Machine Learning Intro

CPS 271

Ron Parr

Contact Information

- Professor
  - Ron Parr
  - D209 LSRC, parr@cs.duke.edu, 660-6537
  - Office hours: M 10-11, Th 9-10
- TA
  - Susanna Ricco
  - D214 LSRC, sricco@cs.duke.edu, 660-6513
  - Office hours: M 3-4, W 1-2

About Me

- My eighth year at Duke
- Bachelor's degree in philosophy
  - Philosophy of mind
- Ph.D. in computer science
  - Hierarchical planning under uncertainty
- Current interests:
  - Planning under uncertainty
  - Reinforcement learning
  - Probabilistic reasoning
  - Game theory
  - Robotics

Requirements

- Some programming skills
  - Write small programs without drama
  - Useful for project
- Prequesites
  - Short proofs
  - Basic algorithmic concepts
    - Complexity - O()
    - Analysis of algorithms
  - Math
    - Some calculus
    - Linear Algebra
  - Basic probability and statistics

Class Mechanics

- Reading
  - Textbook: **Pattern Recognition and Machine Learning**, Christopher M. Bishop
  - Supplementary handouts and papers (learning theory, reinforcement learning)
- Homeworks: 10%
  - Typically assigned on Tuesdays; due next Thursday
  - Discussion OK, write-up must be your own
- Project: 30%
  - Up to 2 person collaboration
- Midterm: 30%
  - Closed book, in class, no collaboration
- Final: 30%
  - Closed book, finals week, no collaboration

Some other ML Books

- *Neural Networks for Pattern Recognition*, Christopher M. Bishop
- *The Elements of Statistical Learning*, Trevor Hastie, Robert Tibshirani, & Jerome Friedman
- *Machine Learning*, Tom M. Mitchell

Major Topics Covered

- Foundations
  - Probability*
  - Statistics
  - Regularization & Learning Theory
- Types of learning
  - Supervised Learning
    - Linear classifiers
    - Neural networks*
    - Support Vector Machines and Kernel Methods
  - Reinforcement Learning*
    - Decision Theory
    - MDPs
    - RL Algorithms
  - Unsupervised Learning
    - Clustering
    - PCA
    - Bayes nets* learning
    - HMM* learning

(* = some overlap with 270)

Major Topics *Not* Covered

- Genetic algorithms
- Biological basis of learning

Project

- Typical: Apply ML to a problem of interest
  - Demonstrate success/failure of ML
  - Provide insight into what worked, what didn't
  - Evaluate alternatives
- Atypical but still possible
  - Propose, evaluate a new learning method
  - Prove new theoretical results

Why Study Learning?

- Considered a hallmark of intelligence
- Viewed as way to reduce programming burden
- Many algorithms assume parameters that are difficult to determine exactly a priori

Examples

- SPAM classification
- Distinguish healthy/diseased tissue
  - Skin cancers
  - Colon Cancers
- Financial events
  - Predict good/bad credit risks
  - Predict price changes
- Drilling sites likely to have oil
- Document categorization
- Predict if mushrooms are safe to eat
- Learn to play games
- Learn to control systems
  - Fly Helicopter
  - Optimize OS components
- Public database of learning problems:
  - http://www.ics.uci.edu/~mlearn/MLSummary.html

Who Does Machine Learning?

- In AI
  - Core AI topic (AAAI, IJCAI)
  - Specialized communities (ICML, NIPS)
- Databases (data mining - KDD)
- Used in (CS):
  - Vision
  - Systems
  - Comp. Bio
- Statistics

Who Does Machine Learning (@Duke)

- CS:
  - Faculty: Pankaj Agarwal, Vince Conitzer, Alex Hartemink, Kamesh Munagala, Ron Parr, Carlo Tomasi, Jun Yang
- ISDS (everybody, but especially):
  - Scott Schmidler, Sayan Mukherjee
- IGSP:
  - Terry Furey, Uwe Ohler
- Engineering:
  - Larry Carin, Silvia Ferrari, Rebecca Willett

Who Hires in Machine Learning?

