CS 4033/5033 - Machine Learning Fundamentals

Dimitris Diochnos
The University of Oklahoma, School of Computer Science

Spring 2024

Time and Location

The course will meet at Sarkeys Energy Ctr N0202. Mondays, Wednesdays, Fridays, 11:30am-12:20pm.

Website: https://www.diochnos.com/teaching/CS4033-5033/2024S/index.php

Canvas: Homework assignments, project submission, and potentially other reading material.

Instructor

Dimitris Diochnos, 230 Devon Energy Hall (DEH), diochnos@ou.edu .

Teaching Assistant

The teaching assistant for the course will be **Naeem Shahabi Sani** (`shahabi`). You can send him an email directly if you append `@ou.edu` to the username shown inside the parentheses next to his name.

Office Hours

Office hours will be held on:

- **Diochnos** at **DEH 230: Mondays** and **Wednesdays** between **2:00pm-3:00pm**,
- **Naeem** at **DEH 115: Tuesdays** and **Thursdays** between **11:00am-12:00pm**.

Alternatively, we can meet **by appointment**, but please send an email to arrange such a meeting.

A Guide on Sending Emails to Faculty and Teaching Assistants. Please see: http://bit.ly/2bbU7Ms.

Prerequisite Background (Prior Programming Experience is Assumed)

1. CS 2413 (Data Structures)
2. CS 2813 or MATH 2513 (Discrete Mathematics)
3. MATH 4753 or ISE 3293 or MATH 4743 (A course on Statistics)
4. MATH 3333 (Linear Algebra)
5. MATH 3113 or MATH 3413 (further mathematical maturity which can be exhibited by ODEs or another advanced mathematical course).

Course Overview

By taking this course, you will learn about current techniques that enable machines to learn and adapt their behavior over time and to new situations. You will take a small step into creating intelligent machines for the future by choosing and completing a team-based semester-long machine learning project. You will also gain experience at teamwork and at presenting your work professionally through the project and homework. The syllabus is continuously updated and subject to change.

Learning Objectives

The general/overall learning objectives are:

- Be able to explain the different types of ML methods and articulate why they are different and what types of problems each is aiming to solve
- Be able to implement any of the basic techniques in ML
- Select the ML solution best-suited for a novel domain and justify your choice
- Synthesize one area of machine learning in depth and apply it to a novel application
- Implement and evaluate the effectiveness of your ML method applied to a novel application
- Communicate ideas clearly to a variety of audiences both in oral and written form
- Function effectively in a team

The specific topics we will cover (not necessarily in this order) include:

Supervised Learning

- Classification and Regression
- Nearest neighbors and naive Bayes
- Linear models for classification and regression
- Neural networks and decision trees
- Support vector machines

Reinforcement Learning

- The RL problem
- Dynamic programming
- Temporal difference learning
- Function approximation

General Techniques/Concepts and Other Content

- Ensemble methods including bagging, random forests, and boosting methods
- Overfitting and regularization
- Model assessment, complex performance measures
- Clustering

Course Catalog Description

Topics include decision trees, relational learning, neural networks, Bayesian learning, reinforcement learning, multiple-instance learning, feature selection, learning appropriate representations, clustering, and kernel methods. No student may earn credit for both 4033 and 5033.

Schedule of Classes

The syllabus is continuously updated and subject to change. We will cover the material at a pace that is comforable. Our **first meeting** is on **Wed, Jan 17, 2024** and our **last meeting** is on **Fri, May 3, 2024**. A **rough outline** for the course, which is subject to change slightly depending on our pace, is:

Table 1: Tentative Lecture Schedule

Date	**Topics**
Week 1	Module 0: About this Course Module 1: Introduction to Reinforcement Learning
Week 2	Module 1 (cont'd): Introduction to Reinforcement Learning
Week 3	Module 2: Dynamic Programming Module 3: Model-Free Methods and Prediction
Week 4	Module 3 (cont'd): Model-Free Prediction Module 4: Model-Free Methods and Control
Week 5	Module 4 (cont'd): Model-Free Control Module 5: Value-Function Approximation
Week 6	Module 6: Introduction to Supervised Learning
Week 7	Module 7: Linear Models
Week 8	Module 7 (cont'd): Linear Models
Week 9	**Spring Break**
Week 10	Module 8: Model Selection, Regularization, and Model Assessment
Week 11	Module 9: Metrics Beyond Accuracy
Week 12	Module 10: Neural Networks
Week 13	Module 11: Decision Trees
Week 14	Module 12: Ensemble Learning
Week 15	Module 13: Support Vector Machines
Week 16	Module 14: Elements of Unsupervised Learning Advertisement for Computational Learning Theory

