

<http://www.cs.washington.edu/education/courses/cse546/14au/>

What's learning? Point Estimation

Machine Learning – CSE546
Carlos Guestrin
University of Washington
September 25, 2014

©2005-2014 Carlos Guestrin

1

What is Machine Learning ?

©2005-2014 Carlos Guestrin

2

Machine Learning

Study of algorithms that

- improve their performance
- at some task
- with experience



©2005-2014 Carlos Guestrin

3

$X \rightarrow \begin{cases} \text{spam, not spam} \\ \text{sports, politics, enter...} \end{cases}$

Classification

from data to discrete classes

©2005-2014 Carlos Guestrin

4

Spam filtering

sexy ..> S3xy ..> S3xyS. :>

data
prediction

Osman Khan to Carlos show details Jan 16 (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: behorney, bcc: 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

X

f


f(x) → not spam

f(x) → not Spam

f(x) → spam

5

Text classification



→ Company home page

VS

Personal home page

VS

Univeristy home page

VS

...

©2005-2014 Carlos Guestrin

6

Object detection

(Prof. H. Schneiderman)

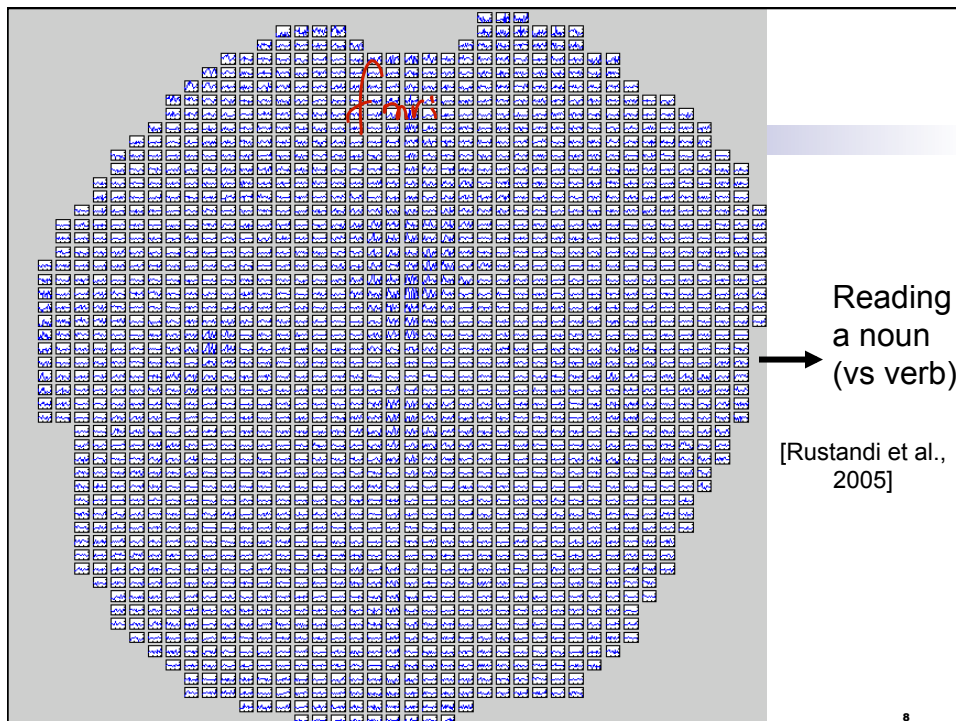


Example training images for each orientation



