Department of Civil and Environmental Engineering
University of Massachusetts Amherst

CEE 590STA: Machine Learning: Foundations and Applications

Fall 2023 Course Syllabus

Contents

1 Personnel and Logistics

1.1 Instructor

Name: Jimi Oke[1]
Email: jboke@umass.edu[2]
Office: 214D Marston Hall

1.2 Venue and Class Time

Days & Times:
Lectures: M/W: 2:30pm–3:45pm; Marston 211
Office Hours: Tu: 3:00pm–4:00pm; Th: 11:30am–12:30pm; Marston 214D/Zoom[3]

2 Course Information

2.1 Description

This course counts for 3 credits. It introduces the theory and applications of core concepts in data mining and machine learning from an engineering perspective. Key topics include:

- fundamentals of data analysis and regression
- classification (support vector machines, decision trees)
- linear model selection and regularization
- nonlinear regression methods
- tree-based/ensemble methods: bagging, boosting, random forests
- support vector machines
- unsupervised learning (dimensionality reduction, clustering)
- neural networks
- time series analysis

Applications to various subdisciplines will be highlighted, especially in transportation, environmental, structural and industrial engineering. Hands-on programming in Python (R will also be supported) throughout the course will enable students to analyze and train models on real-world datasets. Through this course, students will understand the potential of machine learning in civil, environmental and industrial engineering, among other disciplines, as well as learn to create and train models from data to solve challenging problems.

[1] Approximate pronunciation of last name in IPA: /ɔˈkɛ or "aw-KEH"
[2] Please allow up to 48 hours for a response to your email. Be sure to put "CEE590STA" in the subject to ensure a prompt response.
[3] `https://umass-amherst.zoom.us/j/97617298245`)

2.2 Objectives

- Understand the importance of data analytics and machine learning in engineering
- Create and train models for various problem structures and applications
- Understand the theory behind fundamental algorithms and apply them

2.3 Outcomes

In this course, you will:

- Understand the fundamental concepts of model fitting and linear regression methods
- Recognize when to use classification approaches instead of regression
- Learn how to address uncertainty in modeling via cross-validation and bootstrapping
- Understand core methods for model selection and regularization
- Learn nonlinear regression approaches and recognize when to apply them over linear methods
- Develop a working knowledge of ensemble learning methods via decision trees, bagging, random forests and boosting
- Understand and apply support vector machines
- Survey unsupervised learning methods with a focus on principal components analysis, k-means clustering and hierarchical agglomerative clustering
- Gain a basic understanding of neural networks and their applications to regression, classification, pattern recognition, etc
- Use big data and machine learning concepts to solve real-world engineering problems
- Learn to use Python or similar programming language (e.g. R) to perform data analytics and machine learning tasks

2.4 Texts

The primary text for this course is:

- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2017). *An Introduction to Statistical Learning with Applications in Python/R*. Springer, New York. (The text, among other resources is freely available from the author at `https://www.statlearning.com/`. Abbreviated as **ISL** in the schedule and lecture notes.)

Supplementary texts:

- Hastie, T., Tibshirani, R., & Friedman, J. (2017). *The Elements of Statistical Learning: Data Mining, Inference and Prediction*. Springer, New York, NY. Second Edition. (This text, among other resources, is also freely available from the authors at `https://web.stanford.edu/~hastie/ElemStatLearn/`. Abbreviated as **ESL** in the schedule and lecture notes.)

- Goodfellow, I., Bengio, Y. & Courville, A. (2016). *Deep Learning*, MIT Press. (This text is freely available at https://www.deeplearningbook.org/. Abbreviated as **DL** in the schedule and lecture notes.)
- Hyndman, R. J. & Athanasopoulos, G. (2018). *Forecasting: Principles and Practice.* OTexts. (This text is freely available at https://otexts.com/fpp2/. Abbreviated as **FPP** in the schedule and lecture notes.)

Any other recommended or required reading will be provided on Moodle.

2.5 Prerequisites

CEE 244, CEE 260/MIE 273 and MATH 233; or permission of the instructor based on prior knowledge of undergraduate-level probability, statistics, multivariate calculus and programming (in any language). Python/R will be taught in class. Matlab may also be supported.

3 Policies and Values

I will use slides in the classroom, and annotate them electronically when possible. These slides will be available to you prior to the lecture. I will endeavor to foster an equitable and inclusive learning environment that will spark your curiosity and challenge you learn actively. I strongly urge you to come to class prepared, having done the reading, ready to reflect on your homework or problem set and engage with new material. I will ask frequent questions of you, and will also expect you to ask as many questions as possible. Further specifics on class policies and values are as follows.

