Lecture 1a. StatML Welcome

COMP90051 Statistical Machine Learning

Sem2 2020
Lecturer: Ben Rubinstein

THE UNIVERSITY OF MELBOURNE

This lecture

- **About COMP90051**
- Review: Probability theory
- Review: Linear algebra
- Review: Sequences and limits

Subject objectives

- Develop an appreciation for the role of statistical ML, advanced foundations and applications
- Gain an understanding of a representative selection of ML techniques – *how ML works*
- Be able to design, implement and evaluate ML systems
- Become a discerning ML consumer

Subject content

30%+ new content

- The subject will cover topics from

  Foundations of statistical learning, linear models, non-linear bases, regularised linear regression, generalisation theory, kernel methods, deep neural nets, multi-armed bandits, Bayesian learning, probabilistic models

- Theory in lectures; hands-on experience with range of toolkits in workshop pracs and projects
- vs COMP90049: much depth, much rigor, so wow

Subject staff / Contact hours

Contacting staff	*Discussion board first; then combined staff email* **comp90051-2020s2-staff@lists.unimelb.edu.au**
Lecturer & Coordinator	Ben Rubinstein Associate Prof, Computing & Information Systems Associate Dean (Research), Melbourne School of Engineering *Statistical Machine Learning, ML + Privacy/Security/Databases*
Lecturer	Qiuhong Ke Lecturer, Computing & Information Systems *Computer Vision, ML, Deep Learning*
Tutors:	Neil Marchant (Head Tutor) Justin Tan, Jun Wang, Rui Zhang. *See Canvas for latest list and contact details.*
Zoom Contact:	*Weekly, please attend: 2nd Lecture (live discussion), 1 Workshop*
Pre-recorded Lectures:	*Posted to Canvas for you to view safely at home.* Strongly recommend that you keep up, weekly. (viz. quizzes)

About me (Ben)

- PhD 2010 – Berkeley, USA
- 4 years in industry research
  - Silicon Valley: Google Research, Yahoo! Research, Intel Labs, Microsoft Research
  - Australia: IBM Research
  - Patented & Published, Developed & Tested, Recruited
- Impact: Xbox, Bing (MS), Firefox (Mozilla), Kaggle, ABS, Medicare and Myki data privacy
- Interests: machine learning theory; adversarial ML; differential privacy; statistical record linkage

Advanced ML: Expected Background

- Why a challenge: Diverse math + CS + coding
- ML: COMP90049 either 2020s1 “new” or earlier (we’ll review gaps throughout semester)
- Alg & complexity: big-oh, termination; basic data structures & algorithms; solid coding ideally experience in Python

…and more…

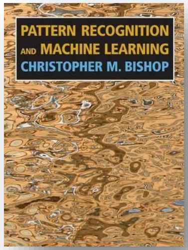
Advanced ML: Expected Background

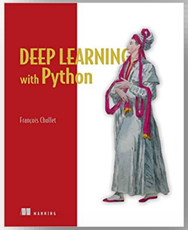
…and more…

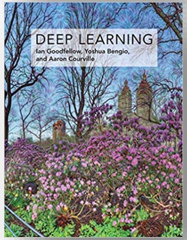
- Maths: Review next videos, but ideally seen most before
  "Matrix **A** *is symmetric & positive definite, hence its eigenvalues…"*
- Probability theory: probability calculus; discrete/continuous distributions; multivariate; exponential families; Bayes rule
- Sequences: sequences, limits, supremum
- Linear algebra: vector inner products & norms; orthonormal bases; matrix operations, inverses, eigenvectors/values
- Calculus & optimisation: partial derivatives; gradient descent; convexity; Lagrange multipliers

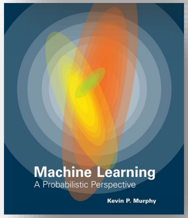
Textbooks

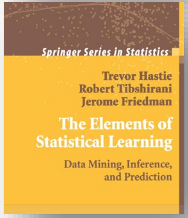
- We don’t have only one reference. We prefer to pick good bits from several. We may also supplement with other readings as we go.
- All are available free online or through the library digitally. See the Canvas lecture outline for links. Therefore, no need to buy.
- Primarily we refer to (good all rounder): Bishop (2007) *Pattern Recognition and Machine Learning*
- Practical Deep Nets: Chollet (2017) *Deep learning with Python*
- More deep learning detail: Goodfellow, Bengio, Courville (2016) *Deep learning*
- For more on PGMs/Bayesian inference: Murphy (2012) *Machine Learning: A Probabilistic Perspective*
- For reference on frequentist ideas, SVMs, lasso, etc.: Hastie, Tibshirani, Friedman (2001) *The Elements of Statistical Learning: Data Mining, Inference and Prediction*











Assessment

- Assessment components
  - Two projects – one group (w4-7), one individual (w9-11)
    - Each (30%)
    - Each has ~3 weeks to complete
  - Final Exam (40%)

- 50% hurdles applied to both **exam** and **combined project**

- Ungraded semi-weekly **quizzes**. Completion expected that week, please

Summary

- **About COMP90051**
- Review: Probability theory
- Review: Linear algebra
- Review: Sequences and limits

Homework week #1: Watch all week 1 recordings.
Jupyter notebooks setup and launch (at home)