

<http://www.cs.washington.edu/education/courses/cse546/13au/>

# What's learning? Point Estimation

Machine Learning – CSE546  
Carlos Guestrin  
University of Washington  
September 25, 2013

©2005-2013 Carlos Guestrin

1

## What is Machine Learning ?

©2005-2013 Carlos Guestrin

2

# Machine Learning

Study of algorithms that

- improve their performance
- at some task
- with experience



## Classification

from data to discrete classes

# Spam filtering

data prediction

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, Miche 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 rk Spam X

Jaquelyn Halley to nherlein, bcc: thehorney, bcc: anc show details 9:52 PM (1 hour ago) Reply

==== Natural WeightLOSS Solution ====

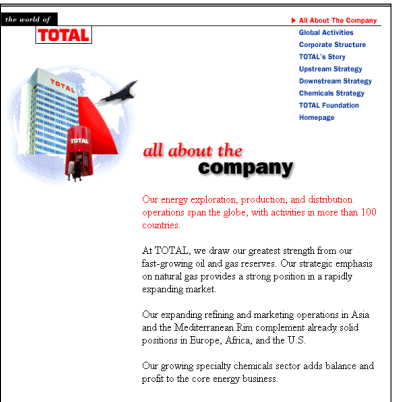
Vital Acai is a natural WeightLOSS product that Enables people to lose weight and cleansing their bodies faster than most other products on the market.

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

5

# Text classification



The screenshot shows the TOTAL company homepage. It features the TOTAL logo, a navigation menu with links like 'Global Activities', 'Corporate Structure', and 'TOTAL & Story'. The main content area has the heading 'all about the company' and several paragraphs of text describing the company's operations and strengths.

Company home page  
VS  
Personal home page  
VS  
Univeristy home page  
VS  
...

©2005-2013 Carlos Guestrin 6

# Object detection

(Prof. H. Schneiderman)

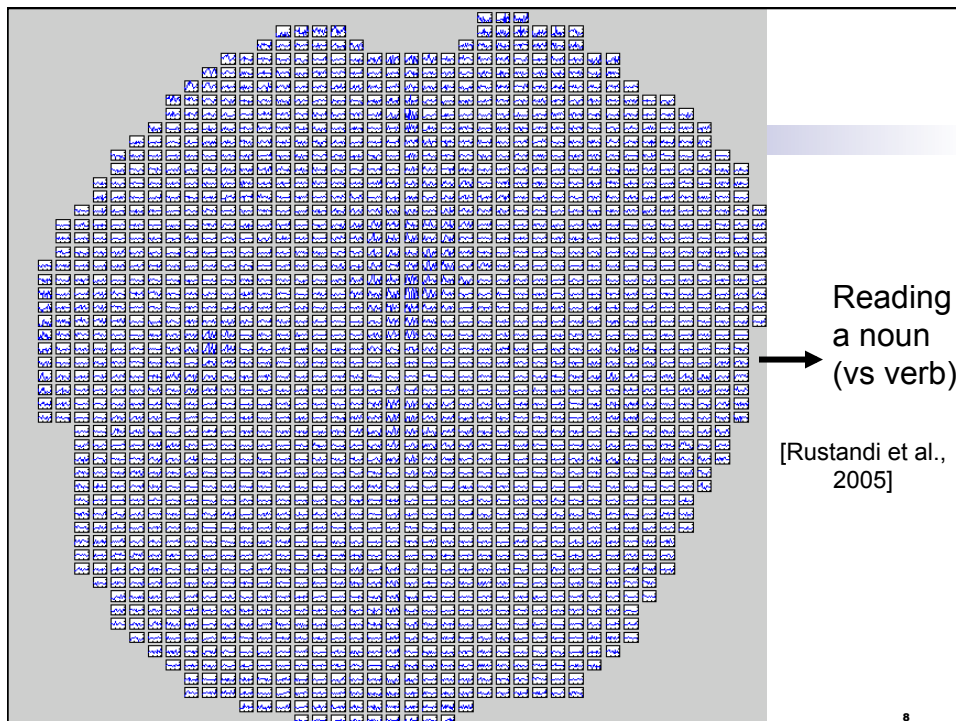


Example training images for each orientation



©2005-2013 Carlos Guestrin

7



8

# Weather prediction



©2005-2013 Carlos Guestrin

9

# The classification pipeline

## Training

```
class Weather {
public:
  Weather(const int temp): m_temp(temp) {}
  int temp() const { return m_temp; }
private:
  int m_temp;
};
```

```
int main() {
  Weather w(10);
  std::cout << w.temp() << endl;
  return 0;
}
```

