Syllabus (2025 W1)

2025-10-14

Table of contents

Course Info . . . 1
Course Objectives . . . 2
    Learning Outcomes . . . 2
Textbooks . . . 3
Lectures . . . 3
Course Assessment Opportunities . . . 4
    Effort-Based Component . . . 4
    Summative Assessment . . . 9
Health Issues and Considerations . . . 9
University policies . . . 10
    Academic honesty and standards . . . 10
    Academic Concessions . . . 11
    Missed final exam . . . 11
    Take care of yourself . . . 11

Course Info

Dates:
Term 2025 W1, 04 Sep - 04 Dec 2025

Instructor:

Geoff Pleiss
Website: https://geoffpleiss.com/
Email: geoff.pleiss@stat.ubc.ca
Slack: @geoff

Office hours:
See Canvas for times and locations.

Course webpage:
WWW: https://ubc-stat.github.io/stat-406/
GitHub: https://github.com/stat-406-2025/
Canvas: https://canvas.ubc.ca/courses/147492/

Lectures/Labs:
See Canvas for times and locations.

Textbooks:
[ISLR] [ESL]

Prerequisite:
STAT 306 or CPSC 340

Course Objectives

This course introduces machine learning from a statistical perspective. Beginning with linear models and building beyond STAT 306 content, we will progress to an in-depth coverage of classic and modern learning methods with greater mathematical and statistical depth than CPSC 340.

Learning Outcomes

The primary aim of this course is to develop deep statistical intuitions about different learning methods and the connections between these methods. If we are successful, by the end of the course you will be able to:

1. Formulate and analyze machine learning algorithms from a statistical perspective
2. Categorize learning methods through multiple criteria (supervised versus unsupervised, linear versus nonlinear, parametric versus nonparametric, high bias versus high variance)
3. Draw connections between any two learning algorithms and paradigms covered in this course
4. Reason about which methods (regularization, nonlinearities, ensembles) are most appropriate for a given situation using mathematical tools like the bias/variance tradeoff and curse of dimensionality
5. Apply these methods correctly, troubleshoot common pitfalls, and identify probable steps for improving model performance
6. Implement and utilize models in R using best coding practices

What This Course Is Not

This course is not a practitioner's introduction to machine learning. While you will gain practical skills in the course and you will be equipped to run and troubleshoot models, the emphasis is not on the latest and greatest models, software packages, using high performance compute infrastructure, or data wrangling. There is no project component where you get to build a model of your choice on your own dataset. If this is the content you are looking for, I suggest CPSC 330 or any number of tutorials that exist on the internet.

This course is also not a comprehensive introduction to all of the latest and greatest ML methods. Methods evolve constantly, and there are too many of them to introduce in one course. I suggest advanced topics courses in computer science or statistics if you want a flavour of new methods. The foundational methods introduced in this course are building blocks used by all machine learning algorithms. Mastering these concepts will dramatically improve your ability to understand new methods, and it will also help you cut through the BS components of modern ML/AI systems that don't actually do anything too fancy.

Building fluency with tools and processes is an important part of becoming a statistician or machine learning practitioner, as is being up to date with methods. It is equally important to be able to analyze and reason about the methods and tools that you're using. The latter is the niche filled by this course.

Textbooks

There are short required readings before each lecture. You will not be quizzed or graded on the readings, but going over the material before seeing it in lecture will accelerate your learning of the material.

Required

An Introduction to Statistical Learning, James, Witten, Hastie, Tibshirani, 2013, Springer, New York. (denoted [ISLR])

Available **free** online: https://www.statlearning.com

Optional (But Excellent)

The Elements of Statistical Learning, Hastie, Tibshirani, Friedman, 2009, Second Edition, Springer, New York. (denoted [ESL])

Also available **free** online: https://web.stanford.edu/~hastie/ElemStatLearn/

This second book is a more advanced treatment of a superset of the topics we will cover. All readings from [ESL] are optional.

