

Welcome to CSC 294

Contents

- [Course Logistics](#)
- [Course Structure](#)
- [Resources Supporting CSC 294](#)
- [Academic Integrity](#)
- [Assessment](#)

I am so excited that you'll be joining this educational journey through the computational aspects of machine learning! You are now, just by being here, a machine learning explorer and practitioner. This course will help us frame out exactly what that means and also what *machine learning* is.

Note

Your are now a machine learning practitioner and explorer!

Official Course Description

An introduction to machine learning from a programming perspective. Students will develop an understanding of the basic machine learning concepts (including underfitting/overfitting, measures of model complexity, training/test set splitting, and cross-validation), but with an explicit focus on machine learning systems design (including evaluating algorithmic complexity and development of programming architecture) and on machine learning at scale. Principles of supervised and unsupervised learning will be demonstrated via an array of machine learning methods including decision trees, k-nearest neighbors, ensemble methods, and neural networks/deep learning as well as dimension reduction, clustering and recommender systems. Students will implement classic machine learning techniques, including gradient descent.

Motivating Questions

There are a few questions motivating our course work:

- What is Machine Learning?
- What role does computer science play in machine learning?
- What habits of mind do we need to develop to become machine learning practitioners?

Course Learning Objectives

By the end of the course, students will be able to...

1. Detail differences between supervised and unsupervised learning tasks and methods, as well as discuss the issues when dealing with large scale data
2. Implement a variety of machine learning algorithms in python and assess their efficacy
3. Compare models, and assess the efficacy of machine learning algorithms and results using evaluation metrics and in terms of the context of the data's domain
4. Develop an appreciation for ethical implications of machine learning algorithms
5. Work iteratively and reflectively to apply machine learning techniques to a data set of interest with informative documentation, written for a variety of audiences

Course Philosophy

This course is designed to be a first course in Machine Learning, but one that also prepares students to develop and implement their own machine learning pipelines using standard industry practices. Completing such a task is not as prescriptive nor as straightforward as completing a standard college course. There are turns and twists, failures and incremental progress, moments of confusion and bursts of clarity. Success in this course will be measured iteratively, placing emphasis on student's ownership of the learning process and a student's consistency of effort.

Stretch Zone

Borrowing language from the 2019 Google J-term course, when learning in a classroom setting, at any moment, we exist in one of three zones: safe, stretch, or strain. A student in the safe zone is completely comfortable, and the course material is completely within what the student already knows. The strain zone is the complete opposite of the safe zone. A student in the strain zone feels completely overwhelmed with the material to the point where they can't find any meaningful connections between the current material and what the student already knows.

The material in this course is intellectually challenging. Several of the topics are ones that took me a long time to fully understand their subtleties. The goal for this course is to spend the vast majority of our time in the stretch zone. In this zone, the material is just beyond our current knowledge base and while we are a bit uncomfortable with the material, we can see possible paths to tie the new material into our existing knowledge. In other words, in this class, our brains should be stretching, but not straining.

At the end of each week, if you have not been surprised or challenged, or if you have no questions about the material, this could mean that you are in the safe or strain zones. If you find yourself exasperated or negative about the class, you may be in the strain zone. If you find yourself using words like "boring" or "easy", you are likely in the safe zone. If you are in either zone, please get in touch with me: in class, on slack, in an appointment, or send me an email. I want you to get the most out of this course, but I cannot make adjustments to the course if I do not have firsthand knowledge of what is going on.

Syllabus

This webpage and the associated links together function as the syllabus. Everything that you need to know about the course is either here or linked from here.

The course syllabus is the most important document underlying the culture of our course and our classroom. I view my syllabus as a sacred document that both introduces and governs the course. In taking this view, I work to detail as much about the course as possible from the big picture ideas to the minute details of course policies. I believe that the first activity a student should do when beginning a course is to carefully read and examine the syllabus. In keeping with that belief, there are a few tasks in this document that will help us build our course community.

Note

Please read the syllabus this week and complete these tasks to help shape our community. If you have questions about the syllabus, please ask them!

This web version of the syllabus has a menu and search bar on the left, but the syllabus is also available as one [pdf document](#).

