Lecture 1: Introduction and Overview

COMP90049
Introduction to Machine Learning

Semester 2, 2020

Hadi Khorshidi, CIS

The presentation adapted from the slides prepared by Lea Frermann, CIS

Roadmap

This lecture

- Warm-up
- Housekeeping COMP90049
- Machine Learning

Warm-up

What is learning?

Think about this concept

What is learning?

Task or Goal

- acquiring knowledge
- acquiring skills
- getting experienced

Learning Algorithm

- learning by instruction (e.g., maths // academic: lectures)
- learning by experience (e.g., riding the bike // academic: projects)
- learning by observation / imitation (e.g., (first) language learning)

Data

- from data (books, videos, interaction, play)
  ... to skills (maths, dancing, cooking, talking)

THE UNIVERSITY OF MELBOURNE

What is machine learning?

Some proposed definitions...

"The computer automatically learns something"

"Statistics, plus marketing"

" ... how to construct computer programs that automatically improve with experience .... A computer program is said to learn from experience ... if its performance ... improves with experience... "
Mitchell [1997, pp. xv-17]

What is machine learning?

"We are drowning in information, but we are starved for knowledge"

John Naisbitt, Megatrends

Our definition of Machine Learning

automatic extraction of **valid, novel, useful and comprehensible knowledge** (rules, regularities, patterns, constraints, models, ...) from arbitrary sets of data

What is machine learning?

Learning what?

- **Task** to accomplish a goal, e.g.,
  - Assign continuous values to inputs (essay $\rightarrow$ grade)
  - Group inputs into known classes (email $\rightarrow$ $\{$spam, no-spam$\}$)
  - Understand regularities in the data

Learning from what?

- **Data**
- Where do the data come from? Is it reliable? Representative?

How do we learn?

- define a **model** that explains how to get from input to output
- derive a **learning algorithm** to find the best model parameters

How do we know learning is happening?

- The algorithm improves at its task with exposure to more data
- We need to be able to **evaluate** performance objectively

About COMP90049

About me (Hadi)

- PhD 2016 – Monash University
- 1.5 years industry
- 2 years research fellow
- Interests:
  1. Machine learning (ML) and Optimisation,
     E.g., Medical problems: health service patterns, risk of surgery,
     E.g., Imbalanced data: synthetic over-sampling
  2. Mathematical modeling,
     E.g.: Simulate the processes and find the optimal solutions
  3. Uncertainty capturing and quantification,
     E.g.: Uncertain data and missing values

About Lida

- PhD 2017 – University of Melbourne
- 3 years of research in academia
- Interests:
  1. Graph Mining and Social Network Analysis,
     E.g., Personalisation through Role Discovery in Social Networks,
     E.g., Anomaly Detection in Dynamic Networks,
     E.g., Sentiment Analysis in Twitter datasets
  2. Measurement Analysis in Information Retrieval and specifically Search Engines Evaluation

THE UNIVERSITY OF MELBOURNE

Vital Information

Who?

Lecturer 1	Hadi Khorshidi DMD 812, hadi.khorshidi@unimelb.edu.au Research Fellow, Computing & Information Systems
Coordinator & Lecturer 2	Lida Rashidi DMD 813, rashidi.l@unimelb.edu.au Lecturer & Postdoc, Computing & Information Systems
Head Tutor	Hasti Samadi hasti.samadi@unimelb.edu.au
Tutors	Tahrima Hashem, Pei-Yun Sun, Kazi Adnan, Oscar Correa, Hasti Samadi

Vital Information

When and Where?

Lecture 1	Wed 16:15-17:15 Q&A sessions
Lecture 2	Thu 14:15-15:15 Q&A sessions
Recorded Lectures	LMS under Lecture Capture Hadi, Lida, Lea Frermann and Qiuhong Ke
Workshops	Online

Workshops start from week 2

Vital Information

Lectures

- Theory
- Derivation of ML algorithms from scratch
- Motivation and context
- Some coding demos in Python

Workshops

- Practical exercises
- Working through numerical examples
- Revising theoretical concepts from the lectures

