COMPSCI762: Advanced Machine Learning

Introduction

Jörg Simon Wicker

The University of Auckland

THE UNIVERSITY OF AUCKLAND
Te Whare Wānanga o Tāmaki Makaurau
NEW ZEALAND

SCIENCE
SCHOOL OF COMPUTER SCIENCE
MACHINE LEARNING

Introduction

Introduction

- Jörg Simon Wicker
  - `j.wicker@auckland.ac.nz`
  - Office hours: Thursday 4pm-5pm (changed!)
    - Online on request, send me a mail and we can meet on Zoom
  - Room: 303-526
  - `https://wicker.nz`
  - `https://ml.auckland.ac.nz`
- My name:
  - To write: ö = **oe** ⇒ Joerg
  - Pretty close: *work* with a *y* instead of a *w*
  - Feel free to use my middle name (Simon)! (My partner does the same)

Introduction

- Diplom (M.Sc.) in Bioinformatics
  - LMU & TUM in Munich
  - *Bioinformatics*
- PhD in Computer Science
  - TUM, *Large Classifier Systems in Bio- and Cheminformatics*
- PostDoc
  - Johannes Gutenberg University Mainz
- Founding CEO, now CTO of enviPath
- since August 2017
  - (Senior) Lecturer at UoA – Machine Learning, Cheminformatics, Computational Sustainability

Introduction

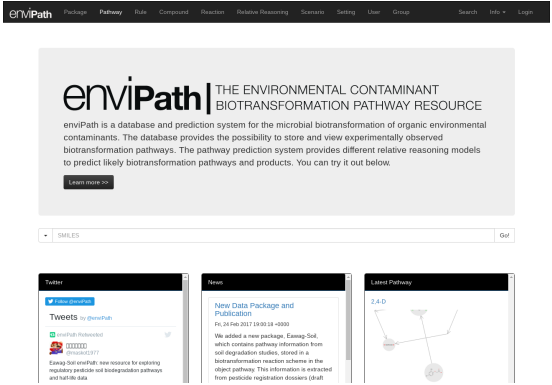
- Diplom (M.Sc.) in Bioinformatics
  - LMU & TUM in Munich
  - *Bioinformatics*
- PhD in Computer Science
  - TUM, *Large Classifier Systems in Bio- and Cheminformatics*
- PostDoc
  - Johannes Gutenberg University Mainz
- Founding CEO, now CTO of enviPath
- since August 2017
  - (Senior) Lecturer at UoA – Machine Learning, Cheminformatics, Computational Sustainability

Introduction

- Diplom (M.Sc.) in Bioinformatics
  - LMU & TUM in Munich
  - *Bioinformatics*
- PhD in Computer Science
  - TUM, *Large Classifier Systems in Bio- and Cheminformatics*
- PostDoc
  - Johannes Gutenberg University Mainz
- Founding CEO, now CTO of enviPath
- since August 2017
  - (Senior) Lecturer at UoA – Machine Learning, Cheminformatics, Computational Sustainability

Introduction

- Diplom (M.Sc.) in Bioinformatics
  - LMU & TUM in Munich
  - *Bioinformatics*
- PhD in Computer Science
  - TUM, *Large Classifier Systems in Bio- and Cheminformatics*
- PostDoc
  - Johannes Gutenberg University Mainz
- Founding CEO, now CTO of enviPath
- since August 2017
  - (Senior) Lecturer at UoA – Machine Learning, Cheminformatics, Computational Sustainability



Introduction

- Diplom (M.Sc.) in Bioinformatics
  - LMU & TUM in Munich
  - *Bioinformatics*
- PhD in Computer Science
  - TUM, *Large Classifier Systems in Bio- and Cheminformatics*
- PostDoc
  - Johannes Gutenberg University Mainz
- Founding CEO, now CTO of enviPath
- since August 2017
  - (Senior) Lecturer at UoA – Machine Learning, Cheminformatics, Computational Sustainability

My research

- Generally, Machine Learning and its interdisciplinary application
  - Computational Sustainability
  - Many projects in chemistry
  - Adversarial Learning
  - Bias
  - Multi-Label Learning
  - Privacy
  - Boolean Matrix Decomposition
  - Time Series
  - ...

UoA Machine Learning Group

https://ml.auckland.ac.nz

UoA Machine Learning Group – Activities

- UoA Machine Learning Seminars
  - International speakers, usually online
  - Currently scheduled for this year, exact times tba
- Milo student seminars
  - Bi-weekly Thursday 2pm
  - Student presentations of their research and discussions about recent papers
- Social events
  - Informal meeting of anyone interested in ML
  - ML lunch
    - Tuesdays 12pm
  - ML Social Drinks
    - Thursdays 5pm
- Great way to connect to researchers and students interested in ML – feel free to join, have a chat with me if you want to know more

`https://ml.auckland.ac.nz/events`

Lecture

What I think you should take out of this course...

- Machine Learning is a continuously changing field
  - Libraries, programming languages, algorithms change
- Important to understand and know the core concepts
- Don't rely on just one paradigm
- There is no "best" model
- The theory is important to understand to avoid that you lie to yourself
- You need to be able to learn concepts, not know everything about every concept

Today we will cover...

What is Machine Learning and Data Mining?

Definition of Machine Learning

Deep Learning

Applications

Partially based on Slides from University of British Columbia

What is Machine Learning and Data Mining?

First...

- What is Machine Learning? What is Data Mining?

Age of Data / Big Data / ...

