

**Syllabus**

**School of Professional Studies**

**DATA 622 Introduction to Machine Learning**

**Instructor Name:** Sabrina Khan

**Instructor Email Address** [sabrina.khan@sps.cuny.edu](mailto:sabrina.khan@sps.cuny.edu)

[sabrina.khan@live.com](mailto:sabrina.khan@live.com)

**Degree Program:** M.S. in Data Science

**Credits:** 3 graduate credits

**Prerequisites:** 605, 606

**Type of Course:** Online. Asynchronous

**Meetup:** Weekly, on Thursdays at 6:00pm EST, on Zoom, beginning Sept 2. Attendance is optional.

**Office hours:** Wednesdays 12pm -1pm EST

Feel free to reach out to get in touch via email and I will accommodate your schedule. My cell number is 929-245-0383, which may be easier to use if you want to text.

**Course Description:**

This course aims to develop basic understanding of foundational machine learning techniques. This course will aim to have a smooth transition from statistical modeling methods to machine learning methods. In this course, we emphasize on computational techniques. We will primarily use R as our programming environment. All assignments are expected to be to submitted in as R-Markdown documents. However, if you are more comfortable using Python, please let me know and we can make some adjustments.

**Learning Outcomes:**

1. Develop deep understanding of introductory machine learning algorithms
2. Prepare datasets for machine learning algorithms and conduct modeling exercises on given datasets.
3. Develop understanding to be able to identify problems that require supervised or unsupervised methods.
4. Develop proficiency of technical terminology expected of a Data Science practitioner.

**Assignments and Grading:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Assignments** | **Possible score** | **Total points** | **% of the final grade** |
| Homework#1 | 50 | 50 | 5% |
| Homework#2 | 150 | 150 | 15% |
| Homework #3 and #4 assignments  (must done in group except for the MCQ section- if any) | 200 each | 100 x 2 = 400 | 20% each; 40% in total |
| 1 project  (must be done in group) | 250 | 250 | 25% |
| Weekly discussion:  Each week, you will enter your submission on the discussion prompt AND will respond to two other entries by your peers | 10 each | 10 x 15 = 150 | 1% each; 15% in total |
|  | Total | 1000 | 100% |

**Letter Grade distribution:**

Please see: <https://sps.cuny.edu/about/policies/academic-and-student-policies/grading-policies-graduate>

**Tentative Schedule:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Week** | **Week of** | **Topics** | **Key Task** | **Due on (mostly on Fridays, 1159 pm EST)** |
| 1 | 25-Aug | Introduction to 622, Intro to Machine Learning | Discussion 1 | 29-Aug |
| 2 | 30-Aug | Review week: Linear & Logistic Regression | Discussion 2 | 3-Sep |
| 3 | 6-Sep | Classification: Discriminant Analysis (LDA, QDA) | Discussion 3 Homework # 1 due | 10-Sep |
| 4 | 13-Sep | Classification: kNN, Naïve Bayes | Discussion 4 | 17-Sep |
| 5 | 20-Sep | Tree based methods: Decision Trees | Discussion 5  Homework #2 due | 24-Sep |
| 6 | 27-Sep | Tree based methods: Bagging, Random Forests, Boosting | Discussion 6 | 1-Oct |
| 7 | 4-Oct | Tree based methods: Bagging, Random Forests, Boosting (continued) | Discussion 7 | 8-Oct |
| 8 | 11-Oct | Support Vector Machines | Discussion 8  Homework #3 due | 15-Oct |
| 9 | 18-Oct | Support Vector Machines | Discussion 9 | 22-Oct |
| 10 | 25-Oct | Unsupervised Learning: Clustering | Discussion 10 | 29-Oct |
| 11 | 1-Nov | Unsupervised Learning: PCA | Discussion 11 | 5-Nov |
| 12 | 8-Nov | Bias-Variance Tradeoff | Discussion 12  Homework #4 due | 12-Nov |
| 13 | 15-Nov | Resampling Methods & Model Selection | Discussion 13 | 19-Nov |
| 14 | 22-Nov | Thanksgiving Break | - | -- |
| 15 | 29-Nov | Introduction to neural network | Discussion 14 | 3-Dec |
| 16 | 6-Dec | AI Ethics/Foundation Models | Discussion 15  **Final Project due** | 10-Dec |

**Required Texts and Materials:**

TITLE: An Introduction to Statistical Learning

AUTHORS: Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani

URL: <https://www.statlearning.com/> (booksite, pdf,data,R etc)

<https://www.ime.unicamp.br/~dias/Intoduction%20to%20Statistical%20Learning.pdf> (Text we are following)

<https://www.dataschool.io/15-hours-of-expert-machine-learning-videos/> (Lecture videos and slides)

TITLE: Elements of Statistical Learning

AUTHORS: Jerome Friedman, Trevor Hastie, Robert Tibshirani

URL: <https://web.stanford.edu/~hastie/ElemStatLearn//printings/ESLII_print10.pdf>

(Text we are following)

**Optional**

TITLE: Learning from Data

AUTHORS: Yaser S. Abu-Mostafa, Malik Magdon-Ismail, Hsuan-Tien Lin

URL: <https://www.amazon.com/gp/product/1600490069>

(Doesn’t seem like there is a (legal) free copy available).

More resources will be shared throughout the semester.

**LATE SUBMISSIONS:**

Unless you have made an arrangement with the instructor ahead of time, late submissions are not acceptable.

**ACCESSIBILITY AND ACCOMMODATIONS**

The CUNY School of Professional Studies is firmly committed to making higher education accessible to students with disabilities by removing architectural barriers and providing programs and support services necessary for them to benefit from the instruction and resources of the University. Early planning is essential for many of the

resources and accommodations provided. Please see: <http://sps.cuny.edu/student_services/disabilityservices.html>

**ONLINE ETIQUETTE AND ANTI-HARASSMENT POLICY**

The University strictly prohibits the use of University online resources or facilities, including Blackboard, for the purpose of harassment of any individual or for the posting of any material that is scandalous, libelous, offensive or otherwise against the University’s policies. Please see: <http://media.sps.cuny.edu/filestore/8/4/9_d018dae29d76f89/849_3c7d075b32c268e.pdf>

**ACADEMIC INTEGRITY**

Academic dishonesty is unacceptable and will not be tolerated. Cheating, forgery, plagiarism and collusion in dishonest acts undermine the educational mission of the City University of New York and the students' personal and intellectual growth. Please see:

<http://media.sps.cuny.edu/filestore/8/3/9_dea303d5822ab91/839_1753cee9c9d90e9.pdf>

**STUDENT SUPPORT SERVICES**

If you need any additional help, please visit Student Support Services:

<http://sps.cuny.edu/student_resources/>