Information Theory in Machine Learning

ICS 636

Professor: Susanna Still

P.O.S.T 303F

Office hours: by appointment.

email: sstill@hawaii.edu

Learning and Adaptation

- Most intelligent systems show signs of learning
- Most biological, "alive" systems utilize adaptation.

Challenges:

- Understand the principles of learning
- Use this understanding to build learning machines

Information theory serves as the basis.

Information Theory and Computer Science

- Questions about information processing, learning and adaptation in the animal and the machine were driving the early development of computer science as a discipline, as some of the main people were interested in these questions:
  - Alan Touring
  - John von Neumann
  - Norbert Wiener, who coined the term “cybernetics”.
  - Claude Shannon

Machine Learning is finding many applications:

- Some Application Areas:
- Bioinformatics
- Computer Vision
- Robotics
- Graphics
- Speech
- Financial analysis
- E-commerce
- Medicine
- Computer games
- Multimedia

- Examples:
- microarray data
- object recognition
- decision making
- realistic simulations
- recognition, identification
- option pricing
- data mining
- diagnostics, drug design
- adaptive opponents
- retrieval across databases

Machine learning is crucial in robotics and AI:

- It is often easier to build a learning system than to hand-code a program that works. Examples: robot that roams around and learns a map of the environment; spam filters.
- Typical tasks that require learning:
  - Speech, handwriting and object recognition
  - Intelligent user interfaces
  - Motor behavior
- A true AI requires learning on many levels!

- To function, an AI system has to implement many aspects of animal/human learning and behavior.
- Motor control and motor learning.
- Object recognition
- Navigation
- Deal with noisy sensory data
- Concept learning (cognition)
- etc...

- How do we do it?
- Sensory Input: Huge data streams; highly complex
- Nervous system: Strong data compression on all levels (starting at first stages of sensory preprocessing, e.g. in the retina)
- Learning of concepts such as objects and words (grouping sensory input)
- Learning of more abstract concepts
- Communication
- Interactive learning and control

Machine learning is crucial for data analysis:

- Huge and very complex data sets too large to analyze by hand; e.g. data from complex systems such as ecological systems.

- No human expert; e.g. DNA analysis.

- Human can solve task, but can not explain how; e.g. character recognition.

- Desired function changes frequently; e.g. stock price prediction from trading data.

Different kinds of learning

- Supervised -- there is a teacher signal
- Unsupervised -- no teacher signal

- Example: cancer tissue (UC Irvine ML repository)

diameter	perimeter	texture	...	outcome	time	training example
13.71	20.83	90.2	...	Recurrence	77	1
13	21.82	87.5	...	Normal	119	2
12.46	24.04	83.97	...	Normal	76	3

- Binary classification:
  Predicting Yes/No label; here: cancer recurrent?
- Regression: Predicting a continuous variable; here: time to recurrence.

Different kinds of learning

- Based on the role of the learner: active learning vs. passive learning
- Most animal learning is active!
- But most results in learning theory address passive learning. Complications for active learning. E. g. data are not i.i.d.
- Based on the problem definition: Reinforcement learning: a reward signal is present (a behavioral goal is specified)

Statistics and Signal Processing

- Concerned with similar questions, maily for the purpose of data analysis.
- Find structure in data.
- Transform data such that structure emerges.
- Problem: many arbitrary ad hoc assumptions

Information Theory

- Serves as a FOUNDATION for the study of information processing systems.
- Foundation for many signal processing tools, famous example: Linear Filters, Wiener-Volterra Filters.
- Helps to overcome arbitrariness!
  FOCUS OF THIS COURSE.
  - Famous example: statistical mechanics (Jaynes, 1957)
  - Other examples (discussed in this class):
    - Unsupervised Learning / Cluster Analysis
    - Time series prediction
    - Interactive, behavioral learning

Course format

- Lectures to introduce a subject
- Followed by
  - discussions of research papers
  - free discussions of course content and research problems
  - hands on research projects
- Students have to be motivated to do:
  - Reading
  - Thinking
  - Research

Outline of Course Content

- Four main sections:
- Basics of Information Theory.
- Applications of Information theory to:
  - Unsupervised Learning / Cluster Analysis.
  - Time Series Prediction.
  - Interactive Learning.

- Motivation: Unsupervised Learning / Cluster Analysis -- an important example.
- Basics of Information Theory
- Using Information theory as a foundation for cluster analysis -- overcoming arbitrariness in cluster analysis.
- Detour: Information theory and Statistical mechanics. Paper: Jaynes, 1957.
- Discussion of other papers (student choice).
- Free discussion and Research Project Opportunity.

- Time series analysis.
- Lectures:
  - Linear filters, Linear Predictive Coding(LPC).
  - Using information theory to build optimally predictive filters.
  - Optimally predictive estimation and inference.
  - Connections to “Computational mechanics”
  - Learning the causal architecture of dynamical systems
- Papers (student choice) and discussions.
- Research Project Opportunity.

- Interactive Learning.
- Lectures:
  - Extending optimally predictive filters to deal with interactions form the learner.
  - Curiosity as a behavioral paradigm.
  - Modeling explorative behavior.
  - Learning the causal architecture of dynamical systems that are coupled to the observer through actions.
- Papers (student choice) and discussions.
- Research Project Opportunity.

Grading

- Final project (report) 50% of final grade.
- Final Exam 50% of final grade.
- Extra Credit:
  - Presenting papers (student's choice). Extra Credit.
  - In each section, there are opportunities for research projects. Students can do as many as they wish, for extra credit.