©2005-2014 Carlos Guestrin

7



8

$$X \rightarrow Y^e + R$$

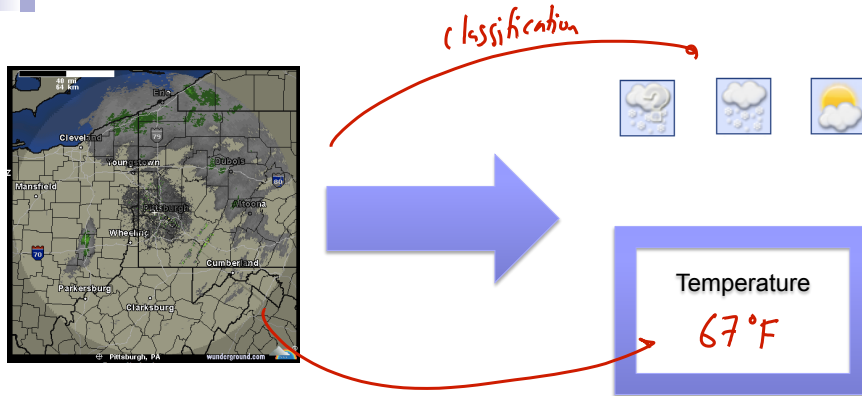
Regression

predicting a numeric value

Stock market



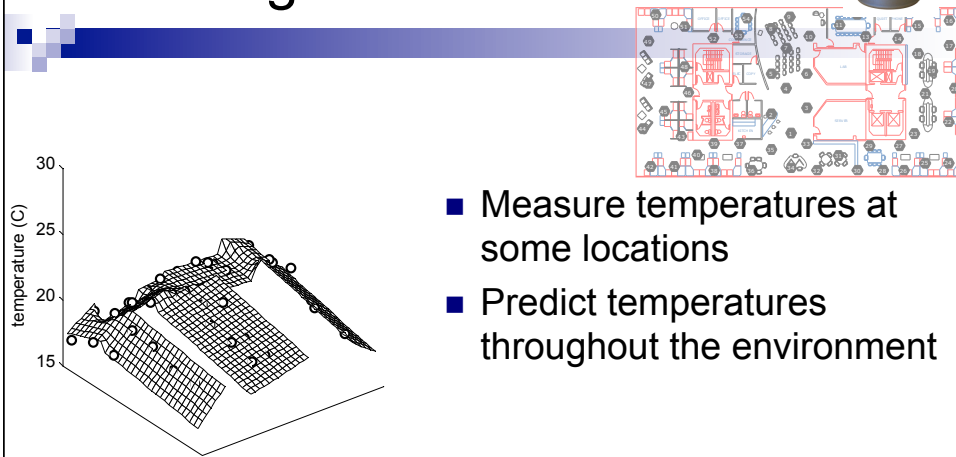
Weather prediction revisited



©2005-2014 Carlos Guestrin

13

Modeling sensor data



[Guestrin et al. '04]

©2005-2014 Carlos Guestrin

14

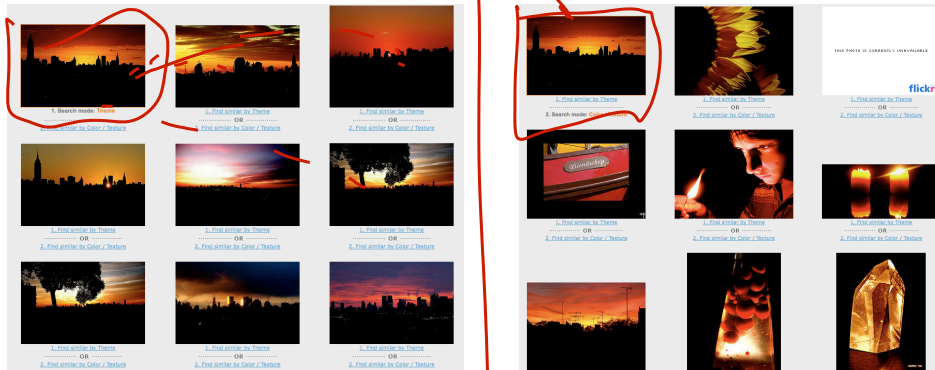
Unsupervised

Similarity

finding data

Given image, find similar images

same input



Sun sets

color

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.
\$3.99 new from \$47.95 | \$ 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 opinion with us](#)
[Write a review](#)
[Write a review](#)
[Write a review](#)
[Write a review](#)
[Write a review](#)

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 [Cancel](#) [View](#) | [Change location](#) [View](#)

[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 XML Training](#)
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 Connected 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 Programming with the Processing Language](#) by Daniel Shiffman
★★★★★ (7) \$44.05

