Computational Structures in Data Science

Lecture #1: Welcome to CS88!

January 19, 2018

http://inst.eecs.berkeley.edu/~cs88

Goals today

• Introduce you to
  – the field
  – the course
  – the team

• Answer your questions

• Big Ideas:
  – Abstraction
  – Data Type

Data Science

Nearly every field of discovery is transitioning from “data poor” to “data rich”

Greatest Artifact of Human Civilization …

A Connected World

in the United States, it is reported that by 2018 there will be more than 490,000 data science positions available, but only 200,000 qualified people to fill the roles. The average age of a graduate class of data science students is 23 years old. With approximately only 110 universities offering data science studies, the growing market will continue to pressure the supply in the US.

Data Science

Nearly every field of discovery is transitioning from “data poor” to “data rich”

Greatest Artifact of Human Civilization …

A Connected World

in the United States, it is reported that by 2018 there will be more than 490,000 data science positions available, but only 200,000 qualified people to fill the roles. The average age of a graduate class of data science students is 23 years old. With approximately only 110 universities offering data science studies, the growing market will continue to pressure the supply in the US.

Data Science

Nearly every field of discovery is transitioning from “data poor” to “data rich”

Greatest Artifact of Human Civilization …

A Connected World

in the United States, it is reported that by 2018 there will be more than 490,000 data science positions available, but only 200,000 qualified people to fill the roles. The average age of a graduate class of data science students is 23 years old. With approximately only 110 universities offering data science studies, the growing market will continue to pressure the supply in the US.

Data Science

Nearly every field of discovery is transitioning from “data poor” to “data rich”

Greatest Artifact of Human Civilization …

A Connected World

in the United States, it is reported that by 2018 there will be more than 490,000 data science positions available, but only 200,000 qualified people to fill the roles. The average age of a graduate class of data science students is 23 years old. With approximately only 110 universities offering data science studies, the growing market will continue to pressure the supply in the US.
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

Pathways

```
c8
 c8 cs88
 c8 cs88 cs47a cs61b *** CS major
```

```
c8 cs61a
 cs61a
```

How does CS88 relate to CS61A ?

- **CS61A**
  - Intro Programming & Tools
  - CS Concepts and Techniques
  - Interpretation
- **CS Concepts in a Computational Approach**
  - Thinking w/ Data
  - Statistics
- **CS88**
  - Working w/ Data
  - **CS Concepts and Techniques**
  - Intro Programming & Tools

Course Structure

- 1 Lecture + 1 Lab/Discussion on Wednesday (!!!)
- Lecture introduces concepts (quickly)
- Lab provides concrete detail hands-on
- Homework (10) cements your understanding
  - Out Monday, Due Sunday
- Projects (3) put your understanding to work in building complete applications
  - Maps
  - Hangman
  - Open Projects!
- Readings: [http://composingprograms.com](http://composingprograms.com)
  - Same as cs61a

CS88 Team

- Prof. Gerald Friedland
  - fractor@eecs.berkeley.edu
- Meghna Dasgupta
  - Meghna.dasgupta@berkeley.edu
- Andrew Tan
  - Email: andrewtan@berkeley.edu
- James Stier
  - Email: jamesstier@berkeley.edu
- Jefferey Hame
  - Email: jefferey@berkeley.edu
CS88 Team - me

• Dr. Gerald Friedland (fractor@berkeley.edu)
  – 424 Sather Daj Hall (CITRIS)
  – http://www.gerald-friedland.org
  – Office hours: Mo 1-2 @ 424 SDH
  – Before/after class

• Adjunct Assistant Professor, EECS UC Berkeley
• Principal Data Scientist, Lawrence Livermore National Labs

Course Culture

• Learning
• Community
• Respect
• Collaboration
• Peer Instruction

Pro-student Grading Policies

• EPA
  – Rewards good behavior
  – Effort
    » E.g., Office hours, doing every single lab, hw, reading Piazza pages
  – Participation
    » E.g., Raising hand in lecture discussion, asking questions on Piazza
    » Altruism
      » E.g., helping other students in lab, answering questions on Piazza

• You have 2 “Slip Days”
  – You use them to extend due date, 1 slip day for 1 day extension
  – You can use them one at a time or all at once or in any combination
  – They follow you around when you pair up (you are counted individually)
    » E.g., A has 2, B has 0. Project is late by 1 day. A uses 1, B is 1 day late

Abstraction

• Detail removal
  – “The act or process 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”
**Experiment**

**Where are you from?**

Possible Answers:
- China
- 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 layer
  - And the abstraction layers are getting deeper by the day!

**Abstraction in CS: Data Type**

- What's this?

Real (or ideal) world

Computer representation

42
Data Types and Operations

- Set of elements
  - with some internal representation
    - e.g., Integers, Floats, Booleans, Strings, ...
- Set of operations on elements of the type
  - e.g., +, *, /, %, //, **
  - ==, <, >, <=, >=
- Properties
  - Commutative, Associative, ... , Closure (???)
- Expressions are valid well-defined sets of operations on elements that produce a value of a type

Questions

- What’s the difference between ‘==’ and ‘=’?

Lab and HW this week

- Lab will get you to where you have a program development environment
  - Even on your computer
- HW will give practice and explain subtleties of types, operators, and expressions
  - In a program development environment