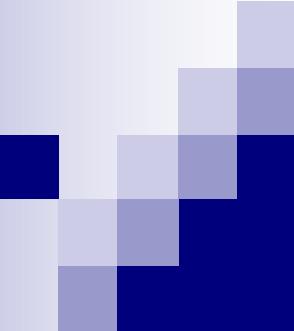


<http://www.cs.washington.edu/education/courses/cse546/16au/>



What's learning? Point Estimation

Machine Learning – CSE546

Sham Kakade

University of Washington

September 28, 2016

©2016 Sham Kakade

1



What is Machine Learning ?

©2016 Sham Kakade

2

Machine Learning

- Study of algorithms that
 - improve their performance
 - at some task
 - with experience



©2016 Sham Kakade

3

Classification

from data to discrete classes

©2016 Sham Kakade

4

Spam filtering

data

prediction

Natural _LoseWeight SuperFood Endorsed by Oprah Winfrey, Free Trial 1 bottle,
pay only \$5.95 for shipping mfw rlk [Spam](#) | [X](#)

★ Jaquelyn Halley to nherlein, bcc: thehorney, bcc: anç 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

©2016 Sham Kakade

5

Text classification



→ Company home page

VS

Personal home page

VS

University home page

VS

...

©2016 Sham Kakade

6

Object detection

(Prof. H. Schneiderman)

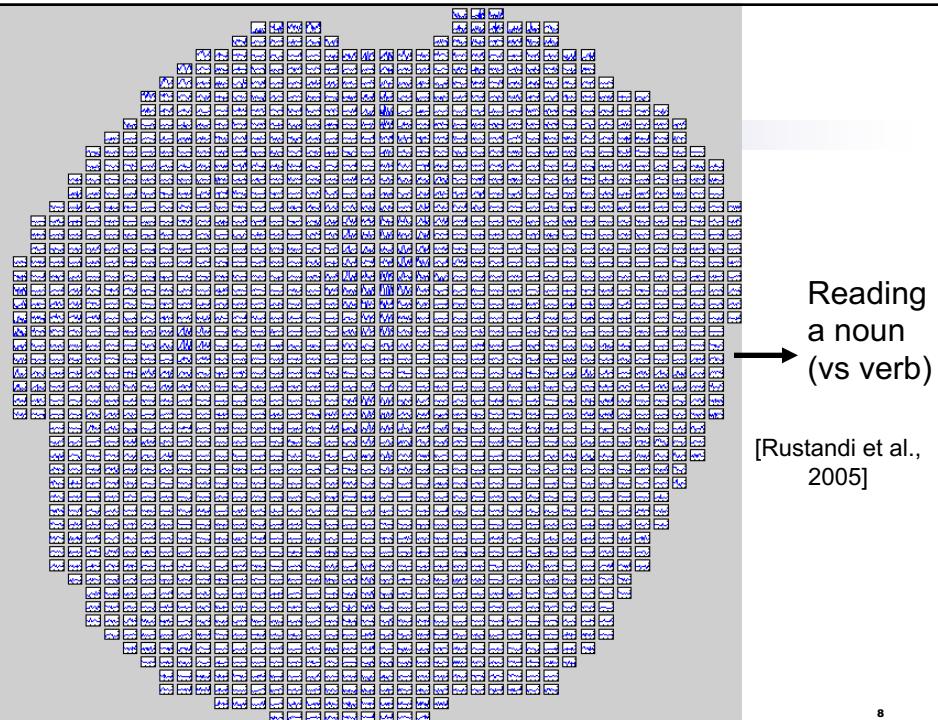


Example training images
for each orientation



©2016 Sham Kakade

7



8

Weather prediction



©2016 Sham Kakade

9

The classification pipeline

Training

Carrie Wears It Carries
source good
link
Carrie Guevara writes:
Let's try to chat on Friday a little to coordinate and more on Sunday in person?
Carrie

Natural Weight Loss Product Endorsed by Oprah Winfrey. Free Trial 1 bottle, pay only \$3.95 for shipping and tax.
Join now! [Join now!](#) [View details](#) [\\$3.95 / 1 hour ago](#) [#1000](#) [#1000](#)

Vital Aire is a natural Weight Loss product that Enables people to lose weight and cleansing their bodies faster than most other products on the market. It is a natural product that can help you lose weight that you might not be aware of. These benefits have helped people who have been using Vital Aire daily to Achieve goals and reach new heights in there dieting that they never thought possible.

