Lecture 1:
Welcome to CS88!
In The News

Job Screening Service Halts Facial Analysis of Applicants
But it’s still using intonation and behavior to assist with hiring decisions.

HireVue, a leading provider of software for vetting job candidates based on an algorithmic assessment, said Tuesday it is killing off a controversial feature of its software: analyzing a person’s facial expressions in a video to discern certain characteristics.

Job seekers screened by HireVue sit in front of a webcam and answer questions. Their behavior, intonation, and speech is fed to an algorithm that assigns certain traits and qualities.

Goals today

• Introduce you to
  – the field
  – the course
  – the team
• Answer your questions

• Big Ideas:
  – Abstraction
  – Data Type
CS88 Team – Gerald and Michael

• Michael Ball
  – ball@Berkeley.edu – You’re best off by using Ed! 😊
  – 625 Soda Hall / Berkeley.zoom.us / my apartment
  – http://michaelball.co – I don’t update this much...
    ➞ It was great procrastination when I was a CS student.
  – Office hours: tentatively Wednesday early afternoon.

• Things I do:
  – Intro CS Research (Tools, curriculum)
  – Training TAs
  – Building Educational Software (Gradescope)
  – Tools for web accessibility
CS88 Team – Gerald and Michael

• Gerald Friedland
  - fractor@Berkeley.edu – You’re best off by using Ed! 😊
  - 424 SD Hall / Berkeley.zoom.us
  - https://www.icsi.berkeley.edu/~fractor – I don’t update this much either...
  - Office hours: tentatively Monday 2pm

• Things I do:
  - Multimedia Computing
  - Machine Learning Research/Data Science Measurements
  - Discussion Group at Berkeley Institute for Data Science (BIDS)
Head Teaching Assistants

Shreya Kannan
Office Hours: Friday 10 am - 12 pm
shreya.kannan@berkeley.edu
Hi everyone, I'm super thrilled to meet y'all this semester! In no particular order, I prefer musicals over concerts, cold over warm weather, dancing over singing, and sushi over tacos. Feel free to talk to me about great restaurants, cool study spots, and of course, CS 88 :)

Vandana Ganesh
Office Hours: Monday 10 am - 12 pm
vandanag@berkeley.edu
Hey everyone, I am super excited to meet all of you virtually! This is my 3rd time TAing for 88, but my 5th time being involved with it and I look forward to a great semester. A bit about me- I enjoy going on walks, attempting painting/drawing (recently), and going through my favorite tv shows, movies, and books. Feel free to come to me with recommendations and definitely feel free to reach out about CS 88!

Teaching Assistants

Lukas Chang
Office Hours: Thursday 1 - 2 pm
lukasc@berkeley.edu
Hi everyone! I’m excited to work with you all this upcoming semester! I’m Lukas, a 2nd year CS major from the bay area. This is my first semester TAing, but my second teaching CS88 material. A little about me—in my free time I love making music though I kind of suck at playing piano, and I also like going on nightly walks. I hope I can share my love for CS with all of you!

Nicholas Ng
Office Hours: Tuesday 2 - 3 pm
nickng88@berkeley.edu
Hi! I’m Nicholas, and I’m a third year Computer Science Major from the San Gabriel Valley. I like playing video games (I probably spent too much of my summer playing Valorant), and every once in a while I’ll go outside for a run (but not really though). I look forward to meeting everyone!

Sophia Qin
Office Hours: Tuesday 9 - 10 am
sophia.qin@berkeley.edu
Hi! I’m Sophia, a fourth-year EECS major from sunny SoCal. In my free time, I love experimenting with new dessert recipes, watching sunsets, and bingeting the Office. Feel free to reach out to me about anything! :)

Tommy Joseph
Office Hours: Tuesday 1 - 2 pm
tommy11jc@berkeley.edu
Hi, I’m a second-year CS major from Southern California. I enjoy experimenting with pasta recipes and playing sports that involve rackets or paddles. I look forward to meeting you all!
Hi everyone! I’m Kevin, a 4th year Data Science major and CS minor from Irvine, So Cal. I am so excited to be tutoring in this class again as I really like the content in this class. While I’m not analyzing data or programming/coding, I like to appreciate the humanities in general. Specifically, I love music, especially classical music or just any classics in general. I also love learning languages and enjoy talking/typing in French. Feel free to talk to me in any of these things, in either English, French (but speak SLOWLY please!), or even Python print statements and strings!