Vital Information

Subject Materials and Communication

- All materials will be made available through LMS (Canvas)
- **Discussion board:** first point of access for any content-related questions. Also: separate forum specific for remote students
- Back-up: email Hasti
- Back-up 2: email lecturer

Subject Content

- Topics include: classification, clustering, semi-supervised learning, association rule mining, anomaly detection, optimisation, neural networks
- All from a theoretical and practical perspective
- Refreshers on maths and programming basics
- Theory in the lectures (some coding demos)
- Hands-on experience in workshops and projects
- Guest lecture 1: academic writing skills
- Guest lecture 2: ML in the industry

Expected Background

Programming concepts

- We will be using **Python** and **Jupyter Notebooks**
- Basic familiarity with libraries (numpy, scikit-learn, scipy)
- You have to be able to write code to process your data, apply different algorithms, and evaluate the output
- Implementation itself is secondary (this is not a programming class!)

Expected Background

Programming concepts

- We will be using **Python** and **Jupyter Notebooks**
- Basic familiarity with libraries (numpy, scikit-learn, scipy)
- You have to be able to write code to process your data, apply different algorithms, and evaluate the output
- Implementation itself is secondary (this is not a programming class!)

Mathematical concepts

- formal maths notation
- basic probability, statistics, calculus, geometry, linear algebra
- (why?)

What Level of Maths are we Talking?

$$\ln \frac{P(y = \texttt{true}|x)}{1 - P(y = \texttt{true}|x)} = w \cdot f$$

$$\frac{P(y = \texttt{true}|x)}{1 - P(y = \texttt{true}|x)} = e^{w \cdot f}$$

$$P(y = \texttt{true}|x) = e^{w \cdot f} - e^{w \cdot f} P(y = \texttt{true}|x)$$

$$P(y = \texttt{true}|x) + e^{w \cdot f} P(y = \texttt{true}|x) = e^{w \cdot f}$$

$$P(y = \texttt{true}|x) = h(x) = \frac{e^{w \cdot f}}{1 + e^{w \cdot f}} = \frac{1}{1 + e^{-w \cdot f}}$$

$$P(y = \texttt{false}|x) = \frac{1}{1 + e^{w \cdot f}} = \frac{e^{-w \cdot f}}{1 + e^{-w \cdot f}}$$

What Level of Maths are we Talking?

$$P(y = 1|x; \beta) = h_\beta(x)$$

$$P(y = 0|x; \beta) = 1 - h_\beta(x)$$

$$\rightarrow P(y|x; \beta) = (h_\beta(x))^y * (1 - h_\beta(x))^{1-y}$$

$$\underset{\beta}{\operatorname{argmax}} \prod_{i=1}^{n} P(y_i|x_i; \beta)$$

$$= \underset{\beta}{\operatorname{argmax}} \prod_{i=1}^{n} (h_\beta(x_i))^{y_i} * (1 - h_\beta(x_i))^{1-y_i}$$

$$= \underset{\beta}{\operatorname{argmax}} \sum_{i=1}^{n} y_i \log h_\beta(x_i) + (1 - y_i) \log(1 - h_\beta(x_i))$$

Some Recommended Textbooks

We won't be following any one specifically, but they are all good for background

1. Jacob Eisenstein. Natural Language Processing. MIT Press (2019)
   `http://cseweb.ucsd.edu/~nnakashole/teaching/eisenstein-nov18.pdf`
2. Marc Peter Deisenroth, A Aldo Faisal, and Cheng Soon Ong. Mathematics for Machine Learning. Cambridge University Press (forthcoming)
   `https://mml-book.github.io/book/mml-book.pdf`
3. Chris Bishop. Pattern Rechognition and Machine Learning. Springer (2009)
   `http://users.isr.ist.utl.pt/~wurmd/Livros/school/Bishop%20-%20Pattern%20Recognition%20And%20Machine%20Learning%20-%20Springer%20%202006.pdf`

Intended Learning Outcomes

- Understand elementary mathematical concepts used in machine learning
- Derive machine learning models from first principles
- Design, implement, and evaluate machine learning systems for real-world problems
- Identify the correct machine learning model for a given real-world problem

Assessment

Project 1

- Worth 20%
- Release week 3, due week 6
- Application and evaluation of ML techniques on a data set. Coding and conceptual questions.

