SFU SIMON FRASER UNIVERSITY
THINKING OF THE WORLD

STAT 452/652
Statistical Learning and Prediction

Lecture 0: Course Overview

- All course elements are remote, asynchronous recordings (you view lecture materials when you want to).

Course Overview:

- Relatively gentle *introduction* to the essential elements of statistical learning
- Very little math and no advanced computing

- Prerequisites:
- A basic understanding of statistical methods, including regression, ANOVA, tests, and confidence intervals, equivalent to at least two undergraduate courses in Statistics.
  - Your background does not need to be mathematical. The equivalent of STAT 302/602 or STAT 305/605 is sufficient. But you must have had some experience with multiple regression.

- Prerequisites:
- At least some experience with the R programming language and a willingness to self-learn secondary details. Code templates will be provided, requiring some data management and function manipulation. Students with no previous R experience can manage the class, but with potentially much greater difficulty. They would be advised to spend time prior to the start of classes learning the basics of R using any of the available online resources.

VERY APPROXIMATE COURSE PLAN

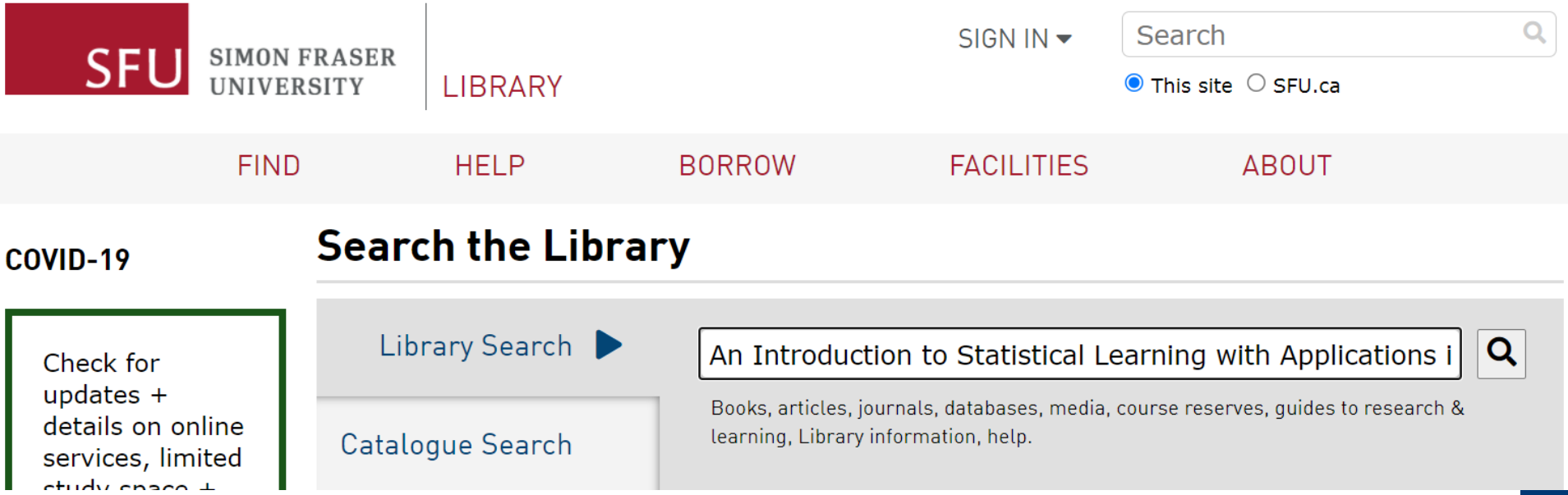
(Plan is likely to be adjusted a little)

Week Ending	Lectures to be covered	Assessments
11-Sep	Syll, Intro, Review Linear Reg	
18-Sep	Review of Linear Reg	Linear Reg 1
25-Sep	Measuring error	Linear Reg 2
02-Oct	Measuring Error	Measuring Error 1
09-Oct	Variable Selection	Measuring Error 2
16-Oct	Variable Selection	Var Select 1
23-Oct	Nonlinear Reg	Var Select 2
30-Oct	Trees, Ensembles	Nonlin Reg
06-Nov	Classification Intro, Linear	Trees, Ensembles
13-Nov	Classification Nonparametric	Project 1
20-Nov	Classification SVM, ROC	Classification 1
27-Nov	Unsupervised PCA, Clustering intro	Classification 2
04-Dec	Clustering	Project 2
11-Dec	Review	
18-Dec		Unsupervised Extra Credit

- *Textbook:*
- *An Introduction to Statistical Learning with Applications in R* (ISLR) by Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani (2013)

