**BST 261: Data Science II**

**Spring 2018**

**9:45am -11:15am Mondays & Wednesdays**

**Kresge 200**

**Instructor**

Heather Mattie

Instructor of Data Science

Harvard T.H. Chan School of Public Health

Building 1 Room 421A, 655 Huntington Ave, Boston MA 02115

hemattie@hsph.harvard.edu

(617) 432-5308

Office hour: TBD or by appointment, Building 1 Room 421A

**Teaching Assistant**

Stephanie Chan

stephaniechan@fas.harvard.edu

Office hour: TBD

Fill out this form to report preferred office hour days and times: <https://docs.google.com/forms/d/e/1FAIpQLSdRiqeLyEBIg7PQLdq8dOlnf3xvim1hulOmFS6NMNJxTM-EZg/viewform?usp=sf_link>

**Course Description**

This course is an introduction to deep learning, a branch of machine learning concerned with the construction, development, and application of neural networks. Deep learning algorithms extract layered high-level representations of data in a way that maximizes performance on a given task. We will cover a range of topics including basic neural networks, convolutional networks, and recurrent networks, and applications to problem domains like computer vision and speech recognition.

**Course Structure**

You will learn these concepts through data analysis homework assignments and a group project. Grades will be based on:

* 2 homework assignments (70%)
* 1 final group project (30%)

All work must be submitted in the form of a Jupyter notebook on Canvas. Students are able to work on their assignments through the JupyterHub installation on the course Canvas site. Deadlines are given in the course schedule at the end of the syllabus. Groups of no more than 4 students are allowed for the group project. All groups must use a data set of their choice that is not available (preloaded) in Keras. Each group will also give a short (<10 mins) presentation on the last day of class. More project details will be posted on Canvas and the course GitHub by the fourth week of the course.

**Late Day Policy**

Each student is given six late days for homework at the beginning of the course. A late day extends the individual homework deadline by 24 hours without penalty. No more than two late days may be used on any one assignment. Assignments handed in more than 48 hours after the original deadline will not be graded. We do not accept any homework more than 48 hours after the original deadline. Late days are intended to give you flexibility: you can use them for any reason, no questions asked. You don’t get any bonus points for not using your late days. Also, you can only use late days for the individual homework deadlines. All other deadlines (e.g., project milestones) are hard.

Although each student is only given a total of 6 late days, we will be accepting homework from students that pass this limit. However, we will be deducting 2 points for each extra late day. For example, if you have already used all of your late days for the course, we will deduct 2 points for assignments <24 hours late, and 4 points for assignments 24-48 hours late.

**Prerequisites**

Must have linear algebra, multivariable calculus, and statistics knowledge, be proficient in Python, and taken BST 260 or equivalent.

**Textbooks**

Most of the course content will come from the following books. Specific chapters and sections will be suggested for reading, but will not be required.

1. Deep Learning, Goodfellow and Bengio, 2016

Freely available online at <http://www.deeplearningbook.org/>

1. Deep Learning with Python, Chollet, 2017

The first few chapters are available at <https://www.manning.com/books/deep-learning-with-python>

**Course Website and Communication**

**Canvas:** The Canvas site is an important learning tool for this course where students will access course materials, **submit course assignments** and share other resources with the class. Course announcements will be posted on the site and students will be required to check the course website on a weekly basis.

**GitHub:** All Jupyter notebooks used for this course (in-class examples, homework assignments, group project) are available on the course GitHub repository here: <https://github.com/hmattie/BST261Spring2018>

**Note that assignments will not be submitted to this repository, but in Canvas**.

**Slack:** Students may utilize the discussion boards in Canvas, or ask questions on the course Slack workspace: <https://join.slack.com/t/bst261fall2018/signup>

**Harvard Chan Policies and Expectations**

**Inclusivity Statement**

Diversity and inclusiveness are fundamental to public health education and practice. It is a requirement that you have an open mind and respect differences of all kinds. I share responsibility with you for creating a learning climate that is hospitable to all perspectives and cultures; please contact me if you have any concerns or suggestions.

**Academic Integrity**

Each student in this course is expected to abide by the Harvard University and the Harvard. T.H. Chan School of Public Health Codes of Academic Integrity. All work submitted to meet course requirements is expected to be a student’s own work. In the preparation of work submitted to meet course requirements, students should always take great care to distinguish their own ideas and knowledge from information derived from sources.

Students must assume that collaboration in the completion of assignments is prohibited unless explicitly specified. Students must acknowledge any collaboration and its extent in all submitted work. This requirement applies to collaboration on editing as well as collaboration on substance.

Should academic misconduct occur, the student(s) may be subject to disciplinary action as outlined in the Student Handbook. See the Student Handbook for additional policies related to academic integrity and disciplinary actions.

**Accommodations for Students with Disabilities**

Harvard University provides academic accommodations to students with disabilities. Any requests for academic accommodations should ideally be made before the first week of the semester, except for unusual circumstances, so arrangements can be made. Students must register with the Local Disability Coordinator in the Office for Student Affairs to verify their eligibility for appropriate accommodations. Contact the OSA [**studentaffairs@hsph.harvard.edu**](mailto:studentaffairs@hsph.harvard.edu) in all cases, including temporary disabilities.

**Course Evaluations**

Constructive feedback from students is a valuable resource for improving teaching. The feedback should be specific, focused and respectful. It should also address aspects of the course and teaching that are positive as well as those which need improvement.

Completion of the evaluation is a requirement for each course. Your grade will not be available until you submit the evaluation. In addition, registration for future terms will be blocked until you have completed evaluations for courses in prior terms.

**Course Schedule Outline**

**This is a TENTATIVE schedule for the course.**

|  |  |  |
| --- | --- | --- |
| **Date** | **Topics** | **Milestones** |
| March 19 | Introduction to course  Review of linear algebra |  |
| March 21 | Review of Python  Linear algebra operations in Python  Review of machine learning |  |
| March 26 | A first look at a neural network  Introduction to deep learning  Deep feedforward networks | Homework #1 Assigned |
| March 28 | Deep feedforward networks continued |  |
| April 2 | Deep feedforward networks continued |  |
| April 4 | A universal deep learning workflow |  |
| April 9 | Convolutional networks | Homework #1 Due  Homework #2 Assigned |
| April 11 | Convolutional networks continued |  |
| April 16 | Convolutional networks continued | Group project proposal due |
| April 18 | Recurrent and recursive networks |  |
| April 23 | Recurrent and recursive networks continued |  |
| April 25 | Recurrent and recursive networks continued |  |
| April 30 | Autoencoders | Homework #2 Due |
| May 2 | Boltzmann machines |  |
| May 7 | Deep Belief Networks |  |
| May 9 | Group projects | Group project presentation |