Philosophy 29300: Philosophical Fundamentals of Machine Learning

Bruce Rushing

Fall, 2024

E-mail: bmrushin@purdue.edu
Office Hours: Thursday 0900–1000
Office: Pfendler Hall of Agriculture 103
CRN: 29553
Modality: Face-to-Face

Web: **brightspace**
Class Hours: Tuesday, Thursday 1030–1145
Class Room: Pfendler Hall of Agriculture 103
Course Credit Hours: 3.0
Prerequisites: None

Course Description

This course is designed as an applied survey course for current methods in machine learning and the philosophical applications of those methods. The goal is for students to have familiarity with modern methods in machine learning including mathematical fundamentals, decision theory, statistical learning theory, regression, classification, unsupervised learning, neural networks, reinforcement learning, language models, and generative methods. Emphasis will be on practical applications through modern software libraries including `numpy`, `pandas`, `scipy`, `scikit-learn`, `pytorch`, `transformers`. Students will also become acquainted with the philosophical ramifications of machine learning in fields such as philosophy of science, philosophy of mind, philosophy of politics, and ethics. Why should someone be concerned about the philosophy of machine learning? There are a couple reasons. First, machine learning has had a tremendous impact on the sciences and societies and economies of countries across the globe. A better understanding of how machine learning works will allow students to critically evaluate claims about the present and future of AI in the world. Second, philosophy helps us deliberate over the best methods we have for making decisions and what we value. Applying philosophy to AI will help us make better informed decisions about the trustworthiness of machine learning systems and how they affect our values.

The first half of this course will focus on giving the student introductory knowledge on the foundations of contemporary techniques in machine learning. The second half of this course will look at the philosophical implications of these techniques.

Required Materials

There are no required textbooks: all necessary readings will be provided in class. However, the recommended textbooks for this course include:

- Probabilistic Machine Learning: An Introduction by Kevin P. Murphy (Murphy I).
- Probabilistic Machine Learning: Advanced Topics by Kevin P. Murphy (Murphy II).

I highly recommend you acquire the Murphy textbook as it is an excellent reference, and relevant readings will be included for each week of the machine learning materials. Philosophical readings will be assigned and provided by the instructor.

Prerequisites/Corequisites

It is required that students have familiarity with at least one programming language, with a preference for Python. It is also recommended—though not required—that students have taken at least one college-level math course, such as Calculus, Linear Algebra, or a statistics course.

Course Objectives

The learning outcomes for this course are for students to successfully:

1. Be able to: perform matrix multiplication, define and identify a vector space, translate a linear transformation into a coordinate matrix, calculate gradient of a loss function, explain gradient descent and backpropagation, define and verify a probability function, compute a posterior probability through Bayes rule, verify conditional independence of random variables, calculate the two lower moments of probability distributions, identify and calculate probabilities for a Bernoulli/Binomial, Categorical/Multinomial, and univariate Normal distributions, understand and explain Bayesian statistical decision theory.
2. Build and troubleshoot at least three of the following machine learning algorithms in Python: k-nearest neighbor, naive bayes, linear regression, logistic regression, decision tree, random forest, convolutional neural network, k-means clustering, expectation maximization, principal components analysis, word2vec, transformer.
3. Apply machine learning methods to at least one of the following topics in the philosophy of artificial intelligence: philosophy of mind, philosophy of science, philosophy of politics, philosophy of ethics, and AI Safety.

Course Structure

Class Structure

Classes will consist of two one hour and twenty minute lectures given on the designated class days. Those lectures will occur in Pfendler Hall of Agriculture 103. Lectures will be punctuated by opportunities for students to answer practice problems for attendance.

Assessments

The primary form of assessment will consist of homework assignments, a midterm, and a final group project. Attendance will be taken via the use of practice problems.

Homework

There will be four homework assignments worth **40** points each. The rough schedule for homework assignments will have them be assigned every two weeks and due the following week. Homework assignments, except for assignment 1, will be released on Thursday and due on Tuesday 1.5 weeks later at 23:59 EDT (please see the schedule for the dates). Students should upload their homework assignments to Gradescope.

