Syllabus

This page will serve as the syllabus for the Spring 2024 version of CS 307. Please read it carefully. You should become familiar with these policies. To do so, you will likely need to return to the syllabus several times throughout the semester. After the start of the semester, this document may continue to be updated. Any such changes will be announced.

We recommend making a dedicated effort to read the syllabus a total of **five** times throughout the course:

1. At the beginning of the semester!
2. Shortly before **Quiz 01**
3. Shortly before **Quiz 02**
4. Shortly before **Quiz 03**
5. After receiving your final letter grade

Why do this? Because you're likely enrolled in several other courses, and recalling the syllabus details of each is difficult! However, at times, being fully aware of course policies can be a huge benefit. We believe that quick refreshers in advance of the quizzes will keep you well informed about the course details, such as letter grade breakdown, that you're likely to care about, but not think about much during the semester. The final read-through at the end of the semester is mostly to prompt you to specifically check the grading policies, and check that your letter grade was calculated correctly. More on that below.

Course Name and Number

- CS 307 - Modeling and Learning in Data Science
- Lecture Section: AL1
- Discussion Section: AL1

Location and Time

The Spring 2024 version of the course is **in-person**.

- **Lecture:** Wednesday and Friday, 12:30 PM - 01:45 PM, 2036 Campus Instructional Facility
- **Discussion:** Monday, 12:30 PM - 01:45 PM, 2036 Campus Instructional Facility

Course Staff

Please refer to the course staff by their given names. For example, your instructor is named Dave[1]. If you refer to the staff as "Professor" or "TA," we might refer to you as "student," which seems odd.

Instructor

- **David Dalpiaz** [Email] [Office Hours]

Teaching Assistants

- **Lahari Anne** [Email] [Office Hours]

Learning Objectives

After this course, students are expected to be able to:

- *Identify* supervised (regression and classification) and unsupervised (clustering) learning problems and their subtasks.
- *Understand* the bias-variance tradeoff and its relationship to model complexity, overfitting, and generalization.
- *Validate* and *select* machine learning models and their parameters using techniques such as cross-validation.
- *Prepare* and *process* data for use with machine learning methods.
- *Formulate* practical, real-world problems as machine learning problems.
- *Evaluate* effectiveness of machine learning methods when used as a tool for data analysis or as a component of a system.
- *Implement* simple machine learning methods from scratch using Python's `numpy`.
- *Apply* machine learning methods to real data using frameworks such as Python's `scikit-learn` and `pytorch`.

Course Content

Course Description

Course Catalog: Introduction to the use of classical approaches in data modeling and machine learning in the context of solving data-centric problems. A broad coverage of fundamental models is presented, including linear models, unsupervised learning, supervised learning, and deep learning. A significant emphasis is placed on the application of the models in Python and the interpretability of the results.

The above description is based on the Illinois Course Catalog. This version of the course may deviate slightly from this description. The course website will provide an overview of the course content and schedule.

Topics

Tentative subjects include:

- *Basics:* Supervised and Unsupervised Learning, Parametric versus Nonparametric Methods, Bias-Variance Trade-Off, Cross-Validation, No Free Lunch, Model Selection and Evaluation
- *Regression:* Linear Regression, Decision Trees, KNN
- *Classification:* Logistic Regression, Decision Trees, KNN, LDA, QDA, Naive Bayes
- *Extensions*: Regularization (Ridge, Lasso, Elastic Net), Ensemble Learning (Bagging, Boosting, Random Forests)
- *Unsupervised:* PCA, K-Means Clustering, Hierarchical Clustering, Mixture Models, EM Algorithm

The [Content](#) page will also list topics and their approximate ordering.

Textbooks

There is no required textbook for CS 307. Instead, course content will be distributed through a combination of lectures, notes, and additional (freely available) resources.

Prerequisites

The stated prerequisite for CS 307 is [STAT 207](#) and a linear algebra course, preferably one of MATH 225, MATH 227, MATH 257, MATH 415, MATH 416, or ASRM 406. Students will be expected to have experience with probability, statistics, and Python programming as taught in STAT 107 and STAT 207. Comparable experiences may be acceptable, but do consider speaking with an advisor if you find yourself in that situation.

Attendance

Attendance is encouraged but not required. Except for discussion (where attendance is strongly encourage, but again not mandatory), we will record and post lecture video. Attendance does not directly have an effect on your grade.

> **Do not come to class if you are sick, COVID, flu, RSV, or otherwise.** While it can be nice to meet in-person, again, it is *not required*. Attendance does not directly have an effect on your grade. Even if you're just a bit sick, say with a runny nose, stay home!
>
> - If you're sick, your need rest!
> - If you're sick, you risk getting the rest of the class sick.
>
> In life, you won't always have the ability and opportunity to stay home and rest when you are sick. In this course, that option is available and encouraged. Use it. Dave won't be offended, even if he has to lecture to an empty room to produce a recording.