3.1 Assessments and grading

There will be no grading on a curve. Consistent with this, after drop date, students who remain in this class are not in jeopardy of seeing their grades change due to the change in class composition. The breakdown is provided in Table 1.

Table 1 Course components and grade breakdown

Assessment	**Value (%)**
Problem Sets (5)	50
Midterm Exam 1	15
Midterm Exam 2	15
Project	20

Individual grades will be based on the scale shown below:

Grade	Range (%)
A	93-100%
A−	90–92
B+	87–89
B	83–86
B−	80-82
C+	77–79
C	73-76
C−	70-72*
D	60-69*
F	$\leq$ 59

Note: Graduate students cannot earn grades of C−, D or D+, so scores lower than 73% are Failing for Graduate students.

3.2 Problem sets

Five problem sets will be assigned. Submission will be online (PDFs and other supporting code; or Jupyter notebooks) via Moodle. Each will be worth 10% of your total grade. *Late problem sets will automatically attract a 25% penalty and will not be accepted more than 4 days beyond the due date* (excepting prior permission).

3.3 Midterms

There will be 2 take-home midterms, which will be open-resource. Previous exams may be available for practice.

3.4 Programming

Some lectures will incorporate engineering applications of machine learning concepts using Python. Problem sets will also involve some coding in Python. I recommend installing JupyterLab. You are welcome to use other languages/platforms such as R/RStudio or Matlab for your assignments. However, I cannot guarantee the same level of support for Matlab in particular.

3.5 Computing resource

Having a laptop is not a requirement for this course. However, if you own one and are able to bring it to the classroom, it may improve your learning experience during the programming segments of the lecture.

3.6 Project

The term project will be worth 20% of your total grade. You are encouraged to start thinking about the concepts and methods you would like to investigate further in a real-world setting. I will ask you to submit a project proposal (individually or with a partner or two of your choice) that applies two of the modeling approaches covered in class to a relevant problem. This may be related to your own research as well. Further guidance will be provided midway through the semester. The final exam time will be devoted to in-class presentations of each project.

3.7 Attendance and participation

You are expected to show up to every class (either virtually or in-person), in the absence of any emergencies or illness (please email me ahead of time if any situations arise).

3.8 Academic Honesty Policy Statement

Since the integrity of the academic enterprise of any institution of higher education requires honesty in scholarship and research, academic honesty is required of all students at the University of Massachusetts Amherst. Academic dishonesty including but not limited to cheating, fabrication, plagiarism, and facilitating dishonesty, is prohibited in all programs of the University. Appropriate sanctions may be imposed on any student who has committed an act of academic dishonesty. Instructors should take reasonable steps to address academic misconduct. Any person who has reason to believe that a student has committed academic dishonesty should bring such information to the attention of the appropriate course instructor as soon as possible. Instances of academic dishonesty not related to a specific course should be brought to the attention of the appropriate department Head or Chair. The procedures outlined below are intended to provide an efficient and orderly process by which action may be taken if it appears that academic dishonesty has occurred and by which students may appeal such actions. Since student are expected to be familiar with this policy and the commonly accepted standards of academic integrity, ignorance of such standards is not normally sufficient evidence of lack of intent. For more information about what constitutes academic dishonesty, please see the Dean of Students' website: `https://www.umass.edu/honesty/`

3.9 Disability Statement

The University of Massachusetts Amherst is committed to making reasonable, effective and appropriate accommodations to meet the needs of students with disabilities and help create a barrier-free campus. If you are in need of accommodation for a documented disability, register with Disability Services to have an accommodation letter sent to your faculty. It is your responsibility to initiate these services and to communicate with faculty ahead of time to manage accommodations in a timely manner. For more information, consult the Disability Services website at `http://www.umass.edu/disability/`.

3.10 Title IX Statement

In accordance with Title IX of the Education Amendments of 1972 that prohibits gender-based discrimination in educational settings that receive federal funds, the University of Massachusetts Amherst is committed to providing a safe learning environment for all students, free from all forms of discrimination, including sexual assault, sexual harassment, domestic violence, dating violence, stalking, and retaliation. This includes interactions in person or online through digital platforms and social media. Title IX also protects against discrimination on the basis of pregnancy, childbirth, false pregnancy, miscarriage, abortion, or related conditions, including recovery. There are resources here on campus to support you. A summary of the available Title IX resources (confidential and non-confidential) can be found at the following link: `https://www.umass.edu/titleix/resources`. You do not need to make a formal report to access them. If you need immediate support, you are not alone. Free and confidential support is available 24 hours a day/7 days a week/365 days a year at the SASA Hotline 413-545-0800.

