

Syllabus

Term 2022 Winter 1: 07 Sep - 07 Dec 2022

COVID information TL;DR

When you come to class, we encourage you to wear a mask.
If you think you're sick, stay home no matter what.
More details below.

Course info

Instructor:

Daniel McDonald

Office: Earth Sciences Building 3106

Website: <https://dajmcdon.github.io/>

Email: daniel@stat.ubc.ca

Slack: [@prof-mcdonald](https://stat-406-2022w.slack.com/team/U040BDRBCBH)(<https://stat-406-2022w.slack.com/team/U040BDRBCBH>)

Office hours:

Daniel: Tuesday 1400-1500 room TBA Also: Thursday 1100-1200 on Zoom only
(this one may be cancelled some weeks, link on Canvas)

Course webpage:

WWW: <https://ubc-stat.github.io/stat-406/>

Github: <https://github.com/stat-406-2022w>

See also Canvas

Lectures:

Tue/Thu 0800h - 0930h

(In person) Earth Sciences Building (ESB) 1012

Textbooks:

[ISLR]

[ESL]

Prerequisite:

STAT 306 or CPSC 340

Course objectives

This is a course in statistical learning methods. Based on the theory of linear models covered in Stat 306, this course will focus on applying many techniques of data analysis methods to interesting datasets.

The course combines analysis with methodology and computational aspects. It treats both the “art” of understanding unfamiliar data and the “science” of analyzing that data in terms of statistical properties. The focus will be on practical aspects of methodology and intuition to help students develop tools for selecting appropriate methods and approaches to problems in their own lives.

This is not a “how to program” course, nor a “tour of machine learning methods”. Rather, this course is about how to understand some ML methods. STAT 306 tends to give background in many of the tools of understanding as well as working with already-written R packages. On the other hand, CPSC 340 focuses introduces many methods with a focus on “from-scratch” implementation (in Julia or Python). This course will try to bridge this gap. Depending on which course you took, you may be more or less skilled in some aspects than in others. That’s ok and expected.

Learning outcomes

1. assess the prediction properties of the supervised learning methods covered in class;
2. correctly use regularization to improve predictions from linear models, and also to identify important explanatory variables;
3. explain the practical difference between predictions obtained with parametric and non-parametric methods, and decide in specific applications which approach should be used;
4. select and construct appropriate ensembles to obtain improved predictions in different contexts;
5. select sensible clustering methods and correctly interpret their output;
6. correctly utilize and interpret principal components and other dimension reduction techniques;
7. employ reasonable coding practices and understand basic R syntax and function;
8. write reports and use proper version control; engage with standard software

Textbooks

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. If you want to learn more and understand the material more deeply, this is the book for you. All readings from [ESL] are optional.

Course assessment opportunities

Homework assignments (up to 60)

Labs (up to 20)

Clickers (up to 10)

Total: up to 70

Final Exam: 30

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 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.

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.

It's important here to recognize just how important active participation in these activities is. You learn by doing, and this is your opportunity to learn in a low-stakes environment. One thing you'll learn, for example, is that all animals urinate in 21 seconds.¹

Assignments

There will be 5 assignments. These are submitted via pull request similar to the labs but to the `homework-<username>` repo. Each assignment is worth up to 12 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 there after.

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 graphs.

After receiving a mark and feedback, if you score less than 10, you may make corrections for up to 50% of lost marks. 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.

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.

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.

These rules also apply to getting help from other people such as friends not in the course (try the problem first, discuss strategies, not step-by-step solutions,

¹A careful reading of the linked paper with the provocative title "Law of Urination: all mammals empty their bladders over the same duration" reveals that the authors actually mean something far less precise. In fact, their claim is more accurately stated as "mammals over 3kg in body weight urinate in 21 seconds with a standard deviation of 13 seconds". But the accurate characterization is far less publicity-worthy.

acknowledge those from whom you received help).

You may not use homework help websites, Stack Overflow, and so on under any circumstances.

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.

But external sources must be used to **support** your solution, not to **obtain** your solution. You may **not** use them to * Find solutions to the specific problems assigned as homework (in words or in code)—you must independently solve the problem assigned, not translate a solution presented online or elsewhere. * Find course materials or solutions from this or similar courses from previous years, or * Copy text or code to use in your submissions without attribution.

If you use code from online or other sources, 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.

Clickers

These are short multiple choice and True / False questions. They happen in class. For each question, correct answers are worth 4, incorrect answers are worth 2. You get zero points for not answering.

At the end of the term, I'll take the total number of clicker questions and multiply by 3 to find the total number possible. If your average is 3 or higher, you get 10 points. If your average, x , is between 1 and 3, you get $5(x-1)$ points.

If your average is less than 1, you get 0 points in this component. In addition your final grade in this course will be reduced by 1 letter grade.

This means that if you did everything else and get a perfect score on the final exam, you will get a 79.

DON'T DO THIS!!