Lectures

This will be taught primarily via lecture, with interactive activities, questions, and discussions interspersed throughout. **Attendance is strongly encouraged.**

Developing intuitions for this material takes time and effort, and the lectures provide an excellent setting and opportunity for you to ponder, struggle, and engage with the material.

When you attend lecture, I ask that you minimize the use of phones and computers other than for participation and note taking. I know that it is easy to be distracted by homework, social media, or k-pop videos.[^The lecture recordings capture the audience, so I get to see the fun things you're looking at on your laptops!]

If you aren't able to provide your full attention during class, especially during the participatory activities, I kindly ask that you stay home instead. Your grade won't suffer (I don't track attendance) and your health and academic performance will benefit from the extra sleep.

Here are the benefits for being participatory during lecture:

- **You'll put in the work to develop deep intuitions.** We'll work through derivations during lecture that you won't see on labs or assignments. Participating in these derivations gives you a different mode for absorbing the material.

- **You get to ask questions.** This material is my bread and butter, it's my area of research, and I have lots of insights that I'm excited to share! I'll also answer questions over slack and during office hours, but your best bet to get good answers from me is during class.
- **You'll get participation points.** See below for a way to boost your grade by participating during lecture.
- **You'll get a better letter of recommendation.** If you want to apply to graduate school and would consider me for a letter of recommendation, I should know something about you beyond your letter grade. Asking and answering questions during lecture and office hours gives me good material to write about.

Course Assessment Opportunities

Effort-Based Component

The effort-based assessments are your main tools to develop the deep intuitions about ML that we're trying to achieve in this course. They will ask you to implement and troubleshoot models, wrestle with complex mathematical ideas, and ponder strange phenomena. Think of these assessments like eating your vegetables, doing a workout, or practicing an instrument. The work is hard, but it builds a deeper understanding about statistics and ML methods that you can't get through other means—helping you learn to formulate algorithms from a statistical perspective, connect different learning methods, and figure out which approaches work best in different situations.

Component	Possible Points
Participation	5
Clickers	10
Labs	20
Homeworks	40
Total	`min(60, Participation + Clickers + Labs + Homeworks)`

Generative AI Policy

Chatbots and agentic tools are quite adept with the content that we'll cover. I used them to help me develop materials for this course (mostly to flesh out/polish ideas rather than to generate content from scratch). I believe they have the potential to aid in your learning, but also the potential to be a tempting shortcut that inhibits your potential to grow from this course.

To get the most out of this course, I **strongly** recommend using generative AI as little as possible. If you feel the need to use it, use it to supplement and strengthen your thinking (e.g. code completion, rewriting bullet points into paragraphs, etc.) rather than to replace your thinking (e.g. feeding the entire assignment into a chatbot or agentic coding tool). Here's why:

- The goal of this course is developing deep intuitions about challenging material. Developing intuitions requires thinking, meditating, and struggling with content.
- Anyone can use Claude Code, Copilot, or Cursor to develop simple ML models and pipelines. If you want to stand out as a job or graduate student candidate, you need to demonstrate the ability to think beyond what these tools can do.

- There are enough questions with course-specific subtleties where a chatbot will produce a subpar or wrong answer. Even with really good prompting, you likely won't score above a 7 or 8 on assignments (which is the median effort grade anyways).
- You won't be able to use any AI tools during the midterm or final. There likely won't be any coding questions, but you don't want to be too dependent on these tools when it comes to proving your knowledge.

If you use any generative AI tool on homeworks or labs, I ask that you self report how you used the tools and how they contributed to your learning/ability to understand the assignment. (There will be a box to fill in at the top of every lab and assignment.) Include every prompt that you used. Please be as honest as possible, even if you did just get ChatGPT to answer every question for you. Your responses will help me structure the material better especially as these tools evolve.

Participation

I believe that active participation during lecture is the best way to learn the material. To that end, I've crafted lectures to be entertaining (especially for 8am!), interactive, and engaging for all learning styles. You can help me (and yourself!) by asking questions, working through course exercises with your neighbours, and sharing insights.