Course Acknowledgements

Parts of this course - including slides, activities, and notes - will be from a variety of sources. Materials will be appropriately attributed and will be used in keeping with copyright and fair use laws.

Course Logistics

Class Meeting Times

The course meets twice a week on Mondays and Wednesdays in Ford 342 from 10:50am to 12:05pm. In supporting our course community, students should make every effort to be at our course meetings on time.

Note

For accessibility reasons, we kindly ask that you refrain from wearing any scented products in class.

Attending Lecture

We are still in the midst of a global pandemic. In keeping with the [Guidelines](#) laid out in Smith's Culture of Care, if you are ill and/or have *any* COVID symptoms, please do **not** come to *in-person* class. Instead, please log in onto our gather site. The link can be found on our Moodle site and our slack space.

You do not need to email me to ask permission to come to class in gather or in person. However, if you are not able to be in-person for 3 consecutive meetings, then we need to check in.

Similarly, do not come to student hours nor appointments if you are ill and/or have *any* COVID symptoms. Please use the gather to attend student hours virtually, and a zoom link can be provided for any in-person appointment.

Warning

Do **NOT** come to in-person class or student hours, if you are ill and/or have any COVID symptoms.

Failure to respect this policy will result in an email to both the class dean and your advisor.

Pivoting Class Meetings

If I am unwell and/or experiencing any symptoms of COVID, class will be held on gather. If this happens, I will send a slack message on the #general channel using the @everyone mention. Using @everyone should send an email to your inbox.

If the college remains remote for the whole semester, the course should be able to remain largely unchanged. The structure of this course is similar to one that I used in Spring 2021 for *Computational Machine Learning*. That course was fully online, and students had positive comments on this format.

Communication

In addition to our synchronous meetings, our class will make of electronic communication, including our class slack, email, and Moodle messages. These methods of communication represent differing levels of formality and collaboration:

- **Slack:** Our slack site is the **primary form** of course communication. It allows for us to share where we are stuck with a reading, idea, or assignment, and where we can add helpful hints, request study groups, or share interesting news articles. Slack is much less formal than email, and salutations and signatures are not required. If you need to ask your instructor a question, this is the best place to do it either over a public channel or through direct message. **Please make a daily practice of checking slack as the primary form of communication for the course**
- **Email:** Email communication is more formal than slack. It should be used in the cases that 1) concern something personal including **accommodations forms**, 2) require attachments, or 3) involve a number of people (who should be copied on the email). We also will use email to confirm individual appointments. Emails to the instructor should include a salutation and a signature.
- **Slack Mentions:** When I need to communicate with the whole class on a time-sensitive manner, I will use the @everybody or @channel functionality in slack. Please check that your slack settings will notify you either via slack or email. If you are using the email notifications, please be sure that your email does not treat these messages as spam.
- **Slack Emojis:** At times, I will ask you to emoji a message to show that you have read the message. This is a quick way for you to signal that you read it. If there's something that you have a question about in a "please emoji" message, please use the threading function for that message or send me a direct message on slack.

Note

Please sign up for [our slack site](#), add a profile picture, and introduce yourself with your name and a fun-fact about yourself on the #intros channel. After signing up, please determine how and where you will get notifications from slack.

Please note that on a typical workday, I leave my office just after 4:00pm. This means that emails and slack messages sent after 4:00pm will likely be received on the next business day. Similarly e-communication sent on the weekend will likely not be received until Monday.

Course Resources

Machine learning is a rapidly evolving field, and a course focusing on computation (instead of theory) is less common. So, our course will gather together a number of readings and resources from different sources. The below list of course materials are the minimum that I believe that you need to be successful in the course. If you feel that there is something critical missing from this list, please let me know.

- Variety of book chapters noted on the detailed course schedule and available at the library
- Bound notebook for notes, ideas, and scratch work
- Highlighters in at least 2 colors
- Pens or pencils in at least 3 colors
- Recommended Texts:
 - [Python for Data Analysis](#) by McKinney
 - [Python Machine Learning](#) by Raschka and Mirjalili

Note

Both of our recommended texts are available at the library as E-Books. Please note that if you are viewing the sources online, you may be limiting another person from using it. Instead, you can download sections (or whole books) for a period of time directly from the library.

