COMP90051 **Statistical Machine Learning**

Semester 2, 2015

Lecturer: Ben Rubinstein

1. Why Learn, This Subject

THE UNIVERSITY OF MELBOURNE

Why Learn Learning?

Motivation

- *“We are drowning in data,*
  but we are starved for knowledge”
  - John Naisbitt, *Megatrends*
- Data = raw information
- Knowledge = patterns or models behind the data

Solution: Machine Learning

- Hypothesis: pre-existing data repositories contain a lot of potentially valuable information
- Mission of learning: find it
- Definition of learning:

  (semi-)automatic extraction of **valid**, **novel**, **useful** and **comprehensible** knowledge – in the form of rules, regularities, patterns, constraints or models – from arbitrary sets of data

Draws on Many Disciplines

- Artificial Intelligence
- Statistics
- Continuous optimisation
- Databases
- Information Retrieval
- Communications/information theory
- Signal Processing
- Computer Science Theory
- Philosophy
- Psychology and neurobiology

…

Applications of ML are Deep and Prevalent

- Online ad selection and placement
- Risk management in finance, insurance, security
- High-frequency trading
- Medical diagnosis
- Mining and natural resources
- Malware analysis
- Drug discovery
- Search engines

...

Jobs

Many companies across all industries hire ML experts:

Data Scientist
Analytics Expert
Business Analyst
Statistician
Software Engineer
Researcher
…

All roads lead to...

- (Maths) Statistics
  - 16th century: Probability
  - 18th century: Early Bayesian work
  - 20th century: Rise of frequentists
  - 1980s: Boost to Bayesians from MCMC
- (CS+) AI → ML, etc.
  - 1940s Perceptrons
  - 1960s-70s Logic programming
  - 1980s Backprop, PAC, Bayes nets
  - 2000s+ SVMs, Graphical models, Transfer learning, Relational learning, Online learning, Deep learning, ...
- (EE) Signal processing
  - Roots in numerical, WWII
  - 1960-70s FFT, filter design
  - 1980-90s sensing, imaging
- (CS) Complementary ideas
  - Databases: Data mining, statistical DBs, probabilistic DBs
  - IR: PageRank
  - Systems: MapReduce, Hadoop

Terminology

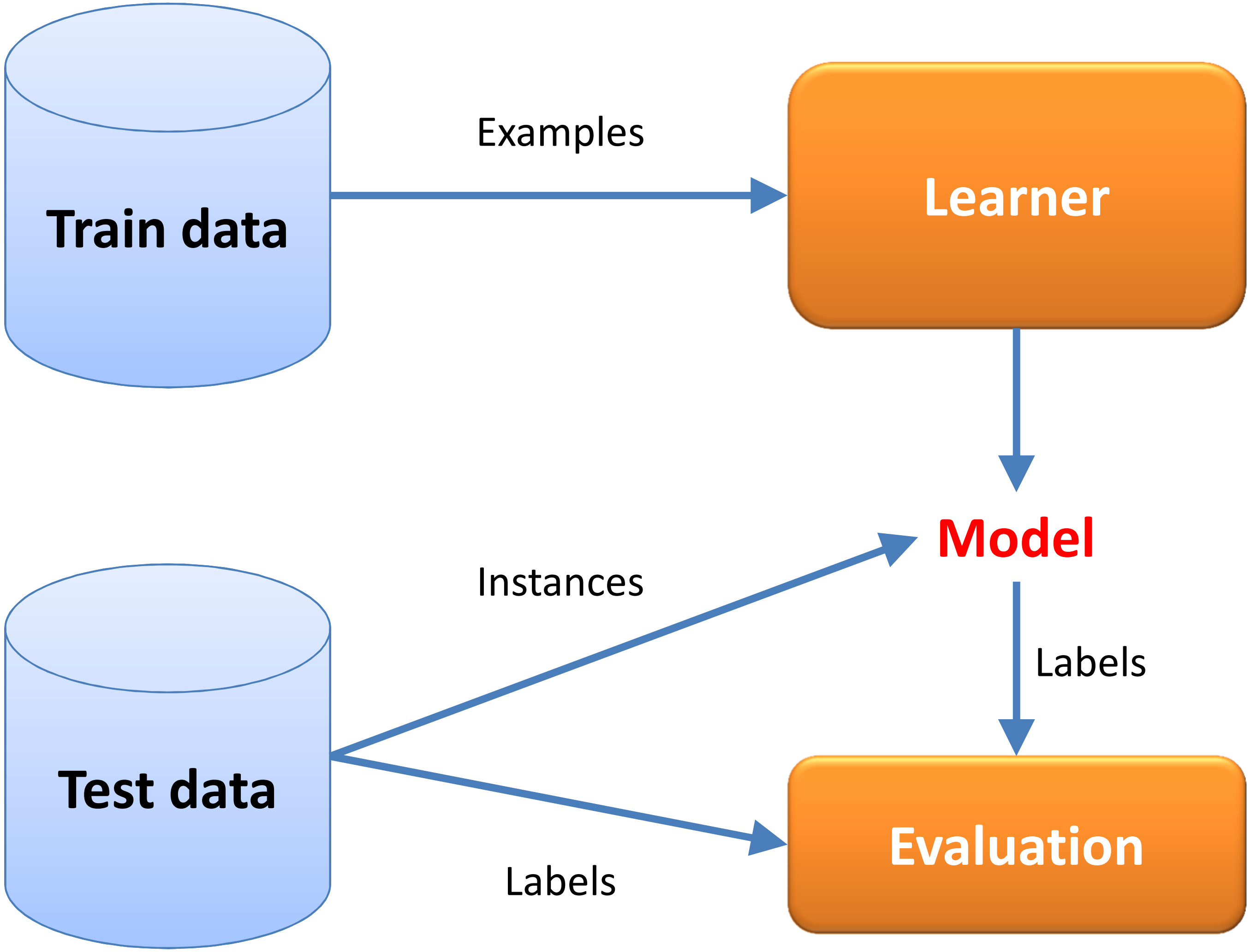
- Input to a machine learning system can consist of
  - Instance: measurements about individual entities/objects
    a loan application
  - Attribute: component of the instances
    the applicant's salary, number of dependents, etc.
  - Label: an outcome that is categorical, numeric, etc.
    forfeit vs. paid off
  - Examples: instance coupled with label
    <(100k, 3), "forfeit">
  - Models: discovered relationship between attributes, label