Course Communication

We will use several forms of communication for this course. The website will be the one-stop-shop for all course information. Course announcements will be sent via email. Be sure you are regularly checking your `@illinois.edu` email account[2].

If you would like to communicate with the course staff, our preferred methods of communication, in order, are:

1. Office Hours
2. Discussion Forum (Ed)
3. Email

Email should largely be reserved for private matters. As much as possible, we would appreciate you asking questions about the course where we can respond so that other students benefit from your questions! It's cliche to say, but if you have a question, someone else is probably thinking it!

Office Hours

The office hours schedule can be found of the [home page](#).

Office hours are by far our preferred forum for discussing individual, specific questions. In office hours, our response time will be literally instant. Also, since we are both present in the same physical location (or together on Zoom), follow-up is both expected, and easy. Using asynchronous forms of communication such as the discussion forum or email will have a slower response rate and a much lower communication bandwidth. In other words, please come to office hours!

> When joining office hours, be prepared to share your screen, virtually or otherwise! Assuming you are having a technical or coding issue, the easiest way for us to assist is for you to **show** us the problem.

> We understand that sometimes asking a question in office hours can be intimidating. Students also find being asked questions, especially questions to which they do not know the answer, to be intimidating. Try your best to avoid these feelings! We are not asking you questions to make you feel bad. We are asking you questions to better understand how we can help. If we don't know what you don't know, we don't know what we need to teach you! If you ask Dave a question, there is a 99% chance he will respond with a question. Questions will often continue until you answer one incorrectly! The sooner you do, the sooner you learn!

Office hours will be a rather informal meeting. As such, if the instructor and a student are engaged in causal conversation not directly related to a pressing matter in CS 307, like a homework question, please just jump into the conversation and interrupt! If office hours are "busy" the instructor may institute an informal queuing system, but the hope is to keep office hours more relaxed and informal.

If you would like to schedule a private meeting outside of regular office hours, please send an email suggesting two possible times, on two different days.[3] We have a preference for time-slots directly adjacent to current office hours. *Please also indicate a brief agenda for the meeting.* Requests to schedule a meeting at a time less than 24 hours in the future are unlikely to be granted. Request without an agenda will be denied. Please give us at least some sense of what you would like to discuss.

Discussion Forum

This course will use **Ed** as our discussion forum.

> Ed access and login information was sent via email. If you registered late and did not see such an email, please be patient! We will send the login information several times, including after the last day to add the course.

Please register your account with your University email.[4]

The course staff will attempt to check Ed at least once a day during the week, thus you can often expect a response within 24 hours, except for weekends. If you need a quicker response, you should consider office hours as an alternative.

> The course staff would strongly prefer the use of Ed to GroupMe or similar services not officially supported by the course. The course staff feels that a GroupMe may exclude some students, whereas all are welcome on Ed. Services like GroupMe also exclude course staff, which while understandable, tends to promote an adversarial relationship between students and their instructors. We're all on the same team, so let's act like it.

Private posts have been disabled. Any private matters should be discussed over email where your identity is known. Some anonymous posting is disabled. You may post anonymously to your classmates, but not the course staff. The course staff will know the identity of all posters.

Additional Ed policy can be found in a pinned post on Ed.

Email Policy

CS 307 will follow a **strict** email policy. Instead of email, consider using the discussion forum! Any quick, non-private communication should take place there.

If you'd like to email the instructor or course staff, consider the following:

- Is your question about course administration? If so, **have you read the syllabus?** If your question is easily answered in the syllabus, we will either refer you to the syllabus, or ignore your email.
- Is your question about part of an assignment? First and foremost: You should ask it in office hours. After that, consider the discussion board. As a last resort, use email, but there is a good chance you will be re-directed to the discussion board.

If you choose to send an email, you must adhere to the following three rules. If you do not, your email will be considered less import than other emails which follow the rules and response time will be slower.

- All email must originate from an `@illinois.edu` email address.[5]
- Your subject line must begin with *exactly* the following: [CS 307]
- After the above, put a single space, followed by a **useful** but *short* description of your message.

```
## good
[CS 307] Grade feedback question


## bad
## improper format
## non-descriptive subject
[cs307] hi


## bad
## improper format
[CS307] Grade feedback question


## bad
## improper format
## subject too long
```

```
## information found in syllabus or website
[CS 307]when is the exam and what is covered on the exam?
```

If your email is sent between 9:00 AM Monday and 11:59 PM Thursday, *and you follow the above directions*, we will try our best to respond within 24 hours. Questions about an assessment sent the same day the assessment is due will likely not receive a response before the assessment is due. Plan accordingly.

