ME 539: Introduction to Scientific Machine Learning

Course Information

CRNs:

- 13685 (lecture): Meets T/R 12:00 – 1:15 PM ET, ARMS B071,
- 22465 and 13686 (async-online): BrightSpace https://purdue.brightspace.com/d2l/home/839270

Instructor(s)

Ilias Bilionis

- Associate Professor of Mechanical Engineering, Purdue University
- E: ibilion@purdue.edu
- Research site: Predictive Science Lab

Teaching Assistant(s)

- Kairui Hao, hao55@purdue.edu
- Ting-Chun Kuo, kuo63@purdue.edu
- Atharva Hans, hans1@purdue.edu
- Zhengwei Chen, chen3413@purdue.edu
- Alexander Alberts, albert31@purdue.edu
- Murali Krishnan Rajasekharan Pillai, mrajase@purdue.edu

Audience

Advanced undergraduate students, graduate students, and professionals interested in applications of data analytics to engineering problems.

Course Description

This course provides an introduction to data science for individuals with no prior knowledge of data science or machine learning. The course starts with an extensive review of probability theory as the language of uncertainty, discusses Monte Carlo sampling for uncertainty propagation, covers the basics of supervised (Bayesian generalized linear regression, logistic regression, Gaussian processes, deep neural networks, convolutional neural networks), unsupervised learning (k-means clustering, principal component analysis, Gaussian mixtures) and state space models (Kalman filters). The course also reviews the state-of-the-art in physics-informed deep learning and ends with a discussion of automated Bayesian inference using probabilistic programming (Markov chain Monte Carlo, sequential Monte Carlo, and variational inference). Throughout the course, the instructor follows a probabilistic perspective that highlights the first principles behind the presented methods with the ultimate goal of teaching the student how to create and fit their own models.

Prerequisites

- Working knowledge of multivariate calculus and basic linear algebra
- Basic Python knowledge
- Knowledge of probability and numerical methods for engineers would be helpful, but not required.

Course Learning Outcomes

After completing this course, you will be able to:

- Represent uncertainty in parameters in engineering or scientific models using probability theory.
- Propagate uncertainty through physical models to quantify the induced uncertainty in quantities of interest.
- Solve basic supervised learning tasks, such as: regression, classification, and filtering.
- Solve basic unsupervised learning tasks, such as: clustering, dimensionality reduction, and density estimation.
- Create new models that encode physical information and other causal assumptions.
- Calibrate arbitrary models using data.
- Apply various Python coding skills.
- Load and visualize data sets in Jupyter notebooks.
- Visualize uncertainty in Jupyter notebooks.
- Recognize basic Python software (e.g., Pandas, numpy, scipy, scikit-learn) and advanced Python software (e.g., pymc3, pytorch, pyro, Tensorflow) commonly used in data analytics.

Office Hours and Course Communications

- **Piazza Discussion Forum:**
  We will use Piazza as the discussion platform in this class. Here you can discuss course-related topics, ask and answer questions, or share concerns with your peers.
- **Office Hours:**
  We will use Zoom for office hours. Office hour schedule will be posted in the course.

Required Texts and Tools

Online Textbook
https://predictivesciencelab.github.io/data-analytics-se/index.html

Jupyter Notebook
Jupyter notebooks are interactive documents that can simultaneously contain text, mathematics, images, and executable code. The executable code can be in many programming languages (e.g., R, Matlab), but we are only going to use Python in this course. The course uses Jupyter notebooks for the following content: Reading Activities, Hands-on Activities, and Homework Assignments. The rationale behind this choice is that it allows the student to focus on the mathematical methods rather than the programming and it ensures the reproducibility of the course content. Of course, understanding the code in Jupyter notebooks does require knowledge of Python, albeit it does not require knowing how to structure and call Python code from the command line. Jupyter notebooks can be run either on the students' personal computers (instructions vary with operating system and can be found here) or in several cloud computing resources. The recommended method for this class is to use Google Colab which is available free of charge and requires only a standard Google account. The activity links included in the course will take you automatically to a copy of the latest version of the corresponding Jupyter notebook which you can then save and edit on your Google Drive.

