DATA 311: Machine Learning

THE UNIVERSITY OF BRITISH COLUMBIA

Computer Science, Mathematics, Physics and Statistics
Irving K. Barber Faculty of Science
Okanagan Campus

Winter Term 1, 2024

2024-09-03

Course Information

Instructor

Name: Dr. Irene Vrbik[1] (she/her)
email: irene.vrbik@ubc.ca
Office Hour: Wed 2:30 p.m - 3:30 p.m
Office: SCI 104

Course TAs

Linda Okpanchi
Babak Fathollahi Dehkordi

Class Schedule

Location: LIB Floor 3 - Room 305
Time: Tue Thu | 3:30 p.m. - 5:00 p.m.

---

[1]see how to pronounce my name on my website: https://irene.vrbik.ok.ubc.ca/about/

Course Description

Official Calendar Description:

DATA_O 311 (3) Machine Learning: Regression, classification, resampling, model selection and validation, fundamental properties of matrices, dimension reduction, tree-based methods, unsupervised learning. [3-2-0].Prerequisite: Either (a) one of STAT 205, STAT 230 or (b) a score more than 75% in one of APSC 254, BIOL 202, PSYO 373; and one of COSC 111, APSC 177.

Course Objectives

Course Overview

The course is designed to introduce students to classical machine learning methods for regression and classification with an emphasis on model validation (i.e. it is not enough to fit a model, students should be able to estimate how good the resulting model is). By taking this course, students will gain experience in applying machine learning algorithms in R and develop skills for effectively communicating a proper interpretation of the results.

Learning Outcomes

At the end of this course, students should be able to:

1. build a model and validate it
2. understand fundamental proofs for techniques that rely on matrix algebra
3. compute linear regression and apply hypothesis testing
4. perform logistic regression and discriminant analysis
5. apply the K-fold cross-validation methods
6. apply the LASSO and ridge regression methods
7. apply bagging and boosting on tree-based methods
8. apply some methods of unsupervised learning (e.g. principal components, or k-means clustering).
9. manipulate data sets in R including applying the above methods
10. create reproducible documents that embed R code, text, figures, and more.

Course Format

The course will be made up of 3 hours of lecture per week plus 2 hours of weekly laboratory (labs).

Lecture Format

Lectures will be given in-person. Slide decks will be posted to https://irene.vrbik.ok.ubc.ca/quarto/machine-learning/schedule.html prior to our scheduled lecture time. Slides might be supplemented with handwritten material which I will upload to Canvas after lecture. Lectures may also include discussions which you will only gain access to by attending lectures. While statistical software (i.e. R code and output) will be discussed during lecture, practical skills and applications of topics are covered primarily in computer labs.

My office hours are stated in the Course Information. During that time students are encouraged to stop in ask questions/discuss anything about course. If you are unable to make those times, please reach out to me to schedule an alternative appointment.

Lab format

All students must be registered for a lab (held weekly unless otherwise specified). Please check your registration to determine your lab section and time. Labs are structured as walk-though tutorials which help to develop the practical skills of performing machine learning in R. You may work through the lab material on your own time and/or work through them during your scheduled lab. To ensure that TAs are not overloaded during a single lab, please do not attend labs for which you are not registered.

Labs sessions will be hosted by your TA in persons. While they are primarily there to provide guidance on carrying out analyses in R, they additionally provide the opportunity to meet other students from class, ask questions and/or discuss concepts from lecture, and receive assistance on assignments. Thus, labs will act as addition "office hours" held by your TAs. While labs are not mandatory (i.e. attendance will not be taken) you are highly encouraged to attend. Do not skip going through this material as lab content will be fair game for testing on midterms and the final exam.

Marking and Evaluation

Table 1: Weighting scheme of final grade calculation.

Grade Item	Percentage of Grade
Assignments	20%
Midterm 1 (Thursday Oct 10)	20%
Midterm 2 (Tuesday Nov 19)	20%
Final Exam	40%

Final grades will be based on the evaluations listed above in the weighting scheme above. The final grades will be assigned according to the standardized grading system outlined in the UBC

. Faculties, departments, and schools reserve the right to scale grades in order to maintain equity among sections and conformity to University, faculty, department, or school norms. Students should therefore note that an unofficial grade given by an instructor might be changed by the faculty, department, or school. Grades are not official until they appear on a student's academic record.