Project Milestones (Tentative Schedule)

Student Milestone	**Date**	**Instructor Milestone**
RL Project Proposal	End of week 3 Beginning of week 4	 Feedback on RL Project Proposal
RL Project Checkpoint	End of week 6 Beginning of week 7	 Feedback on RL Project Checkpoint
RL Project is Due	End of week 9 (end of spring break)	
SL Project Proposal	End of week 10 Beginning of week 11	 Feedback on SL Project Proposal
SL Project Checkpoint	End of week 13 Beginning of week 14	 Feedback on SL Project Checkpoint
SL Project is Due	End of week 16	

This class has no final exam even if it is scheduled (per the academic calendar) to take place during the finals week.

Textbook, Notes, and Related Reading Material

Required Material

The main references for our course are the following two books.

- *Reinforcement Learning: An Introduction* by Sutton and Barto [11]. The book is available online for free at http://incompleteideas.net/book/the-book-2nd.html.
- *Machine Learning* by Tom Mitchell [7]. The book is out of print. Despite its age, in my opinion, this is the best introductory book for traditional concepts of machine learning. Fortunately, the book is available online for free at http://www.cs.cmu.edu/~tom/mlbook.html. For a more modern treatment and insight into machine learning we will also be using the book called *An Introduction to Statistical Learning* that is mentioned below.
- *An Introduction to Statistical Learning (with Applications in Python)* by Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani [4]. The book is available online for free at https://www.statlearning.com.

Optional Material

Optional resources will be listed on Canvas, as well as on the class website:
`https://www.diochnos.com/teaching/CS4033-5033/2024S/index.php`.

Other Books of Interest

The above required books have all the information that you may want for this class (and even more than that)! This is why they are the required books for our class. Having said that, I have appreciated several books on machine learning throughout the years and I really like the exposition of certain topics in these other books.

- *Elements of Statistical Learning* by Hastie, Tibshirani, and Friedman [3]. The book is available online for free at http://web.stanford.edu/~hastie/ElemStatLearn/. This book used to be my primary reference for the course but I have now decided to substitute it with [4] which I believe is better suited for an introductory course in machine learning.
- *Mining of Massive Datasets* [6] by Jure Leskovec, Anand Rajaraman, and Jeffrey David Ullman. The book is available online for free at http://www.mmds.org.
- *Algorithms for Reinforcement Learning* [12] by Csaba Szepesvari. The book is available online for free at https://sites.ualberta.ca/~szepesva/rlbook.html.

If you appreciate theoretical justifications more, then the following two books have lots of technical explanations to some of the concepts that we will see in this class.

- *Foundations of Machine Learning* [8] by Mehryar Mohri, Afshin Rostamizadeh, and Ameet Talwalkar. The book is available for free at https://cs.nyu.edu/~mohri/mlbook/.
- *Understanding Machine Learning: From Theory to Algorithms* [10] by Shai Shalev-Shwartz and Shai Ben-David. The book is available for free at https://www.cs.huji.ac.il/w~shais/UnderstandingMachineLearning/.

A good book for linear models, which is however not available for free online, is *Learning From Data* [1] by Yaser S. Abu-Mostafa, Malik Magdon-Ismail, and Hsuan-Tien Lin. Another good book, which is also not available for free and is also out of print, is Tom Mitchell's book *Machine Learning* [7].

