Stats 5: Seminar in Data Science

Lecture 2, Winter 2017

Professor Padhraic Smyth
Departments of Computer Science and Statistics
University of California, Irvine
Class Organization

- Meet weekly for 50 minute seminar

- 8 guest speakers, weeks 2 through 9
  - You are encouraged to ask questions during the talks
  - You will be asked to submit answers to some brief questions for each talk

- Required to attend and submit reports for at least 7 of the 8 guest speakers to pass the class

- Intro and wrap-up talks in weeks 1 and 10

- Class Web site is at [www.ics.uci.edu/~smyth/courses/stats5](http://www.ics.uci.edu/~smyth/courses/stats5)
  - Slides and related materials will be posted during the quarter
## Schedule of Speakers

<table>
<thead>
<tr>
<th>Date</th>
<th>Speaker</th>
<th>Department</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jan 10</td>
<td>Padhraic Smyth</td>
<td>Computer Science</td>
<td>Introduction to Data Science</td>
</tr>
<tr>
<td>Jan 17</td>
<td>Sameer Singh</td>
<td>Computer Science</td>
<td>Machine Learning for Text</td>
</tr>
<tr>
<td>Jan 24</td>
<td>Charless Fowlkes</td>
<td>Computer Science</td>
<td>Computer Vision using Machine Learning</td>
</tr>
<tr>
<td>Jan 31</td>
<td>Pierre Baldi</td>
<td>Computer Science</td>
<td>Deep Learning and Neural Networks</td>
</tr>
<tr>
<td>Feb 7</td>
<td>Hernando Ombao</td>
<td>Statistics</td>
<td>Brain Signal Analysis: Statistics, Computing and Visualization</td>
</tr>
<tr>
<td>Feb 14</td>
<td>Zhoaxia Yu</td>
<td>Statistics</td>
<td>Multivariate Data Analysis</td>
</tr>
<tr>
<td>Feb 21</td>
<td>Ramesh Jain</td>
<td>Computer Science</td>
<td>Event and Web Data</td>
</tr>
<tr>
<td>Feb 28</td>
<td>James Randerson</td>
<td>Earth Systems Science</td>
<td>Data-Driven Climate Science</td>
</tr>
<tr>
<td>Mar 7</td>
<td>Matthew Harding</td>
<td>Economics</td>
<td>Big Data and Economics</td>
</tr>
<tr>
<td>Mar 14</td>
<td>Padhraic Smyth</td>
<td>Computer Science</td>
<td>The Future of Data Science</td>
</tr>
</tbody>
</table>
Weekly Reports

• For each speaker fill out a brief report about their lecture

• You are required to attend and submit reports for at least 7 out of 8 guest lectures to pass the class

• Reports will be available on the class Web site as text files

• Today’s lecture (Sameer Singh)
  – Deadline = 11pm tonight

• Remaining 7 guest lectures
  – Deadline = noon (after class)

• Submit to appropriate EEE Dropbox
Report Template

Stats 5: Seminar in Data Science: Winter 2017

Seminar Speaker: Professor Sameer Singh, Computer Science
Date: January 17th 2017

Your First Name: ________
Your Last Name: ________

By submitting this form you are verifying that you attended this lecture on the date above.

1. Describe in 1 or 2 sentences the key aspects of this speaker’s presentation:

.........

2. Mention 3 new ideas (one sentence each) that you heard about during this seminar.

.........
The UCI Data Science Major
Department of Statistics

Undergraduate Major in Data Science

The Data Science Major prepares students for a career in data analysis, combining foundational statistical concepts with computational principles from computer science. In the first two years of the program students will take core courses in both the Statistics and Computer Science Departments, providing a strong foundation in the principles of each field. In the 3rd and 4th years of the program, students will take more specialized courses, on topics such as design of algorithms, machine learning, information visualization, and Bayesian statistics. A major component of this degree is the final year capstone project course, a 2-quarter course that teaches students how to apply statistical and computational principles to solve large-scale real-world data analysis problems.