You earn participation points by answering questions or contributing to discussions. I'll use the Agora platform to monitor hand raises and select students. Every time you raise your hand in the Agora app, you earn 1 point. When you're randomly called on from raised hands, you also gain a point. Your final participation score is calculated as `min(5, 5 * num_points / num_classes_with_agora)`. In essence, if you raise your hand at least once every class, you'll receive the maximum participation grade.

Note on attendance. I will not be tracking attendance, so you do not need to alert me if you need to skip the occasional lecture. You can still get a perfect effort-based grade even if you do not participate (though I sincerely hope you choose to do so!)

Clickers

Throughout lecture, I will ask short multiple choice and True/False questions, which you will answer using the iClicker app. For each question, correct answers are worth `4`, incorrect answers are worth `2`. You get `0` points for not answering.

Suppose there are `N` total clicker questions, and you have `x` points. Your final score for the clicker component is `max(0, min(5 * x / N - 5, 10))`.

Note that if your average is less than 1, you get 0 points in this component.

Labs

These are intended to keep you on track. They are to be submitted via pull requests in your personal `labs-<username>` repo (see the computing tab for descriptions on how to do this).

Labs typically have a few questions for you to answer or code to implement. These are designed to be done *during lab* periods, but you can do them on your own as well. These are worth 2 points each up to a maximum of 20 points. They are due at 2300 on the day of your assigned lab section.

If you attend lab, you may share a submission with another student (with acknowledgement on the PR). If you do not attend lab, you must work on your own (subject to the collaboration instructions for Assignments below).

Rules.

You must submit via PR by the deadline. Your PR must include **at least 3 commits**. After lab 2, failure to include at least 3 commits will result in a maximum score of 1.

> Tip
>
> If you attend your lab section, you may work in pairs, submitting a single document to one of your Repos. Be sure to put both names on the document, and mention the collaboration on your PR. You still have until 11pm to submit.

Marking.

The overriding theme here is "if you put in the effort, you'll get all the points." Grading scheme:

- 2 if basically all correct
- 1 if complete but with some major errors, or mostly complete and mostly correct
- 0 otherwise

You may submit as many labs as you wish up to 20 total points. **There are no appeals on grades.**

Homeworks

There will be 4 homework assignments. These are submitted via pull request similar to the labs but to the `homework-<username>` repo. Each assignment is worth up to 10 points. They are due by 2300 on the deadline. You must make **at least 5 commits**. Failure to have at least 5 commits will result in a 25% deduction on HW1 and a 50% deduction thereafter. **No exceptions.**

Assignments are typically lightly marked. The median last year was 8/10. But they are not easy. Nor are they short. They often involve a combination of coding, writing, description, and production of statistical graphics.

After receiving a mark and feedback, if you score less than 7, you may make corrections to bring your total to 7. This means, if you fix everything that you did wrong, you get 7. Not 10. The

revision must be submitted within 1 week of getting your mark. Only 1 revision per assignment. The TA decision is final. Note that the TAs will only regrade parts you missed, but if you somehow make it worse, they can deduct more points.

The revision allowance applies only if you got 3 or more points of "content" deductions. If you missed 3 points for content and 2 more for "penalties" (like insufficient commits, code that runs off the side of the page, etc), then you are ineligible.

Policy on collaboration on assignments

Discussing assignments with your classmates is allowed and encouraged, but it is important that every student get practice working on these problems. This means that **all the work you turn in must be your own**. The general policy on homework collaboration is:

1. You must first make a serious effort to solve the problem.
2. If you are stuck after doing so, you may ask for help from another student. You may discuss strategies to solve the problem, but you may not look at their code, nor may they spell out the solution to you step-by-step.
3. Once you have gotten help, you must write your own solution individually. You must disclose, in your GitHub pull request, the names of anyone from whom you got help.
4. This also applies in reverse: if someone approaches you for help, you must not provide it unless they have already attempted to solve the problem, and you may not share your code or spell out the solution step-by-step.

> ⚠ Warning
>
> Adherence to the above policy means that identical answers, or nearly identical answers, cannot occur. Thus, such occurrences are violations of the Course's Academic honesty policy.