* Rapid Weight Loss
* Natural Weight Loss
* Better Mood and Attitude
* Helps Detoxify Your Body
* Cleanse and Detoxify Your Body
* Detoxify Your Colon
* A Natural Colon Cleanser

Testing

Welcome to New Media Installation: Art that Learns
Carrie Guevara to 1911e-announces, Owner, Mucho [\(view profile\)](#) 2:15 PM (3 hours ago) [#1000](#) [#1000](#)

Welcome to New Media Installation:Art that Learns
The class will start November
Meeting every Saturday at 1pm, even if you are on the West Up
The classes are held in Schreyer Hall C303, and will be Tue, Thu 1:30-2:45 PM
By now you should be subscribed to our course mailing list: [DE113announces@psu.edu](#)
You can contact the professor by emailing: [DE113announces@psu.edu](#)

©2016 Sham Kakade

10

Regression

predicting a numeric value

©2016 Sham Kakade

11

Stock market

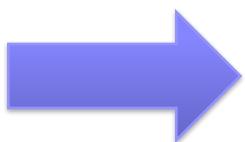
Jan 12, 2009 : ^DJI 8,473.9697



©2016 Sham Kakade

12

Weather prediction revisited

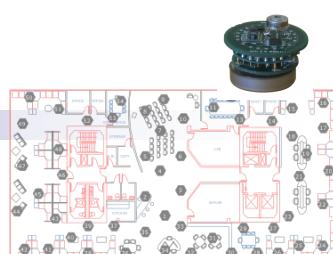
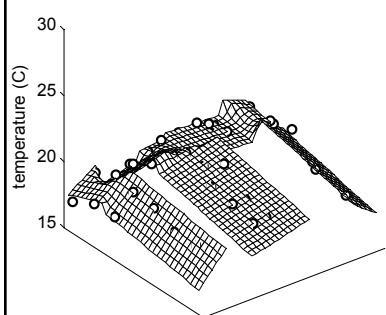


Temperature

©2016 Sham Kakade

13

Modeling sensor data



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

©2016 Sham Kakade

14

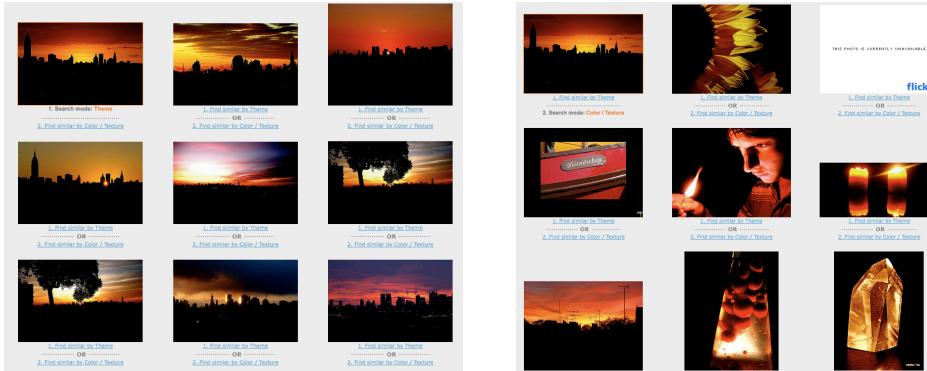
Similarity

finding data

©2016 Sham Kakade

15

Given image, find similar images



©2016 Sham Kakade

<http://www.tiltomo.com/> 16

Similar products

The screenshot shows a product page for 'Processing: A Programming Handbook for Visual Designers and Artists (Hardcover)'. The page includes the book cover, author information (Casey Reas, Ben Fry, John Maeda), customer reviews (4.5 stars from 13 reviews), and purchase options (new from \$47.95, used from \$43.56). It also features a 'Get Free Two-Day Shipping' offer and related education services in Pittsburgh.

