Carnegie Mellon University

17-634: Applied Machine Learning

Class Time: Monday and Wednesday, 11:50-1:10
Location: 3SC 265
Semester: A3, Spring 2022, 6 units

Instructor(s): Prof. Travis Breaux
Office Hours: Online, Mon, 1:30-2:30 pm, or by appointment
Email: tdbreaux@andrew.cmu.edu

Autonomous and intelligent systems increasingly rely on automated decision making based on statistical models used for classification or prediction. The practical application of machine learning requires understanding the underlying theoretical assumptions behind a wide variety of statistical models, how to analyze the performance of such models, and how to integrate models into data processing pipelines. This course introduces students through a broad survey of supervised and unsupervised machine learning methods, including fundamentals in the setup and evaluation of linear and non-linear statistical models.

Learning Objectives. After completing this course, you will be able to:

- Design experiments with the most common statistical learning models
- Implement common models using contemporary tools and frameworks
- Recognize key terms in the discussion of statistical learning

Assessments. Students learn more by applying and explaining ideas to others, thus, the course requires the following activities:

- **Homework assignments**, including questions to help you focus on important points in the readings and to exercise particular skills
- **Class participation**, to enrich the discussion with your insight, relevant experience, critical questions, and analysis of the material. The quality of contribution is more important than the quantity.
- **Final Exam**, to demonstrate your cumulative knowledge on practical examples.

Assessment	Final Grade %
Homework assignments	70%
Final exam	20%
Class participation	10%

Grade	Percentage Interval
A	90-100%
C	70-79%
D	60-69%
R (F)	59% or below

Recommended Textbooks

Machine learning rests at the intersection of computer science and statistics, and thus students should be familiar with a variety of notations to describe models and their optimization. Readings have been selected to emphasize simplicity of presentation, and currency in topics. **All readings are linked from the course schedule, which is in the Modules section.**

- A general textbook that provides broad coverage across all models discussed in class, and exceptional mathematical depth is The Elements of Statistical Learning by Hastie, Tibshirani, and Friedman, 2nd edition, 2016.
- Select research papers, which are linked in the course schedule.

Course and Grading Policies

- **Late-work policy**: All work is expected to be handed in at the indicated due date and time. For fairness to the whole class, no late submissions will be accepted for the group work. In the first week of classes, you should receive your course schedules; please use those to plan ahead.

  Each student is allowed one late submission for the individual homework assignments. You should immediately notify the course TA(s) before the submission deadline that you will submit late. Late work must be submitted as soon as circumstances allow, ordinarily within 24 hours of the due date. If you have any questions you should raise them immediately rather than waiting for conflicts to arise.

- **Participation policy**. Class participation will be graded by in-class engagement, including asking relevant questions based on a critical review of required readings, on lectures, and on comments made by your peers. The lack of attendance, and the use of mobile devices, including phones and laptops, which regular engagement will count against your participation grade.

Learning Disabilities. If you have a documented learning disability, please notify the instructor during the first week of class.

Academic Integrity. Honesty and transparency are important to good scholarship. Plagiarism and cheating, however, are serious academic offenses with serious consequences. If you are discovered engaging in either behavior in this course, you will earn a failing grade on the assignment in question, and further disciplinary action may be taken. For a clear description of what counts as plagiarism, cheating, and/or the use of unauthorized sources, please see the University's Policy on Academic Integrity.

If you have any questions regarding plagiarism or cheating, please ask the instructor as soon as possible to avoid any misunderstandings.

Student Wellness. As a student, you may experience a range of challenges that can interfere with learning, such as strained relationships, increased anxiety, substance use, feeling down, difficulty concentrating and/or lack of motivation. These mental health concerns or stressful events may diminish your academic performance and/or reduce your ability to participate in daily activities. CMU services are available, and treatment does work. You can learn more about confidential mental health services available on campus at: http://www.cmu.edu/counseling/. Support is always available (24/7) from Counseling and Psychological Services: 412-268-2922

Course Schedule

The following schedule provides a general overview of topics and readings. Please refer to the syllabus online in Canvas for specific lecture topics, reading assignments and due dates.

Class	Date	Topic and Lectures
1	1/19	Data Pipelines and Learning Frameworks **Before Class**: Getting Starting, Install Anaconda
2	1/24	Random Variables, Bayes Rule, MLE and MAP **Readings**: Ch2, Estimating Probabilities by Mitchell, 2017.
3	1/26	Generative and Discriminative Classifiers: Naive Bayes and Logistic Regression **Readings**: Ch3, Generative and Discriminative Classifiers by Mitchell, 2017.
4	1/31	Evaluating Statistical Models **Readings**: None
5	2/2	Data Preparation, Imputation and Vectorization **Readings**: Assumptions and Conventional Methods, in Missing Data by Allen, 2022
6	2/7	Bias and Variance: Linear Regression and Nearest Neighbor **Optional Readings**: 7.1 Bias, Variance and Model Complexity through 7.3 Example: Bias-Variance Tradeoff in Hastie et al., 2016
7	2/9	Error Analysis and Fairness
8	2/14	Bagging, Boosting and Ensemble Methods **Optional Readings:** 8.7 Bagging, and 10.1 Boosting in Hastie et al., 2016.
9	2/16	Kernel Methods and Support Vector Machines **Optional Readings**: 12.2 and 12.3 Support Vector Classifiers in Hastie et al., 2016

10	2/21	Machine Learning in Practice. Guest Lecturer Isaac Faber **Readings**: Invaluable Data Science Lessons to Learn from the Failure of Zillow's Flipping Business.
11	2/23	Clustering and Topic Modeling **Readings**: Latent Dirichlet Allocation by Bei et al., 2002. Optional Readings: Automatic Evaluation of Topic Coherence by Newman et al., 2010; Exploring the Space of Topic Coherence Measures by Roder et al., 2015.
12	2/28	Topic Interpretation **Readings**: Reading the Tea Leaves by Change et al., 2009.
13	3/2	Corpora Development and Crowdsourcing