Project 2

- Worth 40%
- Release week 7, due week 10
- Open-ended research project. Coding and research paper.

Final exam

- Worth 40%
- **Hurdle requirement:** you have to pass the exam.

No mid-semester test.

A Note on Lectures

Cons

- passive and disengaging
- a lazy way to teaching
- ‘it’s almost unethical to be lecturing’ (A. Bajak. 2014. Science)

Pros

- potentially engaging and inspiring
- social, community building
- mindful, attention building, mental workout (M. Worthen. 2015. The New York Times)

Pragmatics

- unavoidable with 500+ sized classes
- let’s make the most of it!

A Note on Lectures

You (students)

- attend and participate
- communicate with your peers
- ask questions and give feedback

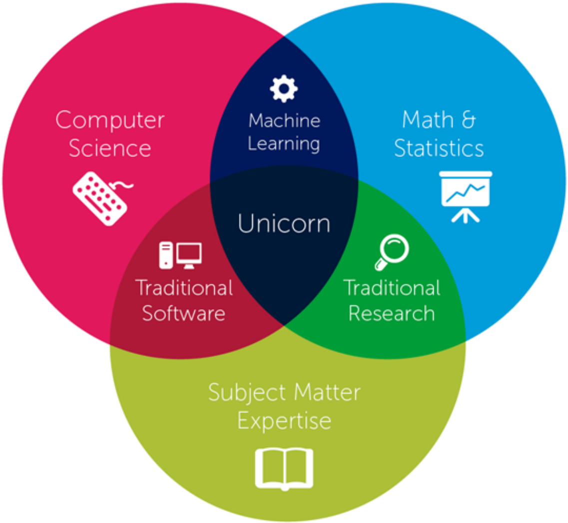
Me (Lecturer)

- aims to be engaging
- quizzes, demos, coding

(You may want to check out this study on the effect of lecture capture on achievements.)

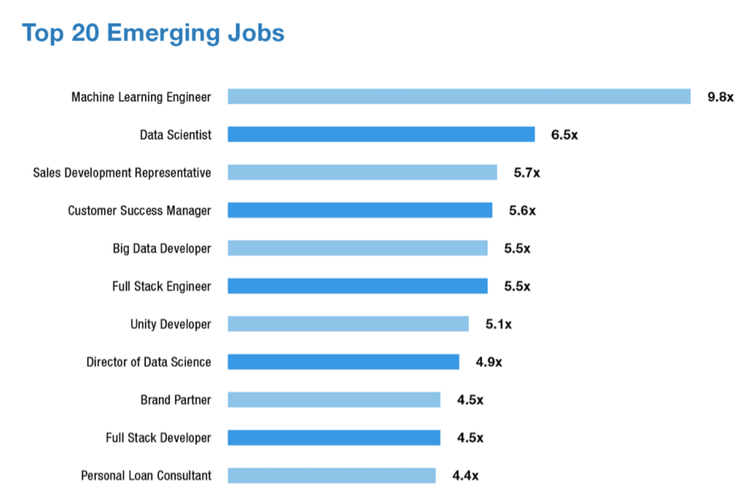
What and Why of Machine Learning?

What is Machine Learning?



Relevance

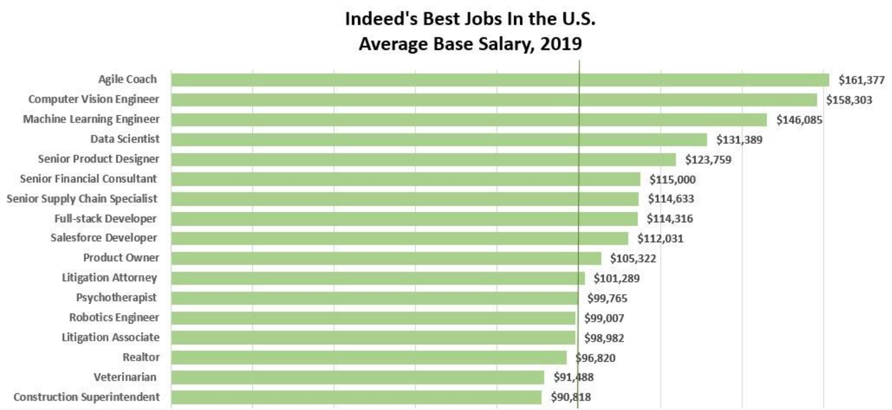
(you're sitting in the right class!)