Hi, I’m Minnie and I’m a third year Civil Engineering major and EECS+Sustainability minor from San Gabriel Valley. When I’m lazy, I enjoy watching all kinds of movies/shows (from cheesy rom-coms to horrors and action) and when I’m not feeling like a couch potato, I love to do physical activities like hiking or basketball. Excited to meet everyone!
Course Structure

• 2 lectures, 1 lab each week
• Lecture introduces concepts (quickly!), answers why questions.
• Lab provides concrete detail hands-on
• Homework (12) cements your understanding
• Projects (2) put your understanding to work in building complete applications
  – Maps
  – Ants vs Some Bees

• Readings: http://composingprograms.com
  – Same as cs61a
Class Format

• Mon and Weds Lectures:
  – Some lectures will be pre-recorded, class time will be largely demo/Q&A
  – Each lecture has a series of short self-check questions

• Labs are paced throughout the week. See the Ed post to pick a time.

• Labs are HANDS ON – get help as you’re trying the lab.
Class Format: Assignments

• Lecture Quizzes, 1 point, max 20.
  – 1 per lecture, due in 1 week. (Half credit after)

• Lab Work: 4 points, 12 labs, 1 drop
  – Start them during lab! You can probably finish some labs in 2 hours. Will be Python + some interactive questions.
  – Out Tues, due next Tues Night.

• Homework: 8 points, 12 HW, 1 drop
  – Start early!
  – Out Weds, Due Next Friday Night
Class Format: Assignments

• **Projects: 100 points between 2 projects**
  - Start early! "Checkpoint" assignments

• **Slip Days: 8 total**
  - Use up to 3 on any assignment
  - We apply the in the order that’s most beneficial!
    » i.e. use them on projects if you need!
  - Can be used for homework, labs, projects, but not project checkpoints.

• **Slip Days take care of nearly all, but not all special circumstances!**

• What if you go over slip days?
  - 25% deduction for each day over. Mathematically you can still earn 25% if you turn in something 3 days late.
Class Format: Exams

• 1 midterm and 1 final exam

• Midterm 2 hours, “Mid march”, sometime before Spring Break; 2 hours.

• Exam will be during the slot assigned by campus.

• We will be proctoring via Zoom. Exact policies coming soon, but, essentially, you’ll record yourself at home.

• Open book, but no collaboration.
Nearly every field of discovery is transitioning from “data poor” to “data rich”
A National Challenge

Increasingly US jobs require data science and analytics skills. Can we meet the demand? The current shortage of skills in the national job pool demonstrates that business-as-usual strategies won’t satisfy the growing need. If we are to unlock the promise and potential of data and all the technologies that depend on it, employers and educators will have to transform.

By 2021, 69% of employers expect candidates with DSA skills to get preference for jobs in their organizations. Only 23% of college and university leaders say their graduates will have those skills.
Greatest Artifact of Human Civilization ...
### WORLD INTERNET USAGE AND POPULATION STATISTICS
**DEC 31, 2017 - Update**

<table>
<thead>
<tr>
<th>World Regions</th>
<th>Population (2018 Est.)</th>
<th>Population % of World</th>
<th>Internet Users 31 Dec 2017</th>
<th>Penetration Rate (% Pop.)</th>
<th>Growth 2000-2018</th>
</tr>
</thead>
<tbody>
<tr>
<td>Africa</td>
<td>1,287,914,329</td>
<td>16.9 %</td>
<td>453,329,534</td>
<td>35.2 %</td>
<td>9,941 %</td>
</tr>
<tr>
<td>Asia</td>
<td>4,207,588,157</td>
<td>55.1 %</td>
<td>2,023,630,194</td>
<td>48.1 %</td>
<td>1,670 %</td>
</tr>
<tr>
<td>Europe</td>
<td>827,650,849</td>
<td>10.8 %</td>
<td>704,833,752</td>
<td>85.2 %</td>
<td>570 %</td>
</tr>
<tr>
<td>Latin America / Caribbean</td>
<td>652,047,996</td>
<td>8.5 %</td>
<td>437,001,277</td>
<td>67.0 %</td>
<td>2,318 %</td>
</tr>
<tr>
<td>Middle East</td>
<td>254,438,981</td>
<td>3.3 %</td>
<td>164,037,259</td>
<td>64.5 %</td>
<td>4,893 %</td>
</tr>
<tr>
<td>North America</td>
<td>363,844,662</td>
<td>4.8 %</td>
<td>345,660,847</td>
<td>95.0 %</td>
<td>219 %</td>
</tr>
<tr>
<td>Oceania / Australia</td>
<td>41,273,454</td>
<td>0.6 %</td>
<td>28,439,277</td>
<td>68.9 %</td>
<td>273 %</td>
</tr>
<tr>
<td><strong>WORLD TOTAL</strong></td>
<td><strong>7,634,758,428</strong></td>
<td><strong>100.0 %</strong></td>
<td><strong>4,156,932,140</strong></td>
<td><strong>54.4 %</strong></td>
<td><strong>1,052 %</strong></td>
</tr>
</tbody>
</table>