This may sound harsh, but think about what is required for such a penalty. You'd have to skip more than 50% of class meetings and get every question wrong when you are in class. This is an in-person course. It is not possible to get an A without attending class on a regular basis.

To compensate, I will do my best to post recordings of lectures. Past experience has shown 2 things:

1. You learn better by attending class than by skipping and "watching".
2. Sometimes the technology messes up. So there's no guarantee that these will be available.

The purpose is to let you occasionally miss class for any reason with minimal consequences. See also below. If for some reason you need to miss longer stretches of time, please contact me or discuss your situation with your Academic Advisor.

Your score on Labs, HW, and Clickers

The total you can accumulate across these 3 components is 70 points. But you can get there however you want. The total available is 90 points. The rest is up to you. But with choice, comes responsibility.

Rules: * Nothing dropped. * No extensions. * If you miss a lab or a HW deadline, then you miss it. * Make up for missed work somewhere else. * If you isolate due to Covid, 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.

We're not going to police this stuff. You don't need to let me know. There is no reason that every single person enrolled in this course shouldn't get > 70 in this class.

Illustrative scenarios:

* Doing 80% on 5 homeworks, coming to class and getting 50% correct, get 2 points on 6 labs gets you 70 points. * Doing 90% on 5 homeworks, getting 50% correct on all the clickers, averaging 1.7 on all the labs gets you 70 points. * Going to all the labs and getting 100%, 94% on 4 homeworks, plus being wrong on every clicker gets you 70 points

Choose your own adventure. Note that the biggest barrier to getting to 70 is skipping the assignments.

Final exam

30 points

- All multiple choice, T/F, matching.
- The clickers are the best preparation.

- Questions may ask you to understand or find mistakes in code.
- No writing code.

The Final is very hard. By definition, it cannot be effort-based.

It is intended to separate those who really understand the material from those who don't. Last year, the median was 50%. But if you put in the work and get 50%, you get an 85. **If you put in the work and skip the final, you get a 70**

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

Health issues and considerations

Covid Safety in the Classroom

Masks. Masks are **recommended**. For our in-person meetings in this class, it is important that all of us feel as comfortable as possible engaging in class activities while sharing an indoor space. Masks are a primary tool to make it harder for Covid-19 to find a new host. Please note that there are some people who cannot wear a mask. These individuals are equally welcome in our class.

Vaccination. If you have not yet had a chance to get vaccinated against Covid-19, vaccines are available to you, free, and on campus. See <http://www.vch.ca/covid-19/covid-19-vaccine> for help finding an appointment. The higher the rate of vaccination in our community overall, the lower the chance of spreading this virus. You are an important part of the UBC community. Please arrange to get vaccinated if you have not already done so.

Your personal health

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, other).

- Do not come to class if you have Covid symptoms, have recently tested positive for Covid, or are required to quarantine. You can check this website to find out if you should self-isolate or self-monitor: <http://www.bccdc.ca/health-info/diseases-conditions/covid-19/self-isolation#Who>.
- Your precautions will help reduce risk and keep everyone safer. In this class, the marking scheme is intended to provide flexibility so that you can prioritize your health and still be able to succeed. All work can be completed outside of class with reasonable time allowances.
- **If you do miss class because of illness:**
 - Make a connection early in the term to another student or a group of 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 on here and on Canvas. We will post all the slides, readings, and recordings for each class day.

- Use Slack for help.
- Come to virtual office hours.
- See the marking scheme for reassurance about what flexibility you have. No part of your final grade will be directly impacted by missing class.
- **If you are sick on 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.

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

Several commercial services have approached students regarding selling class notes/study guides to their classmates. Please be advised that selling a faculty member's notes/study guides individually or on behalf of one of these services using UBC email or Canvas, violates both UBC information technology and UBC

intellectual property policy. Selling the faculty member's notes/study guides to fellow students in this course is not permitted. Violations of this policy will be considered violations of UBC Academic Honesty and Standards and will be reported to the Dean of Science as a violation of course rules. Sanctions for academic misconduct may include a failing grade on the assignment for which the notes/study guides are being sold, a reduction in your final course grade, a failing grade in the course, among other possibilities. Similarly, contracting with any service that results in an individual other than the enrolled student providing assistance on quizzes or exams or posing as an enrolled student is considered a violation of UBC's academic honesty standards.

Some of the problems that are assigned are similar or identical to those assigned in previous years by me or other instructors for this or other courses. Using proofs or code from anywhere other than the textbooks (with attribution), this year's course notes (with attribution), or the course website is not only considered cheating (as described above), it is easily detectable cheating. Such behavior is strictly forbidden.

In previous years, I have caught students cheating on the exams or assignments. I did not enforce any penalty because the action did not help. Cheating, in my experience, occurs because students don't understand the material, so the result is usually a failing grade even before I impose any penalty and report the incident to the Dean's office. I carefully structure exams and assignments to make it so that I can catch these issues. I **will** catch you, and it does not help. Do your own work, and use the TA and me as resources. If you are struggling, we are here to help.

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 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.

A dated PDF is available at this link.