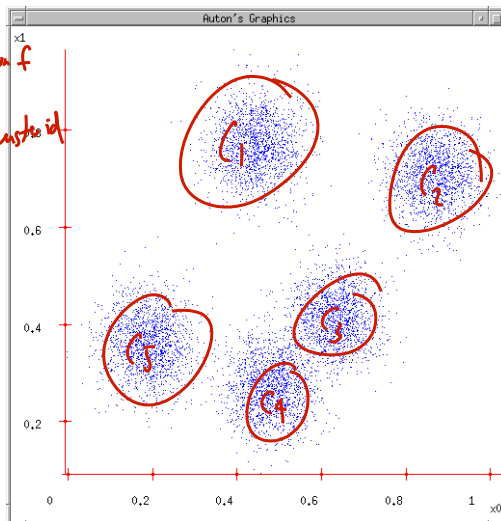
©2005-2014 Carlos Guestrin 17

Clustering

discovering structure in data

Clustering Data: Group similar things

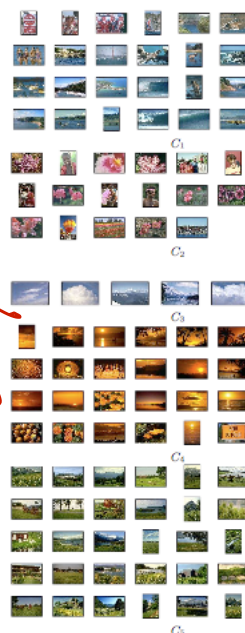
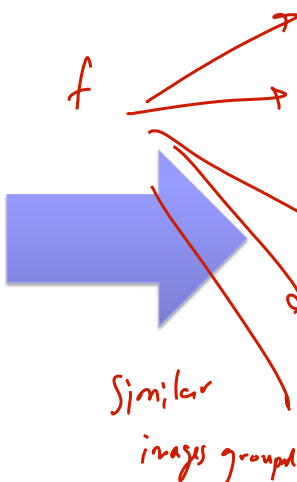
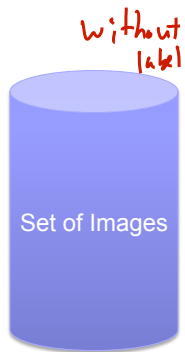
Learning a function f
 f : location a cluster
diff
Supervised learning
training data
had labeled
classes



©2005-2014 Carlos Guestrin

19

Clustering images



©2005-2014 Carlos Guestrin

[Goldberger et al.]₂₀

Clustering web search results

The screenshot shows the Clusty search interface. At the top, there are navigation links for 'news', 'images', 'wikipedia', 'blogs', 'jobs', and 'more'. A search bar contains the word 'race', and a 'Search' button is next to it. Below the search bar, a sidebar on the left lists various clusters under the heading 'All Results (238)'. The 'Human' cluster is selected, showing a list of sub-clusters: 'Classification Of Human (2)', 'Statement, Evolved (2)', 'Other Topics (4)', 'Weekend (8)', 'Ethnicity And Race (7)', 'Race for the Cure (8)', and 'Race Information (8)'. The main content area displays a list of search results for the 'Human' cluster, with the first result being 'Race (classification of human beings) - Wikipedia, the free ...'. The results include snippets of text from various sources, such as Wikipedia, Answers.com, and Dopefish.com. A 'Find in clusters' search box is located at the bottom left of the results area. The footer of the page contains the copyright information '©2005-2014 Carlos Guestrin' and the page number '21'.

Embedding

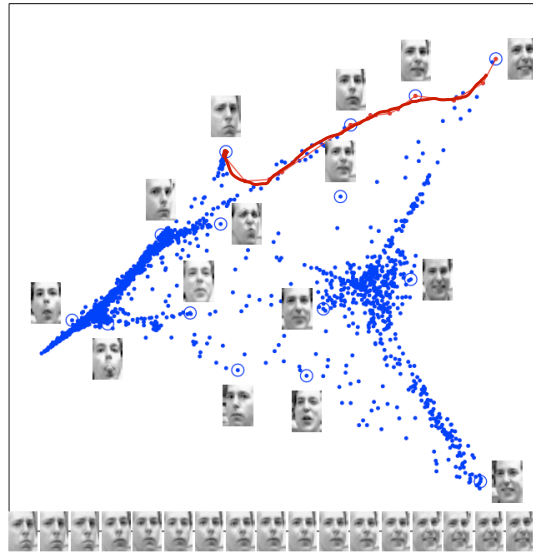
visualizing data

Embedding images

1000,000 dims
 $X \rightarrow Y$ *20 dims*

Images have thousands or millions of pixels.

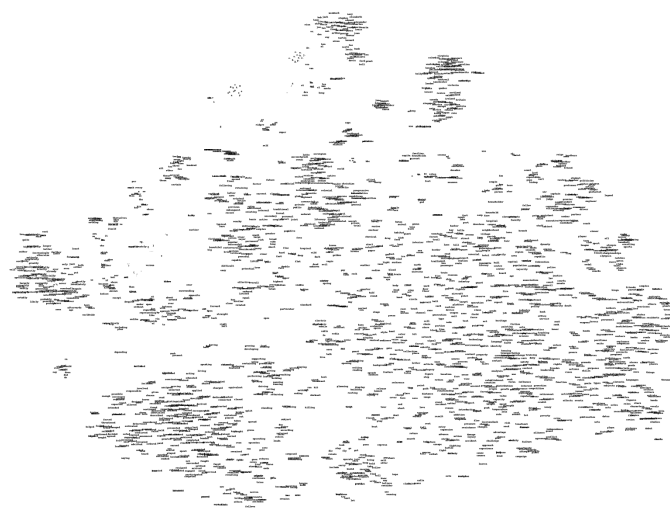
Can we give each image a coordinate, such that similar images are near each other?