Access to the Jupyter Notebooks Repository
As stated earlier, the recommended method for using the Jupyter notebooks of this class is to use Google Colab. The links to all the activities will take you directly to a Google Colab copy of the Jupyter notebook. If you want to use any alternative method (e.g., your personal computer or anything else), you will need access to the Jupyter Notebook repository for the class. Select this link for a Git version control repository.

(If you have no idea what Git is, this is a good tutorial.) All the activities are inside the "activities" folder and they are named using the following convention: "activity_type_<lecture_number>.<activity_type_count>.ipynb." For example, "hands-on-06.3.ipynb" is the third hands-on activity of Lecture 6.

Grading

This course will be graded based on the following criteria:

Assessment Type	Description	% of Final Grade
Homework	There will be seven (7) homework assignments. The homework assignments will be both theoretical (e.g., prove this, derive that) and computational (e.g., use this data to fit that model, create and fit a model for this situation). The assignments will be in the form of a Jupyter notebook with empty space reserved for your writing or coding. If you wish, you can do the writing by hand (instead of the latex required by Jupyter notebooks), scan it and submit a single PDF. Submissions should be made through Gradescope.	90%
Midterm	Midterm exam will include lectures 1 – 12. • On-campus students will take the exam during scheduled class time. • Distance students will be given options to accommodate schedules.	10%

Grading Scale

Your course grade will be based on the following grading scale:

A+ = > 95%	B+ = 80-85%	C+ = 67-70%	D+ = 57-60%	F = < 50%
A = 88-95%	B = 73–80%	C = 62-67%	D = 52-57%	
A- = 85–88%	B- = 70-73%	C- = 60-62%	D- = 50-52%	

The instructor reserves the right to change the grading scale in a way that favors students.

Discussion Guidelines

Please follow the Discussion Guidelines when contributing to discussions in this course. Here are a few of the key points you should remember:

- Do not use offensive language. Present ideas appropriately.
- Be cautious in using Internet language. For example, do not capitalize all letters since this suggests shouting.
- Avoid using vernacular or slang language. This could possibly lead to misinterpretation.
- Do not hesitate to ask for feedback.
- Be concise and to the point.
- Think and edit before you push the "Send" button.

Accessibility Information

Purdue University strives to make learning experiences as accessible as possible. If you anticipate or experience physical or academic barriers based on disability, you are welcome to let me know so that we can discuss options. You are also encouraged to contact the Disability Resource Center at: drc@purdue.edu or by phone: 765-494-1247

- Google Accessibility
- Jupyter Notebook a11y toolbar (from Microsoft research)

Mental Health/Wellness Statement:

If you find yourself beginning to feel some stress, anxiety and/or feeling slightly overwhelmed, try WellTrack**.** Sign in and find information and tools at your fingertips, available to you at any time.

If you need support and information about options and resources, please contact or see the Office of the Dean of Students. Call 765-494-1747. Hours of operation are M-F, 8 am- 5 pm.

If you find yourself struggling to find a healthy balance between academics, social life, stress, etc., sign up for free one-on-one virtual or in-person sessions with a Purdue Wellness Coach at RecWell. Student coaches can help you navigate through barriers and challenges toward your goals throughout the semester. Sign up is free and can be done on BoilerConnect.

If you're struggling and need mental health services: Purdue University is committed to advancing the mental health and well-being of its students. If you or someone you know is feeling overwhelmed, depressed, and/or in need of mental health support, services are available. For help, such individuals should contact Counseling and Psychological Services (CAPS) at 765-494-6995 during and after hours, on weekends and holidays, or by going to the CAPS office on the second floor of the Purdue University Student Health Center (PUSH) during business hours. The CAPS website also offers resources specific to situations such as COVID-19.

Academic Integrity

Academic integrity is one of the highest values that Purdue University holds. Individuals are encouraged to alert university officials to potential breaches of this value by either emailing or by calling 765-494-8778. While information may be submitted anonymously, the more information that is submitted provides the greatest opportunity for the university to investigate the concern.