You can always, of course, ask me for help on Slack. And public Slack **questions** are allowed and encouraged.

You may also use external sources (books, websites, papers, …) to

- Look up programming language documentation, find useful packages, find explanations for error messages, or remind yourself about the syntax for some feature. I do this all the time in the real world. Wikipedia is your friend.
- Read about general approaches to solving specific problems (e.g. a guide to dynamic programming or a tutorial on unit testing in your programming language), or
- Clarify material from the course notes or assignments.

If you use code from online or other sources (including generative AI), you must include code comments identifying the source. It must be clear what code you wrote and what code is from other sources. This rule also applies to text, images, and any other material you submit.

Please talk to me if you have any questions about this policy. Any form of plagiarism or cheating will result in sanctions to be determined by me, including grade penalties (such as negative points for the assignment or reductions in letter grade) or course failure. I am obliged to report violations to the appropriate University authorities. See also the text below.

Your score on HW, Labs, and Clickers

The total you can accumulate across these 3 components is 60 points. But you can get there however you want. The total available is 75 points. The rest is up to you. But with choice, comes responsibility.

Rules:

- Nothing dropped.
- No extensions.
- If you miss a lab or a homework deadline, then you miss it.
- Make up for missed work somewhere else.
- If you get sick, fine. You miss a few clickers and maybe a lab (though you can do it remotely).
- If you have a job interview and can't complete an assignment on time, then skip it.

I'm not going to police this stuff. You don't need to let me know if you miss an assignment. There is no reason that every single person enrolled in this course shouldn't get > 65 in this class.

Illustrative scenarios:

- Doing 80% on 4 homeworks (32 points), getting 5 clicker points, completing 9 labs with perfect scores, and gaining 5 participation points gets you 60 points.
- Doing 90% on 4 homeworks (36 points), getting 7 clicker points, completing 6 labs with perfect scores, and gaining 5 participation points gets you 60 points.
- Getting full homeworks (40 points) and full labs (20 points), with 0 clicker and 0 participation points gets you 60 points.

Choose your own adventure. Note that the biggest barrier to getting to 60 is skipping the assignments.

Late policy

Late lab/homework submissions will not be accepted.

More specifically, any submission that we receive after grading has commenced will receive a 0. We likely won't start grading at 11:01pm on the due date, so don't worry if you're a few minutes late. On the other hand, don't even bother submitting if you've missed the deadline by a few days. This policy may seem harsh, but remember that there are many paths to a full 65 on the effort-based grade. If you miss one assignment, focus on doing well on the other labs/assignments and you might still end up with an A in the course.

Summative Assessment

Midterm Exam

10 points, in class, on Canvas.

- All multiple choice, True/False, matching.
- The clickers are the best preparation.
- Questions may ask you to understand or find mistakes in code.
- No writing code.

Final Exam

30 points, hand written.

The midterm and final are very hard. It is intended to separate those who really understand the material from those who don't. Last year, the median grade on the final was 50%.

You can still end up with a very good grade even if you don't do well on the exams. If you put in the work (do all the effort points) and the median grade on the midterm and final (50%), you'll get an 80. If you put in the work (do all the effort points) and skip the midterm and final, you get a 60. You do not have to pass the final to pass the course. You don't even have to take the final.

The point of this scheme is for those who work hard to do well. But only those who really understand the material will get 90+.

Health Issues and Considerations

> ⚠ Warning
>
> If you are sick, it's important that you stay home – no matter what you think you may be sick with (e.g., cold, flu, covid, other).

- Sleep is one of the most important factors for your health, and I realize that an 8am class makes good sleep challenging. Work on developing good habits that will enable you to attend class while staying well-rested. However, it may be better to sacrifice in-person attendance than your sleep, especially if you would be so sleep-deprived that you won't be able to benefit from class. I hope you make responsible decisions.
- Your health precautions help reduce risk and keep everyone safer. In this class, the marking scheme provides flexibility so that you can prioritize your health and still succeed. All work can be completed outside of class with reasonable time allowances.
- **If you do miss class because of illness:**
  - Make connections early in the term with other students in the class. You can help each other by sharing notes. If you don't yet know anyone in the class, post on the discussion forum to connect with other students.