Passing Criterion

To pass the course, a student must:

1. Achieve a final grade of 50% or higher, as determined by the weighting scheme table **and**
2. Obtain a passing grade (50% or more) on at least one of the in-person assessments (midterm 1, midterm 2 or final exam).

Note: If a student does not meet criterion 2, their final grade will be capped at 45, regardless of their overall score.

Grade Items

Midterms

There will be two (2) synchronous closed-book midterms given during the term, held in class on Thursday Oct 10 and Tuesday Nov 19. The midterms are not cumulative; the second midterm will cover only the material presented after the cutoff for the first midterm. Specific details on the material covered in each midterm will be provided closer to the dates and may vary depending on the pace of the course.

Final Exam

The examination period begins Monday, December 9 and ends Friday, December 20. The comprehensive final exam is cumulative, covering all the material presented throughout the course. The final exam will be in-person and closed-book.

Assignments

There will be approximately four (4) assignments. Assignments will incorporate material covered during lab as well as lecture. Answers will be submitted electronically through Canvas. Students will be required to create and submit a fully reproducible document using Quarto (.qmd).

Missing/Late Grade Items

Assignments

Assignments must be submitted electronically through Canvas. Late assignments will incur a 10% deduction for each day (including weekends) past the due date. Assignments more than 2 days (48 hours) late will not be accepted. Any assignment not submitted will receive a grade of 0%.

Midterms

If a sufficient excuse is provided (e.g., a medical condition supported by documentation from a doctor), the weight of a missed midterm will be transferred to the final exam. **No** make-up tests will be offered.

Final Examinations

Except in the case of examination clashes and hardships (three or more formal examinations scheduled within a 24-hour period) or unforeseen events, students will be permitted to apply for out-of-time final examinations only if they are representing the University, the province, or the country in a competition or performance; serving in the Canadian military; observing a religious rite; working to support themselves or their family; or caring for a family member. Unforeseen events include (but may not be limited to) the following: ill health or other personal challenges that arise during a term and changes in the requirements of an ongoing job. Further information on Academic Concession can be found under Policies and Regulation in the Okanagan Academic Calendar (see Academic Concession)

Course Material and Tools

We will be using UBCO's Learning Management System (LMS) Canvas: https://canvas.ubc.ca/. It is recommended that you log in daily to check for announcements, participate in discussions, access course materials, submit assignments, and review upcoming deadlines. You may also review and adjust your notification preferences for this course (refer to the supporting documentation for instructions).

Textbook

Our primary source of reference will be (Gareth et al. 2013). A free downloadable version can be found at https://www.statlearning.com/. Some additional content will be available from (Hastie 2009).

Software

Our course will exclusively be using R (see https://www.r-project.org/). I strongly recommend that you use RStudio for running R (see https://posit.co/products/open-source/rstudio/).

Tentative Course Schedule

Below is a tentative weekly schedule for the course. The topics are subject to change depending on the pace of coverage. A lecture breakdown by topic (along with suggested readings) will be updated as we progress. You can find it here: https://irene.vrbik.ok.ubc.ca/quarto/machine-learning/schedule.html.

learning/schedule.html.

Table 2: Tentative course schedule

Week	Topic
1	Introduction: R and Rstudio; Notation and Terminology
2	Regression models: Linear regression; model assessment
3	Extension to linear regression models (interaction, categorical predictors, polynomial regression); KNN regression
4	Classification models: polynomial regression; model assessment
5	Classification models II: Bayes Classifier, KNN Classification and Discriminant Analysis
6	Clustering: distance measures (Euclidean, Gowers...), Hierarchical clustering
7	Cross validation and bootstrap
8	Trees: Classification and Regression trees, bagging and random forests, boosting
9	Ridge Regression and the LASSO
10	Dimensionality reduction with PCA
11	Principal Component Analysis (PCA) regression and PLS
12	Gaussian Mixture Models (GMM) and Neural Networks
13	Review

Please note these important Dates and Deadlines:

- Start: Tuesday, September 3
- Finish: Friday, December 6
- Midterm Break: November 11 - 15
- Teaching Days: 62
- Exams Start: Monday, December 9

- Exams Finish: Friday, December 20

There will be no class, office hours, or labs during the **Midterm Break** nor on the following Statutory holidays:

- Monday, September 30: National Day for Truth and Reconciliation
- Monday, October 14: Thanksgiving Day

If you observe any other holidays not listed above, please feel free to contact me directly if you believe they may conflict with the outlined course structure.