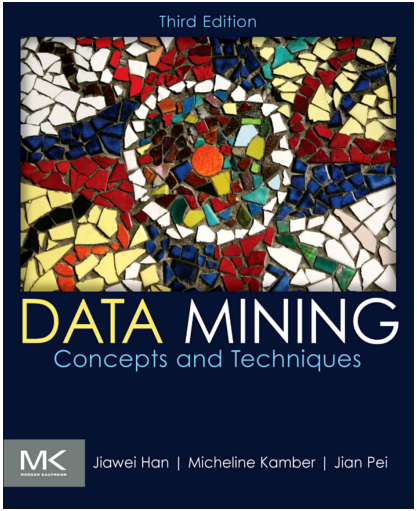
- We are collecting and storing data at unprecedented rate
  - Youtube, Facebook, news sites
  - Credit card transactions and Amazon purchases
  - Transportation data (Google Maps, Uber, Auckland Transport)
  - Gene expression data, protein interaction assays, chemistry pathways

Age of Data / Big Data / ...

- What can you with this data?
  - Too much data to search through it manually?
- But there is valuable information in the data
  - How can we use it for fun, profit, and/or the greater good?
- **Machine Learning** and **Data Mining** are key tools we use to make sense of large data sets!

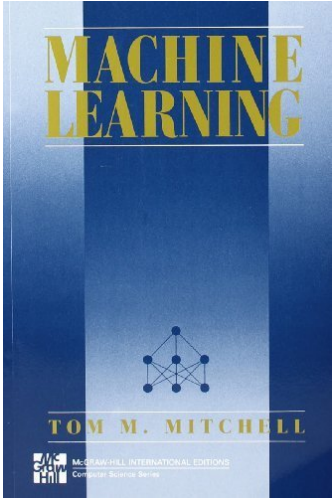
Data Mining

- Automatically extract useful knowledge from large data sets
- Often to help human decision making



Machine Learning

- Using computers to automatically detect patterns in data and use these to make predictions or decisions
- Most useful when
  - We want to automate something a human can do
  - We want to do things a human can't do (looking through 1TB of data)



Definition of Machine Learning

Definitions

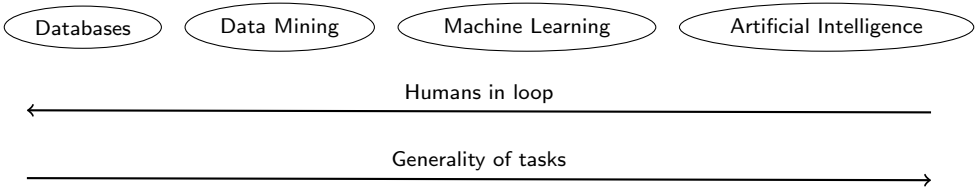
- *"Machine Learning is the field of study that gives the computer the ability to learn without being explicitly programmed."* – Arthur Samuel (1959)
- *"A computer is said to* **learn** *from experience E with respect to some class of task T and performance measure P, if its performance at tasks in T, as measured by P, improved with experience E"* – Tom Mitchell

Example Problem – Self-Driving Cars

- T: Driving on public highway using vision sensors
- P: Average distance traveled before an error / crash
- E: A sequence of images and steering commands recorded while observing a human driver

Machine Learning vs. Data Mining

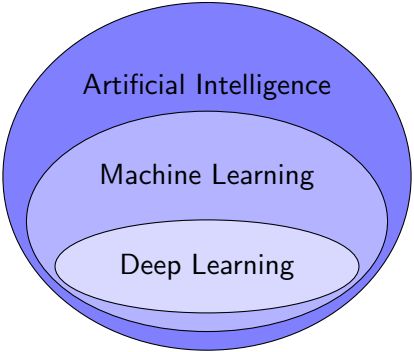
- Data Mining and Machine Learning are very similar:
  - Data Mining often viewed as closer to databases
  - Machine Learning often viewed as closer to AI
- Both are similar to statistics, but more emphasis on
  - Large data sets and computation
  - Predictions (instead of descriptions)
  - Flexible models (that work on many problems)



Deep Learning

How about AI and Deep Learning?

- Traditionally, we've viewed ML as a subset of AI
  - And "Deep Learning" as a subset of Machine Learning



Applications

Applications

- Which applications of Machine Learning do you know?

Applications

- Spam Filtering
- Credit card fraud detection
- Product recommendation
- Motion capture
- OCR
- Speech recognition
- Face recognition
- Object detection
- Sports analytics

- Medical imaging
- Self-driving cars
- Image annotation
- Discovering new cancer subtypes
- Protein folding prediction
- Drug discovery
- Chess / Go / Starcraft / ...
- ...

Summary

- There is a lot you can do with a bit of statistics and a lot of data / computation
- Major recent progress in fields like speech recognition. computer vision, biology, ...
- Things are changing a lot on the timescale of 3-5 years
- ML investment bubble
- **But** it is important to know the **limitations** of what you are doing
  - *"The combination of some data and an aching desire for an answer does not ensure that a reasonable answer can be extracted from a given body of data"* – John Tukey
  - A huge number of people applying ML are just "overfitting"
    - Or don't understand the assumptions needed for them to work
    - Their methods do not work when they are released "into the wild"

Literature

- Machine Learning – Tom Mitchell
- Pattern Recognition and Machine Learning – Christopher Bishop
- Data Mining – Jiawei Han, Micheline Kamber, Jian Pei
- Data Mining – Ian Witten, Eibe Frank, Mark Hall, Christopher Pal

THE UNIVERSITY OF AUCKLAND
Te Whare Wānanga o Tāmaki Makaurau
NEW ZEALAND

SCIENCE
SCHOOL OF COMPUTER SCIENCE
MACHINE LEARNING

Thank you for your attention!

https://ml.auckland.ac.nz
https://wicker.nz