Processing: A Programming Handbook for Visual Designers and Artists (Hardcover)
by [Casey Reas](#) (Author), [Ben Fry](#) (Author), [John Maeda](#) (Foreword)
4.5 out of 5 stars (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](#)
[Leave a review](#)
Don't have a Kindle? [Get your kindle](#)

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

[Learn HTML Coding](#)
[www.Full Sail.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.objectcallos.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 \$43.99 | Visualizing Data: Exploring and Explaining Data... by Ben Fry \$26.39 | Making Things Talk: Practical Methods for Controlling the Physical World with Arduino by Tom Igoe \$19.79 | Physical Computing: Sensing and Controlling the Physical World with Arduino by Tom Igoe \$19.00 | Learning Processing: A Beginner's Guide to Programming Visual Art, Animation, and Interaction by Daniel Shiffman \$44.95 |

©2016 Sham Kakade

17

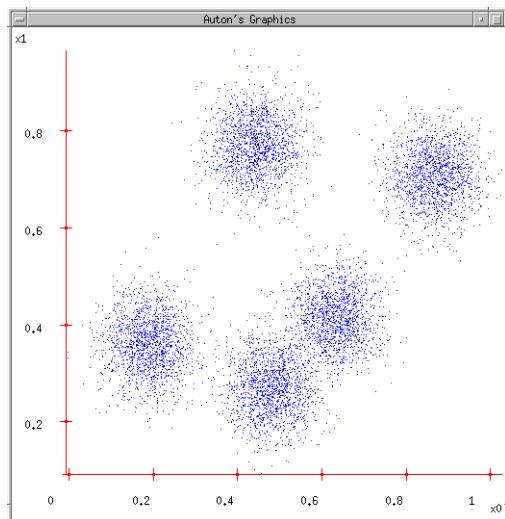
Clustering

discovering structure in data

©2016 Sham Kakade

18

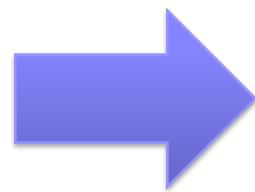
Clustering Data: Group similar things



©2016 Sham Kakade

19

Clustering images



[Goldberger et al.]₂₀

©2016 Sham Kakade

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', 'more', and a search bar containing the word 'race'. Below the search bar are buttons for 'Search' and 'advanced preferences'. The main area is titled 'Cluster Human contains 8 documents.' On the left, there's a sidebar with tabs for 'clusters', 'sources', and 'sites'. Under the 'clusters' tab, there's a tree view of search results categorized by topic: 'All Results (238)', 'Car (28)', 'Race cars (7)', 'Photos, Races Scheduled (6)', 'Game (4)', 'Track (3)', 'Nascar (2)', 'Equipment And Safety (2)', 'Other Topics (7)', 'Photos (22)', 'Game (14)', 'Definition (13)', 'Team (18)', and 'Human (8)'. Under 'Human', there are further sub-categories: 'Classification Of Human (2)', 'Statement, Evolved (2)', 'Other Topics (4)', 'Weekend (8)', 'Ethnicity And Race (7)', 'Race for the Cure (8)', 'Race Information (8)', and 'more [all clusters]'. At the bottom of the sidebar, there's a 'Find in clusters:' input field and a 'Find' button. To the right of the sidebar, the main content area lists 8 search results, each with a title, a snippet of the page content, and small icons for sharing or viewing. The results include links to Wikipedia pages on race classification, a free encyclopedia on race, Human Rights Watch, Amazon.com, the AAPA statement on biological aspects of race, and a definition from Answers.com. There's also a link to Dopefish.com. The footer of the page includes the text '©2016 Sham Kakade' and the number '21'.