I highly encourage that students work together on the homework assignments. If you do work with your classmates, you are still required to write up the homework separately and you must list who you worked with on the first page of your homework submission. *If you do not write up your homework assignment separately, you will receive only* ***50%*** *credit on that assignment.*

Midterm

The midterm will be in person and will be worth **150** points. It will be held on **17 October**, and it will cover all technical material covered in the first eight weeks of class.

Final Project

The final project is a group project between two students that will focus on applying the machine learning methods learned in the first half of the course. It will be worth **150** points total, with **50** points coming from the presentation and **100** points from the final paper.

The assignment will consist of the students selecting one of several topics provided by the instructor. Undergraduate students will then form groups of pairs to work on the topics while graduate students will work on topics individually. Students will have to run experiments and present the results of their work in the final two weeks of class with a **10 minute** presentation. In addition, students will have to write either a short, **4 page** IEEE-style explanation of their experiments, results, and applications to the question in the project. The due date for the paper is scheduled day of the final exam XXX.

Grading Policy

The typical Purdue grading scale will be used. I reserve the right to curve the scale dependent on overall class scores at the end of the quarter. Any curve will only ever make it easier to obtain a certain letter grade. The grade will count the assessments using the following proportions:

- **32%** of your grade will be determined by 4 homework assignments (8% each).
- **8%** of your grade will be determined by attendance.
- **30%** of your grade will be determined by the midterm exam.
- **30%** of your grade will be determined by the final project.

You will have opportunities to earn extra credit by attending talks and providing feedback on the course. Expect more details in class.

Course Policies

Learning Resources, Technology, and Texts

Textbook Requirements: See required texts for the textbook requirements for the course. The course website and my personal website (brucemrushing.com) will post all assignments.

Large Language Model (LLM) Policy: students are allowed to consult LLMs through services like ChatGPT, Claude, and Gemini. However, if you do consult them for homework assignments or final project, you should include them in your list of collaborators, and you should document exactly where they were used and how you used them. I recommend you avoid using them for solutions to the problems for a number of reasons. I have tested them and they do not always get the problems reliably correct; any mistake due to them will be on you. But on the problems where they get things right, it will shortchange you as a learning opportunity. That being said, they are an important tool, especially in software development, so you are allowed to use them. Just use them wisely and document your usage extensively.

Hardware Requirements: most experiments will be conducted in class with VRAI lab hardware. Homework problems are tailored so that they run effectively on any fairly modern computer. If you need lab equipment for homework, please contact me or the TA for access to the lab.

Tutoring Support: all tutoring support is through you and a private part.

During Class

Students are expected to be respectful to both me as well as their class mates. Students will obey the Purdue Code of Student Conduct. Obey the golden rule here: treat others as you would have yourself be treated and do not be a knucklehead. When not speaking, students should have their mikes muted. It is highly encouraged that students turn on their cameras but it is not required. If you have a question, feel free to interrupt me or post it in chat.

Attendance Policy

Attendance of the lectures is mandatory and attendance will be taken with questions in class.

This course follows the University Academic Regulations regarding class attendance, which state that students are expected to be present for every meeting of the classes in which they are enrolled. Attendance will be taken at the beginning of each class and lateness will be noted. When conflicts or absences can be anticipated, such as for many University-sponsored activities and religious observations, you should inform me of the situation as far in advance as possible. For unanticipated or emergency absences when advance notification to is not possible, contact me as soon as possible by email or phone. For absences that do not fall under excused absence regulations (see below), this course follows the following procedures:

1. Do not come to class if you are feeling ill, but DO email me at bmrushin@purdue.edu, with the subject line: CRN 29553 absence. I do not need details about your symptoms. Just let me know you are feeling ill and cannot come to class. If it is an emergency situation, please follow the University regulations on emergent medical care (see below).
2. Unless it falls under the University excused absence regulations (see below), any work due should be submitted on time via our course Brightspace.
3. If that day's class involves assessed work such as a test or presentation, you and I will plan if and how you can make up the work, following the assignment guidelines. This plan must be done before the next class period, so again, email me immediately when you know that you will miss class.
4. The most important consideration in any absence is how it will affect your achievement of the assignment objectives and the course learning outcomes. For cases that fall under excused absence regulations, you or your representative should contact or go to the Office of the Dean of Students (ODOS) website to complete appropriate forms for instructor notification. Under academic regulations, excused absences may be granted by ODOS for cases of grief/bereavement, military service, jury duty, parenting leave, or emergent medical care. The processes are detailed, so plan ahead.

