STAT 7240

Advanced Topics in Statistics 1
(*An Introduction to Statistical Learning*)

Fall 2015

Calendar Description

This course is a survey of statistical learning methods and will cover 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. Students completing this course should be able 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.

Prerequisite The only prerequisite is STAT 4100. I will assume that you are familiar with regression analysis and are comfortable with basic probability, statistics, linear algebra, and R programming.

Instructor

Instructor: Dr. Mohammad Jafari Jozani
Office: 325 Machray Hall,
Phone: 272-1563
E-mail: $M\_jafari\_jozani@umanitoba.ca$
Office hours: Mondays & Wednesdays, 10:30 a.m. to 12:00 a.m. or by appointment.

Mark Breakdown

Homework and assignments	20%
Term Test (October 30th, 2015, from 3:00–5:00, Room 316 Machray Hall)	30%
Final Examination (180 minutes, time and place to be announced)	50%

Tentative Topics

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

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

Supervised Learning

- Linear regression: Univariate, multivariate and multiple linear regression. Regression models with qualitative variables. Comparison of regression with K-nearest neighbours.

- Regularized Regression and comparing linear model selections and regularization. Best subset selection. Stepwise selection. Shrinkage method.

- Model Selection and validation. Ridge regression. Regression in High dimensions. Subset selections, etc.

- Resampling Techniques such as the Cross-Validation and Bootstrap.

- Classification using the logistic regression, Linear Discriminant analysis and Quadratic Discriminant analysis methods.

- Moving beyond linearity by working with Polynomial regression and regression Splines.

-(if time permits) Trees and Boosting

Unsupervised Learning

- Clustering methods such as K-means clustering, Hierarchical clustering, understanding dendrograms and different similarity and dissimilarity measures.

- Dimension Reduction techniques. Principal components

- Correlation Analysis

Grade cut-offs

The following are the minimum percentage grades required to receive each of the various letter grades: A+ (90%), A (80%), B+ (75%), B (70%), C+ (65%), C (60%), D (50%).

There is an additional requirement for obtaining a C or a D in the course: to obtain a grade of C or better, you must obtain at least 50% on the final examination; to obtain a D you must obtain at least 40% on the final examination.

Homeworks

Homeworks and Assignments are due at the start of class (time will be announced). Homeworks submitted late will be severely penalized. Homeworks submitted after the solutions are posted or after the graded assignments are return 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:

- Assignments are to be done on 8.5×11 paper, writing on one side only.
- Assignments are to be stapled.
- Write your name at the top of each page.
- Assignments that involve R programming should be accompanied with the R codes.
- 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.
- Assignments that are well presented and properly typesetted in Latex will get up to a 5% bonus.

Tests

The term test and the final examination are closed book. Statistical tables will be provided if required. A non-programmable calculator might be necessary (graphing calculators are not permitted). However, other electronic devices, such as cell phones and MP3, are strictly prohibited.

If you miss the terms test, you will be assigned a mark of zero, unless reasons and acceptable evidence are provided. A make-up test will not be scheduled. Students who miss the term test 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.

Required Text Books

I will have my own notes. However, most of the lecture notes are based on the following text-books that are also available online for free download.

(1) **An Introduction to Statistical Learning with Applications in R**
by: Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani
Springer Texts in Statistics
ISBN 978-1-416-7138-7
2013.

(2) **The Elements of Statistical Learning** (2nd Edition)
by: Trevor Hastie, Robert Tibshiani and Jerome Friedman
Springer Series in Statistics
2009.

Voluntary Withdrawal

Note that the voluntary withdrawal date is November 18, 2014.

Academic Dishonesty

It is important that you understand what constitutes academic dishonesty and that you are familiar with the very serious consequences. Links to resources that that describe academic dishonesty (including plagiarism, cheating, inappropriate collaboration and examination impersonation) can be found at:
`http://www.umanitoba.ca/faculties/science/student/webdisciplinedocuments.html`
or through the Faculty of Science home page at:
`http://umanitoba.ca/faculties/science.`
Typical penalties imposed within the Faculty of Science for academic dishonesty are also described. See also:
`http://crscalprod1.cc.umanitoba.ca/Catalog/ViewCatalog.aspx?pageid=viewcatalog&catalogid=60&chapterid=227&topicgroupid=4056&loaduseredits=False.`

2014-2015 REGISTRATION ADVISORY

Important Note from the Dean of Science:

It is your responsibility to ensure that you are entitled to be registered in this course. This means that you have:

- the appropriate prerequisites, as noted in the calendar description, or have permission from the instructor to waive these prerequisites;
- 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. For example, STAT 1000 cannot be held for credit with STAT 2220.

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 in which you are registered.