Admissions

Freshman Applicants: See the Undergraduate Admissions section.

Transfer Applicants: Junior-level applicants who satisfactorily complete course requirements will be given preference for admission. Applicants must satisfy the following requirements:

1. Completion of one year of college level mathematics (calculus or discrete math) and one semester of college level statistics.
2. Completion of one year of transferable Computer Science courses*, at least one of these should involve concepts such as those found in the Python and C++ programming languages, or another high-level programming language.
Is the Data Science Major a good match for you?

• Are you interested in computing?
  – Enjoy working with algorithms, programming, machine learning,…

• Do you have a good mathematics background?
  – Comfortable with mathematical ideas and concepts?
  – Interested in applying mathematical ideas to real-world problems?

• Enthusiastic about analyzing data?
  – Enjoy working with data? exploring, visualizing, modeling, understanding

• Seeking a career that has broad and flexible options?

If your answers are YES, the DS Major is for you 😊
What Classes do students take in the DS Major?

**Statistics**
- Stats 120 ABC: Intro to Prob and Stats
- Stats 68: Exploratory Data Analysis
- Stats 110-112: Statistical Methods
- CS 178: Machine Learning
  (Stats 140: Multivariate Statistics)

**Computing**
- ICS 46: Data Structures
- IFMTX 43: Intro to Software Engineering
- CS 122A: Intro to Data Management
- CS 161: Design and Analysis of Algorithms
  (CS 131: Parallel and Distributed Computing)
  (CS 172: Neural Networks/Deep Learning)

**Applications**
- Stats 170AB: Data Science Capstone Project
- INF 143: Information Visualization
  (INF 131: Human Computer Interaction)
  (CS 121: Information Retrieval)
  (CS 122B: Project in Databases/Web Applications)
  (Summer internships, e.g., junior year)

(Sample electives shown in parentheses)
Sample Course of Study in the Major

Years 1 and 2: foundational courses in computer science, mathematics, statistics, including statistical computing

### 2015-16, First Year: 41 units

<table>
<thead>
<tr>
<th>Fall</th>
<th>12</th>
<th>Winter</th>
<th>13</th>
<th>Spring</th>
<th>16</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICS 31</td>
<td>4</td>
<td>ICS 32</td>
<td>4</td>
<td>ICS 33</td>
<td>4</td>
</tr>
<tr>
<td>Math 2A</td>
<td>4</td>
<td>Math 2B</td>
<td>4</td>
<td>Math 2D</td>
<td>4</td>
</tr>
<tr>
<td>Writing 39A</td>
<td>4</td>
<td>Writing 39B</td>
<td>4</td>
<td>Stats 7</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Stats 5</td>
<td>1</td>
<td>Writing 39C</td>
<td>4</td>
</tr>
</tbody>
</table>

### 2016-17, Second Year: 46 units

<table>
<thead>
<tr>
<th>Fall</th>
<th>16</th>
<th>Winter</th>
<th>14</th>
<th>Spring</th>
<th>16</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICS 6B</td>
<td>4</td>
<td>ICS 45C</td>
<td>4</td>
<td>Stats 68</td>
<td>4</td>
</tr>
<tr>
<td>Math 3A</td>
<td>4</td>
<td>ICS 51</td>
<td>6</td>
<td>Stats 120C</td>
<td>4</td>
</tr>
<tr>
<td>Stats 120A</td>
<td>4</td>
<td>Stats 120B</td>
<td>4</td>
<td>ICS 46</td>
<td>4</td>
</tr>
<tr>
<td>GE III</td>
<td>4</td>
<td></td>
<td></td>
<td>ICS 6D</td>
<td>4</td>
</tr>
</tbody>
</table>
Years 3 and 4: more emphasis and specialization in data science topics such as machine learning, databases, visualization, advanced statistics

