

# STAT 4600: Statistical Learning, Winter 2019 (T02)

## Tentative Course Outline

---

### Course Details

---

<b>Course Title &amp; Number:</b>	Statistical Learning
<b>Credit Hours:</b>	3
<b>Class Schedule:</b>	11:30 AM–12:20 PM, MWF
<b>Location for Lectures:</b>	202 St. John's
<b>Location for Tutorials:</b>	Statistics Lab (311 Machray Hall)
<b>Course Material:</b>	All course materials are posted on UMLearn (D2L) web
<b>Pre-Requisites:</b>	Prerequisite: STAT 3800.

---

### Instructor Contact Information

---

<b>Instructor:</b>	Mohammad Jafari Jozani
<b>Office:</b>	365 Machray Hall
<b>Office Hours &amp; Availability:</b>	Mondays/Wednesdays 10:15–11:15 or by appointment (call or email to confirm).
<b>Office Phone Number:</b>	204-272-1563
<b>E-mail:</b>	m.jafari_jozani@umanitoba.ca
	I will only respond to e-mail from UMLearn ID's
	When feasible, I normally return a call or an email within 24 hours.

---

### Tutorial

---

<b>Instructor:</b>	Pramoda Jayasinghe and Keith Uzelmann
<b>Office:</b>	349 and 347 Machray Hall
<b>E-mail:</b>	jayasinp@myumanitoba.ca, uzelmank@myumanitoba.ca
<b>Lab:</b>	Fridays 2:30 PM–4:00 PM

Attendance of tutorials is not mandatory but STRONGLY recommended. During the tutorials, the TA will solve some selected theoretical problems as well as some using R and answer your questions. If needed, a tutorial may be replaced by a lecture. The first LAB is dedicated to teaching the basics of R and R-markdown.

---

### General Course Information and Course Registration

---

This course is a survey of statistical learning methods and covers major techniques and concepts for both supervised and unsupervised learning. Students will learn how and when to apply statistical learning techniques, their comparative strengths and weaknesses, and how to critically evaluate the performance of learning algorithms. The goal is to

- (i) apply basic statistical learning methods and perform exploratory analysis,
- (ii) properly select statistical learning models,
- (iii) implement these methods using the R programming language,
- (iv) correctly assess model fit and error,
- (v) build an ensemble of learning algorithms.

---

## Course Registration

---

Not to be held with STAT 3800. It is **your responsibility** to ensure that you are entitled to be registered in this course. This means that you:

1. have the appropriate prerequisites, as noted in the calendar description, or have an appropriate permission from the instructor to waive these prerequisites;
2. have not previously taken, or are concurrently registered in, this course and another that has been identified as "not to be held with" in the course description.

The registration system may have allowed you to register in this course, but it is **your responsibility to check**. If you are not entitled to be in this course, you will be withdrawn, or the course may not be used in your degree program. There will be no fee adjustment. This is not appealable. Please be sure to read the course description for this and every course for which you are registered.

---

## Textbook, Readings, Materials

---

**Textbook:** I will have my own course notes. However, I will be using the following textbooks (in particular the first one) for this class. Lecture notes will be available through the UM Learn system (see below). Other references will be suggested during the course if required.

1. *An Introduction to Statistical Learning with Applications in R*. Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani. Springer Texts in Statistics. ISBN 978-1-416-7138-7. New York (2013). E-book is available for download here <https://www-bcf.usc.edu/gareth/ISL/ISLR%20First%20Printing.pdf>.
2. *The Elements of Statistical Learning (2nd Edition)*. Trevor Hastie, Robert Tibshirani and Jerome Friedman. Springer Series in Statistics. (2009). E-book is available for download using <https://web.stanford.edu/hastie/Papers/ESLII.pdf>.

**Readings:** In order to prepare for class, I will normally ask you to read about the topics to be covered before the lecture. I am not expecting you to learn the material on your own, only to familiarize yourself with the main ideas and vocabulary so that the lectures are easier to follow. Do not get bogged down in formulae or minute details. If you come across something that is confusing or troubling, don't despair. If your questions are not resolved during the lecture, please ask. As you work on the problem sets, it will be helpful to re-read the material on a more detailed level.