SFU SIMON FRASER UNIVERSITY
THINKING OF THE WORLD

- go to lib.sfu.ca



SFU SIMON FRASER UNIVERSITY
THINKING OF THE WORLD

DATABASES

Competition Policy International : the Global Resource for Antitrust and Competition Policy

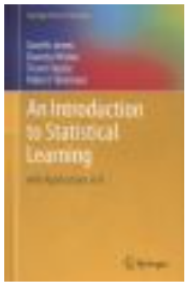
Statistical Abstract of the United States

PEEL in Practice

35 results; see all in Library databases

Details

Authors Gareth James author
Published 2013
ISBN 1-4614-7137-0



Description
Introduction -- Statistical Learning -- Linear Regression -- Classification -- Resampling Methods -- Linear Model Selection and Regularization -- Moving Beyond Linearity -- Tree-Based Methods -- Support Vector Machines --

sfu-primo.hosted.exlibrisgroup.com/.../fulldisplay?adaptor...

BOOKS + CHAPTERS

An introduction to statistical learning: With applications in R

Type Book
Authors James, Gareth ; Witten, Daniela ; Hastie, Trevor ; Tibshirani, Robert
Published 2013
Availability Full text available

An introduction to statistical learning : with applications in R / Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani.

Type Book
Published 2013
Availability Available W.A.C. Bennett Library Stacks (QA 276 I585 2013)

An Introduction to Statistical Learning with Applications in R / by Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani.

Type Book
Authors Gareth James author
Published 2013
Availability Online access

LIBRARY INFORMATION + RESEARCH GUIDES

R: Downloading, learning about, and support for R software

Digital Humanities research guide: Tools & methods

Resources: Digital Humanities Innovation Lab (DHIL)

8 results; see all in Library web pages

MORE OPTIONS

Search Google Scholar

Consult Wikipedia: An Introduction to Statistical Learning with Applications in R

Search all SFU webpages

ASK A LIBRARIAN

Didn't find what you were looking for? Ask a librarian.

Grades....

- **Grade Inflation: Colleges With the Easiest and Hardest Grades**

- https://www.cbsnews.com/news/grade-inflation-colleges-with-the-easiest-and-hardest-grades/

- SFU among the top 10 hardest universities in North America.

- **East**
- Boston University
- MIT
- Princeton
- Rensselaer Polytechnic Institute
- **Midwest/South**
- Auburn University
- Florida International University
- Hampden-Sydney College
- Purdue University
- Roanoke College
- Southern Polytechnic State
- University of Houston
- Virginia Commonwealth University
- **West**
- Cal State University-Fullerton
- Harvey Mudd College
- Reed College
- Simon Fraser University (Canada)

Grading:

- Exam will be open book (hooray!).
- One midterm (30%), one final (35%). ***
- Two projects (10% + 15%)
- Three assignments (10%)

*** Grad student from 652 will have a midterm (30%) and a final report (35%) instead.

Each week I will post:

- Video (Lecture notes)
- Pdf (Lecture notes)
- Video (Tutorial)
- Pdf (Tutorial)

At the end of the lecture note, there are usually some exercises.

- Most of these exercise are optional. Solutions will be provided.
- A few of them will be assignments and there won't be solution posted until the assignments have been graded.

In addition, I will also provide a *review video*

- I will go through the lecture note and let you know :
  - My expectation for you
  - Examples of questions to be expected in the midterm/final.
  - What kinds of questions will NOT be in the midterm/final

Occasionally, I will make some fun videos about the following:

- Data science in general
- Communication skill
- Tips on job seeking
- Interview skill

- Marking Errors:
- TAs try to be consistent during marking but sometimes mistakes still happen.
- Also, students sometimes disagree with the proportion of marks granted for partially correct solutions.

- As an allowance for mistakes that happen over the course of the term, all students in the course will receive 2% bonus marks added in at the end of the term. In exchange, students will not request extra marks on assignments and midterms.
- If a student truly feels that a huge error has been made, they may forfeit these 2 free percentage points and request that their appeal be considered. The 2 points will NOT be returned to the student, regardless of the outcome of the appeal.

- Lets talk about your project 2:

- Project will be in the form of a competition.
- Students will provide predicted values for sets of data that I provide, and write a short report on the process that they took to create the predictions.
- *Grades will be assigned partly competitively, based on the performance of students' prediction models.*

- Students in 652 will do another project in lieu of a final exam.
- They will be responsible for locating an appropriate topic in data science and will provide a written report on their research.
- Topics will be need to be approved by me in advance.