Supervised vs Unsupervised Learning

- Training data: used to construct models

	Training data	Model used for
Supervised learning	Labelled	Predict labels on new instances
Unsupervised learning	Unlabelled	Cluster related instances; Understand attribute relationships

Architecture of a Supervised Learner



Evaluation (Supervised Learners)

- How you measure quality depends on your problem!
- Typical process
  - Pick an evaluation metric comparing label vs prediction
  - Procure an independent, labelled test set
  - “Average” the evaluation metric over the test set
- Example evaluation metrics
  - Accuracy, Contingency table, Precision-Recall, ROC curves
- When data poor, cross-validate

About this Subject

Vital Statistics

Lecturers:	Ben Rubinstein (DMD7.21, brubinstein@unimelb.edu.au)
Weeks 1-5	Senior Lecturer, Dept Computing & Information Systems
	Statistical Machine Learning, Databases, Security/Privacy
	Andrey Kan (andrey.kan@unimelb.edu.au)
Weeks 6-12	Research Fellow, Walter and Eliza Hall Institute
	ML, Computational immunology, Medical image analysis
Tutors:	Andrew Bennett a.bennett2@student.unimelb.edu.au
	Yuxuan Li yuxuan.li@unimelb.edu.au
Contact:	*Weekly you should attend 2x Lectures, 1x Workshop*
Office Hours	*Wednesdays 12-1pm, 7.02 DMD Building*
Website:	http://www.lms.unimelb.edu.au

About Ben

- PhD EECS 2010 – UC Berkeley
- 4 years in industry research
  - Silicon Valley: Google Research, Yahoo! Research, Intel Labs, Microsoft Research
  - Australia: IBM Research
  - Patented & Published, Developed & Tested, Recruited
  - Impacted: Xbox, Bing (MS), Firefox (Mozilla), Kaggle
- Expertise: Basic research in machine learning; information integration; adversarial learning/privacy

Subject Content

- The subject will cover topics from

  Foundations of statistical learning, frequentist vs Bayesian approaches, linear models, ensemble methods, trees and forest models, probabilistic graphical models (HMMs, LDA, MRF), kernel approaches, neural networks, cluster analysis, dimensionality reduction, regularisation, social network analysis, semisupervised and active learning