---

## Topics To Be Covered

---

Here is the outline of the course material (not necessarily in the same order that I will be teaching in the class), which is subject to change, depending on time and class interests.

### 1. Introduction

- Brief overview of statistical learning concepts such as supervised and unsupervised learning with several examples (Regression versus Classification problems, supervised versus unsupervised learning, trade-off between prediction accuracy and model interpretability)
- Introducing methods for assessing the model accuracy
- A brief introduction to R programming (in the Lab)

### 2. Supervised Learning

- An overview of linear regression: Univariate, multivariate and multiple linear regression.
- Regression versus K-nearest neighbours.
- Moving beyond linearity by working with Polynomial regression and regression Splines.
- Ridge regression, regularized regression and some related topics.
- Perceptron and Support Vector Machine
- Resampling Techniques such as the Cross-Validation and Bootstrap.
- Classification using the logistic regression, LDA, QDA, finite mixture models and Naive Bayes approaches.
- (if time permits) Trees, Boosting and Neural Networks

### 3. Unsupervised Learning

- Clustering methods such as K-means clustering, Hierarchical clustering
- Understanding dendrograms and different similarity and dissimilarity measures.
- Dimension Reduction techniques such as principal component analysis

---

## Course Technology

---

**Course web-page:** Course materials will be made available through the University of Manitoba's UM Learn system ([umanitoba.ca/d21](http://umanitoba.ca/d21)).

**Software:** We will extensively be making use of the R statistical software. R is freely available for Linux, Macintosh and Windows from *The Comprehensive R Archive Network* at <http://cran.r-project.org/>. Please download and install.

**Other Technology:** It is the general University of Manitoba policy that all technology resources are to be used in a responsible, efficient, ethical and legal manner. Students should restrict their use of technology to those approved by the instructor and/or University of Manitoba Accessibility Services for *educational purposes only*. Electronic messaging, e-mail, social networking, gaming, etc. should be avoided during class time. Cell phones should be turned off. If a student is on call for emergencies, his/her cell phone should be on vibrate mode and the student should leave the classroom before using it.

---

## Important Dates

---

These dates are tentative and subject to change at the discretion of the instructor and/or based on the learning needs of the students but such changes are subject to Section 2.8 of the ROASS Procedure.

Date	Information
January 7, 2019	Classes Begin
March 1st, 2019	<b>Midterm Test</b> (311 Machray Hall: 2:30pm–4:00pm)
March 20, 2019	Last Day for VW
April 8, 2019	<b>End of Classes</b>

---

## Assignments, Examinations & Grading

---

**Assignments:** Assignments worth 40% of your final grade. Assignments are due at the start of the class (time will be announced). Each assignment involves three parts (theory, application and simulation). Assignments submitted late will be severely penalized. Assignments submitted after the solutions are posted or after the graded assignments are returned to students will not be marked and receive a grade of 0. Obviously, exceptions can be made to the above policy if special/exceptional circumstances warrant them (e.g., serious illness).

Students are encouraged to discuss and work together on the solutions to the assignments. However, each student must hand in his or her own copy of each assignment with personalized solutions, including comments, discussions and interpretations. Note that actions will be taken against students who are found guilty of acts of academic dishonesty.

Your assignments should conform to the following standards:

- Theoretical part of the assignments are to be done on  $8.5 \times 11$  paper, writing on one side only and they are to be stapled.
- Write your name at the top of each page.
- Applied I and simulation parts of each assignment that involve R programming should be accompanied with the R codes and results should be reproducible. I do encourage you to use R-markdown to hand in your R assignments.
- Revise your assignments so they are reasonably free of grammatical and typographical errors.
- Make sure each step in your solutions is well justified: I mark what is written on paper and should not have to guess what you mean.
- Messy or unreadable assignments will be returned with a mark of zero.