Expectations

Your responsibilities to this class, and your education as a whole, include regular attendance and active participation. You are responsible for helping to create a classroom environment where everyone can learn. At a basic level, this means respecting your classmates and the instructor, and treating them with the courtesy you expect to receive in return.

Inappropriate classroom behavior includes, but is not limited to:

- Disrupting the classroom atmosphere
- Engaging in non-class activities
- Talking on a cell phone
- Inappropriate use of profanity during discussions
- Using abusive or disrespectful language toward the instructor, other students, or about individuals or groups

Academic Integrity

The academic enterprise is founded on honesty, civility, and integrity. As members of this enterprise, all students are expected to know, understand, and follow the codes of conduct regarding academic integrity. At the most basic level, this means submitting only original work done by you and acknowledging all sources of information or ideas and attributing them to others as required. This also means you should not cheat, copy, or mislead others about what is your work; nor should you help others to do the same. For example, it is prohibited to: share your past assignments and answers with other students; work with other students on an assignment when an instructor has not expressly given permission; or spread information through word of mouth, social media, websites, or other channels that subverts the fair evaluation of a class exercise, or assessment. Learn more through the Academic Integrity website.

Academic Misconduct

Violations of academic integrity (i.e., academic misconduct) lead to the breakdown of the academic enterprise, and therefore serious consequences arise and harsh sanctions are imposed. For example, incidences of plagiarism or cheating may result in a mark of zero on the assignment or exam and more serious consequences may apply if the matter is referred for consideration for academic discipline. Careful records are kept to monitor and prevent recurrences. Any instance of cheating or taking credit for someone else's work, whether intentionally or unintentionally, can and often will result in at minimum a grade of zero for the assignment, and these cases will be reported to the Head of the Department and Associate Dean Academic of the Faculty.

A note on collaboration: While collaboration with peers is encouraged, submitting work that you do not fully understand or cannot explain will be considered a violation of academic integrity. Any form of academic dishonesty, including plagiarism or unapproved sharing of work, will be handled according to UBC's academic integrity policies.

Use of generative artificial intelligence (AI):

Students are permitted to use artificial intelligence tools, including generative AI, to gather information, review concepts or to help produce assignments. However, students are ultimately accountable for the work they submit, and any content generated or supported by an artificial intelligence tool must be cited appropriately. Use of AI tools is not permitted during midterm exams and final exams in this course. Learn more through the Generative AI website. Below are some general guidelines:

1. **Understanding and Originality**: Any code, solutions, or explanations provided by AI tools must be reviewed, understood, and revised by you. Submitting work that you do not fully understand or cannot explain will be considered a violation of academic integrity. You should be able to discuss and defend the solutions as if you wrote them independently.
2. **Attribution**: If you use AI tools to help generate code or content for assignments, you must clearly indicate this in your submission. A simple acknowledgment at the top of your assignment, such as "Parts of this assignment were assisted by AI tools (e.g., ChatGPT)," will suffice.
3. **Responsible Use**: Using AI tools to enhance your understanding is acceptable, but relying solely on them without making an effort to learn the material independently will be detrimental to your success in this course. Your goal should be to develop a deep understanding of machine learning principles that you can apply beyond the classroom.

By adhering to these guidelines, you will ensure that your learning experience is both ethical and effective, setting you up for success in your studies and future career.

Grievances and Complaints Procedures

A student who has a complaint related to this course should attempt to resolve the matter with the instructor first. Students may talk first to someone other than the instructor if they do not feel, for whatever reason, that they can directly approach the instructor. If the complaint is not resolved to the student's satisfaction, the student should e-mail the Department Head Dr. Raymond Lawrence (ramon.lawrence@ubc.ca).