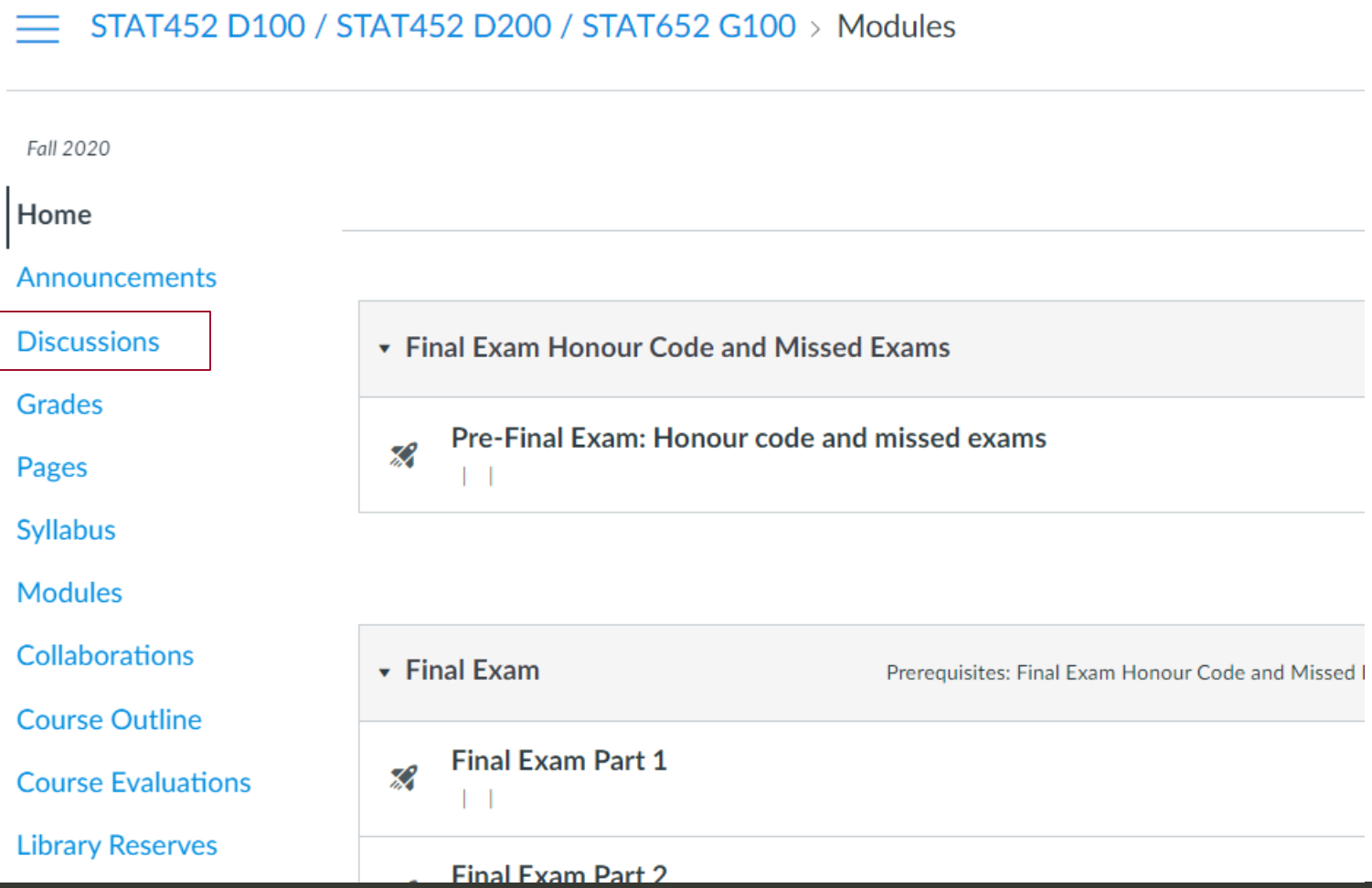
- Lets talk about your project 1:
- Project will be in the form of writing paragraph of answers to interview questions.
- I will find 50 data science interview questions related to the course.
- Each of you will be answering 3 of these questions.
- We will select the best answers and compile the answers into an single answer set for everyone's future use.

About Questions and Answer:

- Try to ask good question that enable us to help you better. I will make a small video on how to ask good questions.

- Me and TAs will have regular office hour (in-person or by zoom) to be announced later.

My preferred way : Canvas Discussion page



My preferred way : Canvas Discussion page

- I am not able to answer questions on course content through emails. There are too many students to make these possible.
- Only email me if you have a personal request (e.g., medical leave) you want to discuss in private.
- Office hour or post your question on Canvas Discussion page

My preferred way : Canvas Discussion page

- Question and answer will be public. Everyone can learn from them.
- I'm not very good at giving "spontaneous" response. I typically need time to think and prepare an answer.
- On average, people absorb and process information 2x faster by reading and writing.

Good luck with the course. It's a good one.