Instructor Information

The instructor for this course is Katherine Kinnaird. My office is room 218 Bass Hall. My email is kkinnaird@smith.edu, but the best way to reach me is on slack.

Student hours

My student hours are time blocked in my calendar for you! Please drop by either in person or virtually.

Note

Student hours are:

- Wednesdays from 3pm to 4pm
- Thursdays from 11am to noon
- By appointment

Below is an incomplete list of great reasons to hangout in gather before or after class or to make an appointment:

- You haven't had the chance to chat with me yet
- You want to see what my home office (ie. my basement) looks like without having to split your focus with learning
- You have a question on a reading, assignment, or activity
- You heard that that my office mascot is a sheep
- You want to share about the course, your time at Smith, or about yourself
- You read something about machine learning, computer science, statistics, or math that you have questions about
- You are thinking about what you want to do after Smith

Appointments

You are welcome and encouraged to make individual appointments with me. To make an appointment, please check my [appointment calendar](#) for appointment slots that are set aside each week.

Note

To book an appointment with me, go to my [appointment calendar](#)

Course Structure

Each week will be a mix of reading about machine learning topics, discussing and debating concepts, implementing algorithms and ideas in code, and constant reflection about our learning and professional development. All best by dates are posted on the Detailed Course Schedule.

Note

The [Detailed Course Schedule](#) lists the best by dates for the course. To view it, please log into your **Smith** account.

Assignments

The work in this course can be broadly broken down into four types that build on each other:

- **Reading and Prep work for class** - This is your first contact with the material. Just like meeting a person for the first time, you should pay attention to the big picture pieces, but not fuss over learning each detail perfectly. Instead, note questions you want to ask in your engagement journal the next time you come into contact with the material
- **In class Labs** - The course is designed to be largely hands on. Each class meeting will be mostly working on labs and talking about the labs. The labs are in jupyter notebooks so that we can play with small pieces of code that don't impact other pieces, take tons of notes, and make mistakes! This is our second contact with the material, and just like meeting a person for the second time, you should have follow-up questions about the material
- **Homework** - Once we've read about the material and worked on highly structured labs, it is time to practice deploying concepts from end to end in the usual python setting (i.e. not in a notebook). Homework is one step beyond the labs in difficulty because only the last step (i.e. the result) is asked for. You might want to refer to your labs to do these assignments and note the steps that you got hung up on.
- **Projects** - After three different interactions with each concept, it is time to start experimenting with mixing them and adding a bit of creativity. The projects in this course are very open-ended. The primary goal of these projects is to take the concepts from class and do something that stretches your knowledge.

Engagement Journal

As part of this course, you are being asked to keep an engagement journal on Google Drive. This engagement journal aims to be a tether point to bring together the ideas that we encounter in our readings and in our labs. Check out this [google doc](#) to start your engagement journal

You should use these journals in a way that makes sense to you. For some, that means doing one entry at the end of the week. For others, that means adding a little after each reading and activity. For some, this means deleting the questions that you don't answer. For others, this might mean answering every question. Remember that the ultimate goal of the engagement journal is to help you engage with the course. You may have to experiment with how you do this, so that you get the maximal benefit from this tool.

Labs

The labs are designed to be completed in class with some work out of class to practice the ideas from our readings. The labs pull together a number of resources and require careful reading of the labs in addition to the code blocks. In addition to the interactive jupyter notebook format, a pdf version of each lab will be provided so that one can write notes by hand (or with a digital pen) on the labs directly.

Homework

Homework assignments are designed to stretch your new machine learning knowledge. There are six homework assignments throughout the semester, and they pull together ideas from a number of the labs.

Projects

There are 4 projects for the semester. The primary goal of these projects is for you to apply the concepts from this class to a topic that interests you. These are the most creative part of the course and as such, these projects have the least structure. It is highly encouraged that you start these projects early.

Weekly flow for the course

Each week, you will have readings to prepare for class as well as two labs. Spread over the semester are six homework assignments and two projects. The course will operate as follows:

Day	Course Preparation and/or Activity	Best by Date
Monday	Lab in class	Lab from previous Wednesday
Wednesday	Lab in class	Lab from Monday
Friday	(no class)	Homework or Project

Note

The [Detailed Course Schedule](#) lists the best by dates for the course. To view it, please log into your **Smith** account.