**Exams:** There will be one mid-term exam, worth 30% of your final grade. The tentative date is March 1st, 2019 from 2:30 pm to 4:00 pm. Tentative location for the exams is 311 Machray Hall (Statistics Lab) but this might be **subject to change**. The midterm test and the final examination are closed book. No formula sheet will be provided. A non-programmable calculator might be necessary (graphing calculators are not permitted). However, other electronic devices, such as Laptops, cell phones and MP3, are strictly prohibited.

Item	Percent
Assignments	50%
Mid-term Test	30%
Final Exam	20%
Total	100%

**Note:** There will not be any makeup (deferred) mid-term exam for this course. If you miss the mid-term exam, you will be assigned a mark of zero for the test, unless you **have a valid excuse**, and you **notify me within 48 hours of the scheduled exam**. Students who miss any of the term tests with legitimate reasons will have the midterm weight added to the final examination. The Final Exam will be held on a date to be selected later by the Department of Statistics and will be 3 hours in duration. Students who miss midterm tests, with or without valid documentation, will be reported to the Dean's office as having incomplete term work. This could have repercussions on their ability to write a deferred exam for the course, should such a deferral be requested.

**Grading:** The following are the minimum percentage grades required to receive each of the various letter grades.

Letter Grade	A+	A	B+	B	C+	C	D	F
Final % mark	[90, 100]	[80, 90)	[75, 80)	[70, 75)	[60, 65)	[65, 70)	[50, 60)	[0, 50)

---

### Class Communications

---

The University requires all students to activate an official U of M email account, which should be used for all communications between yourself and the university (including all your instructors). All these email communications should comply with the University's policy on electronic communication with students, which can be found at: [http://umanitoba.ca/admin/governance/governing\\_documents/community/electronic\\_communication\\_with\\_students\\_policy.html](http://umanitoba.ca/admin/governance/governing_documents/community/electronic_communication_with_students_policy.html)

---

### Using Copyrighted Material

---

Please respect copyright and we will use copyrighted content in this course. All course notes, assignments, tests, exams, practice exams and solutions are the intellectual property of your instructor or the Department of Statistics. Reproduction or distribution of these materials is strictly forbidden without their consent. For more information, see the University's Copyright Office website at <http://umanitoba.ca/copyright/orcontactum.copyright@umanitoba.ca>.

---

### Recording Class Lectures

---

Mohammad Jafari Jozani and the University of Manitoba hold copyright over the course materials, presentations and lectures which form part of this course. No audio or video recording of lectures or presentations is allowed in any format, openly or surreptitiously, in whole or in part without permission of Mohammad Jafari Jozani. Course materials (both paper and digital) are for the participant's private study and research. More details are available online at: <http://umanitoba.ca/copyright/>

---

### Student Accessibility Services

---

If you are a student with a disability, please contact Student Accessibility Services (SAS) for academic accommodation supports and services such as note-taking, interpreting, assistive technology and exam accommodations. Students who have, or think they may have, a disability (e.g. mental illness, learning, medical, hearing, injury-related, visual) are invited to contact SAS to arrange a confidential consultation.

Student Accessibility Services, <http://umanitoba.ca/student/saa/accessibility/>  
 520 University Centre, (204) 474-7423, [Student.accessibility@umanitoba.ca](mailto:Student.accessibility@umanitoba.ca)

---

## Academic Integrity

---

It is important that you understand what constitutes academic dishonesty and that you are familiar with the very serious consequences. Links to resources that describe academic dishonesty (including plagiarism, cheating, inappropriate collaboration and examination impersonation, as well as typical penalties) can be found at:

<http://umanitoba.ca/faculties/science/undergrad/resources/webdisciplinedocuments.html>

or

<http://umanitoba.ca/faculties/science/undergrad/resources/webdisciplinedocuments.html>

---

## ROASS Schedule A

---

Schedule "A" of the *Responsibilities of Academic Staff with regards to Students (ROASS)* policies of the University of Manitoba lists resources and policies for students. It is important that you familiarize yourself with these resources and policies. This document will be posted to the Department of Statistics web page and to the UM Learn system.

<http://umanitoba.ca/science/statistics/files/pages/2016/09/Schedule-A-ROASS-Statistics.pdf>