Year 3: sample program

<table>
<thead>
<tr>
<th>Fall</th>
<th>Winter</th>
<th>Spring</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stats 110, Statistical Methods for Data Analysis I</td>
<td>Stats 111, Statistical Methods for Data Analysis II</td>
<td>Stats 112, Statistical Methods for Data Analysis III</td>
</tr>
<tr>
<td>CS 161, Design and Analysis of Algorithms</td>
<td>CS 178, Machine Learning and Data-Mining</td>
<td>CS 122A, Introduction to Data Management</td>
</tr>
<tr>
<td>In4matx 43, Introduction to Software Engineering</td>
<td>ICS 139W, Critical Writing on Information Technology</td>
<td>In4matx 143, Information Visualization</td>
</tr>
<tr>
<td>GE IV/VIII,</td>
<td>GE III/VII,</td>
<td>GE VI,</td>
</tr>
</tbody>
</table>

Year 4: two-quarter capstone “data-intensive” project, + statistics and CS electives
Change of Major Requirements for Data Science Major

- Cumulative UC GPA: 2.7 or higher.
- 3.0 or higher average GPA and no grade lower than a C for ICS 31, ICS 32, and one of the following: Math 2A, Math 2B, Math 2D, ICS 6B, or ICS 6D.
- Students with more than 60 units will be reviewed on a case-by-case basis and may not be admitted to the major.
- Students will not be able to complete the degree in Data Science prior to Spring 2018.

- If you are a freshman, contact Professor Zhaoxia Yu at zhaoxia@ics.uci.edu to inquire about possibly getting a waiver to change into the major.
What can I do with a Data Science Major?

- Careers in “Data-Oriented” Companies and Organizations
  - Computing/internet companies: Google, Amazon, Facebook, IBM,….
  - Engineering companies: Intel, Samsung, Boeing, ….
  - Finance/insurance companies
  - Medical/pharmaceutical companies
  - Government/national labs: NASA, NIST, DoD, …. 
  - Many many more……

- Option to specialize with a Graduate Degrees (MS or PhD)
  - Computer Science: specialize in a topic such as machine learning, databases, etc
  - Statistics: specialize in a statistical topic, e.g., computational statistics
  - MS/PhD degrees lead to a wide variety of careers
Jobs for Data Scientists?

Percentage of job posts with the term “data scientist”, from indeed.com
Data Scientist jobs

My recent searches
Data Scientist - Los Angeles, CA
» clear searches

Sort by: relevance - date

Salary Estimate
$45,000+ (17246)
$60,000+ (13735)
$75,000+ (10312)
$90,000+ (6687)
$105,000+ (3972)

Company
Location
Job Type

what: Data Scientist
where: 
job title, keywords or company

Tip: Enter your city or zip code in the "where" box to show results in your area.

New! Join Indeed Prime - Get offers from great tech companies

Jobs 1 to 10 of 19,855

Senior Data Scientist/Applied Researcher
eBay Inc. ★★★★★ 597 reviews - San Jose, CA
Conceptualize, code, deploy, and iterate on designs from prototypes all the way through to production systems. Analyze petabytes of real-world performance data...
21 days ago - email
Sponsored

Data Scientist
Tremor Video - Boston, MA
Passion for “playing” with tons of data and supporting scientific experiments to improve and validate the performance of algorithms....
12 days ago - email
Sponsored

Machine Learning Algorithm Developer
Lucidyne Technologies, Inc. - Corvallis, OR
The position requires either a PhD degree, Masters degree or equivalent work experience in machine learning, with a focus on machine learning or a related field...
Easily apply
30+ days ago - email
Sponsored

Machine Learning Scientist/Architect - Polygraph required
Resolute Technologies, LLC - Hanover, MD
$190,000 a year
Develop on-sensor machine learning analytics to support multiple operational scenarios. An understanding of SIGINT processing systems, data flows, data formats,...
Easily apply
10 hours ago - save job - email - more...
Data Scientist
6sense - San Francisco, CA
Our data scientists are not optimizing software; We work with big data at scale, advanced machine learning and predictive modeling to find buyers and predict...

