Syllabus: EEE 549 Statistical Machine Learning: Theory to Practice (Fall 2024)

Instructor: Lalitha Sankar

Contact Info: email: `lsankar@asu.edu` Office: GWC 436

Office Hours: TBD

Meeting Info: Tue, Thu 1:30pm–2:45pm, Location: STPV-324

Syllabus Disclaimer The syllabus is a statement of intent and serves as an implicit agreement between the instructor and the student. Every effort will be made to avoid changing the syllabus but the possibility exists that unforeseen events will make changes necessary.

Course Description: Machine learning explores the design, analysis, and construction of algorithms that can learn from data and make inferences or predictions about future outcomes. The focus is on a **methodical approach** that will highlight the role of statistical and computational methods in analysis of data. This course includes a **near equal dose of theory and practice** with the goal of providing a thorough grounding in the fundamental methodologies and algorithms in machine learning. The focus will be a methodical way of learning that begins from the theoretical underpinnings of machine learning focused broadly on two distinct types of learning methods, namely supervised and unsupervised learning. Within each type, various well-studied and formulated approaches will be studied.

The aim of this course is to introduce students across engineering to basic data science concepts and algorithms in a rigorous manner. A desired outcome is for students to be able to learn to distinguish between different algorithms and determine which methods are ideal for a problem setting at hand.

Prerequisites: Linear algebra and **an elementary** knowledge of probability/statistics (all relevant probability concepts will be covered in class)

Co-requisite: EEE554 or equivalent (email Professor for clarification, if needed)

Course outcome: students should be able to apply ML techniques to a variety of engineering problems.

Course Topics:

- Introduction to machine learning, review of probability and linear algebra
- Supervised learning; linear regression; Bias and variance
- Weighted least squares; logistic regression
- Perceptron and general linear models
- Support Vector Machines, Kernels
- p-values, hypothesis testing, Bayesian inference
- Unsupervised learning: k-means, expectation maximization
- Principal and independent component analysis
- Decision trees, boosting, bagging
- learning theory, deep learning

Textbook and Reference Materials: (both are free online)

[HTF] *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Trevor Hastie, Robert Tibshirani, Jerome Friedman.

[Murphy] *Machine Learning: A Probabilistic Perspective*, Kevin Murphy.

Grading:	Homework Assignments	40 %
	Quiz (every class)	05 %
	Midterm	20 %
	Class/Slack Participation	05 %
	Final Project	30 %
	Homework 0 (review)	10 % **(bonus)**

Homework: There will be a homework assignment due every week. Homework will include analytical problems as well as computer exercises (Python preferred).

In-class Quizzes: Every week there will be a (very) short multiple-choice quiz via Canvas.

Midterm Exams: In class, Midterm I: TBD – late Sep/early October

Final Exam: Phase I submission: mid-to-late October; Final Report due: Dec 10, 2024.

Homework Expectations:

1. Homework assignments will be posted on Canvas via the Gradescope tool.
2. Homework assignments will be graded using Gradescope. So, upload your written/completed assignments via Gradescope.
3. Gradescope makes it easier to understand the grading rubric (consistent across all students).
4. Handwritten solutions should be collated as a PDF and uploaded. Detailed instructions will be shared in every HW assignment.
5. Show all work for complete credit.
6. You may work together on the homework, but copying is unacceptable.
7. Homework assignments should be uploaded via Gradescope at or before the assigned time.
8. **Late or Missed Assignments**: Notify the instructor *before* an assignment is due if an urgent situation arises and you are unable to submit the assignment on time. Assignments submitted after the due date will be penalized 50% per day, unless prior arrangements have been made.
9. Late assignments will be allowed without penalty to the grade in the case of (1) a university-sanction event, (2) religious holidays, (3) work performed in the line-of-duty, and (4) illness, quarantine, or self-isolation related to illness as documented by a health professional.
10. Homework assignments will contain both theoretical and applied problems; applied problems are to be solved using Python or a similar software.
11. Slack channel can be used to engage in technical discussions related to homework assignments.

Academic Integrity:

If a student copies their homework or other assignments from another student, both parties will be held responsible. You are urged to review the ABOR Student Code of Conduct and disciplinary procedures.
http://students.asu.edu/files/StudentCodeofConduct.pdf
http://students.asu.edu/files/StudentDisciplinaryProceduresChapter5.pdf

Academic honesty is expected of all students in all examinations, papers, laboratory work, academic transactions, and records. The possible sanctions include, but are not limited to appropriate grade penalties, course failure (indicated on the transcript as a grade of E), loss of registration privileges, disqualification and dismissal. For more information, click here.

The use of generative software such as ChatGPT is not prohibited but clarifying your prompts and your efforts beyond that is crucial.

Clear demonstration of copying will lead to a complete loss of points on the relevant homework and a direct reporting to the academic integrity office.

You may also find these reference materials useful throughout the semester.