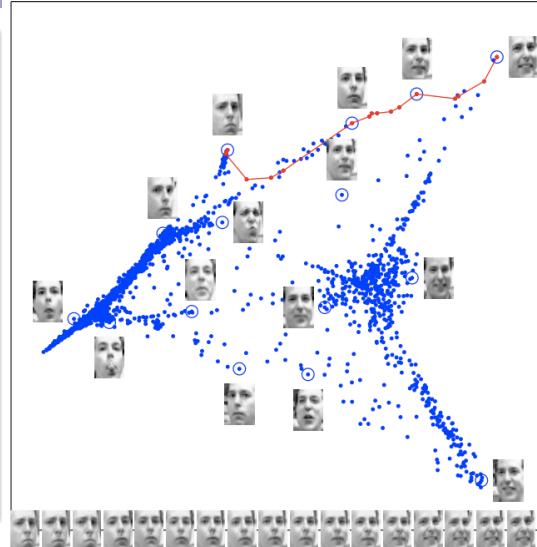
Embedding

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?



©2016 Sham Kakade

[Saul & Roweis '03] 23

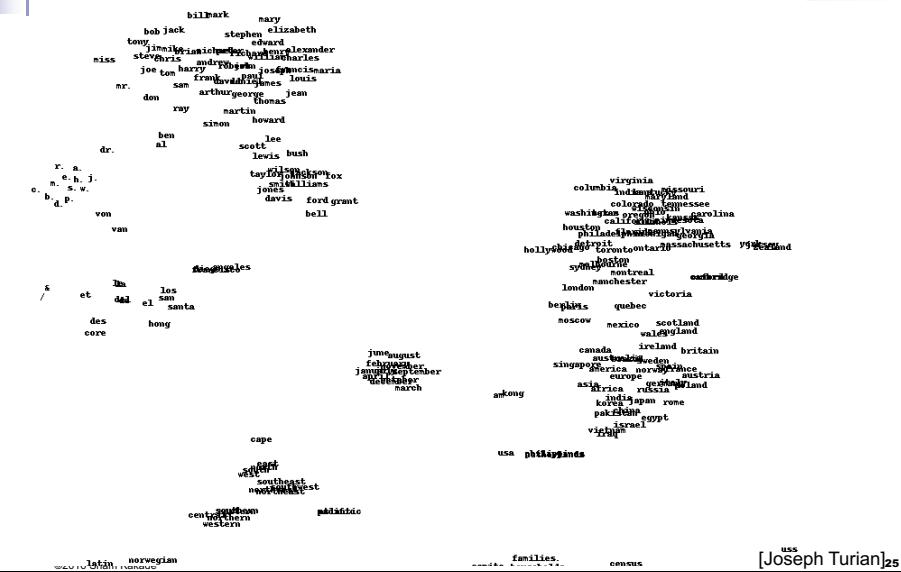
Embedding words

©2016 Sham Kakade

[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]

©2016 Sham Kakade

27

Impact

What are the biggest successes?

©2016 Sham Kakade

28

Successes

- Speech Recognition
 - SIRI, Alexa, etc.
- Computer vision
 - ImageNet
- Alpha-Go
 - Game playing
 - Go was 'solved' with ML/AI
- And more:
 - Natural language processing
 - Robotics (self-driving cars?)
 - Medical analysis
 - Computational biology

©2016 Sham Kakade

29

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

©2016 Sham Kakade

30

Logistics

©2016 Sham Kakade

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, logistic regression, optimization, nearest-neighbor, decision trees, boosting, perceptron, overfitting, regularization, dimensionality reduction, PCA, error bounds, SVMs, kernels, margin bounds, K-means, EM, mixture models, HMMs, graphical models, deep learning, reinforcement learning...
- Covers algorithms, theory and applications
- **It's going to be fun and hard work.**

©2016 Sham Kakade

32

Prerequisites

- Linear algebra:
 - SVDs, eigenvectors, matrix multiplication
- Probabilities
 - Distributions, densities, marginalization...
- Basic statistics
 - Moments, typical distributions, regression...
- Algorithms
 - Dynamic programming, basic data structures, complexity...
- Programming
 - Python will be very useful
- We provide some background, but the class will be fast paced
- Ability to deal with “abstract mathematical concepts”

©2016 Sham Kakade

33

Recitations & Python

- We'll run an ***optional*** recitations:
 - Time/Location
- 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

©2016 Sham Kakade

34

Staff

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

- **Dae Hyun Lee**
Office hours: TBD

- **Angli Liu**
Office hours: TBD

- **Alon Milchgrub**
Office hours: TBD

Communication Channels

- Announcements on Canvas.
- Use the Discussion board!
 - 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 and grading use:
 - cse546-instructors@cs.washington.edu
- Office hours limited to knowledge based questions. Use email for all grading questions.