  - Consult the class resources here and on Canvas. We will post all slides, readings, and recordings for each class.
  - Use Slack for help.
  - Come to virtual office hours.
  - See the marking scheme for reassurance about your flexibility options. No part of your final grade will be directly impacted by missing class.

- **If you are sick on midterm day or final exam day, do not attend the exam.** You must follow up with your home faculty's advising office to apply for deferred standing. Students who are granted deferred standing write the final exam at a later date. If you're a Science student, you must apply for deferred standing (an academic concession) through Science Advising no later than 48 hours after the missed final exam/assignment. Learn more and find the application online. For additional information about academic concessions, see the UBC policy here.

> **Note**
>
> Please talk with me if you have any concerns or ask me if you are worried about falling behind.

University policies

UBC provides resources to support student learning and to maintain healthy lifestyles but recognizes that sometimes crises arise and so there are additional resources to access including those for survivors of sexual violence. UBC values respect for the person and ideas of all members of the academic community. Harassment and discrimination are not tolerated nor is suppression of academic freedom. UBC provides appropriate accommodation for students with disabilities and for religious, spiritual and cultural observances. UBC values academic honesty and students are expected to acknowledge the ideas generated by others and to uphold the highest academic standards in all of their actions. Details of the policies and how to access support are available here.

Academic honesty and standards

UBC Vancouver Statement

Academic honesty is essential to the continued functioning of the University of British Columbia as an institution of higher learning and research. All UBC students are expected to behave as honest and responsible members of an academic community. Breach of those expectations or failure to follow the appropriate policies, principles, rules, and guidelines of the University with respect to academic honesty may result in disciplinary action.

For the full statement, please see the 2022/23 Vancouver Academic Calendar

Course specific

While course materials are freely available online, selling or distributing homework solutions, lab answers, or exam questions to other students or commercial services is strictly prohibited and violates UBC's academic integrity policies. Violations will be reported to the Dean of Science and may result in serious academic consequences, including course failure.

I have caught students cheating on exams in previous years. In my experience, cheating typically stems from students not understanding the material, which usually results in a failing grade even before any penalties are imposed and the incident is reported to the Dean's office. Please do your own work and utilize the TAs and me as resources. We are here to help if you are struggling.

> **Caution**
>
> If I suspect cheating, your case will be forwarded to the Dean's office. No questions asked.

Academic Concessions

These are handled according to UBC policy. Please see

- UBC student services
- UBC Vancouver Academic Calendar
- Faculty of Science Concessions

Missed final exam

Students who miss the final exam must report to their Faculty advising office within 72 hours of the missed exam, and must supply supporting documentation. Only your Faculty Advising office can grant deferred standing in a course. You must also notify your instructor prior to (if possible) or immediately after the exam. Your instructor will let you know when you are expected to write your deferred exam. Deferred exams will ONLY be provided to students who have applied for and received deferred standing from their Faculty.

Take care of yourself

Course work at this level can be intense, and I encourage you to take care of yourself. Do your best to maintain a healthy lifestyle this semester by eating well, exercising, avoiding drugs and alcohol, getting enough sleep and taking some time to relax. This will help you achieve your goals and cope with stress. I struggle with these issues too, and I try hard to set aside time for things that make me happy (cooking, playing/listening to music, exercise, going for walks).

All of us benefit from support during times of struggle. If you are having any problems or concerns, do not hesitate to speak with me. There are also many resources available on campus that can provide help and support. Asking for support sooner rather than later is almost always a good idea.

If you or anyone you know experiences any academic stress, difficult life events, or feelings like anxiety or depression, I strongly encourage you to seek support. UBC Science has resources on their website. UBC Counseling Services is here to help: call 604 822 3811 or visit their website. Consider also reaching out to a friend, faculty member, or family member you trust to help get you the support you need.