Additional References on Machine Learning (and related topics)

1. **Supplementary Books and References:** For a gentler introduction to machine learning; **free online**:
   - A Course in Machine Learning by Hal Daume III
   - Pattern Recognition and Machine Learning, Christopher Bishop.
   - Computer Age Statistical Inference: Algorithms, Evidence and Data Science, Bradley Efron, Trevor Hastie.
   - Crib sheet of math for ML by Iain Murray
   - Understanding Machine Learning: From Theory to Algorithms, Shai Shalev-Shwartz, Shai Ben-David. An introduction to theoretical machine learning.
   - Foundations of Data Science, by Avrim Blum, John Hopcroft and Ravi Kannan. This freely available pdf has nice chapters on machine learning (chapter 5), clustering (chapter 7) and SVD (chapter 3).
2. **Linear Algebra and Matrix Analysis**
   - These wonderful videos by 3blue1brown provide a gentle and highly intuitive overview of linear algebra.
   - Linear Algebra Review and Reference by Zico Kolter and Chuong Do (free). Light refresher for linear algebra and matrix calculus if you're a bit rusty.
   - Linear Algebra, David Cherney, Tom Denton, Rohit Thomas and Andrew Waldron (free). Introductory linear algebra text.
   - Matrix Analysis Horn and Johnson. A great reference from elementary to advanced material.
3. **Probability and Statistics**
   - Probability Review by Arian Maleki and Tom Do. (From Andrew Ng's machine learning class.)
   - Sankar's slides (on canvas) on probability (Module 0 on canvas)
   - All of Statistics, Larry Wasserman. Chapters 1-5 are a great probability refresher and the book is a good reference for statistics.
   - A First Course in Probability, Sheldon Ross. Elementary concepts (previous editions are a couple bucks on Amazon)
   - Sal Academy and 3blue1brown videos on probability are good starting points
4. **Optimization**
   - Numerical Optimization, Nocedal, Wright. Practical algorithms and advice for general optimization problems.
   - Convex Optimization: Algorithms and Complexity, Sébastien Bubeck. Elegant proofs for the most popular optimization procedures used in machine learning.
5. **Python**

- www.learnpython.org "Whether you are an experienced programmer or not, this website is intended for everyone who wishes to learn the Python programming language."
- Convex Optimization: Algorithms and Complexity, Sébastien Bubeck. Elegant proofs for the most popular optimization procedures used in machine learning.
- NumPy for Matlab users
- Resources to install Conda are available with the official documentation: https://docs.conda.io/en/latest/

6. **Latex**
   - Learn Latex in 30 minutes
   - Overleaf. An online Latex editor.
   - TexPad for both OSX and Windows
   - Standalone Latex editor on your local machine
   - Detexify LaTeX handwritten symbol recognition
   - Latex Math symbols

Remote/Asynchronous Attendance

The primary mode of class activity is in-person, and so students are expected to attend class when possible. However, if you are sick, feeling ill, or have had a possible exposure to infectious disease, **DO NOT COME TO CLASS**! Unfortunately, live remote participation is no longer an option; however, email the instructor with supporting documentation to obtain links to the recorded lectures. Recordings of class sessions **will not** in general be posted to Canvas but shared as needed individually.

Even in the case of an exam, if you are feeling ill, stay home. Email the instructor as soon as possible to arrange a make-up exam.

Student Accessibility

Students who need disability accommodations in this class but have not registered with Student Accessibility and Inclusive Learning Services (SAILS) should contact SAILS immediately. Information and instructions for requesting an accommodation are available at https://eoss.asu.edu/accessibility.

Policy Against Threatening Behavior

All incidents and allegations of violent or threatening conduct by an ASU student (whether on-or off campus) must be reported to the ASU Police Department (ASU PD) and the Office of the Dean of Students. If either office determines that the behavior poses or has posed a serious threat to personal safety or to the welfare of the campus, the student will not be permitted to return to campus or reside in any ASU residence hall until an appropriate threat assessment has been completed and, if necessary, conditions for return are imposed. ASU PD, the Office of the Dean of Students, and other appropriate offices will coordinate the assessment in light of the relevant circumstances. For more information please visit https://eoss.asu.edu/dos/srr/PoliciesAndProcedures and https://eoss.asu.edu/dos/safety/ThreateningBehavior.

Reporting Title IX Violations

Title IX is a federal law that provides that no person be excluded on the basis of sex from participation in, be denied benefits of, or be subjected to discrimination under any education program or activity. Both Title IX and university policy make clear that sexual violence and harassment based on sex is prohibited. An individual who believes they have been subjected to sexual violence or harassed on the basis of sex can seek support, including counseling and academic support, from the university. If you or someone you

know has been harassed on the basis of sex or sexually assaulted, you can find information and resources at https://sexualviolenceprevention.asu.edu/faqs.

Policy on Sexual Discrimination

Arizona State University is committed to providing an environment free of discrimination, harassment, or retaliation for the entire university community, including all students, faculty members, staff employees, and guests. ASU expressly prohibits discrimination, harassment, and retaliation by employees, students, contractors, or agents of the university based on any protected status: race, color, religion, sex, national origin, age, disability, veteran status, sexual orientation, gender identity, and genetic information.

As a mandated reporter, I am obligated to report any information I become aware of regarding alleged acts of sexual discrimination, including sexual violence and dating violence. ASU Counseling Services, https://eoss.asu.edu/counseling, is available if you wish discuss any concerns confidentially and privately.

Copyrighted Materials

Students must refrain from uploading to any course shell, discussion board, or website used by the course instructor or other course forum, material that is not the student's original work, unless the students first comply with all applicable copyright laws; faculty members reserve the right to delete materials on the grounds of suspected copyright infringement.