Code Discussion

If your question is technical in nature, there are several steps you can take to insure a speedy response on Ed.

First and foremost, you should ask "Google"[6] before you ask the course staff. Take the error message you obtained and search it with Google. The ability to solve problems this way is an extremely valuable skill, possibly one of the most important you should learn (but are not taught) during your academic career. Make a legitimate effort to solve the problem on your own. You won't always be able to, and if you can't, post on Ed. (Or better yet, stop by office hours.)

If you need to ask the course staff, include the following in your discussion forum post:

- **All** code that is required to re-create the error.
- Staff should be able to run your code, *without any modification*, and obtain the same error or output.
- The exact error message received.

> Do not use screenshots of code and error messages to communicate about them. Copy-paste them so that others can copy-paste them as well. Posts containing screenshots of code will likely be deleted.

In this course, for everything expect quizzes, we greatly prefer over-sharing to under-sharing code. We would rather everyone learn from others' "mistakes" than have everyone experience the same issues over and over again. However, if you simply try to copy and paste other students' code to get through the homework, you will likely fail the quizzes. The course staff reserves the right to change this policy if we feel it is being abused.

Because this is our first semester assigning MPs, we will likely have discussions on the proper way to share and discuss code related to the MPs. But until we say otherwise, assume you should over-share rather than under-sharing.

Course Staff Emails

Role	Name	Email
Instructor	David Dalpiaz	dalpiaz2@illinois.edu
Teaching Assistant	Lahari Anne	lanne2@illinois.edu

Assessments

CS 307 will use four types of assessments: homework, labs, machine problems (MPs), and quizzes.

With the exception of quizzes, all course assignments are due at 11:59 PM, Central (Champaign) time, on the listed due date.

- Homework and MPs are due on **Tuesdays**.
- Labs are due on **Thursdays**.

In general, we will attempt to release assignments at least one week before their first deadline.

Homework

Throughout the semester, there will be a total of ten **homework** assignments, administered through the [PrairieLearn system](#).

> To access the course's PrairieLearn content, simply navigate to `prairielearn.org` and add CS 307 Spring 2024.

Additional information and instructions can be found on the homework policy page of the course website:

- [Homework Policy](#)

Labs

There will be a total of ten **labs** throughout the semester, submitted through a combination of [PrairieLearn](#) and [Canvas](#). Each lab will have two components, a *model* and a *report*. Additional information and instructions can be found on the lab policy page of the course website:

- [Lab Policy](#)

Machine Problems

There will be a total of two **machine problems (MPs)** throughout the semester, administered through [PrairieLearn](#). Additional information and instructions can be found on the MP policy page of the course website:

- [Machine Problem Policy](#)

Quizzes

There will be three quizzes throughout the semester taken at the [Computer-Based Testing Facility (CBTF)](#). Additional information (including dates and times) and instructions can be found on the quiz policy page of the course website:

- [Quiz Policy](#)

Deadlines

Except for the quizzes, all deadlines are at 11:59 PM, Champaign local time, on the listed day. All deadlines are listed on the [landing page](#) of this website.

> Homework 00, Lab 00, and MP 00 are **not** part of your grade. They are only for practice.

Course Technology

Use of [Python](#) is required to complete the course. [Visual Studio Code](#) will be our supported IDE, but alternative tools may be used as a substitute.

> A detailed guide for setting up your local machine can be found on the **Getting Started** page.

Learning Management

A mixture of Canvas, Ed, PrairieLearn, and PrairieTest will be used for Learning Management.

- [Ed](#) - Discussion Forum
- [PrairieLearn](#) - Homework, Lab (Models), MPs, Quizzes
- [PrairieTest](#) - Quiz Scheduling
- [Canvas](#) - Lab (Reports)

Grading

Assessment Weights

Assessment	Percentage
Homework	30
Lab (Models)	15
Lab (Reports)	5
MP	10
Quiz	40

Sub-scores for each category are the average of the assignments for that category. All assignments are equally weighted within a category. Grade information for each individual assignment can be found on the platform used to submit the assignment.[7]

> Homework 00, Lab 00, and MP 00 are **not** part of your grade. They are only for practice.

Because Quiz 01 will be your first experience in the CBTF for CS 307, we will allow the average of your Quiz 02 and Quiz 03 scores to **replace** your Quiz 01 score. In other words:

```
quiz01 = max(quiz01, (quiz02 + quiz03) / 2)
```

No other replacements will be allowed or considered.