©2005-2014 Carlos Guestrin

[Saul & Roweis '03] 23

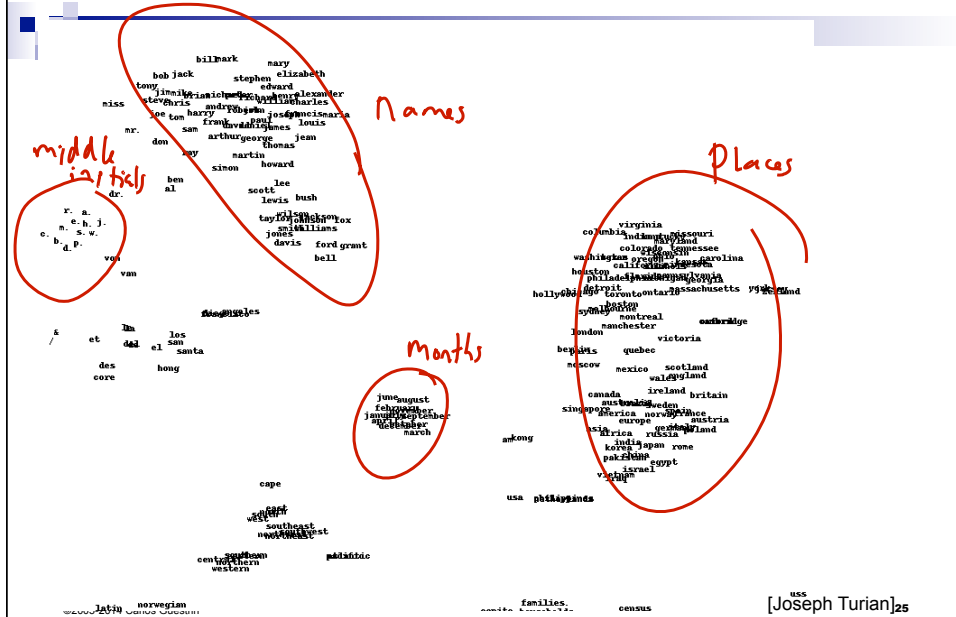
Embedding words



©2005-2014 Carlos Guestrin

[Joseph Turian] 24

Embedding words (zoom in)



Supervised classification ← MANY → unsupervised clustering

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.
©2005-2014 Carlos Guestrin

Automatically Discovered and Labeled Actions

shout
(JACK) (shouts) ()

smile
(Kate) (smiles) ()

follow
(Kate) (follows) (Jack)

sit down
(Locke) (sits down) ()

wake
(Sawyer) (wakes up) ()

swim
(Sawyer) (turns) (swimming)

grab
(Kate) (grabs) (case)

kiss
(Shannon) (kisses) (ear)

open door
(door) (opens) ()

point
(JACK) (points) ()

©2005-2014 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

©2005-2014 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 😊**

©2005-2014 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
 - Python 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:
 - Time/Location ~~TBD~~ EEB 045
Wednesdays @ 5-6:20pm
- We are recommending Python for homeworks!
 - There are many resources to get started with Python online
 - We'll run an **optional** tutorial:
 - First recitation: next week

Staff

- Three Great TAs: Great resource for learning, interact with them!
 - **April Shen**
Office hours:

 - **Harley Montgomery**
Office hours:

 - **Tianqi Chen**
Office hours:

 - Prof: **Carlos Guestrin**
Office hours: Thursdays 12:30-1:30pm (starting next week)

©2005-2014 Carlos Guestrin

35

Communication Channels

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

©2005-2014 Carlos Guestrin

36

Homeworks

- Homeworks are hard, start early ☺
- Due in the beginning of class
- 33% subtracted per late day *without penalty*
- 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