The Purdue Honor Pledge

"As a boilermaker pursuing academic excellence, I pledge to be honest and true in all that I do. Accountable together - we are Purdue"

Nondiscrimination Statement

Purdue University is committed to maintaining a community which recognizes and values the inherent worth and dignity of every person; fosters tolerance, sensitivity, understanding, and mutual respect among its members; and encourages each individual to strive to reach his or her own potential. In pursuit of its goal of academic excellence, the University seeks to develop and nurture diversity. The University believes that diversity among its many members strengthens the institution, stimulates creativity, promotes the exchange of ideas, and enriches campus life. Link to Purdue's nondiscrimination policy statement.

Generative AI Policy

Generative AI is a form of artificial intelligence that enables machines to generate new content or data that's similar to some given training data. This includes creative tasks like writing, composing music, creating images, and solving problems in mathematics and science.

Some specific examples of generative AI programs that may be utilized within this class include ChatGPT, Co-pilot, Bard, etc. These tools can assist in various tasks related to the coursework.

Policy Guidelines

1. Permitted Use in Class:
    a. Generative AI is allowed throughout the course in homework and the midterm.
    b. Students are encouraged to familiarize themselves with generative AI, reflecting its importance in our professional lives.
2. Allowable Queries to Generative AI:
    a. Assistance with code for homework or midterm problems.
    b. Clarifications on class concepts.
    c. Facilitate proofs in mathematical contexts.
    d. Construct paragraph draft from bullet points.
    e. Improve language in written paragraphs.
3. Responsibilities and Cautions:
    a. Please be cautious, as generative AI does not guarantee accuracy.
    b. Students are responsible for the correctness of their submissions and must understand why an answer provided by AI is correct.
    c. Randomly generated submissions without personal effort to understand will lead to point deductions.
4. Emphasizing Personal Effort and Learning:
    a. Resist the temptation to rely solely on generative AI, especially at the last minute.
    b. Striving to understand, engaging with lectures, fellow students, and instructors, and attempting homework by hand is vital to the learning process.
    c. Focus on understanding rather than solely on grades.
5. The Role of Generative AI:
    a. Consider generative AI as a tool to polish your work, avoid repetitive tasks, and enhance productivity.
    b. Embrace generative AI as a means to focus on more advanced aspects of science and math and ask interesting questions.

Conclusion

Generative AI is, in our course, serving as a valuable tool to assist and augment our efforts. However, it should never replace the core educational principles of understanding, personal growth, and engagement with the material.

This policy was generated by GPT 4.0 using bullet points from Prof. Bilionis, and it was further polished using Grammarly Premium.

Diversity and Inclusion Statement

In our discussions, structured and unstructured, we will explore a variety of challenging issues, which can help us enhance our understanding of different experiences and perspectives. This can be challenging, but in overcoming these challenges we find the greatest rewards. While we will design guidelines as a group, everyone should remember the following points:

- We are all in the process of learning about others and their experiences. Please speak with me, anonymously if needed, if something has made you uncomfortable.
- Intention and impact are not always aligned, and we should respect the impact something may have on someone even if it was not the speaker’s intention.
- We all come to the class with a variety of experiences and a range of expertise, we should respect these in others while critically examining them in ourselves.

Course Evaluation

At the end of the course, you will be provided with an opportunity to evaluate this course and your instructor. Your participation is an integral part of this course, and your feedback is vital to improving education at Purdue University. I strongly urge you to participate in the Qualtrics survey.

Disclaimer

This syllabus is subject to change.

Introduction to Scientific Machine Learning Schedule (FALL'23)