Policies on Late Assignments and Regrades

Please email me **before the deadline** of an assignment if you will not be able to submit it on time. Extensions will be granted on a case-by-case basis.

Regrades should be requested on Gradescope. I will only accept regrade requests up to 48 hours after the grade is released.

Academic Integrity and Honesty

Academic misconduct will not be tolerated. This includes, but is not limited to, copying another studentâĂŹs work, writing a friendâĂŹs name on a sign-in sheet, or editing graded assessments to misrepresent your work. All students are expected to adhere to the Purdue Academic Dishonesty Policies. For more information, please visit academic integrity. Information about copyright infringement can be found at use of copyrighted materials.

If you are caught cheating on any assignment or exam, your grade for the course will be an automatic zero and you will be dropped from the course

Accommodations for Disabilities

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.

Nondiscrimination Policy

Purdue University is committed to maintaining a community that 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 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. A hyperlink to PurdueâĂŹs full Nondiscrimination Policy Statement is included in our course Brightspace under University Policies and Statements.

Mental Health Policy

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 a.m.- 5 p.m.

Emergency Preperation

In the event of a major campus emergency, course requirements, deadlines and grading percentages are subject to changes that may be necessitated by a revised semester calendar or other circumstances beyond the instructorâĂŹs control. Relevant changes to this course will be posted onto the course website or can be obtained by contacting the instructors or TAs via email or phone. You are expected to read your @purdue.edu email on a frequent basis. A link to PurdueâĂŹs Information on Emergency Preparation and Planning is located on our Brightspace under "University Policies and Statements." This website covers topics such as Severe Weather Guidance, Emergency Plans, and a place to sign up for the Emergency Warning Notification System. I encourage you to download and review the Emergency Preparedness for Classrooms document.

The first day of class, I will review the Emergency Preparedness plan for our specific classroom, following Purdue's required Emergency Preparedness Briefing. Please make note of items like:

- The location to where we will proceed after evacuating the building if we hear a fire alarm.
- The location of our Shelter in Place in the event of a tornado warning.
- The location of our Shelter in Place in the event of an active threat such as a shooting

Schedule and weekly learning goals

The schedule is tentative and subject to change. The learning goals below should be viewed as the key concepts you should grasp after each week, and also as a study guide before each exam, and at the end of the quarter. Each exam will test on the material that was taught up until 1 week prior to the exam. The applications in the second half of the quarter tend to build on the concepts in the first half of the quarter though, so it is still important to at least review those concepts throughout the semester.

Week 01, 08/19 - 08/23: Linear Algebra and Probability Theory

- Matrix Multiplication
- Vector Spaces.
- Linear Transformations.
- Probability Calculus and Conditional Independence.
- Random Variables and Probability Distributions.
- Summary Statistics
- **Reading:** Murphy I, 7.1-7.2, 7.8, 2.1-2.6
- **Homework 1 Released**

Week 02, 08/26 - 08/30: Decision Theory, k-Nearest Neighbors, and Naive Bayes

- Bayesian statistical decision theory.
- ROC Curves.
- Empirical Risk Minimization.
- Bias-Complexity Tradeoff.
- Naive Bayes.
- k-Nearest Neighbors.
- **Reading:** Murphy I, 5.1, 5.4, 9.3, 16.1
- **Homework 1 Due 08/29**
- **Homework 2 Released**

Week 03, 09/02 - 09/06: Data Preprocessing, Linear Regression, and Logistic Regression

