IS5152 Data-driven decision making

Semester 2, 2018/19
Monday, 6.30-8.30 pm, COM1-204
Instructor: Dr. Rudy Setiono (www.comp.nus.edu.sg/~rudys)
Contact: disrudy@nus.edu.sg
Office: COM2 04-13

IS5152 Data-driven decision making

- **Course objective:** to introduce students to decision making technologies that can support decision making in the financial, operational, marketing and other strategic areas.
- **Description:** Data-driven decision making improves productivity and profitability of businesses. This module teaches students decision making techniques based on data analysis. Various machine learning (ML) techniques for data analysis will be presented. The module also discusses aspects related to building an effective model for decision making such as: (i) methods for data preparation such as feature selection, data reduction and sample selection, (ii) metric for determining a good model, (iii) visualization of model performance, (iv) over-fitting and its avoidance. Examples of practical business decision making problems will be used to illustrate the merits of the ML techniques presented.

IS5152 Data-driven decision making

- **Topics covered:**

  The techniques covered in this course include neural networks for classification/regression/clustering, genetic algorithm for optimization, decision tree methods, support vector machine, data envelopment analysis and data mining.

- Journal articles that present relevant techniques for decision making and/or describe successful application of the existing methods in solving practical problems will be discussed in class.

IS5152 Data-driven decision making

This course requires the students to have some background knowledge in:

- Calculus
- Simple linear algebra
- Basic probability and statistics
- See: notation.pdf

No computer programming skill is required.

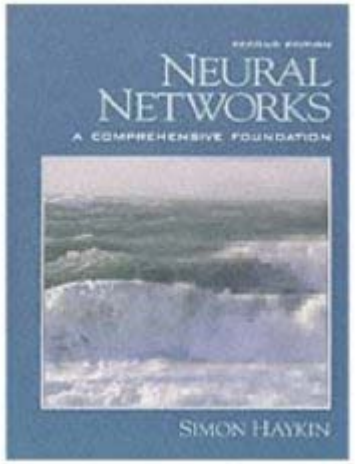
IS5152 Data-driven decision making

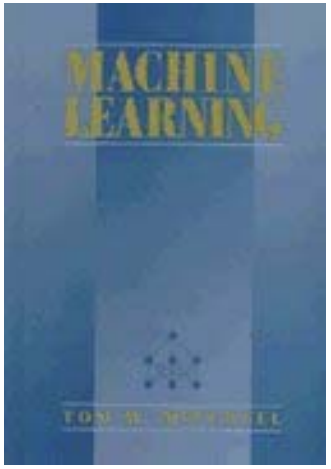
Tentative schedule:

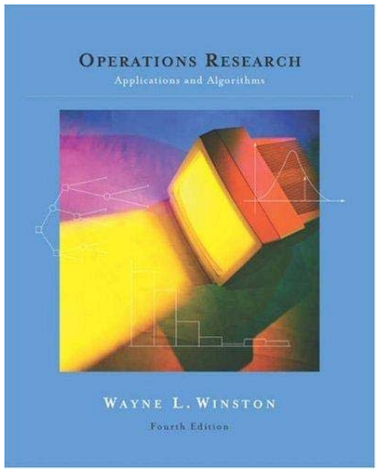
Week 1	**January 17, 2019**	**Introduction and class administration**
Week 2	**January 24, 2019**	**Decision making with decision trees and rules**
Week 3	**January 31, 2019**	**Data preparation and exploration**
Week 4	**February 7, 2019**	**Classification: model development and evaluation**
Week 5	**February 14, 2019**	**Optimization and decision making**
Week 6	**February 21, 2019**	**Data envelopment analysis**
	February 28, 2019	**No lecture. Mid-semester break**
Week 7	**March 7, 2019**	**Support vector machines**
Week 8	**March 14, 2019**	**Neural networks for decision making (Part 1)**
Week 9	**March 21, 2019**	**Neural networks for decision making (Part 2)**
Week 10	**March 28, 2019**	**Decision making with multiple objectives**
Week 11	**April 4, 2019**	**Decision making under uncertainty**
Week 12	**April 11, 2019**	**Genetic algorithm for decision making**
Week 13	**April 18, 2019**	**Course summary, discussion & feedback**

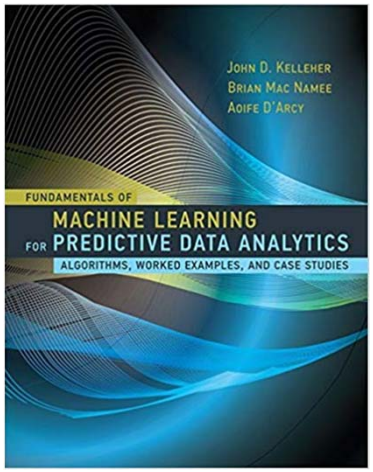
IS5152 Data-driven decision making

References: Available in the RBR sections of Central Library and HSS Business Library. Check IVLE for call/ISBN numbers.









