequence D of α_H Heads and α_T Tails

$$P(\mathcal{D} \mid \theta) = \theta^{\alpha_H}(1-\theta)^{\alpha_T}$$