MONASH University

ETC3555

Statistical Machine Learning

Introduction

24 July 2018

Prerequisites

- ETC3250 (Business analytics)
  - Statistical learning
  - Regression, classification and clustering
  - Model selection
  - Resampling methods
  - Dimension reduction
  - High-dimensional regression
  - Tree-based methods
- FIT3154 (Advanced data analysis)

Statistical Machine Learning = Business Analytics II

Contact details

- Souhaib Ben Taieb
  - Chief Examiner
  - Room E759, Menzies building
  - Email: souhaib.bentaieb@monash.edu
- Cameron Roach
  - Teaching Associate
  - Email: cameron.roach@monash.edu

Communication

- **Website**
  - `https://github.com/bsouhaib/SML2018`
  - Lecture notes
- **Moodle**
  - `https://moodle.vle.monash.edu/course/view.php?id=45443`
  - Forum for asking questions, etc.
  - Assignment submissions
  - **No email please — use the forum**

Learning goals

The learning goals associated with this unit are to:

1. identify and understand the statistical and computational trade-offs in modern data analysis problems;
2. develop computer skills for exploring modern data sets;
3. understand and apply machine learning algorithms to solve modern data analysis problems.

Learning goals

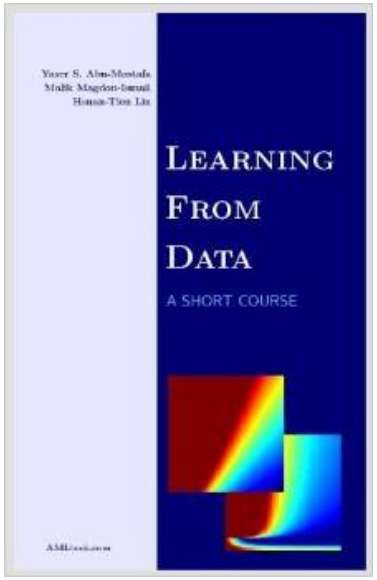
This unit covers the methods and practice of statistical machine learning for modern data analysis problems. Topics covered will include **recommender systems**, **social networks**, **text mining**, **matrix decomposition and completion**, and **sparse multivariate methods**. All computing will be conducted using the **R programming language**.

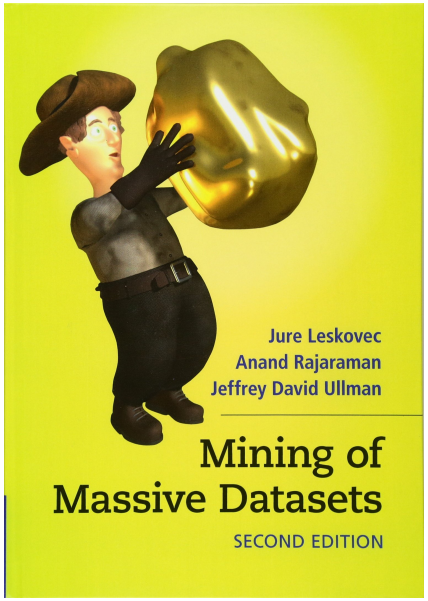
Teaching and learning approach (12 weeks)

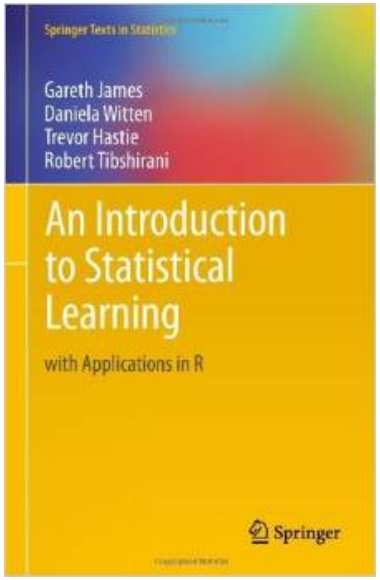
Two 60 minutes lectures (Tues 9am; Wed 10am)

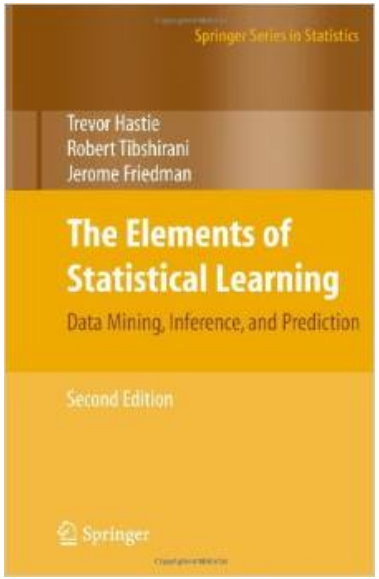
One 90 minutes tutorial (Wednesday 12:30pm)