Workload

According to federal standards, each four-credit course should equate to at least 180 hours of work over the semester. If you are taking 16 credits, that equates to 720 hours of work over the 15 weeks of the semester, from the first day of classes until the end of the final exam period. In the case of this course, you will spend nearly 12 hours each week on my class alone (including our three hours of class meetings per week).^[1]

In considering the work for this course, I believe that the approximate 9 hours per week *outside* of class will breakdown something along these lines:

- 2 hours of class prep – 90 minutes per class meeting reading the book and prep activities
- 2 hours per week of lab wrap up
- 3 hours per week on homework and/or projects
- 2 hours of “flex time” used to supplement any of the above areas and/or go to student hours

Notice that there are 2 hours of flex time. This is to accommodate weeks were you might want to spend more time on an assignment. For example, you might spend extra time working on a project in the weeks leading up to its best by date.

If you find that the time you are spending on this class is a lot more than 9 hours per week *outside* of class or a lot less, let’s check in.

Flexible Accountability

There are only two firm due dates for this course: the last day of classes and the last day of finals. Instead of firm deadlines throughout the semester, we will be using “best by dates.” These best by dates are my recommendation of when you should aim to complete work to continue making progress in the course. If you work best with firm deadlines, then you can treat these best by dates as due dates.

I will be marking assignments in batches as enough work is submitted for each assignment. I will communicate on slack when work has been returned.

Mid-semester Conferences

The week after spring break, we will have mid-semester conferences. I will be meeting with each student for 10 minutes to discuss their progress in the course. Please sign-up for these meetings using the link on the detailed course schedule.

Two firm deadlines

The only two firm deadlines of the course are the last day of classes and the last day of finals:

- **Due on the Last day of classes** – All labs, all homework assignments, and the first two projects
- **Due on the last day of finals** – The last two projects and your engagement journal

There are no extensions to these two deadlines without intervention from a class dean.

^[1] Each of our 75 minute meetings is “counted” as 90 minutes. Since we meet 2 times per week, our “in-class contact time” is counted as 3 hours.

Resources Supporting CSC 294

Here is a list of additional resources supporting our class. Information about your instructor can be found on the [logistics page](#).

Inclusivity for all students

Smith is committed to providing support services and reasonable accommodations to all students with disabilities. Please inform me early in the term if there are aspects of the course that need to be modified to serve your learning, health, or person. You may speak with me after class, during student hours, or during an appointment. To request an accommodation, please register with the Disability Services Office at the beginning of the semester. To begin the process, either call 413-585-2071 or email ods@smith.edu.

Students in need of short-term academic advice or support can contact your class dean in the Dean of the College office. If you require some accommodation for religious or cultural purposes, please do not hesitate to let me know.

Spinelli Center

The Spinelli Center for Quantitative Learning is a resource to support students as they take STEM courses and as they engage in other quantitative work. To see the list of available peer tutors and other resources available at the Spinelli Center, please consult their [website](#).

Jacobson Center

Smith has an additional resource for writing support: the Jacobson Center for Writing, Teaching & Learning make an appointment to take your work to the Jacobson Center on [their website](#). In particular, you may choose to bring your work to Peer Writing Tutors Elisabeth Nesmith or Elina Gordon-Halpern, both SDS majors who tutor for the Jacobson Center. Contact Sara Eddy (seddy@smith.edu) for more information about their schedules or how to make an appointment.

The above language about the Jacobson Center was adapted from language from Sara Eddy. Used with permission

Library

Both of our recommended texts are available at the library as E-Books. They can be found at the following links:

- [Python for Data Analysis](#) by McKinney
- [Python Machine Learning](#) by Raschka and Mirjalili

Note

Please note that for anything denoted as an E-Book at the library, there are limits placed on how many people can view the source. Instead, you can download sections (or whole books) for a period of time directly from the library.

Pedagogical Partner

Kathleen is my pedagogical partner this semester. She took CSC294 last spring and will be helping to redesign some of the labs and the final project. If you would like to provide feedback regarding course content or structure to a fellow student (and anonymously if you wish), please reach out to Kathleen (khablutzel@smith.edu or on slack).