- Data cleaning.
- Rescaling and normalization.
- Maximum Likelihood Estimation and OLS.
- Regularization.
- Binary Logistic Regression and Gradient Descent.
- Categorical Logistic Regression.
- **Reading:** Murphy I, 11.1-4, 10.1-3

Week 04, 09/09 - 09/13: Neural Networks and Decision Trees

- Entropy, KL-Divergence, Mutual Information.
- Decision Trees.
- Bagging, Boosting, and Random Forests.
- Stochastic Gradient Descent and Backpropagation.
- MLPs and Regularization.
- Convolutional Neural Networks.
- **Reading:** Murphy I, 8.4, 13.1-5, 6.1-2, 18.1, 18.3-5
- **Homework 2 Due 09/10**
- **Homework 3 Released**

Week 05, 09/16 - 09/20: Unsupervised Learning and Dimensionality Reduction

- k-Means Clustering.
- Hierarchical Agglomerative Clustering.
- Expectation Maximization.
- Eigenvalues and Eigenvectors.
- Principal Components Analysis.
- Autoencoders.
- **Reading:** Murphy I, 21.1-4, 20.1, 20.3

Week 06, 09/23 - 09/27: Recurrent Networks and Sequence Prediction

- Word2Vec.
- RNN: Vec2Seq.
- RNN: Seq2Vec.
- RNN: Seq2Seq.
- LSTM.
- Autoregression.
- **Reading:** Murphy I, 20.5, 15-1-5
- **Homework 3 Due 09/24**
- **Homework 4 Released**

Week 07, 09/30 - 10/04: Transformers

- Attention.
- Self-attention.
- Multi-headed Attention.
- Positional Encoding.
- Masking.
- Transformers.
- **Reading:** Murphy I, 20.5, 15-1-5

Week 08, 10/07 - 10/11: Reinforcement Learning

- Markov Processes.
- Markov Reward Processes.
- Markov Decision Problems.
- **Reading:** Murphy II, 34.5.
- **Homework 4 Due 10/08**

Week 09, 10/14 - 10/18: Midterm

- Midterm Review.
- **Midterm**.

Week 10, 10/21 - 10/25: ML and Philosophy of Science

- No Free Lunch Theorems and Induction.
- PAC Learning and Occam's Razor.
- **Reading:** Schurz, "No Free Lunch Theorem, Inductive Skepticism, and the Optimality of Meta-induction", and Herrmann, "PAC Learning and Occam's Razor: Probably Approximately Incorrect".

Week 11, 10/28 - 11/01: ML and Philosophy of Mind

- Intelligence as the simplest program to achieve a goal.
- Intelligence as the most efficient program to acquire new skills.
- **Reading:** Legg and Hutter, "Universal Intelligence: A Definition of Machine Intelligence" sections Introduction, Natural Intelligence, A Definition of Machine Intelligence (pp. 391–423), Chollet, "On the Measure of Intelligence" sections I-II.1 (pp. 1–27).

Week 12, 11/04 - 11/08: ML and Ethics

- Bias in algorithmic decisions.
- Duplicating human morality.
- **Reading:** Lazar and Stone, "On the Site of Predictive Justice", Schramowski et al, "Large pre-trained language models contain human-like biases of what is right and wrong to do".

Week 13, 11/11 - 11/15: ML and Philosophy of Politics

- The effect of ML on democracy.
- The right to an explanation.
- **Reading:** Coeckelbergh, "Democracy, epistemic agency, and AI: political epistemology in times of artificial intelligence", Vredenburgh, "The Right to an Explanation".

Week 14, 11/18 - 11/22: ML and Economics

- ML as a replacement for labor.
- ML and comparative advantage.
- **Reading:** Eloundou et al, "GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models", and Smith, "Plentiful, high-paying jobs in the age of AI".

Week 15, 11/25 - 11/29: ML and Superintelligence

- The alignment problem.
- The singularity.
- **Reading:** Ngo et al, "The Alignment Problem from a Deep Learning Perspective", and Thorstead, "Against the Singularity".

Week 16, 12/02 - 12/06: Presentations

Exam Week, 12/09 - 12/13: Presentations

- **Final Project Paper Due**