Not Technical but Interesting Books. Domingos in *The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World* [2], has a very nice (high-level) description, sometimes accompanied by historical anecdotes, on different aspects of machine learning. Valiant's book *Probably Approximately Correct: Nature's Algorithms for Learning and Prospering in a Complex World* [13] provides a good narrative for connections between evolution and learning. Finally, an unfortunate side of the ever-increasing influence that machine learning algorithms have in our daily lives is discussed in *Weapons of math destruction: How big data increases inequality and threatens democracy*, by Cathy O' Neil [9]. Along these lines the book by Michael Kearns and Aaron Roth, titled *The Ethical Algorithm: The Science of Socially Aware Algorithm Design* [5], is also highly recommended.

Grading

You will be learning and practicing many aspects of machine learning. What you get out of a course will depend on what you put into it! In order to give you a fair grade at the end of the semester, I will evaluate you using the following weighting. Grading will be based on the following:

- 50% homework problems,
- 50% semester-long project.

Grades may also be adjusted slightly upward or downward depending on class participation. Grading will be along the lines shown in the table below:

percentage	grade
$\geq 90\%$	A
$\geq 80\%$	B
$\geq 70\%$	C
$\geq 60\%$	D
otherwise	F

Grades may be curved at the end of the semester, but grade thresholds will never be higher than those shown above.

Undergraduates: Students taking 4033 will have shorter homework assignments and a smaller project. The grade thresholds are not required to be the same for the two courses.

Grade Questions: To maintain fairness in grading, the items should be brought to the person who graded it. To maintain fairness, all disagreements about the grading of projects should be brought to our attention within one week of when the item was returned.

Online Grade Summary: Canvas has a grade book that I will use to store all of your grades. It is your responsibility to verify that the grades on Canvas are correct. If an error is found, bring the document to me and I will correct Canvas.

Borderline Grades: Borderline final grades will be decided by your class participation which means that being an active participant in class can push you over a grade boundary.

Final Examination: Because this class contains a semester-long project, there will be no final examination.

Due Dates: To be fair to everyone and to minimize disruption to class, homeworks and projects are due at midnight on the day listed in the schedule.

Projects: Your final project will be due the last week of classes. Per university policy, you may turn this project in prior to pre-finals week if you have completed the project.

Homework Assignments

There will be 6 homework assignments. Your lowest homework grade will be dropped.

The contribution for your grade based on homework will be computed by adding up all the points that you receive from the individual homework assignments and then dividing by the maximum amount of points that you could gather from all these assignments. I expect the assignments to be weighted roughly evenly.

Programming Language. The programming language is Python. Your submissions are expected to run by using the command line on the main file of your submission using a recent Python interpreter (e.g., Python 3.10).

Same is true for your project; it has to be written in Python and that we can compile the main program of your submission using the command line without any issues.

Late Work Policy

You can postpone once your homework submission by 24 hours without any penalty. After the first time that you have a late submission, a 10% (of the maximum possible grade) penalty will be applied for every day that is late – the maximum delay can be 3 days (including the first time that you have a late submission).

We will be using an electronic system (Canvas) for the students' submissions and therefore it is your responsibility to turn in your homework on time. You are allowed to upload multiple copies of your work, so always make sure that you have submitted something.

Collaboration Policy

Collaboration is encouraged for homework and projects. For the projects, you will work within your groups. For the homework, you may form study groups so long as each homework is in your own words. Write your study partners' names on your homework when you turn it in. You may form a group of size two (including yourself) for the project and in one or two rare situations I may allow the formation of a group of three students working on the project, but this will be some last resort measure so that someone will not have to work alone.

Academic Misconduct

Academic misconduct hurts everyone but particularly the student who does not learn the material. All work submitted for an individual grade should be the work of that single individual and not his/her friends. It is fine to ask a fellow student for help as long as that help does not consist of copying any computer code, or solutions to other assignments. Furthermore, per the collaboration policy you need to list your study partners' names on your homework when you turn it in. Students working on joint projects may certainly help one another and are expected to share code within the project group. However, they may not share beyond the group.

General Remarks. Please note the following.

- Do not show another student (or group) a copy of your projects or homework before the submission deadline. The penalties for permitting your work to be copied are the same as the penalties for copying someone else's work.