---

**Internet Usage Statistics**

- Internet Users in the world: 4,003,737,907
- Total number of Websites: 1,906,393,398
- Emails sent today: 124,065,255,834
- Google searches today: 3,026,650,785
- Blog posts written today: 2,858,890
- Tweets sent today: 357,960,955
- Videos viewed today on YouTube: 3,297,002,756
- Photos uploaded today on Instagram: 37,897,179
- Tumbler posts today: 62,202,109
Era of Transformation

Connected

Industrial Revolution

Age of Enlightenment

World 21st Century
A Connected World of Data

• The world’s knowledge at our finger tips
• Digitalization of life, industry and society
• Intimately connected to billions of us, globally
• Explosion of observational instruments
  – Genomics, Microscopy, Astronomical, …
• Vast Computational power to do analytics
• Synthetic design exploration thru simulation
• Machine reading of everything
• Statistical machine learning algorithms to “discover” structure
What if I could … ?

• See the world’s digital footprints?
• Read everything that’s ever been written?
• Take it all in and dive down anywhere as far as the science can take me?
• Learn the physical/chemical/biological/sociological/neurological… models from the data?
• Explore billions of designs and pick the one I want?
• … ?
A Connected World

Internet Users in the world

3,293,151,639

RFC 675 TCP/IP

WWW

ARPANet

3.0 B 11/15

2.0 B 1/26/11

1969

1974

HTTP

1974

1990

2010

UCB CS88 Fa20 L1

2,652,887,737

Google searches today

5,835,884,253

Videos viewed today on YouTube

UC Berkeley

UCP CS88 Fa20 L1

http://cs88.org
Data 8 – Foundations of Data Science

• Computational Thinking + Inferential Thinking in the context of working with real world data

• Introduce you to several computational concepts in a simple data-centered setting
  – Authoring computational documents
  – Tables
  – Within Python3 and “SciPy”
CS88 – Computational Structures in Data Science

- Deeper understanding of the computing concepts introduced in c8
  - Hands-on experience => Foundational Concept
  - How would you create what you use in c8?
- Extend your understanding of the structure of computation
  - What is involved in interpreting the code you write?
  - Deeper CS Concepts: Recursion, Objects, Classes, Higher-order Functions, Declarative programming, ...
  - Managing complexity in creating larger software systems through composition
- Create complete (and fun) applications
- In a data-centric approach
How does CS88 relate to CS61A?

![Diagram showing the relationship between CS61A, DATA8, and CS88 units.

- CS61A: Intro Programming & Tools, CS Concepts and Techniques
- DATA8: Statistics Concepts in a Computational Approach, Intro Programming
- CS88: Working w/ Data, CS Concepts and Techniques & Tools]
Opportunities for students

- CS minor
- CS major

UC Berkeley | Computer Science 88 | Michael Ball | http://cs88.org
The Data Science Major

Individualized Upper Division
30 units

- Computational & Inferential Depth
- Modeling, Learning & Decision Making
- Probability
- Domain Emphasis
- Human Contexts & Ethics
- Electives

Foundational Lower Division

- Data 8: Foundations of Data Science
- Data 100: Principles & Techniques of Data Science
- Mathematics
- Computing
- Domain Emphasis
- College Breadth & Electives

30 units

Modeling, Learning & Decision Making
Probability
Domain Emphasis
Human Contexts & Ethics
Electives
Course Culture

• Learning
• Community
• Respect
• Collaboration
• Peer Instruction
Ed For Class Discussion: Try it!