> So how do **buffer points** work? First, the assignments that have the opportunity for buffer points are:
>
> - Homework
> - Lab (Models)
> - MP

Suppose there were only four homework assignments and a student obtained scores of:

```
[100, 100, 105, 105]
```

These average to 102.5, but with buffer points, sub-scores cannot exceed 100, so their Homework sub-score used for final grade calculations would be 100. Alternatively, suppose their scores were:

```
[95, 95, 105, 105]
```

Here, these average to 100, so their sub-score for Homework would be 100. This demonstrates that buffer points can help you get to 100 if you loose points on some assignments of the same type, but cannot move past 100 and help you make up for point loss on assignments of a different type.

We do not provide a "unified" gradebook and will not provide an instantaneous estimate of your grade during the semester. Why? First, we believe that as data science students, you are more than prepared to manage your own grade data! We'll give you the individual grades for each assignment, then you can use your data science skills to answer whatever questions you might ask. We do not provide an instantaneous estimate of your grade, because statistically, we feel that could be misleading. We are also unlikely to provide your overall percentage at the end of the course. Instead, we will give you all of your individual assignment grades, and your overall letter grade. Then, because you're data science students, and already created a Python function to calculate your grades (right?), you can perform that calculation and check our work!

Grading Scale

	A	B	C	D
Plus	99	87	77	67
Neutral	93	83	73	63
Minus	90	80	70	60

The instructor reserves the right to lower, but not raise, grade cutoffs. However, this policy should not create an expectation that this will happen. Asking for a change in cutoffs will make any change in cutoffs less likely. Grading in the course is not competitive. There is nothing (other than some statistical realities) that would prevent the entire class from receiving a grade of A.

Final letter grades will be posted to Canvas as soon as is reasonably possible.

Grade Disputes

If you feel an assignment was graded incorrectly, you have **one week** from the date you received a grade for the assignment to discuss it with the **instructor**. Do not bring grade disputes to any other course staff such as teaching assistants.

You may not simply ask for a re-grade, but instead must justify to the instructor why the grading was done incorrectly. By disputing any grading, you agree to allow the instructor to review the entire assessment in question for other errors missed during grading. Requests must be sent via email.[8] Grade disputes over trivial points will likely be met with frustration.[9]

After one week, grading is final except for exceptional circumstances.

All grade disputes must be approved by the course instructor. Teaching Assistants do not have authority to modify grades.

Academic Integrity

The official University of Illinois policy related to academic integrity can be found in Article 1, Part 4 of the Student Code. Section 1-402 in particular outlines behavior which is considered an infraction of academic integrity. These sections of the Student Code will be upheld in this course. Any violations will be dealt with in a swift, fair, and strict manner. **In short, do not cheat, it is not worth the risk.** *You are more likely to get caught than you believe.* If you think you may be operating in a gray area, you most likely are.

Under no circumstances should course materials be provided to Course Hero, Chegg, or any similar for-profit website. The course staff will seek the harshest possible academic integrity penalty for any students who do so.

Additional Information

Safety

The university values your safety. Please review the Run-Hide-Fight documentation provided by the Division of Public Safety.

Disability Accommodations

To obtain disability-related academic adjustments or auxiliary aids, students with disabilities must contact the course instructor and the Disability Resources and Educational Services (DRES) as soon as possible. To contact DRES, you may visit 1207 S. Oak St., Champaign, call 217-333-4603, email disability@illinois.edu or go to the DRES website.

To ensure appropriate accommodation is provided in a timely manner, please provide your Letter of Accommodation during the first week of class. Letters received after a relevant assessment has been administered will likely cause logistical issues that could result in an inability to accommodate.

The Extended Syllabus

For some thoughts on teaching philosophy, some explanation of policies, and some general tips for success, please see The Extended Syllabus.

Changes

The instructor reserves the right to make any changes he considers academically advisable. Such changes, if any, will be announced. Please note that it is your responsibility to keep track of the course proceedings.

Footnotes

1. I list David in the syllabus since that is my official name that you'll see in University systems, but please call me Dave. ↵
2. If you aren't already, you should get into the habit of checking your University email at least once a day. ↵
3. A total of four suggested times. ↵
4. Accounts registered with an email other than an `@illinois.edu` account will be removed. ↵
5. Depending on the situation, failure to follow this rule may make a response impossible. ↵
6. "Google" here refers to any search engine, and now, in the year 2024, a generative AI could also be used. ↵
7. PrairieLearn for everything except for Lab Reports which are submitted to Canvas. ↵
8. Failure to follow the email policy will result in your request being denied. ↵
9. A grade on a single assignment is not reflective of your overall grade in the course. The generous buffer points should more than make up for a single point deduction on a single assignment. ↵