- Make sure that your computer account is properly protected. Use a good password, and do not give your friends access to your account or your computer system. Do not leave printouts or thumb drives around a laboratory where others might access them.
- Upon the first documented occurrence of academic misconduct, I will report it to the Campus Judicial Coordinator. The procedure to be followed is documented in the University of Oklahoma Academic Misconduct Code[1]. In the unlikely event that I elect to admonish the student, the appeals process is described in http://www.ou.edu/integrity.
- **If you are unsure if something is permitted, consult with me before doing it.**

Project Code. Your project code and writeups must be written exclusively by you or your group. **Use of any downloaded code or code taken from a book (whether documented or undocumented) is considered academic misconduct and will be treated as such.** Exceptions from this policy (such as a project that builds on an existing open-source project) may be granted but you MUST speak with me first.

Chegg and Other Online Tutoring Sources

There are a wide variety of tutoring resources available through paid websites. Many of these sites have students upload assignments and solutions and surreptitiously provide these documents to other students. What appears to be a session with a tutor may be, behind the scenes, the tutor doing a search of their company database of solutions to share. By using these sites you risk being charged with academic misconduct, either by supplying other students with answers they did not author or by receiving someone else's answer that you did not author. Since these companies are not open with students about their practices, you cannot know whether a tutor is providing meaningful support (for example, identifying misunderstandings of content and explaining them) or simply feeding you someone else's solution a bit at a time. The tutor's actions can result in different students submitting answers that are identical, which may be flagged as academic misconduct during grading.

General Policies by the University of Oklahoma

OU is committed to creating a learning environment that meets the needs of its diverse student body. If you anticipate or experience any barriers to learning in this course, please feel welcome to discuss your concerns with me.

Classroom Conduct. Disruptions of class will not be permitted. Examples of disruptive behavior include:

- Allowing a cell phone or pager to repeatedly beep audibly.
- Playing music or computer games during class in such a way that they are visible or audible to other class members.
- Exhibiting erratic or irrational behavior.
- Behavior that distracts the class from the subject matter or discussion.
- Making physical or verbal threats to a faculty member, teaching assistant, or class member.
- Refusal to comply with faculty direction.

In the case of disruptive behavior, I may ask that you leave the classroom and may charge you with a violation of the Student Code of Responsibilities and Conduct.

[1] http://www.ou.edu/studentcode

Class Web Page and Canvas. The main web page for the class is

https://www.diochnos.com/teaching/CS4033-5033/2024S/index.php

Login to the Canvas website using your 4+4 (first four letters of your last name followed by the last four digits of your student number), using your standard OU password. If you have difficulty logging in, call 325-HELP. This software provides a number of useful features, including a list of assignments and announcements, an electronic mailing list, newsgroups, and grade book. All handouts are available from Canvas. You should check the site daily. When I update the site, I will post an announcement telling you what has been added and where it is located.You are responsible for things posted on the site with a 24 hour delay.

Class Evaluations. The College of Engineering utilizes student ratings as one of the bases for evaluating the teaching effectiveness of each of its faculty members. The results of these forms are important data used in the process of awarding tenure, making promotions, and giving salary increases. In addition, the faculty uses these forms to improve their own teaching effectiveness. The original request for the use of these forms came from students, and it is students who eventually benefit most from their use. Please take this task seriously and respond as honestly and precisely as possible, both to the machine-scored items and to the open-ended questions

Class Email Alias. Urgent announcements will be sent through email. It is your responsibility to:

- Have your university supplied email account properly forwarded to the location where you read email.
- Make sure that your email address in Canvas is correct, and forwards email to the place where you read it. I'll send out a test message during the first week of class. If you do not receive this message, it is your responsibility to get the problem resolved immediately.
- Have your email program set up properly so that replying to your email will work correctly the first time. You can send email to yourself and reply to yourself to test this.

If you need assistance in accomplishing any of these tasks, contact 325-HELP.

Newsgroups and Email. The newsgroup on Canvas should be the primary method of communication, outside of class. This allows everyone in the class to benefit from the answer to your question. If you email me a question of general interest, I may post your question and my answer to the newsgroup. Matters of personal interest should be directed to email instead of to the newsgroup, e.g. informing me of an extended personal illness. Posting guidelines for the newsgroup are available on Canvas.