Student Service Resources

Disability Resource Centre

The Disability Resource Centre (DRC) facilitates disability-related accommodations and programming initiatives to that ameliorate barriers for students with disabilities and/or ongoing medical conditions. If you require academic accommodations to achieve the objectives of a course please contact the DRC at:

UNC 215 250.807.8053
Email: drc.questions@ubc.ca
Web: www.students.ok.ubc.ca/drc

Equity and Inclusion Office

Through leadership, vision, and collaborative action, the Equity & Inclusion Office (EIO) develops action strategies in support of efforts to embed equity and inclusion in the daily operations across the campus. The EIO provides education and training from cultivating respectful, inclusive spaces and communities to understanding unconscious/implicit bias and its operation within in campus environments. UBC Policy 3 prohibits discrimination and harassment on the basis of BC's Human Rights Code. If you require assistance related to an issue of equity, educational programs, discrimination or harassment please contact the EIO.

UNC 325H 250.807.9291
Email: equity.ubco@ubc.ca
Web: www.equity.ok.ubc.ca/

Office of the Ombudperson

The Office of the Ombudsperson for Students is an independent, confidential and impartial resource to ensure students are treated fairly. The Ombuds Office helps students navigate campus-related fairness concerns. They work with UBC community members individually and at the systemic level to ensure students are treated fairly and can learn, work and live in a fair, equitable and respectful environment. Ombuds helps students gain clarity on UBC policies

and procedures, explore options, identify next steps, recommend resources, plan strategies and receive objective feedback to promote constructive problem solving. If you require assistance, please feel free to reach out for more information or to arrange an appointment.

UNC 328 250.807.9818
Email: ombuds.office.ok@ubc.ca
Web: www.ombudsoffice.ubc.ca/

Student Learning Hub

The Student Learning Hub is your go-to resource for free math, science, writing, and language learning support. The Hub welcomes undergraduate students from all disciplines and year levels to access a range of supports that include tutoring in math, sciences, languages, and writing, as well as help with academic integrity, study skills and learning strategies. Students are encouraged to visit often and early to build the skills, strategies and behaviors that are essential to being a confident and independent learner. For more information, please visit the Hub's website.

LIB 237 250.807.8491
Email: learning.hub@ubc.ca
Web: www.students.ok.ubc.ca/slh

Sexual Violence Prevention and Response Office (SVPRO)

The Sexual Violence Prevention and Response Office (SVPRO) is a confidential place for those who have been impacted by any form of sexual or gender-based violence, harassment, or harm, regardless of where or when it took place. SVPRO aims to be a safer space for all UBC students, faculty, and staff by respecting each person's unique and multiple identities and experiences. All genders and sexualities are welcome.

Nicola Townhome 120, 1270 International Mews 250.807.8053
Email: svpro@okangan@ubc.ca
Web: www.svpro.ok.ubc.ca/

Wellbeing and Accessibility Services (WAS)

Wellbeing and Accessibility Services (WAS) supports holistic student wellbeing in body, mind, and spirit. Students can access nurses, physicians and counsellors for health care and counselling related to physical health, emotional/mental health and sexual/reproductive health concerns. WAS is also home to the Disability Resource Centre, Spiritual and Multi-Faith Services, and Campus Health and Education. If you require assistance with your health, please contact Wellbeing and Accessibility Services for more information or to book an appointment.

UNC 337 250.807.9270
Email: healthwellness.okanagan@ubc.ca
Web: www.students.ok.ubc.ca/was

Independent Investigations Office

If you or someone you know has experienced sexual assault or some other form of sexual misconduct by a UBC community member and you want the Independent Investigations Office (IIO) at UBC to investigate, please contact the IIO. Investigations are conducted in a trauma informed, confidential and respectful manner in accordance with the principles of procedural fairness. You can report your experience directly to the IIOby calling 604-827-2060.

Web: https://investigationsoffice.ubc.ca/
E-mail: director.of.investigations@ubc.ca

Safewalk

Download the UBC SAFE – Okanagan app. Don't want to walk alone at night? Not too sure how to get somewhere on campus? Call Safewalk at 250.807.9270. For more information visit: https://security.ok.ubc.ca/safewalk/

References

Gareth, James, Witten Daniela, Hastie Trevor, and Tibshirani Robert. 2013. *An Introduction to Statistical Learning: With Applications in r.* 2nd ed. Spinger.

Hastie, Trevor. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* Springer.