- We will gain hands-on experience with all of this via a range of toolkits, workshop pracs, and projects

Subject Objectives

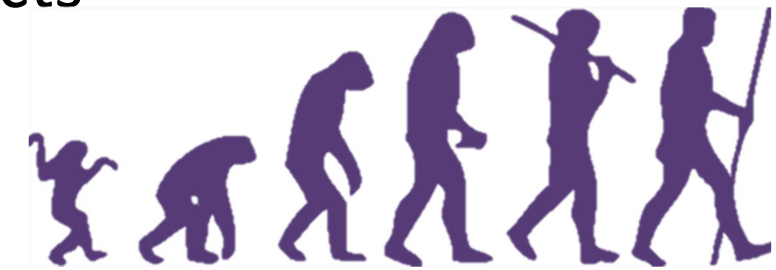
- Develop an appreciation for the role of statistical machine learning, both in terms of foundations and applications
- Gain an understanding of a representative selection of ML techniques
- Be able to design, implement and evaluate ML systems
- Become a discerning ML consumer

Subject Outcomes

- Upon successful completion, students will be able to
  - Describe a range of statistical machine learning algorithms
  - Design, implement and evaluate statistical learning systems to solve real-world problems, based on an appreciation of their relative suitability to different tasks

Evolution of the Subject

- Pre-history: separate CS subjects
  - Machine Learning
  - Neural and Evolutionary Comp

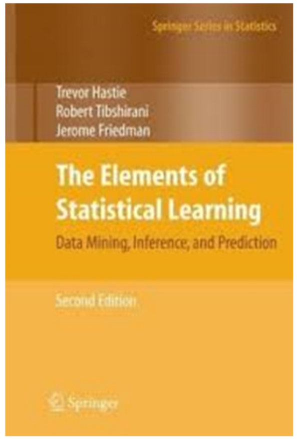


M. Garde

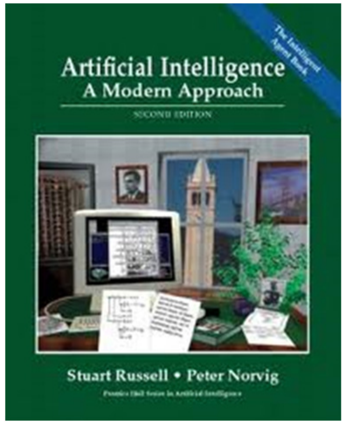
- Modern times: merged
  - Statistical Learning, Evolutionary Algorithms
  - Neural networks belonged, but genetic algorithms didn't
- 2015: the future has arrived!
  - Focus on machine learning
  - Foundations to unify presentation of techniques

Textbooks

- "Required" (more as a reference)
  - Haste, Tibshirani, Friedman (2001) *The Elements of Statistical Learning: Data Mining, Inference and Prediction* [free at http://www-stat.stanford.edu/~tibs/ElemStatLearn]



- Another good reference (more broad less deep)
  - Russell, Norvig (1995) *Artificial Intelligence: A Modern Approach*



Textbooks

- Reference for further general reading
  - Bishop (2007) *Pattern Recognition and Machine Learning*
- Reference for further deeper reading
  - Koller, Friedman (2009) *Probabilistic Graphical Models: Principles and Techniques*
  - Shawe-Taylor, Cristianini (2004) *Kernel Methods for Pattern Analysis*

Assessment

- Assessment components
  - Project 1 – Released week 3; due week 6
  - Midsemester Test (optional/bonus) – Week 7 (in class, Sep 9)
  - Project 2 – Released week 9; due week 11
  - Final Exam
- Breakdown
  - 50% Exam
  - Best of: (10% Test + 20% P1 + 20% P2) OR (25% P1 + 25% P2)

Assumed Knowledge

(Week 1 Workshop Revises COMP90049)

- Programming
  - Required: proficiency programming a high-performance language like Java, python for data processing in projects
  - Ideal: exposure to Matlab or R
- Maths
  - Familiarity with formal notation $\mathbf{Pr}(\boldsymbol{x}) = \sum_{\boldsymbol{y}} \mathbf{Pr}(\boldsymbol{x}, \boldsymbol{y})$
  - Familiarity with probability (Bayes rule, marginalisation)
  - Exposure to optimisation (gradient descent)
- ML: decision trees, naïve Bayes, kNN, kMeans

Summary

- Why study machine learning; Basics
- Machine learning vs. Universe