## Testing

```
int main() {
  Weather w(10);
  std::cout << w.temp() << endl;
  return 0;
}
```

©2005-2013 Carlos Guestrin

10

©2005-2013 Carlos Guestrin

# Regression

predicting a numeric value

11

## Stock market

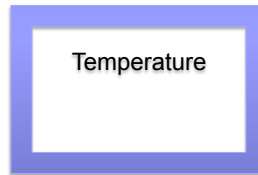
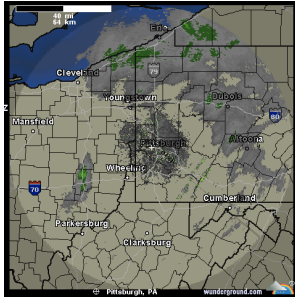
Jan 12, 2009 : ^DJI 8,473.9697

© 2008 Yahoo! Inc.

Volume 4,725,049,856

12

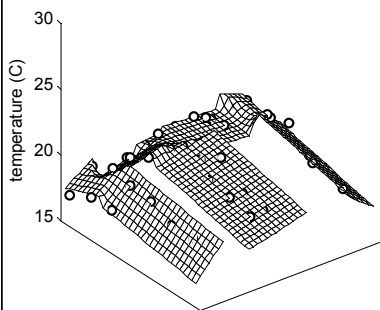
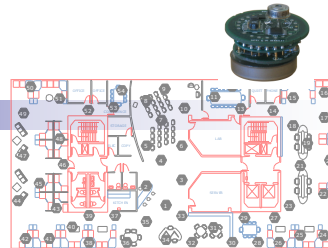
# Weather prediction revisited



©2005-2013 Carlos Guestrin

13

# Modeling sensor data




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

[Guestrin et al. '04]

©2005-2013 Carlos Guestrin

14



# Similarity

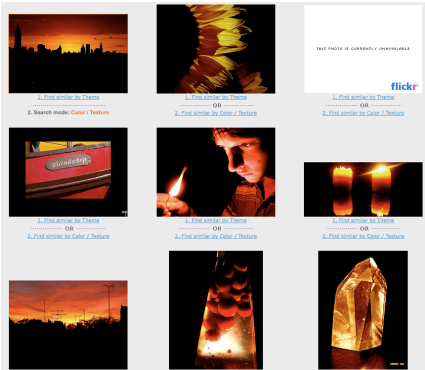
## finding data

©2005-2013 Carlos Guestrin 15

# Given image, find similar images







©2005-2013 Carlos Guestrin <http://www.tiltomo.com/> 16



# Similar products



Processing: A Programming Handbook for Visual Designers and Artists (Hardcover)

by Casey Reas (Author), Ben Fry (Author), John Maciej (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 [\(show more\)](#) | [change location](#)

[Learn HTML Coding](#)

[www.FullSail.edu](http://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 XML Training](#)

[www.objectdatabases.com/course10.asp](http://www.objectdatabases.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 Iza Greenberg  
★★★★★ (7) \$43.99



Visualize Data: Exploring and Explaining Data... by Ben Fry  
★★★★★ (11) \$26.39



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



Physical Computing: Sensing and Responding to the World... by Tom Igoe  
★★★★★ (20) \$19.00



Learning Processing: A Beginner's Guide to Coding... by Daniel Shiffman  
★★★★★ (7) \$44.05

