Introduction to Machine Learning
236756

Lecture 0:
What is Machine Learning?
What and how will we study in this course?

Original Version by Prof. Nati Srebro-Bartom
Modifications by Nir Ailon
What is “Machine Learning”?

“Learning” (in nature): Using past experience to make future decisions or guide future actions

“Machine Learning” as an Engineering Paradigm: Use data and examples, instead of expert knowledge, to automatically create systems that perform complex tasks
OCR—Optical Character Recognition

- Less programming
- Robust and easily adaptive
- Less dependent on expert knowledge
- Performs very well, used in many applications
SPAM Filtering

- Adaptive
- Continuously learn and improve
Face Detection
Face Recognition
Object Recognition
Object Recognition
OrCam

http://www.youtube.com/watch?v=ykDDxWbt5Nw
Machine Translation

- Complex output
- Learn from:  - annotated translations
  - matching text
  - text (in a single language)
  - corrections
Depth from Stereo Images

\[
\frac{b}{f} = \frac{Z}{x_L - x_R} \\
\frac{b}{f} = \frac{Z}{y_L - y_R} \\
\frac{b}{f} = \frac{Z}{Z}
\]

\[
X = \frac{b(x_L + x_R)}{2(x_L - x_R)} \\
Y = \frac{b(y_L + y_R)}{2(x_L - x_R)} \\
Z = \frac{bf}{x_L - x_R}
\]
Machine Learning Everywhere

• OCR (addresses, checks, books, forms, pen input, ...)
• Spam Filtering
• Machine Translation
• Speech Recognition
• Vision:
  • Face detection and face recognition
  • Object detection (search by object, pedestrian detection, ...)
  • Pose estimation in Kineckt
  • Driving assistance systems and self driving cars: pedestrian and hazard detection, staying in-lane, reading road signs, ...
• Control: helicopters, robots, self-driving cars
• Search (which is the relevant page?)
• Advertising and ad placement
• Recommendation systems (what movie will you like?)
• Protein fold prediction (protein sequence: KVFGRCLEAAAMKHRGLDNYRGYSLLGNWVCANFESNFTQATNRTGSDTYGILQRINSRWCNDGTFRPGSRNFLCNIPCSALSSDITASVNCARRIVSDGNMQWAQGARIKCNCKGTDVQA4IRGCR)

(Protein structure diagram)
Generic Learning

Parameterized model of a face

Examples of faces

“Correct” parameters
(distribution of distance between eyes, importance of different parts, etc)

Learning Algorithm

Sample emails

Examples of bicycles

Hebrew and English texts

Recording transliterated audio

Protein sequences and folds

Spam detector

Bike detectors

Face recognizer

Translation system

Speech recognizer

Protein fold predictor
The ability to learn grammars is **hard-wired** into the brain. It is not possible to “learn” linguistic ability—rather, we are born with a brain apparatus specific to language representation.

There exists some “universal” learning algorithm that can learn **anything**: language, vision, speech, etc. The brain is based on it, and we’re working on uncovering it. (Hint: the brain uses neural networks)

There is no “free lunch”: no learning is possible without some prior assumption about the structure of the problem (prior knowledge)
More Data, Less Expert Knowledge

Expert knowledge:
- full specific knowledge
- no specific knowledge

Use data to fit specific model

Expert Systems
(no data at all)

Machine Learning
(Our Course)

more data

no free lunch
“Machine Learning”: Use data and examples, instead of expert knowledge, to automatically create systems that perform complex tasks

- 99% of faces have two eyes
- People with beards buy less nail polish
- …

Does smoking contribute to lung cancer?
- Yes, with p-value = $10^{-72}$

How long ago did cats and dogs diverge?
- About 55 MY, with 95% confidence interval [51,60]

Does smoking contribute to lung cancer? Why? Does this mean it’s a cause? Why? How strong is this evidence? What other factors might be at play? How can we be sure? What do we mean by “strong evidence”? What are the implications of this finding? How can we use this information to improve public health? What are the limitations of this study? How can we replicate these results? What are the ethical considerations involved in studying smoking and lung cancer?

How long ago did cats and dogs diverge? Why is this important? What does the 95% confidence interval tell us? How does this relate to other species? How can we use this information to understand evolution? What other factors might have influenced the divergence of cats and dogs? How can we replicate these results? What are the limitations of this study? How can we use this information to improve our understanding of the natural world?

Knowledge Discovery/Data Mining

Machine Learning

Data

System for Performing Task (e.g. Predictor)

Statistics

(NP → adj NP)

(NP → det N)

(det → ‘the’)

(Rotation time)$^2$ ∝ (avg radius)$^3$
Goal:
- Understand the main concepts underlying machine learning
- Know the prominent methods used in contemporary machine learning
- Learn how to use machine learning correctly

Part I: Fundamentals
- More rigorous, understand concepts through basic theory (although goal is not detailed theory and theorem-proving)
- Emphasis on concepts, less on specific algorithms

Part II: Survey of prominent methods and approaches

We will mostly consider binary classification, but concepts and methods applicable to more complex
Course Structure

• Lectures: Wednesday 10:30-12:30  
  • Nir Ailon  
  • Slides will be available online

• Recitations: Monday 16:30-17:30  
  • Review of math needed for lectures  
  • Examples and further elaboration  
  • Integral part of the course  
  • Notes will be available online

• Self-Serve Demos:  
  • Weekly demos of methods discussed in class or relevant to material  
  • In first few weeks: empirical demonstration of mathematical concepts  
  • Numeric Python; Step-by-step instructions on course website; Assistance from TAs  
  • Should take you up to about an hour  
  • Strongly encouraged, up to 10 points of final grade

• Four Homeworks  
  (1-2 include also experimentation component)
Requirements / Grades

• Homework
  • Mandatory, must get 50/100 on each one to pass course
  • Some homework will include experimentation, mostly using existing code (some coding will be required)

• Self-serve demos
  • Very strongly encouraged
  • Some exam questions based on demos
  • 1 point for each one done on time, up to max 10 points (0.5 point if late)

• Grade= (if at least 50/100 on each homework) max(final, 0.5*final + 0.4*homework + 0.1*demos)

• Detailed information on course website
TAs

• Available for assistance with demos and homeworks

• Head TA: Guy Uziel
Pre-Requisites

• **Introduction to AI**
  • Recommended, but not required
  • Python
  • Some examples

• **Probability**

• **Linear Algebra**
  • Very strongly recommended; Essentially essential
Course Material

• Course Text:
  • Will not follow order, but will base most material on this text and will refer to specific sections and pages
  • Non-circulating copies in library

• On website: slides, recitation notes, additional notes and references, demos