4 Schedule

This course is broadly organized around 9 modules. The schedule (see Table 2 on the next page) may be adapted over the duration of the semester to suit the needs of the class. Assigned readings are provided in Table 3.

4.1 Important dates

- No class on the following dates:
  - Wednesday, November 22 (Thanksgiving Break)
  - Monday, October 16 (Exam I)
  - Wednesday, October 18 (Instructor at conference from Oct 15–18)
  - Wednesday, December 6 (Exam II)
- Last day of class: Wednesday, December 4
- Project presentations: Monday, December 11 (3:30–5:30pm) [Subject to change]

Table 2 Course schedule

Day	Date	L/N	Topic	Assignments
Module 1: Background				
W	Sep 6	1a	Introduction	
M	Sep 11	1b	Supervised Learning Overview	PS1 assigned
W	Sep 13	1c	Data Handling and Visualization	
Module 2: Regression (Linear Methods)				
M	Sep 18	2a	Least Squares Estimation	
W	Sep 20	2b	Subset Selection and Shrinkage Methods	
M	Sep 25		(2b continued)	PS1 due; PS2 assigned
Module 3: Classification (Linear Methods)				
W	Sep 27	3a	Performance Metrics	
M	Oct 2	3b	Logistic Regression	
Module 4: Resampling Methods				
W	Oct 4	3c	Linear Discriminant Analysis	
Tu	Oct 10	4b	Cross-Validation & Bootstrapping	PS2 due
Module 5: Flexible Functional Forms				
W	Oct 11	5a	Polynomial Regression & Basis Functions	
M	Oct 16	E1	Exam I (take-home; no class)	PS3 assigned
M	Oct 23	5b	Splines	
W	Oct 25	5c	Generalized Additive Models	
Module 6: Tree-Based Methods				
M	Oct 30	6a	Decision Trees for Classification and Regression	PS3 due; PS4 assigned
W	Nov 1	6b	Bagging, Random Forests & Boosting	
Module 7: Support Vector Machines				
M	Nov 6	7a	Support Vector Classifier	
W	Nov 8	7b	Support Vector Machines and Kernels	
Module 8: Deep Learning				
M	Nov 13	8a	Feed-forward Neural Networks	PS4 due; PS5 assigned
W	Nov 15	8b	Fitting a Neural Network	
M	Nov 20	8c	Convolutional Neural Networks	
M	Nov 27	8d	Recurrent Neural Networks	PS5 due
Module 9: Unsupervised Learning				
W	Nov 29	9a	Principal Components Analysis	
M	Dec 4	9b	Clustering Analysis	
W	Dec 6	E2	Exam II (take-home; no class)	
M	Dec 11		Project Presentations	

5 Readings

Table 3 Primary readings

Module	Topics	Readings
1	**Background**	
a	Introduction	ISL 1
b	Supervised Learning Overview	ISL 2
c	Data Handling and Visualization	
2	**Regression (Linear Methods)**	
a	Least Squares Estimation	ISL 3
b	Subset Selection and Shrinkage Methods	ISL 6.1–2
3	**Classification (Linear Methods)**	
a	Performance Metrics	ISL 4.1–2
b	Logistic Regression	ISL 4.3
c	Linear Discriminant Analysis	ISLR 4.4–5
4	**Model Assessment and Selection**	
a	Cross-Validation and Bootstrapping	ISL 5
5	**Flexible Functional Forms**	
a	Polynomial Regression & Basis Functions	ISL 7.1–3
b	Splines	ISL 7.4–5
c	Generalized Additive Models & Local Regression	ISL 7.6–7
6	**Tree-Based Methods**	
a	Decision Trees for Classification and Regression	ISL 8.1
b	Bagging, Random Forests & Boosting	ISL 8.2
7	**Support Vector Machines**	
a	Support Vector Classifier	ISL 9.1–2
b	Support Vector Machines	ISL 9.3–5
8	**Deep Learning**	
a	Feed-forward Neural Networks	ISL 10.1–2
b	Fitting Neural Networks	ISL 10.7
c	Convolutional Neural Networks	ISL 10.3
d	Recurrent Neural Networks	ISL 10.5
9	**Unsupervised Learning**	
a	Principal Components Analysis	ISL 12.1–2
b	Clustering Analysis	ISL 12.3