©2005-2013 Carlos Guestrin

17

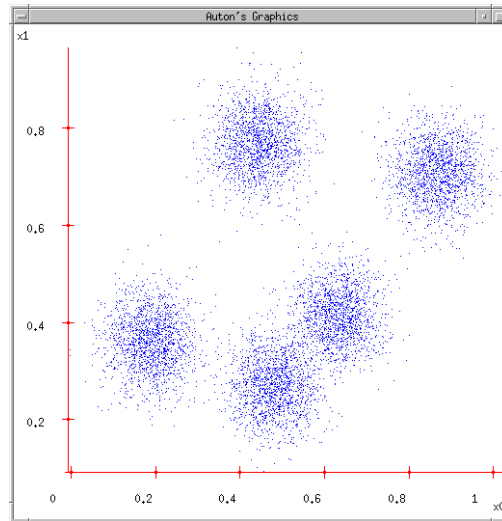
# Clustering

discovering structure in data

©2005-2013 Carlos Guestrin

18

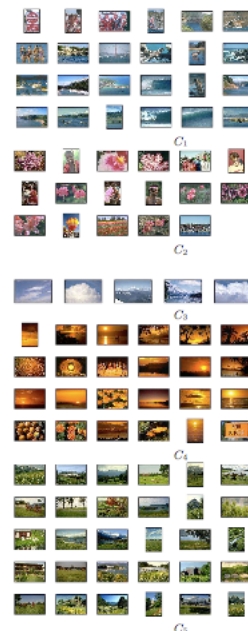
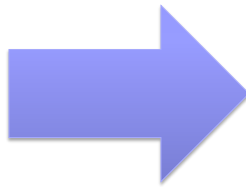
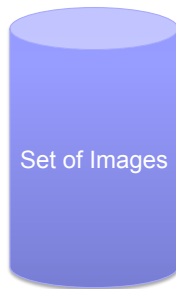
# Clustering Data: Group similar things



©2005-2013 Carlos Guestrin

19

# Clustering images



©2005-2013 Carlos Guestrin

[Goldberger et al.]<sub>20</sub>

# Clustering web search results

The screenshot shows the Clusty search interface. At the top, there's a navigation bar with links for 'web', 'news', 'images', 'wikipedia', 'blogs', 'jobs', and 'more'. Below this is a search bar containing the word 'race' and a search button. The main content area is titled 'Cluster Human contains 8 documents.' and lists seven search results. On the left side, there's a sidebar with 'All Results (238)' and various categories like 'Car (28)', 'Race cars (7)', 'Photos, Races Scheduled (5)', etc. The search results include links to Wikipedia, a free encyclopedia, a racing event, a human rights watch article, an Amazon.com page, an APA statement, an Answers.com definition, and a Dopefish.com site.

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 **rac**es, 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) - [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/background/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, chronicing 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

©2005-2013 Carlos Guestrin 21

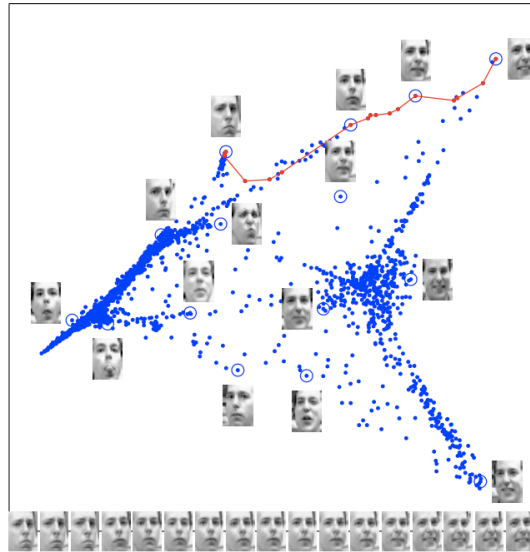
## Embedding

## visualizing data

# Embedding images

Images have thousands or millions of pixels.