Text Books

- Required Textbook:
 - Machine Learning: a Probabilistic Perspective; Kevin Murphy
- Optional Books:
 - Understanding Machine Learning: From Theory to Algorithms; Shai Shalev-Shwartz and Shai Ben-David.
 - 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

©2016 Sham Kakade

37

Grading

- 4 homeworks (65%)
 - First posted today
 - Start early!
 - HW 1,2,4 (15%)
 - Collaboration allowed
 - You must write (and submit) your own code, which we may run.
 - You must write (and understand) your own answers.
 - HW 3 midterm (20%)
 - No collaboration allowed.
- Final project (35%)
 - Full details: see website
 - Projects done individually, or groups of two students

©2016 Sham Kakade

38

HW Policy (SEE WEBSITE)

- Homeworks are hard/long, start early
 - Heavy programming component.
 - They will build on themselves (you will re-use your code).
- 33% subtracted per late day.
- You have 2 LATE DAYS to use for homeworks throughout the quarter
 - Please plan accordingly.
 - No exceptions (aside from university policies).
- All homeworks **must be handed in**, even for zero credit.
- Use Canvas to submit homeworks.
- No collaboration allowed on HW 3
- Collaboration: HW 1,2,4
 - Each student writes (and understands) their own answers.
 - You may **discuss** the questions.
 - 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

©2016 Sham Kakade

39

Projects (35%)

- SEE WEBSITE
- An opportunity/intro for research.
 - encouraged to be related to your research, but must be something new you did this quarter
 - It's Not a project you worked on during the summer, last year, etc.
- Grading:
 - We seek some novel exploration.
 - If you write your own code, great. We take this into account for grading.
 - You may use ML toolkits (e.g. TensorFlow, etc), then we expect more ambitious project (in terms of scope, data, etc).
 - If you use simpler/smaller datasets, then we expect a more involved analysis.
- 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

©2016 Sham Kakade

40

(tentative) project dates (35%)

- Full details in a couple of weeks
- Mon., October 24, 5p: **Project Proposals**
- Mon., November 14, 5p: **Project Milestone**
- Thu., December 8, 9-11:30am: **Poster Session**
- Thu., December 15, 10am: **Project Report**

©2016 Sham Kakade

41

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
- Have fun..

©2016 Sham Kakade

42

A Data Science Job

- Someone asks you a stat/data science question:
 - She 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:
- **She says: Why???**
- You say: Because...

©2016 Sham Kakade

43

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 α_H Heads and α_T Tails

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

©2016 Sham Kakade

44

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

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

©2016 Sham Kakade

45

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

©2016 Sham Kakade

46

How many flips do I need?

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

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

47

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

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

©2016 Sham Kakade

48

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?

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

©2016 Sham Kakade

49

What about continuous variables?

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

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

©2016 Sham Kakade

50

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

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

©2016 Sham Kakade

51

Learning a Gaussian

- Collect a bunch of data

- Hopefully, i.i.d. samples
 - e.g., exam scores

- Learn parameters

- Mean
 - Variance

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

©2016 Sham Kakade

52

MLE for Gaussian

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

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

- Log-likelihood of data:

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

©2016 Sham Kakade

53

Your second learning algorithm: MLE for mean of a Gaussian

- What's MLE for mean?

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

©2016 Sham Kakade

54

MLE for variance

- Again, set derivative to zero:

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

©2016 Sham Kakade

55

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:

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

©2016 Sham Kakade

56

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
- Like everything in life, there is a lot more to learn...
 - Many more facets... Many more nuances...
 - More later...