Source: https://www.springboard.com/blog/machine-learning-engineer-salary-guide/

Relevance

(you're sitting in the right class!)



Source: https://blogs-images.forbes.com/louiscolumbus/files/2019/03/average-base-salary.jpg

Three ingredients for machine learning

... and related questions

... and related questions

1. Data

- Discrete vs continuous vs ...
- Big data vs small data
- Labeled data vs unlabeled data
- Public vs sensitive data

Three ingredients for machine learning

... and related questions

Models

- function mapping from inputs to outputs
- motivated by a data *generating* hypothesis
- probabilistic machine learning models
- geometric machine learning models
- parameters of the function are unknown

Three ingredients for machine learning

... and related questions

Learning

- Improving (on a task) after data is taken into account
- Finding the best model parameters (for a given task)
- Supervised vs. unsupervised learning

ML Example Problem

ML Example: the *Cool/Cute* Classifier

- According to Tim's 2 y.o. son:

Entity	Class	Entity	Class
self	cute	sports car	cool
self as baby	???	tiger	cool
big brother (4 y.o.)	cool	Hello Kitty	cute
big sister (6 y.o.)	cute	spoon	???
Mummy	cute	water	???

- Which class label would we predict for the following entities:

koala, book on ML, train

Yeah yeah, but what's in it for me?

- Scenario 1

  You are an archaeologist in charge of classifying a mountain of fossilized bones, and want to quickly identify any "finds of the century" before sending the bones off to a museum

- Solution:

  Identify bones which are of different size/dimensions/characteristics to others in the sample and/or pre-identified bones

Yeah yeah, but what's in it for me?

- Scenario 1

  You are an archaeologist in charge of classifying a mountain of fossilized bones, and want to quickly identify any "finds of the century" before sending the bones off to a museum

- Solution:

  Identify bones which are of different size/dimensions/characteristics to others in the sample and/or pre-identified bones

CLUSTERING/OUTLIER DETECTION

Yeah yeah, but what's in it for me?

- Scenario 2:

  You are an archaeologist in charge of classifying a mountain of fossilized bones, and want to come up with a consistent way of determining the species and type of each bone which doesn't require specialist skills

- Solution:

  Identify some easily measurable properties of bones (size, shape, number of "lumps", ...) and compare any new bones to a pre-classified database of bones

THE UNIVERSITY OF MELBOURNE

Yeah yeah, but what's in it for me?

- Scenario 2:

  You are an archaeologist in charge of classifying a mountain of fossilized bones, and want to come up with a consistent way of determining the species and type of each bone which doesn't require specialist skills

- Solution:

  Identify some easily measurable properties of bones (size, shape, number of "lumps", ...) and compare any new bones to a pre-classified database of bones

SUPERVISED CLASSIFICATION ;

Yeah yeah, but what's in it for me?

- Scenario 3:

  You are a supermarket manager, wishing to boost sales without increasing expenditure, but with lots of historical purchase data

- Solution:

  Strategically position products to entice consumers to spend more:
  - beer next to chips?
  - beer next to bathroom cleaner?

Yeah yeah, but what's in it for me?

- Scenario 3:

  You are a supermarket manager, wishing to boost sales without increasing expenditure, but with lots of historical purchase data

- Solution:

  Strategically position products to entice consumers to spend more:
  - beer next to chips?
  - beer next to bathroom cleaner?

ASSOCIATION RULES

Yeah yeah, but what's in it for me?

- Scenario 4:

  You are in charge of developing the next "release" of Coca Cola, and want to be able to estimate how well received a given recipe will be

- Solution:

  Carry out taste tests over various "recipes" with varying proportions of sugar, caramel, caffeine, phosphoric acid, coca leaf extract, ... (and any number of "secret" new ingredients), and estimate the function which predicts customer satisfaction from these numbers

Yeah yeah, but what's in it for me?