Religious Holidays. It is the policy of the University to excuse the absences of students that result from religious observances and to provide without penalty for the rescheduling of examinations and additional required classwork that may fall on religious holidays.

Incompletes. The grade of I is intended for the rare circumstance when a student who has been successful in a class has an unexpected event occur shortly before the end of the class. I will not consider giving a student a grade of I unless the following three conditions have been met.

1. It is within two weeks of the end of the semester.
2. The student has a grade of C or better in the class.
3. The reason that the student cannot complete the class is properly documented and compelling.

Accommodation of Disabilities. The University of Oklahoma is committed to providing reasonable accommodation for all students with disabilities. Students with disabilities who require accommodations in this course are requested to speak with the professor as early in the semester as possible. Students with disabilities must be registered with the Office of Disability Services prior to receiving accommodations in this course. The Office of Disability Services is located in 730 College Ave, phone 405/325-3852 or TDD only 405/325-4173.

Adjustments for Pregnancy/Childbirth Related Issues. Should you need modifications or adjustments to your course requirements because of documented pregnancy-related or childbirth-related issues, please contact me as soon as possible to discuss. Generally, modifications will be made where medically necessary and similar in scope to accommodations based on temporary disability. Please see http://www.ou.edu/eoo/faqs/pregnancy-faqs.html for commonly asked questions.

Title IX Resources. For any concerns regarding gender-based discrimination, sexual harassment, sexual misconduct, stalking, or intimate partner violence, the University offers a variety of resources, including advocates on-call 24.7, counseling services, mutual no contact orders, scheduling adjustments and disciplinary sanctions against the perpetrator. Please contact the Sexual Misconduct Office 405-325-2215 (8-5) or the Sexual Assault Response Team 405- 615-0013 (24.7) to learn more or to report an incident.

Add/Drop/Withdrawal Deadlines. Please consult the OU academic calendar (as well as the policies of the School of Engineering) for the following deadlines:

- **Add a course**
- **Drop a course without penalty (course removed from transcript)**
- **Drop a course with a W on transcript**

Acknowledgements

I would like to thank Prof. Amy McGovern whose help has been immense for preparing a first iteration of the syllabus, as well as for preparing the class as a whole.

References

[1] Yaser S. Abu-Mostafa, Malik Magdon-Ismail, and Hsuan-Tien Lin. *Learning From Data*. AMLBook, 2012.

[2] Pedro Domingos. *The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World*. Basic Books, Inc., USA, 2018.

[3] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning*. Springer Series in Statistics. Springer New York Inc., New York, NY, USA, 2001.

[4] Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An Introduction to Statistical Learning: With Applications in R*. Springer Publishing Company, Incorporated, 2 edition, 2021.

[5] Michael Kearns and Aaron Roth. *The Ethical Algorithm: The Science of Socially Aware Algorithm Design*. Oxford University Press, Inc., USA, 2019.

[6] Jure Leskovec, Anand Rajaraman, and Jeffrey David Ullman. *Mining of Massive Datasets*. Cambridge University Press, 3 edition, 2020.

[7] Tom M. Mitchell. *Machine Learning*. McGraw Hill Series in Computer Science. McGraw-Hill, 1997.

[8] Mehryar Mohri, Afshin Rostamizadeh, and Ameet Talwalkar. *Foundations of Machine Learning*. Adaptive Computation and Machine Learning. MIT Press, Cambridge, MA, 2 edition, 2018.

[9] Cathy O'Neil. *Weapons of math destruction: How big data increases inequality and threatens democracy*. Broadway Books, 2017.

[10] Shai Shalev-Shwartz and Shai Ben-David. *Understanding Machine Learning: From Theory to Algorithms*. Cambridge University Press, USA, 2014.

[11] Richard S. Sutton and Andrew G. Barto. *Reinforcement Learning: An Introduction*. The MIT Press, second edition, 2018.

[12] Csaba Szepesvari. *Algorithms for Reinforcement Learning*. Morgan and Claypool Publishers, 2010.

[13] Leslie Valiant. *Probably Approximately Correct: Nature's Algorithms for Learning and Prospering in a Complex World*. Basic Books, Inc., New York, NY, USA, 2013.