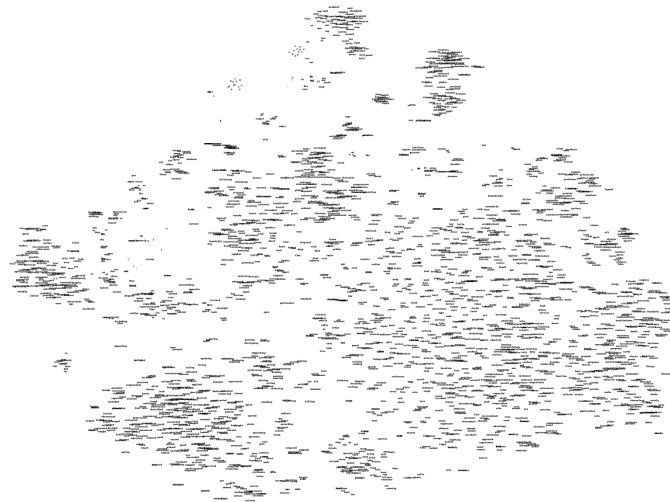
Can we give each image a coordinate, such that similar images are near each other?



©2005-2013 Carlos Guestrin

[Saul & Roweis '03] 23

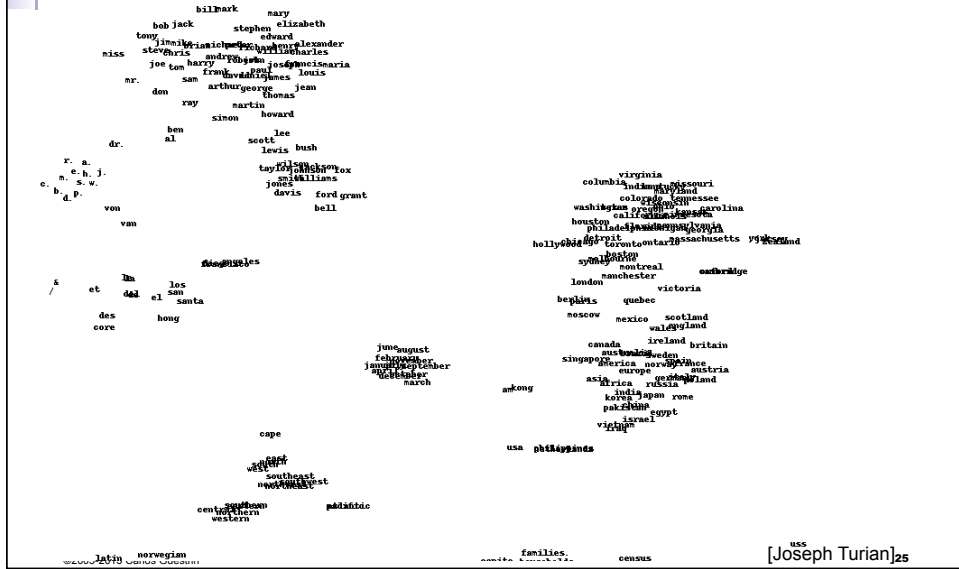
# Embedding words



©2005-2013 Carlos Guestrin

[Joseph Turian] 24

# Embedding words (zoom in)



# 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.)

**SAYID**

**SUN**

**locke**

**HOLDING**

**Taskar et al.**  
©2005-2013 Carlos Guestrin

# Automatically Discovered and Labeled Actions

**shout**  
(JACK) (shouts) ()

**smile**  
(Kate) (smiles) ()

**sit down**  
(Locke) (sits down) ()

**follow**  
(Kate) (follows) (Jack)

**wake**  
(Sawyer) (wakes up) ()

**swim**  
(Sawyer) (turns) (swimming)

**grab**  
(Kate) (grabs) (case)

**kiss**  
(Shannon) (kisses) (ear)

**open door**  
(door) (opens) ()

**point**  
(JACK) (points) ()

©2005-2013 Carlos Guestrin

# 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

©2005-2013 Carlos Guestrin

31

# 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 😊**

©2005-2013 Carlos Guestrin

32



# Prerequisites

- Formally:
  - STAT 341, STAT 391, or equivalent
- 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”

# Recitations & Python

- We'll run an **optional** recitations:
  - Tuesdays @5:30pm
  - Location TBD
- We are recommending Python for homeworks!
  - There are many resources to get started with Python online
  - We'll run an **optional** tutorial:
    - First recitation: Tuesday 10/1 @5:30pm

# Staff

- Three Great TAs: Great resource for learning, interact with them!

