

**Introduction to Data Science   
553.436, 553.636**

**Spring, 2022 (4 credits, EQ)**

**Instructor**

Professor Tamas Budavari, <https://engineering.jhu.edu/ams/faculty/tamas-budavari/>

Email: [budavari@jhu.edu](mailto:budavari@jhu.edu)

Office: Wyman N437

Phone: (410) 516-7914

Office Hours: T 12:15pm–1:00pm, via Zoom

Zoom Link: <https://wse.zoom.us/my/budavari>

**Head Teaching Assistant**

Please direct any concerns about course logistics to the head teaching assistant, preferably via Teams:

Benjamin Weinberg

Email: [bweinbe5@jhu.edu](mailto:bweinbe5@jhu.edu)

Zoom meetings available by appointment.

**Lectures**

Times: MW 12:00pm–1:15pm

Zoom Link:

<https://wse.zoom.us/j/99830919265?pwd=QlJsek5qSStVWmI3bFBYZkM0TDVZQT09>

Zoom Meeting ID: 998 3091 9265

Passcode: 436636

**Sections**

|  |  |  |  |
| --- | --- | --- | --- |
| **Section** | **Time** | **Room** | **TA** |
| EN.553.436(01) | F 9:00am-9:50am | Hodson 216 | Jingyi Wu |
| EN.553.436(02) | F 12:00pm-12:50pm | Hodson 203 | Heena Saqib |
| EN.553.436(03) | F 10:00am-10:50am | Hodson 301 | Mihir Chakravarthi |
| EN.553.636(01) | F 10:00am-10:50am | Hodson 216 | Xuanzhao Dong |
| EN.553.636(02) | F 11:00am-11:50am | Hodson 301 | Qianmin Zhou |
| EN.553.636(03) | F 1:30pm-2:20pm | Hodson 305 | Mona Yang |
| EN.553.636(04) | F 3:00pm-3:50pm | Hodson 203 | Xiaotong Hu |
| EN.553.636(05) | F 9:00am-9:50am | Croft G02 | Mengmeng Liu |

**Teaching Assistant Office Hours**

The office hours of any TA are open to all students regardless of section.

|  |  |  |  |
| --- | --- | --- | --- |
| **TA** | **Email** | **Office Hours** | **Zoom Link** |
| Jingyi Wu | jwu192@jhu.edu | Th 9:00am–11:00am | [https://JHUBlueJays.zoom.us/j/3089298841](https://jhubluejays.zoom.us/j/3089298841) |
| Heena Saqib | hsaqibo1@jhu.edu | W 5:00pm–7:00 pm | [https://JHUBlueJays.zoom.us/j/92446465088](https://jhubluejays.zoom.us/j/92446465088) |
| Mihir Chakravarthi | mchakra9@jhu.edu | T 8:00am–9:00am, 5:00pm–6:00pm | https://JHUBlueJays.zoom.us/my/mihir2000 |
| Xuanzhao Dong | xdong26@jhu.edu | M 3:00pm–5:00pm | [https://JHUBlueJays.zoom.us/j/9258837295](https://jhubluejays.zoom.us/j/9258837295) |
| Qianmin Zhou | qzhou28@jhu.edu | W 2:00pm–4:00 pm | <https://JHUBlueJays.zoom.us/j/4742461121> |
| Mona Yang | [myang83@jh.edu](mailto:myang83@jh.edu) | M 9:00am–11:00 am | [https://JHUBlueJays.zoom.us/j/9851966033](https://jhubluejays.zoom.us/j/9851966033) |
| Xiaotong Hu | [xhu66@jhu.edu](mailto:xhu66@jhu.edu) | W 10:00am–12:00pm | [https://JHUBlueJays.zoom.us/j/5661659369](https://jhubluejays.zoom.us/j/5661659369) |
| Mengmeng Liu | mliu89@jhu.edu | W 9:00am–11:00 am | [https://JHUBlueJays.zoom.us/j/7646614097](https://jhubluejays.zoom.us/j/7646614097) |

**Course Description**

Today the term Data Science is widely used covering a broad range of topics from mathematics and algorithms to actual data analysis and machine learning techniques. This course provides a thorough survey of relevant methods balancing the theory and the application aspects. We will cover several supervised methods for regression and classification, as well as unsupervised methods for clustering and dimensional reduction. In chronological order, the topics will include generalized linear regression, principal component analysis, nearest neighbor and Bayesian classifiers, support vector machines, logistic regression, decision trees, random forests, K-means clustering, Gaussian mixtures and Laplacian eigenmaps. The course uses Python and Jupyter Notebook and includes visualization techniques throughout the semester. Time permitting, an introduction to the Structured Query Language (SQL) is provided toward the end of the semester.

**Recommended Course Background**

EN.553.413 Applied Statistics and Data Analysis

**Topic Outline**

* descriptive statistics
* probability, probability density functions, moments
* sampling from distributions, density estimation
* linear methods for regression and classifications
* principal component analysis
* Bayesian inference, numerical methods
* model selection and averaging
* support vector machines
* neural networks
* cluster analysis, k-means clustering
* spectral methods
* robust statistics
* databases

**Course Technology**

Course communication will be conducted through Microsoft Teams. Lecture notes, lecture recordings, and course assignments will be made available on the course Team. Questions or comments on course content may be communicated in the General channel of the Team. There is a course Blackboard but it is used only by staff for roster management. We do not expect to upload any material to Blackboard. For assignment submission and grading, we will use Gradescope.

**Readings**

The lecture notes are the basis of the course. Useful background for the lecture notes can be found in:

* T. Hastie, R. Tibshirani, and J. Friedman (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer.
* Zeljko Ivezic, Andrew J. Connolly, Jacob T VanderPlas and Alexander Gray (2014). *Statistics, Data Mining and Machine Learning in Astronomy*, Princeton University Press.

**Grading**

Grades will be based on two to four homework assignments (30%), two examinations (50%), and a group project presentation (20%). Examinations are to be completed within a fixed time window on the day of their distribution, with start times at the discretion of individual students.

**Key Dates**

Exam 1: February 23

Exam 2: April 13

Project Presentation: Finals Week

**Ethics**

The strength of the university depends on academic and personal integrity. In this course, you must be honest and truthful. Ethical violations include cheating on exams, plagiarism, reuse of assignments, improper use of the Internet and electronic devices, unauthorized collaboration, alteration of graded assignments, forgery and falsification, lying, facilitating academic dishonesty, and unfair competition. Report any violations you witness to the instructor. You can find more information about university misconduct policies on the web at these sites:

* For undergraduates: <http://e-catalog.jhu.edu/undergrad-students/student-life-policies/>
* For graduate students: <http://e-catalog.jhu.edu/grad-students/graduate-specific-policies/>

**Students with Disabilities**

Any student with a disability who may need accommodations in this class must obtain an accommodation letter from Student Disability Services, 385 Garland, (410) 516-4720, [studentdisabilityservices@jhu.edu](mailto:studentdisabilityservices@jhu.edu) .