Easily apply
3 days ago - save job - email - more...

GBS Entry Level Data Scientist Analytics Co-Op
IBM 4.5 stars - 6,979 reviews - Columbus, OH 43228 (Westland area)
Growth Play Analytics (BigData). You must reside within a reasonable commuting distance - generally 50 miles or less from Columbus, Ohio OR be available to work...
4 hours ago - save job - email - more...

Data Scientist
General Motors 4.2 stars - 1,276 reviews - Austin, TX
Expert level knowledge, development expertise & commiserate experience with Python, Matlab, R, Java and SQL. Data Engineering Team....
11 days ago - email
Sponsored

Data Scientist
National Security Agency 4.5 stars - 63 reviews - Fort George G Meade, MD
$64,923 - $83,774 a year
Applied statistics, calculus, quantitative or statistical methods and techniques, data mining, informatics, data science, programming, computational algorithms,....
7 days ago - email
Sponsored
10 hours ago - save job - email - more...
Do I need a Data Science degree to do Data Science?

• Technically no......many people currently are “data scientists” with backgrounds in quantitative degrees that are not data science
  – Some with statistics, some with computer science, some with a combination
  – Some with other quantitative degrees

• Advantages of the DS major
  – Puts you on the “fast track” to becoming a Data Scientist
  – Ensures that you will know the fundamentals of both
    • Computing
    • Statistics
  – Provides you with skills that are likely to have lasting value (as technology changes)
What are other degree options?

• Computer Science with a Statistics minor?
  – More classes in “systems” aspects of computer science
  – Fewer classes in statistics
  – No capstone data science project class

• Another degree like Math or Economics with a Statistics minor?
  – Far fewer classes in computer science
  – Fewer classes in statistics
  – No capstone data science project class

• Statistics undergraduate degree (e.g., at another UC)?
  – More classes in mathematics and statistics
  – Far fewer classes in computer science
  – No capstone data science project class
## Schedule of Speakers (from class Website)

<table>
<thead>
<tr>
<th>Date</th>
<th>Speaker</th>
<th>Department</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jan 10</td>
<td>Padhraic Smyth</td>
<td>Computer Science</td>
<td>Introduction to Data Science</td>
</tr>
<tr>
<td>Jan 17</td>
<td>Sameer Singh</td>
<td>Computer Science</td>
<td>Machine Learning for Text</td>
</tr>
<tr>
<td>Jan 24</td>
<td>Charless Fowlkes</td>
<td>Computer Science</td>
<td>Computer Vision using Machine Learning</td>
</tr>
<tr>
<td>Jan 31</td>
<td>Pierre Baldi</td>
<td>Computer Science</td>
<td>Deep Learning and Neural Networks</td>
</tr>
<tr>
<td>Feb 7</td>
<td>Hernando Ombao</td>
<td>Statistics</td>
<td>Brain Signal Analysis: Statistics, Computing and Visualization</td>
</tr>
<tr>
<td>Feb 14</td>
<td>Zhoaxia Yu</td>
<td>Statistics</td>
<td>Multivariate Data Analysis</td>
</tr>
<tr>
<td>Feb 21</td>
<td>Ramesh Jain</td>
<td>Computer Science</td>
<td>Event and Web Data</td>
</tr>
<tr>
<td>Feb 28</td>
<td>James Randerson</td>
<td>Earth Systems Science</td>
<td>Data-Driven Climate Science</td>
</tr>
<tr>
<td>Mar 7</td>
<td>Matthew Harding</td>
<td>Economics</td>
<td>Big Data and Economics</td>
</tr>
<tr>
<td>Mar 14</td>
<td>Padhraic Smyth</td>
<td>Computer Science</td>
<td>The Future of Data Science</td>
</tr>
</tbody>
</table>