- Eric Lei**

Office hours: Fridays 1:30-3:30pm



- Marco Ribeiro**

Office hours: Tuesdays 1:30-3:20pm



- Tyler Johnson**

Office hours: Mondays 3-5pm



- Prof: **Carlos Guestrin**

Office hours: Wednesdays 10:30-11:30am

©2005-2013 Carlos Guestrin

35

# Communication Channels

- Only channel for announcements, questions, etc. – Catalyst Group:

- <https://catalyst.uw.edu/gopost/board/tbjohns/34218/>

- 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:

- [cse546-instructors@cs.washington.edu](mailto:cse546-instructors@cs.washington.edu)

©2005-2013 Carlos Guestrin

36

## 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 (35%)**
  - First one goes out 9/30
    - Start early, Start early
- **Final project (30%)**
  - Full details out around 10/9
  - Projects done individually, or groups of two students
- **Midterm (15%)**
  - Wed., 10/30 in class
- **Final (20%)**
  - TBD by registrar

# Homeworks

- Homeworks are hard, start early ☺
- Due in the beginning of class
- 33% subtracted per late day
- You have 3 LATE DAYS to use for homeworks only throughout the quarter
  - Please plan accordingly and after that don't be about deadlines, travel,... ☺
- 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

# Projects

- An opportunity to exercise what you learned and to learn new things
- Individually or groups of two
- Must involve real data
  - Must be data that you have available to you by the time of the project proposals
- Must involve machine learning
- It's encouraged to be related to your research, but must be something new you did this quarter
  - Not a project you worked on during the summer, last year, etc.
  
- Full details in a couple of weeks
  
- Wed., October 23 at 9:00am: **Project Proposals**
- Mon., November 11 at 9:00am: **Project Milestone**
- Wed., December 4, 3-5pm: **Poster Session**
- Mon., December 9 at 9:00am: **Project Report**

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

## Thumbtack – Binomial Distribution

- $P(\text{Heads}) = \theta$ ,  $P(\text{Tails}) = 1 - \theta$
- Flips are i.i.d.:
  - Independent events
  - Identically distributed according to Binomial distribution
- Sequence  $D$  of  $\alpha_H$  Heads and  $\alpha_T$  Tails

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

©2005-2013 Carlos Guestrin

43

## Maximum Likelihood Estimation

- **Data:** Observed set  $D$  of  $\alpha_H$  Heads and  $\alpha_T$  Tails
- **Hypothesis:** Binomial distribution
- Learning  $\theta$  is an optimization problem
  - What's the objective function?
- MLE: Choose  $\theta$  that maximizes the probability of observed data:

$$\begin{aligned}\hat{\theta} &= \arg \max_{\theta} P(D | \theta) \\ &= \arg \max_{\theta} \ln P(D | \theta)\end{aligned}$$

©2005-2013 Carlos Guestrin

44

## Your first learning algorithm

$$\begin{aligned}\hat{\theta} &= \arg \max_{\theta} \ln P(\mathcal{D} | \theta) \\ &= \arg \max_{\theta} \ln \theta^{\alpha_H} (1 - \theta)^{\alpha_T}\end{aligned}$$

- Set derivative to zero:  $\frac{d}{d\theta} \ln P(\mathcal{D} | \theta) = 0$

## How many flips do I need?

$$\hat{\theta}_{MLE} = \frac{\alpha_H}{\alpha_H + \alpha_T}$$

- Billionaire says: I flipped 3 heads and 2 tails.
- You say:  $\theta = 3/5$ , I can prove it!
- He says: What if I flipped 30 heads and 20 tails?
- You say: Same answer, I can prove it!
- **He says: What's better?**
- You say: Humm... The more the merrier???
- He says: Is this why I am paying you the big bucks???

