



DATA 101 (3) Making Predictions with Data

Instructor: Irene Vrbik (SCI 393, irene.vrbik@ubc.ca)

Duration: Winter 2018 Term 2, 3 credits

Lectures: 3 hours/week – Mon & Wed: 3:30 PM to 5:00 PM in SCI-247

Lab: 1 hour/week –

Laboratory 1	Fri	8:00	10:00
Laboratory 2	Mon	17:30	19:30
Laboratory 3	Wed	17:30	19:30
Laboratory 4	Tue	14:00	16:00

Office hours: 2 hours per week TBD or by appointment.

Academic Calendar Entry

DATA 101 (3) Basic Predictive Modelling

An introduction to the techniques and software for handling real-world data. Topics include data cleaning, visualization, simulation, basic modeling, and prediction making. [3-1-0]

Prerequisite: None

Course Format

The course is delivered through 3 hours of lectures and 1 hour of labs. Please check your registration to determine your lab/tutorial section and time. You must register for a lab and attend the one you are registered in.

Course Overview, Content and Objectives

The objective of the course is to introduce a combination of fundamental data science concepts to perform simple analyses of real-world and simulated data. The emphasis is on predictive validation rather than traditional statistical inference, which would involve p -values and confidence intervals.

The course covers introductory aspects of supervised machine learning, such as regression, classification, and feature selection. Explorations in data collection, validity, and privacy are incorporated alongside other topics of the course to foster awareness of the growing importance of these issues.

Students will be introduced to a variety of data sets, highlighting that the same concepts can be applied to several fields of study. Simulations, lab quizzes, and assignments are designed to emphasize the link between the application and the concept.

Learning Outcomes

After completing this course, students will be able to:

- apply analytic techniques to large and small data sets
- organize data using open source software
- use carefully constructed visualizations and appropriate numerical summaries to explore and/or present data.
- explain concepts and applications of simulations
- identify potential privacy and data concerns surrounding a data set
- identify and apply appropriate predictive techniques
- apply predictive validation techniques



Evaluation Criteria and Grading

There will be one midterm (25%), one written cumulative final exam (45%), assignments and labs (30%). To pass the course students must obtain an average of 50% and pass the final exam. Failing the final exam will result in a maximum grade of 45%.

Required Readings and Videos

All the resources for tools are available freely on the web. Lectures will introduce the necessary concepts.

Course Schedule:

Week	Topics	Lab	Outcomes
1	Introduction to Data Science and R	N/A	B, D
2	Objects, Computation and Manipulation	Software Basics I: installation, packages, data frames, arithmetic operators, reading external data, help files	B, E
3	Visualization	Software Basics II: indexing, loops, apply functions, conditionals, logical operators	B, C
4	Numerical Summaries	Graphics: univariate, bivariate, multivariate/conditional plotting; basics of human visual perception and implications for graphic design	A, B, C
5	Simulation: Regression Models	Functions: declaring arguments, storing results, using summary function and other object-oriented methods; big data methods	A, B, C
6	Simulation: Nonlinear Models	Simulations I: simple regression, based on additive simulated errors and noting effects on coefficient estimates and predictions, qq-plots, dangers of extrapolation	A, B, C, D, G
7	Simulation: Tree models	Simulations II: other random mechanisms that might lead to outliers and other forms of contamination; implications for data management and cleaning; nonlinear regression, including quadratic, exponential and some intrinsically nonlinear models.	A, B, C, D
8	Review and Midterm	Simulations III: fitting classification and regression trees with CART using simulated and real data.	A, B, C, D
9	Measuring Errors in Prediction	Scraping: rvest, sourcing public data, html and css selectors	A, B, C, D, E
10	Predictive: Multiple Regression and Diagnostics	Predicting I: predict function, error calculations, such as mean-squared-error, mean-absolute-error, standard error, prediction sums of squares	A, B, C, D, E, F, G
11	Training and Testing	Predicting II: multiple regression, diagnostics, visualization in two and three-dimensions, including an introduction to RGL	A, B, C, D, E, F, G



12	Cross-Validation and Over-fitting	Predicting III: fitting and validation applied to both multiple regression and CART models	A, C, F, G
13	Review and Conclusion	Model Selection: training/testing/validation, predicted error types; feature reduction	A, B, C, D, E, F, G

Final Examinations

The examination period for Term 2 of Winter 2018 is Monday, April 8 to Friday, April 26. 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* <http://www.calendar.ubc.ca/okanagan/index.cfm?tree=3,48,0,0>

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. Violations of academic integrity (i.e., 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 to the President's Advisory Committee on Student Discipline. Careful records are kept in order to monitor and prevent recurrences.

A more detailed description of academic integrity, including the University's policies and procedures, may be found in the Academic Calendar at <http://okanagan.students.ubc.ca/calendar/index.cfm?tree=3,54,111,0>.

UBC Okanagan Disability Resource Centre

The Disability Resource Centre ensures educational equity for students with disabilities, injuries or illness. If you are disabled, have an injury or illness and require academic accommodations to meet the course objectives, please contact Earllene Roberts, the Diversity Advisor for the Disability Resource Centre located in Commons Corner in the University Centre building (UNC 227).

UNC 227A 250.807.9263, email earllene.roberts@ubc.ca, Web: www.ubc.ca/okanagan/students/drc



a place of mind

THE UNIVERSITY OF BRITISH COLUMBIA

UBC Okanagan Ombuds Office

The Ombuds Office offers independent, impartial, and confidential support to students in navigating UBC policies, processes, and resources, as well as guidance in resolving concerns related to fairness.

UNC 227B 250.807.9818

email: ombuds.office.ok@ubc.ca, Web: <http://ombudsoffice.ubc.ca/ubc-okanagan-2/>

UBC Okanagan Equity and Inclusion Office

UBC Okanagan is a place where every student, staff and faculty member should be able to study and work in an environment that is free from discrimination and harassment. UBC prohibits discrimination and harassment on the basis of the following grounds: age, ancestry, colour, family status, marital status, physical or mental disability, place of origin, political belief, race, religion, sex, sexual orientation or unrelated criminal conviction. If you require assistance related to an issue of equity, discrimination or harassment, please contact the Equity and Inclusion Office.

UNC 227C 250.807.9291

email: equity.ubco@ubc.ca Web: www.ubc.ca/okanagan/equity

Health & Wellness

SAFEWALK

Don't want to walk alone at night? Not too sure how to get somewhere on campus? Call Safewalk at 250-807-8076. For more information, see:

<http://www.ubc.ca/okanagan/students/campuslife/safewalk.html>