Academic Integrity

Nothing is held more dear to a researcher than their integrity. To researchers, the most precious commodity is our ideas. While some researchers build tangible items that can be protected by patents, many researchers' products are ideas and theories, which are less likely to be protected under things like patents. Instead, it is academic integrity that governs our research communities, protecting our ideas from being stolen. The system of academic integrity relies on researchers trusting each other to be honest and to act with integrity.

Honor Code

Being at a school with an Honor Code, like Smith, is a special privilege. The Honor Code goes beyond inviting students to act with integrity, instead it welcomes students as equal participants of the learning community, imbuing students with the same level of trust that we extend to our collaborators and colleagues. The Smith College Honor Code, established in 1944, and as stated in the Student Handbook, says:

Smith College expects all students to be honest and committed to the principles of academic and intellectual integrity in their preparation and submission of course work and examinations. Students and faculty at Smith are part of an academic community defined by its commitment to scholarship, which depends on scrupulous and attentive acknowledgement of all sources of information, and honest and respectful use of college resources.

This trust bestowed to students in the Honor Code is the same trust that exists in and among researchers within a research community, the same trust that exists between me and my collaborators. Simply put, at Smith, under the Honor Code, I trust that you will each act with integrity, citing sources when you celebrate others' ideas and noting who you work with when collaborating.

This trust manifests in how assignments are created, in how work is completed, and in how we resolve instances where the trust has been broken. At an Honor Code school, assignments are created knowing that while there are resources online that can offer complete solutions, you are trusted to not seek out such complete resources and that if you do stumble on a solution guide, you will not use it. Under an Honor Code, students are expected to keep careful notes about the resources that they consult and the people that they collaborate with. We get to assume that the work handed in by a student is the creation of that student. Lastly, when there are violations of the Honor Code, the resolution is determined by a committee established by the community and trusted to seek restoration of the whole community's trust through education and action.

I regard the Honor Code with deep and profound respect. Being an educator at an Honor Code school means we begin from a place of trust without any underlying suspicion of our students. Simply put, we - instructors and students - work from the assumption that all are acting in good faith and with the utmost integrity. This is an assumption that cannot be made at a school without an Honor code, and it is why it is a privilege to be both an instructor and a student at Smith under the Honor Code.

Honor Code Practicalities

The work you submit should be your own and created by you, unless explicitly listed as a group assignment. With the exception of the starred problems, I do encourage you to ask for help from your peers or myself when you have questions; however, copying is never allowed. The line between copying and helping is subtle. Below are a few guidelines:

- Do not share nor give your work with other students; instead, offer to discuss the big ideas of the task at hand.
- Do not look at someone else's work (including online solutions); instead, ask if you could talk with them about your ideas and share where you are getting stuck.
- Acknowledge those you talk or work with at the top of every assignment. I will not dock points for getting acknowledged help from others. If you generate a solution or an argument with someone, in addition to acknowledging that person, recreate the solution or argument on your own in your own words.

If you think you may have crossed the line between helping and copying, please talk to me. Do not let me discover that the line was crossed. If a violation of the Honor Code is suspected, the student will be informed and will be given the opportunity to meet with the instructor. As recommended by the Academic Honor Board and in keeping with Smith tradition, the student will be given time to self-report, and after such time, the suspected violation will be reported to the Academic Honor Board by the instructor.

If you are unsure about how Smith's Honor Code applies to our course, please consult with me or a class dean. We are happy to discuss the Honor Code with you.

Creation vs. Curation

In my role as your instructor, I am **curating** resources for you that I feel are best for the teaching and learning of machine learning. In this process of curation, I will give credit to those whose materials I have used and I will only use materials as allowable by copyright and fair use.

In this course, you will be asked to **create** solutions for homework assignments and projects. In these acts of creation, you are demonstrating your most current understanding of the concepts in our course. This course will give you lots of chances to demonstrate what you know and lots of flexibility.

