👤🛡 Syllabus

Term 2021 Winter 1: 07 Sep - 07 Dec 2021

COVID information TL;DR

When you come to class, 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-daniel

Office hours:
Daniel: Thursday 1400-1500

TAs
Xiaoxuan: Friday 1500-1600
Chloe: Wednesday 0800-0900
Xiaoting: Monday 0900-1000

Course webpage:
WWW: https://ubc-stat.github.io/stat-406/
Github: https://learning.github.ubc.ca/STAT-406-101-2021W/
See also Canvas

Lectures:
Tue/Thu 1230h - 1400h UTC-7 Vancouver local time
(In person) Leonard S. Klinck Rm 200

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

15 Worksheets and labs
20 Assignments
10 Mini quizzes
20 Project
35 Final exam

Worksheets and labs

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

Worksheets are simply extracts of the code from the online [course notes](). The code for each topic has two lines removed. You should read the notes and work through the code you have, executing the commands and examining the output. The purpose is for you to engage in *active* rather than *passive* reading. When you discover the missing lines, simply copy the correct lines into your file. That's it. These are simple and worth 1 point each up to a maximum of 20 points.

Labs are more thorough. They typically have a few questions for you to answer or code to implement. These will often be done *during class* periods. But you can do them on your own as well. These are worth 2 points each up to a maximum of 20 points.

Rules To get all 15%, you need to accumulate 30 total points during the semester. This can be done in any combination of worksheet points and labs (respecting the limit). Of course you can (and should!) do more, but this isn't required. You may submit up to 5 points at a time in a pull request and you may submit up to 5 points per week and only 1 PR per week. In your PR, there must be at least as many commits as points. So a PR with 2 worksheets and 1 lab should have at least 4 commits (probably one for each worksheet and 2 for the lab). This will be lightly enforced initially. The last day to open a PR is December 5.

Marking Worksheets get 1 if correct and 0 otherwise. If you get 0, you can't resubmit this check (but you can submit others). Labs get 2 points when correct. Sometimes the graders will ask for changes in their PR review. If they do, implement the corrections, make another commit, push, and re-request the review. Once correct, you receive 2 points. Only one additional review will be allowed per submission. If egregious errors remain (in the opinion of the grader) then you will not receive credit (but you can submit another labs). The overriding theme here is "if you put in the effort, you'll get all the points."

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, you should complete 4. These are submitted via pull request similar to the worksheets. You are allowed *one late class period*. This means that of the 4 assignments you submit, one can be submitted late, up until the next class meeting. This policy (drop 1, submit 1 late) is intended to accommodate any and all issues. It is not a "pad your mark" policy. If you submit the first 4 assignments, we won't mark the fifth. The TAs have a lot on their plate and "marking dropped assignments" takes a lot of work. Both policies are "no questions asked". Simply turn it in late or not at all and we'll take care of the rest. You don't need to justify anything.

Policy on collaboration

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.

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, 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.

Mini quizzes

These are short Canvas multiple choice and True / False questions. There are 5 quizzes, one to go with each Assignment. They're really "part of the assignment": you do them on your own and submit by the accompanying assignment deadline. They're just on Canvas to make grading easier for the TAs. Just like the Assignments, only 4 will count, but because they're on Canvas, your lowest score will get dropped. They will be open for 1 week up to the deadline. You can complete it anytime by the deadline, but you only get **one** chance.

Group/individual project

There will be a project. You may choose to complete it in a group of your choosing or individually. The group can contain at most 4 individuals. The group project has two intermediate checkpoints (see schedule).

Similar policies on plagiarism as for the homework apply.

More details will be made available.

Important considerations

Covid Safety in the Classroom

Masks

Masks are **required** for all indoor classes, as per the BC Public Health Officer orders. 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. For the purposes of this order, the term "masks" refers to medical and non-medical masks that cover our noses and mouths. Masks are a primary tool to make it harder for Covid-19 to find a new host. You will need to wear a medical or non-medical mask for the duration of our class meetings, for your own protection, and the safety and comfort of everyone else in the class. You may be asked to remove your mask briefly for an ID check for an exam, but otherwise, your mask should cover your nose and mouth. **Please do not eat in class. If you need to drink water/coffee/tea/etc, please keep your mask on between sips**. Please note that there are some people who cannot wear a mask. These individuals are equally welcome in our class.

Students who need to request an exemption to the indoor mask mandate must do so based on one of the grounds for exemption detailed in the PHO Order on Face Coverings (COVID-19). Such requests must be made through the Center for Accessibility (info.accessibility@ubc.ca).

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.

Seating in class
To reduce the risk of Covid transmission, please sit in a consistent area of the classroom each day. This will minimize your contacts and will still allow for the pedagogical methods planned for this class to help your learning.

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).

- A daily self-health assessment is **required** before attending campus. Every day, before coming to class, complete the self-assessment for Covid symptoms using this tool: https://bc.thrive.health/covid19/en
- 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.
    - All recordings from last year are available on Canvas.
    - 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 a 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/assignment 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 2020/21 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. 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 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.

Censorship

During this pandemic, the shift to online learning has greatly altered teaching and studying at UBC, including changes to health and safety considerations. Keep in mind that some UBC courses might cover topics that are censored or considered illegal by non-Canadian governments. This may include, but is not limited to, human rights, representative government, defamation, obscenity, gender or sexuality, and historical or current geopolitical controversies. If you are a student living abroad, you will be subject to the laws of your local jurisdiction, and your local authorities might limit your access to course material or take punitive action against you. UBC is strongly committed to academic freedom, but has no control over foreign authorities (please visit this link for an articulation of the values of the University conveyed in the Senate Statement on Academic Freedom). Thus, we recognize that students will have legitimate reason to exercise caution in studying certain subjects. If you have concerns regarding your personal situation, consider postponing taking a course with manifest risks, until you are back on campus or reach out to your academic advisor to find substitute courses. For further information and support, please visit this link.

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.

1. A 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 charactization is far less publicity-worthy. ↩