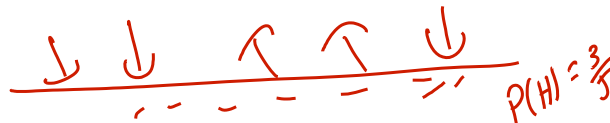
- An opportunity to exercise what you learned and to learn new things
- Individually or groups of two
- Must involve real data
 - Must be data that you have available to you by the time of the project proposals
- Must involve machine learning
- It's encouraged to be related to your research, but must 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
- Tue., October 21 at 10:30am: **Project Proposals**
- Tue., November 11 at 10:30am: **Project Milestone**
- Thu., December 4, 2:30-4:30pm: **Poster Session**
- ~~Fri.~~ Thu., December ~~12~~ 11 at 10:30am: **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(Heads) = θ , P(Tails) = $1-\theta$

$$P(D|\theta) = P(HHTTTH|\theta) = \theta \theta (1-\theta) (1-\theta) \theta$$

$$= \theta^3 (1-\theta)^2$$

independent
identically
distributed

IID

HHTTH

⋮

HHHTT

⋮

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

- Sequence D of α_H Heads and α_T Tails

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

©2005-2014 Carlos Guestrin

43

Maximum Likelihood Estimation

- **Data:** Observed set D of α_H Heads and α_T Tails
- **Hypothesis:** Binomial distribution
- Learning θ is an optimization problem

- What's the objective function?

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

- MLE: Choose θ that maximizes the probability of observed data:

$$\hat{\theta}_{MLE} = \arg \max_{\theta} P(D | \theta)$$

$$= \arg \max_{\theta} \ln P(D | \theta)$$

$\arg \max f(x)$

$\equiv \arg \min -f(x)$

- loss function

©2005-2014 Carlos Guestrin

44

Your first learning algorithm

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

$$\begin{aligned}\ln ab &= b \ln a \\ \ln ab &= \ln a + b \ln 1 \\ \frac{d \ln \theta}{d\theta} &= \frac{1}{\theta}\end{aligned}$$

- Set derivative to zero:

$$\frac{d}{d\theta} \ln P(\mathcal{D} | \theta) = 0$$

$$\frac{d}{d\theta} \ln \theta^{\alpha_H} (1 - \theta)^{\alpha_T} = \frac{d}{d\theta} (\alpha_H \ln \theta + \alpha_T \ln(1 - \theta)) = \alpha_H \frac{d \ln \theta}{d\theta} + \alpha_T \frac{d \ln(1 - \theta)}{d\theta}$$

$$= \frac{\alpha_H}{\theta} - \frac{\alpha_T}{1 - \theta} = 0 \Rightarrow \theta = \frac{\alpha_H}{\alpha_H + \alpha_T} \Rightarrow \frac{3}{5} !!$$

©2005-2014 Carlos Guestrin

45

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: $\hat{\theta} = 3/5$, 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???

©2005-2014 Carlos Guestrin

46

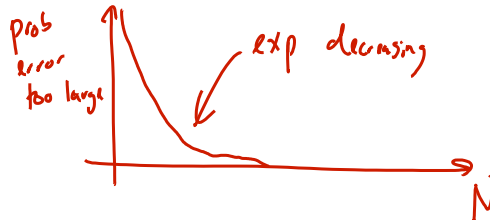
Simple bound (based on Hoeffding's inequality)

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

$$|\theta^* - \hat{\theta}_{MLE}| ?$$

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

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



©2005-2014 Carlos Guestrin

47

PAC Learning

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

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

$$\ln \delta \geq \ln 2 - 2N\epsilon^2$$

$$\Rightarrow N \geq \frac{\ln \frac{2}{\delta}}{2\epsilon^2}$$

$$\text{if } \delta = 0.05, \epsilon = 0.1$$

$$N \geq 184.4 \text{ flips}$$

loose bound

©2005-2014 Carlos Guestrin

48

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...**

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



©2005-2014 Carlos Guestrin

49

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
 - $X \sim N(\mu_X, \sigma_X^2)$
 - $Y \sim N(\mu_Y, \sigma_Y^2)$
 - $Z = X + Y \rightarrow Z \sim N(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$

©2005-2014 Carlos Guestrin

50

Learning a Gaussian

HW v. 98
scores : 16
85
x₁₀ :

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

- Learn parameters
 - Mean
 - Variance

$\hat{\mu} = \frac{1}{N} \sum_{i=1}^N x_i$ ← why? MLE

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

MLE for Gaussian

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

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

$\hat{\mu}_{MLE}, \hat{\sigma}_{MLE}^2 = \underset{\mu, \sigma}{\operatorname{argmax}} P(D | \mu, \sigma) = \underset{\mu, \sigma}{\operatorname{argmax}} \ln P(D | \mu, \sigma)$

- Log-likelihood of data:

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

$\max_{\mu, \sigma} = -N \ln \sigma \sqrt{2\pi} - \sum_{i=1}^N \frac{(x_i - \mu)^2}{2\sigma^2}$