Welcome

Michael Ball  
INSTRUCTOR

Hi everyone,

Welcome to CS88 Fall 2020!!

We're just getting things setup, so you'll find some stuff is less than perfect. Please bear with us! (Bad pun intended. If you're allergic to bad puns I might recommend another course. No hard feelings.)

A Short List Week 1 Tasks:

- Please attend any lab section this week! We will be sending out a welcome survey as well as form to sign up for permanent section times. Labs in CS88 are part lab, part discussion. They're a time to meet peers and your TA. They are challenging, but hopefully interesting and engaging. There's plenty of times to get questions answered!

- Please checkout this short welcome video and let us know how you're feeling about the course.

Welcome to CS88
Computational Structures in Data Science
Fall 2020

8/26/2020
Where will we work?

• Your laptop
  – Using an editor and a terminal

• cs88.org

• datahub.berkeley.edu
  – Not as often, but an option

• us.edstem.org
  – Check out the “Workspaces”,
  – Can write and run (!!) python in you own posts!
Poll: Check In

• How has lab gone so far?

• A. Labs have gone fantastic!
• B. Labs have gone alright...
• C. Labs have gone very well...
• D. I haven’t been to lab yet.
Poll: Check In

• Are you enrolled in Data 8?
  • A. I took if Spring 2020 or earlier
  • B. I took it Fall 2020 ("last semester")
  • C. I’m taking it right now
  • D. I am trying to enroll in Data 8
  • E. I am not taking Data 8
Poll: Check In

• Where are you right now?

• A. I made it to Berkeley!
• B. I’m somewhere in California
• C. I’m somewhere else in the US
• D. I’m somewhere internationally for the semester
• E. I’ve made it to Space where there is no COVID.
Pro-student Grading Policies

• EPA
  - Rewards good behavior
  - Effort
    » E.g., Office hours, doing every single lab, hw, reading Ed posts
  - Participation
    » E.g., Raising hand in lec or discussion, asking questions
  - Altruism
    » E.g., helping other students in lab, answering questions on Ed
Abstraction

• Detail removal
  "The act of leaving out of consideration one or more properties of a complex object so as to attend to others."

• Generalization
  "The process of formulating general concepts by abstracting common properties of instances"

• Technical terms: Compression, Quantization, Clustering, Unsupervised Learning
WHERE ARE YOU FROM?
Where are you from?

Possible Answers:
• Planet Earth
• Europe
• California
• The Bay Area
• San Mateo
• 1947 Center Street, Berkeley, CA
• 37.8693° N, 122.2696° W

All correct but different levels of abstraction!
Abstraction gone wrong!
Detail Removal (in Data Science)

• You’ll want to look at only the interesting data, leave out the details, zoom in/out...
• Abstraction is the idea that you focus on the essence, the cleanest way to map the messy real world to one you can build
• Experts are often brought in to know what to remove and what to keep!

The London Underground 1928 Map & the 1933 map by Harry Beck.
The Power of Abstraction, Everywhere!

• Examples:
  – Functions (e.g., sin x)
  – Hiring contractors
  – Application Programming Interfaces (APIs)
  – Technology (e.g., cars)

• Amazing things are built when these layers
  – And the abstraction layers are getting deeper by the day!

We only need to worry about the interface, or specification, or contract
NOT how (or by whom) it’s built

Above the abstraction line

Abstraction Barrier (Interface)
(the interface, or specification, or contract)

Below the abstraction line

This is where / how / when / by whom it is actually built, which is done according to the interface, specification, or contract.
Abstraction: Pitfalls

• Abstraction is not universal without loss of information (mathematically provable). This means, in the end, the complexity can only be “moved around”

• Abstraction makes us forget how things actually work and can therefore hide bias. Example: AI and hiring decisions.

• Abstraction makes things special and that creates dependencies. Dependencies grow longer and longer over time and can become unmanageable.
Abstraction in CS: Data Type

- What’s this?

Real (or ideal) world

42

Computer representation
Your Tasks

• Lecture 1 Quiz On Gradescope
• Signup Genius form for lab times

Welcome, and Good luck!