## Simple bound (based on Hoeffding's inequality)

- For  $N = \alpha_H + \alpha_T$ , and  $\hat{\theta}_{MLE} = \frac{\alpha_H}{\alpha_H + \alpha_T}$

- Let  $\theta^*$  be the true parameter, for any  $\epsilon > 0$ :

$$P(|\hat{\theta} - \theta^*| \geq \epsilon) \leq 2e^{-2N\epsilon^2}$$

## PAC Learning

- PAC: Probably Approximate Correct
- Billionaire says: I want to know the thumbtack parameter  $\theta$ , within  $\epsilon = 0.1$ , with probability at least  $1 - \delta = 0.95$ . How many flips?

$$P(|\hat{\theta} - \theta^*| \geq \epsilon) \leq 2e^{-2N\epsilon^2}$$



## What about continuous variables?

- Billionaire says: If I am measuring a continuous variable, what can you do for me?
- **You say: Let me tell you about Gaussians...**

$$P(x \mid \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

## Some properties of Gaussians

- affine transformation (multiplying by scalar and adding a constant)
  - $X \sim N(\mu, \sigma^2)$
  - $Y = aX + b \rightarrow Y \sim N(a\mu + b, a^2\sigma^2)$
- Sum of Gaussians
  - $X \sim N(\mu_X, \sigma_X^2)$
  - $Y \sim N(\mu_Y, \sigma_Y^2)$
  - $Z = X + Y \rightarrow Z \sim N(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$

## Learning a Gaussian

- Collect a bunch of data
  - Hopefully, i.i.d. samples
  - e.g., exam scores
- Learn parameters
  - Mean
  - Variance

$$P(x \mid \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

©2005-2013 Carlos Guestrin

51

## MLE for Gaussian

- Prob. of i.i.d. samples  $D=\{x_1, \dots, x_N\}$ :

$$P(D \mid \mu, \sigma) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^N \prod_{i=1}^N e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$

- Log-likelihood of data:

$$\begin{aligned} \ln P(D \mid \mu, \sigma) &= \ln \left[ \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^N \prod_{i=1}^N e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} \right] \\ &= -N \ln \sigma\sqrt{2\pi} - \sum_{i=1}^N \frac{(x_i - \mu)^2}{2\sigma^2} \end{aligned}$$

©2005-2013 Carlos Guestrin

52

## Your second learning algorithm: MLE for mean of a Gaussian

- What's MLE for mean?

$$\frac{d}{d\mu} \ln P(\mathcal{D} | \mu, \sigma) = \frac{d}{d\mu} \left[ -N \ln \sigma \sqrt{2\pi} - \sum_{i=1}^N \frac{(x_i - \mu)^2}{2\sigma^2} \right]$$

## MLE for variance

- Again, set derivative to zero:

$$\begin{aligned} \frac{d}{d\sigma} \ln P(\mathcal{D} | \mu, \sigma) &= \frac{d}{d\sigma} \left[ -N \ln \sigma \sqrt{2\pi} - \sum_{i=1}^N \frac{(x_i - \mu)^2}{2\sigma^2} \right] \\ &= \frac{d}{d\sigma} \left[ -N \ln \sigma \sqrt{2\pi} \right] - \sum_{i=1}^N \frac{d}{d\sigma} \left[ \frac{(x_i - \mu)^2}{2\sigma^2} \right] \end{aligned}$$

# Learning Gaussian parameters

- MLE:

$$\hat{\mu}_{MLE} = \frac{1}{N} \sum_{i=1}^N x_i$$

$$\hat{\sigma}_{MLE}^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \hat{\mu})^2$$

- BTW. MLE for the variance of a Gaussian is **biased**

- Expected result of estimation is **not** true parameter!
- Unbiased variance estimator:

$$\hat{\sigma}_{unbiased}^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \hat{\mu})^2$$

©2005-2013 Carlos Guestrin

55

# What you need to know...

- Learning is...

- Collect some data
  - E.g., thumbtack flips
- Choose a hypothesis class or model
  - E.g., binomial
- Choose a loss function
  - E.g., data likelihood
- Choose an optimization procedure
  - E.g., set derivative to zero to obtain MLE
- Collect the big bucks

- Like everything in life, there is a lot more to learn...

- Many more facets... Many more nuances...
- The fun will continue...

©2005-2013 Carlos Guestrin

56