Week	Section	Lectures	Assignments
1 & 2 8/21 – 9/3	**Introduction** **Review of Probability Theory**	1. Introduction to Predictive Modeling Videos: 1.1 – 1.5 (29:14) 2. Basics of Probability Theory Videos: 2.1 – 2.5 (26:14) 3. Discrete Random Variables Videos: 3.1 – 3.6 (28:09) 4. Continuous Random Variables Videos: 4.1 – 4.6 (42:44)	**Homework 1: Lectures 1 - 4** ○ Due: Monday, 9/4, 11:59 PM ET *(9/5, 03:59 UTC)*
3 & 4 9/4 – 9/17	 **Uncertainty Propagation**	5. Collections of Random Variables Videos: 5.1 – 5.4 (45:43) 6. Random Vectors Videos: 6.1 – 6.5 (36:51) 7. Basic Sampling Videos: 7.1 – 7.4 (26:00) 8. The Monte Carlo Method for Estimating Expectations Videos: 8.1 – 8.4 (22:27)	**Homework 2: Lectures 5 - 8** ○ Due: Monday, 9/18, 11:59 PM ET *(9/19, 03:59 UTC)*
5 & 6 9/18 – 10/1	 **Principles of Bayesian Inference**	9. Monte Carlo Estimates of Various Statistics Videos: 9.1 – 9.4 (43:05) 10. Quantify Uncertainty in Monte Carlo Estimates Videos: 10.1 – 10.4 (26:04) 11. Selecting Prior Information Videos: 11.1 – 11.4 (34:47) 12. Analytical Examples of Bayesian Inference Videos: 12.1 – 12.4 (37:40)	**Homework 3: Lectures 9 - 12** ○ Due: Monday, 10/2, 11:59 PM ET *(10/3, 03:59 UTC)*

Week	Section	Lectures	Assignments
7, 8 & 9 **10/2 – 10/22**	**Supervised Learning: Linear Regression and Logistic Regression** **Unsupervised Learning**	13. Linear Regression Via Least Squares Videos: 13.1 – 13.5 (45:48) 14. Bayesian Linear Regression Videos: 14.1a-c, 14.2 – 14.4 (50:41) 15. Advanced Topics in Bayesian Linear Regression Videos: 15.1 – 15.3 (22:14) 16. Classification Videos: 16.1a-d, 16.2 – 16.5 (1:16:08) 17. Clustering and Density Estimation Videos: 17.1 – 17.4 (46:18)	**October Break: 10/9 – 10/10** **MIDTERM: Lectures 1 – 12** Midterm Exam has been cancelled. **Homework 4: Lectures 13 - 17** ○ Due: Monday, 10/23, 11:59 PM ET *(10/24, 03:59 UTC)*
10 & 11 **10/23 – 11/5**	 **State-Space Models** **Gaussian Process Regression**	18. Dimensionality Reduction Videos: 18.1, 18.2a-c, 18.3 – 18.6 (50:38) 19. State-Space Models – Filtering Basics Videos: 19.1 – 19.7 (46:18) 20. State-Space Models – Kalman Filters Videos: 20.1-20.6 (50:06) 21. Gaussian Process Regression – Priors on Function Spaces Videos: 21.1 – 21.5 (52:30)	**Homework 5: Lectures 18 - 21** Due: Monday, 11/6**,** 11:59 PM ET *(11/7, 04:59 UTC)*
12 & 13 **11/6 – 11/19**	 **Neural Networks**	22. Gaussian Process Regression – Conditioning on Data Videos: 22.1 – 22.4 (55:04) 23. Bayesian Global Optimization Videos: 23.1 – 23.9 (49:05) 24. Deep Neural Networks Videos: 24.1 – 24.5 (1:01:39) 25. Deep Neural Networks Continued Videos: 25.1 – 25.7 (53:24)	**Homework 6: Lectures 22 - 25** • Due: Monday, 11/20, 11:59 PM ET *(11/21, 04:59 UTC)*

Week	Section	Lectures	Assignments
14 & 15 **11/20 – 12/3**	**Advanced Methods for Characterizing Posteriors**	26. Physics-Informed Deep Neural Networks Videos: 26.1 – 26.6 (49:07) 27. Sampling Methods Videos: 27.1 – 27.6 (1:01:43) Example Application I: Model Calibration Videos: New Content to Be Recorded	**Thanksgiving Break: 11/22 – 11/24**
16 **12/4 – 12/10**		28. Variational Inference Videos: 28.1 – 28.6 (4:28) Example application II: Particle Image Velocimetry Videos: New Content to Be Recorded	**Homework 7 (Lectures 26 – 30)** ○ Due: Monday, 12/11, 11:59 PM ET (*12/12, 04:59 UTC)*