I am aware that there is a tension between the curation that I am modeling and the creation that is being asked of you. My goal is to provide the best resources to you, which requires reading and sorting through many sources (curation). Your goal is to learn as much machine learning as deeply as possible, which requires a lot of practice, trying out machine learning ideas, reflecting on your attempts, and trying again (creation).

Assessment

Just like beginning a project, we begin this course with nothing completed yet; similarly, we begin the semester with no points, just yet. As assignments are assigned and returned we will slowly build points towards the 1000 possible points. In keeping with this philosophy, each assignment adds to your total score. In more traditional courses, one might begin with 100% and have their grade slowly chipped down to a final score. Instead in this course, we are building our grades up in concert with our knowledge about machine learning. View each assignment as a reflection of the depth of your machine learning knowledge base.

Grading Practicalities

There will be three kinds of grading for this course: 1) rubrics (supported by the checkpoint system), 2) on completion, and 3) with student input.

1. Homework assignments and projects will be graded under typical rubric grading. Depending on the assignment, detailed rubrics may be issued before the assignment is submitted, while in other cases only general outlines will be given in advance of the assignment.
2. Labs will be graded on completion. Note that completion here means that the work was completed in good faith. The act of turning in something does not guarantee a score of completion. But any work that is clearly done in good faith will be marked as complete and will earn full marks.
3. The class engagement part of the course will be graded with student input.

There are **four** parts of your grade: 1) labs, 2) homework, 3) projects, and 4) course engagement. The 1000 possible points for the course are distributed as follows:

- **Labs - 240 points:** Each lab is worth 12 points and there are 25 labs for a total of 300 possible points.
- **Homework - 260 points:** Each homework is worth 52 points and there are 6 homework assignments for a total of 312 possible points.
- **Projects - 300 points:** Each project is worth 100 points and there are 4 projects for a total of 400 possible points.
- **Course Engagement - 200 points:** This part of your grade with student input. After your mid-semester conference, you will reflect on your engagement with the course so far and make a plan for the remainder of the semester. That plan is what you will then use to assign yourself a grade for the course engagement part of your grade.

Flexibility in grading structure

Note

For the lab, homework, and project parts of your grade, there are many more possible points than what the final grade computation uses

To help you keep track of your grade, there is a [grading matrix](#) on google docs.

Translating to the “real world”

These systems attempt to match my experience of being in the “real world.” Assignments under the first system (ie. labs, homework assignments, and projects), you need to do about 80% of them well to get full credit for 100% in that portion of the grade. This matches my daily life where I can physically get to about 80% of my to-do list. Under the second system, it is often easier to edit a draft than to start from a blank page, and having small deadlines (where the reward is quick feedback) gives structure to larger projects.

While I do believe that each assignment for this course is important, I also believe that the college structure does not always allow you to practice “real world” decisions such as balancing your health with your work. So while I will not tell you to hand in an assignment or to not hand one in, I would love to listen to you practicing making these decisions.

Latework

Instead of firm deadlines throughout the semester, we will be using “best by dates.” These best by dates are my recommendation of when you should aim to complete work to continue making progress in the course. If you work best with firm deadlines, then you can treat these best by dates as due dates.

All of these best by dates are listed on the [Detailed Course Schedule](#). This is a great moment to note these dates your system for keeping track of your work (ie. planner, calendar, etc). (If you want to really stretch here, do this for all major assignments in all your courses.)

In general, later course content builds on previous content, and if one is spending lots of time on an assignment (and thus is well past the recommended “best by” date), then there is less time for the next assignment. If you find yourself routinely working on assignments well past their best by dates, let’s check in.

The only two firm deadlines of the course are the last day of classes and the last day of finals:

- **Due on the Last day of classes** – All labs, all homework assignments, and the first two projects
- **Due on the last day of finals** – The last two projects and your engagement journal

There are no extensions to these two deadlines without intervention from a class dean.

Letter grades:

Traditional letter grades will be assigned using the below listed minimum points:

Letter Grade	Number of points
A	950 - 1000 pts
A-	900 - 949 pts
B	800 - 899 pts
C	700 - 799 pts
D	600 - 699 pts

At the instructor’s discretion, these grade cutoffs may be lowered at the end of the semester, but they will never be raised. Cut-offs not stated here for the pluses and minuses will be determined at the end of the semester.