1. Neural networks: A comprehensive foundation
   Author: Haykin, Simon S
2. Machine Learning
   Author: Mitchell, Tom M
3. Operations research : applications and algorithms
   Author: Winston, Wayne L
4. Fundamentals of Machine Learning for Predictive Data Analytics
   Authors: John D. Kelleher, Brian Mac Namee, Aoife D'Arcy.

IS5152 Data-driven decision making

Grading:

1. **Group project: 30%**
2. **Two assignments: 20%**
3. **Final exam on Monday, 6 May 2019, morning: 50%**

The final exam is an <u>open-book</u> examination.

- Do check IVLE for this course regularly for announcements, updates, etc.
- All lecture materials will be placed in the workbin.

IS5152 Data-driven decision making

Group project:

- Objective: to provide students with an experience in data analysis using one or more Decision Making techniques discussed in class.
- What to do:
  - Identify an interesting problem/topic to test one or more of the techniques for decision making discussed in class.
  - Search/find/collect relevant data.

    Use available software to analyze the data.
- Software will be provided or they can be obtained via the internet.

IS5152 Data-driven decision making

Group project:

- Select a problem domain of your interest and pick a data set from the following sites:
  - UCI Machine Learning Repository.
  - Kaggle: Your Home for Data Science
  - DASL: The Data and Story Library
  - A list from StatSci.org
  - Delve datasets from the University of Toronto
  - Data mining data sets from the University of Wisconsin
- You may use your own/any other data sets.
- When selecting a data set, consider how you would analyze the data:
  - classification
  - regression
  - clustering, others.

IS5152 Data-driven decision making

Tools for data analysis (1):

- **A nice website to visit: KDnuggets**
- **Statistics:** SAS and SPSS (in Com1 PL6) , R for Statistical Computing
- General data mining tools: Weka, RapidMiner
- **Optimization:**
  - Minos from Stanford University, run on Unix machine for LP and QP (please contact me if you need this)
  - A large collection of optimisation software can be found at Netlib repository. Search the data base using the keywords: “quadratic program” or “linear program”

IS5152 Data-driven decision making

Tools for data analysis (2):

Support Vector Machines:

- A comprehensive list
- SVMLight from the University of Dortmund, Germany
- SVMTorch from IDIAP, Switzerland
- LIBSVM from National Taiwan University
- **Highly recommended:**

  Weka from Waikato University, New Zealand

  (Note: Weka also includes other machine learning tools)

IS5152 Data-driven decision making

Tools for data analysis (3):

Decision tree methods:

- C4.5 and C4.5rules on SOC Unix machine: send me email if you need this.
- Download your own copy of C4.5 here.
- C5.0 (for Unix) and See5.0 (for Windows) from RuleQuest.

  Also Cubist (for continuous class/regression)
- J48 from Weka is a Java implementation of C4.5.

IS5152 Data-driven decision making

Tools for data analysis (4):

Neural networks:

- A list of neural network software.
- Netlab software from Aston University.
- Tool for clustering using SOM (Self Organizing Map)ToolBox from the Helsinki University of Technology.
- R package for deep learning h2o

IS5152 Data-driven decision making

Project report:

- 20 page limit, single space, single column, 12 point font size.
- Do not include/submit the data set.
- Dateline for submission: Sunday, 21 April, 2019, 11.59 pm.
- Upload softcopy of your report to the course Workbin/Student Submission in IVLE.

IS5152 Data-driven decision making

Content outline of the report (Note this is just a guideline)

1. **Title:** What is this project all about

2. **Introduction:** explain why this problem is important/interesting. How the data has been collected/obtained. What has been done by others in this problem domain, or on the same data set.

3. **Preliminary analysis:** if the data needs to be preprocessed, cleaned, etc. If visualization or simple statistical analysis would give us insightful information about the data.

4. **Full results from analysis:** explain why some particular tool(s) for data analysis is selected. What other approaches can be taken to improve the results in terms of accuracy or other criteria.

5. **Discussion and Summary:** What valuable information has been obtained about the problem and the data through the analysis. What other steps can be taken to improve the findings or to explore the data further.

6. **List of references, not more than 10.**

7. **Appendix:** Screen shots showing the output from your program/software may be included in the Appendix. No limit on the number of pages in the Appendix.

IS5152 Data-driven decision making

- Form a group consisting of 4 or 5 students. You may join a group via IVLE.
- Send me a brief email that includes the names of your group members and a tentative title/topic/description for your project. Do use the discussion forum if you need to find group members.

  Deadline: Friday, 8 February 2019.

- Make an appointment to see me:
  - 1st meeting to discuss your project proposal between 11 - 16 February (15-30 mins)
  - 2nd meeting to discuss the progress of your work in the week of 11 - 16 March.
  - 3rd meeting to discuss your findings before report submission in the week of 8 – 13 April.