- Scenario 4:

  You are in charge of developing the next "release" of Coca Cola, and want to be able to estimate how well received a given recipe will be

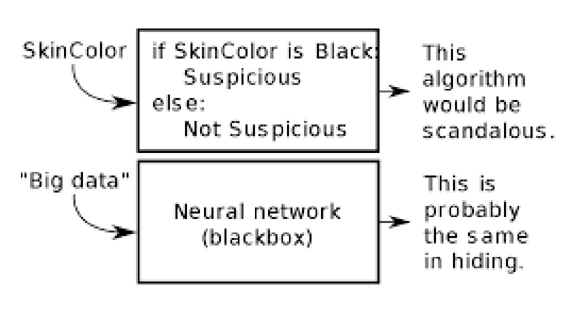
- Solution:

  Carry out taste tests over various "recipes" with varying proportions of sugar, caramel, caffeine, phosphoric acid, coca leaf extract, ... (and any number of "secret" new ingredients), and estimate the function which predicts customer satisfaction from these numbers

  REGRESSION

More Applications

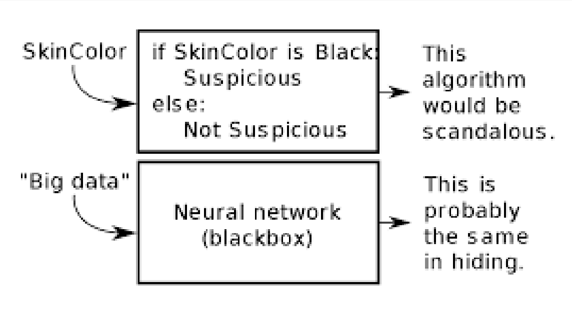
- natural language processing
- image classification
- stock market prediction
- movie recommendation
- web search
- medical diagnoses
- spam / malware detection
- ...



commons.wikimedia.org/wiki/File:Pseudo-algorithm_comparison_for_my_slides_on_machine_learning_ethics.svg

Def 1. **Discrimination**= To make distinctions.

For example, in supervised ML, for a given instance, we might try to discriminate between the various possible classes.



commons.wikimedia.org/wiki/File:Pseudo-algorithm_comparison_for_my_slides_on_machine_learning_ethics.svg

Def 2. **Discrimination**= To make decisions based on prejudice.

Digital computers have no volition, and consequently cannot be prejudiced. **However**, the data may contain information which leads to an application where the ensuing behavior is prejudicial, intentionally or otherwise.

Machine Learning and Ethics i

ML has the potential to *discriminate* [def 2.] people

- some uses of data are unethical, some plainly illegal
  - race & sex in medical applications: OK
  - race & sex in loan applications: unethical
  - race & sex in student applications: ??? (affirmative action vs. racial/sex discrimination)
- legal frameworks are still being defined

THE UNIVERSITY OF MELBOURNE

Machine Learning and Ethics ii

Not everything that *can* be done, *should* be done

- attributes in the data can encode information in an indirect way
  - For example, home address and occupation can be used (perhaps with other seemingly-banal data) to infer age and social standing of an individual
- potential legal exposure due to implicit "knowledge" used by a classifier
- just because you didn't realize doesn't mean that you shouldn't have realized, or at least, made reasonable efforts to check

THE UNIVERSITY OF MELBOURNE

Questions to Ask

- Who is permitted to access the data?
- For what purpose was the data collected?
- What kinds of conclusions are legitimate?
- If our conclusions defy common sense, are there confounding factors?
  - car insurance & young male drivers?
  - car loans & owners of red cars?

Summary

Today

- COMP90049 Overview
- What is machine learning?
- Why is it important? Some use cases.
- What can go wrong?

Next lecture: Concepts in machine learning

References i

Jacob Eisenstein. Natural Language Processing. MIT Press (2019)

Marc Peter Deisenroth, A Aldo Faisal, and Cheng Soon Ong. Mathematics for Machine Learning. Cambridge University Press (forthcoming)

Chris Bishop. Pattern Rechognition and Machine Learning. Springer (2009)

Tom Mitchell. Machine Learning. McGraw-Hill, New York, USA (1997).