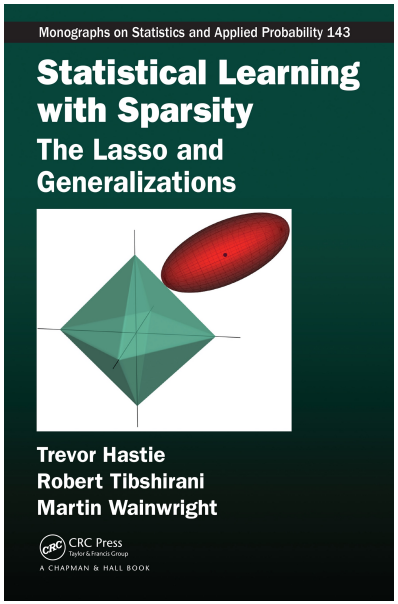
Many references

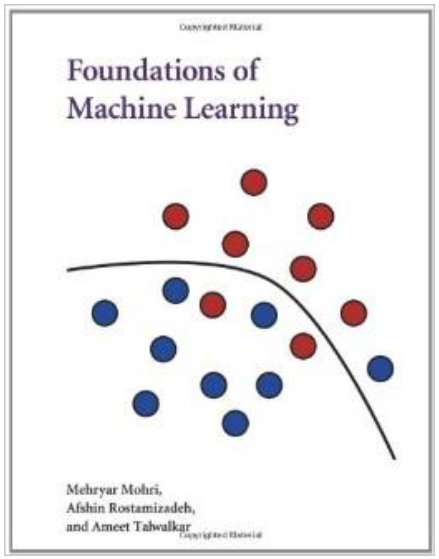


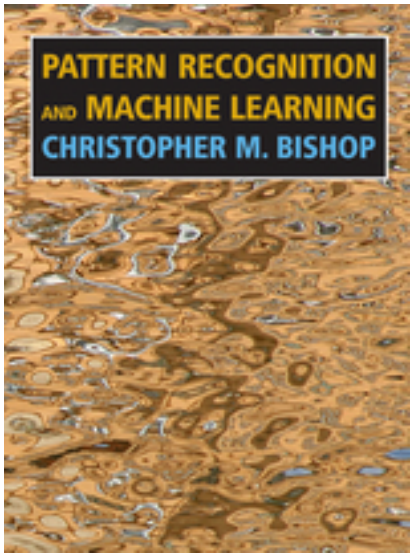


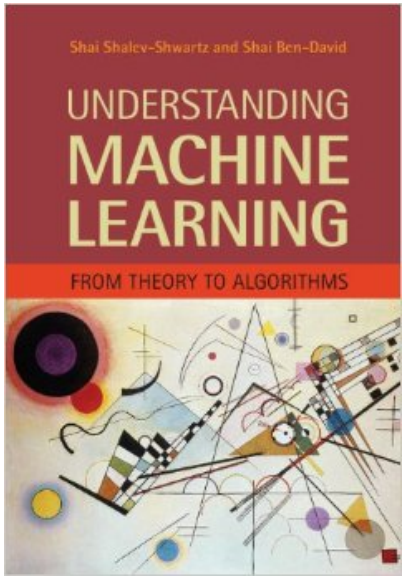












Key references

- Yaser Abu-Mostafa, Malik Magdon-Ismail and Hsuan-Tien Lin (2012). **Learning from data**. AMLBook. `amlbook.com`. Video lectures available.
- Julia Silge and David Robinson (2018). **Text Mining with R: A Tidy Approach**. `www.tidytextmining.com`.
- Jure Leskovec, Anand Rajaraman, Jeff Ullman. **Mining of Massive Datasets**. `www.mmds.org`. Video lectures available.
- James, Witten, Hastie and Tibshirani (2012) **An Introduction to Statistical Learning**. Springer. `www.statlearning.com`. Video lectures available.
- Trevor Hastie, Robert Tibshirani and Jerome Friedman (2017). **The Elements of Statistical Learning**. Springer. `web.stanford.edu/~hastie/ElemStatLearn`.
- Martin Wainwright, Robert Tibshirani et Trevor Hastie (2015). **Statistical Learning with Sparsity**. `web.stanford.edu/~hastie/StatLearnSparsity`.

Outline (subject to change)

Week	Topic
1	Introduction/The learning problem
2	The learning problem
3	Neural networks
4	Deep neural networks
5	Support vector machines
6	Kernel methods
7	Recommender systems and matrix completion
8	Text mining
9	Text mining
	Semester break
10	Social networks
11	Social networks
12	Project presentation

Assessment

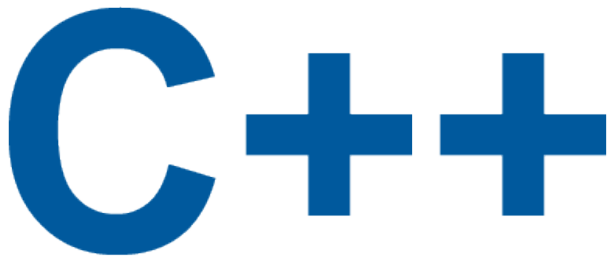
- Final exam (*Open book*) (2 hours): **60%**
- One project due at the end of the semester (group of 2 students): **20%**
- Five fortnightly assignments (Individual): **20%** (4% each)

Task	**Due Date**	**Value**
Final exam	Official exam period	60%
Project	Fri 19 October (Available end of August)	20%
Assignments 1–5	Sunday 11:55pm	20%

Assignments

	Release Date (Wednesday)	Due Date (Sunday next week)
1	1 August	12 August, 11:55pm
2	15 August	26 August, 11:55pm
3	29 August	9 September, 11:55pm
4	12 September	23 September, 11:55pm
5	3 October	14 October, 11:55pm

Programming languages



R

RStudio