- Universities
- Microsoft Research
- Search: Google/Yahoo/Amazon
- Defense contractors
- Some financial institutions (quietly)
- Many startups

- ML viewed as good background for many other tasks (robotics, vision, systems, engineering)

Where to Publish/Find ML Papers

- Conferences:
  - International Conference on Machine Learning (ICML)
  - Neural Information Processing Systems (NIPS)
  - Uncertainty in Artificial Intelligence (UAI)
  - Association for the Advancement of Artificial Intelligence (AAAI)
  - International Joint Conference on Artificial Intelligence (IJCAI)
  - Knowledge Discovering in Databases (KDD)
- Journals
  - Journal of Machine Learning Research (MLJ)
  - Machine Learning Journal (JMLR)
  - Journal of Artificial Intelligence Research (JAIR)
  - Artificial Intelligence Journal (AIJ)
  - Pattern Analysis and Machine Intelligence (PAMI)
- Things I've probably left out
  - Venues related to applications
  - Statistics venues
  - Engineering & Signal Processing Venues

What is Machine Learning

- Learning Element
  - The thing that learns

- Performance Element
  - Objective measure of progress

- Learning is simply an increase in the ability of the learning element over time to achieve the task specified by the performance element

ML vs. Statistics?

- Machine learning is:
  - Younger
  - More empirical
  - More algorithmic
  - (arguably) More practical
  - (arguably) More decision theoretic

- Statistics is:
  - More mature
  - (arguably) More formal and rigorous

ML vs. Data Mining

- Machine Learning is:
  - (Arguably) more formal
  - (Arguably) more task driven/decision theoretic

- Data Mining is:
  - More constrained by size of data set
  - More closely tied to database techniques

Types of Learning

- Inductive Learning
  - Acquiring new information that previously was not available
  - Learning concepts
- Speedup learning
  - Learning to do something you already "know" faster or better

Feedback in Learning

- Supervised Learning
  - Given examples of correct behavior
- Unsupervised Learning
  - No external notion of what is correct
  - Is this well-defined?
- Reinforcement Learning
  - Indirect indication of effectiveness

Learning Methodology

- Distinction between training and testing is crucial
- Correct performance on training set is just memorization!
- Researcher should ***never*** look at the test data
- Raises some troubling issues for "benchmark" learning problems

Computational Learning Theory

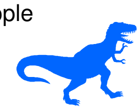
- Formal study of what can be learned from data
- Closely related to ML, but also to CS theory
- Assumptions:
  - Training examples must be representative
  - Algorithm needn't *always* work, but should scale well
- Goals:
  - Algorithms that have a low error rate with high probability
  - Good characterization of how performance scales

Example: Supervised Learning

- Classical framework
- Target concept, e.g., green
- Learner is presented with labeled instances
  - True: Green cones, green cubes, green spheres
  - False: Red cones, red cubes, red spheres, blue cones, blue cubes, blue spheres
- Learner must correctly identify the target concept from the training data

Performance Measure

- Training set won't have all possible objects
- Test set will contain novel objects
  - Blue cylinders, yellow tetrahedra
- To learn successfully, learner must have good performance when confronted w/novel objects
  - This is what we would expect from people
  - A blue Broccolisaurus is still blue



Why Learning Is Tricky

- Suppose we have seen:
  - Red tetrahedron(f), Blue sphere(t), Blue cone(t), green cube(f)
- Possible concepts:
  - Blue
  - (Blue Sphere) or (Blue Cone)
  - Objects a prime number from start
  - Objects with a circular cross-section
- What if some data are mislabeled?

Learning and Representation

- Learning is very sensitive to representation
- Every learning algorithm can be viewed as a *search through a space of concepts*
- Space of concepts determines
  - Difficulty of task
  - Appropriate algorithm
  - Restricting too aggressively can trivialize problem
  - Failure to restrict (or regularize) can trivialize the problem
- Example Space: Conjunctions of colors and shapes
  - Eliminates primes and (possibly) cross sections

Management of the Hypothesis Space

- Ockham's Razor:
  - All things being equal, favor the simplest consistent hypothesis
  - Guiding principle of science,e.g., Einstein:

    'In my opinion the theory here is the logically simplest relativistic field theory that is at all possible. But this does not mean that nature might not obey a more complex theory. More complex theories have frequently been proposed… In my view, such more complicated systems and their combinations should be considered only if there exist physical-empirical reasons to do so.'
- Ockham's razor is not provably correct, but
  - Computational learning theory shows us that the more choices we have, the more data we need to distinguish reliably among these choices
  - Well known trade off between bias and variance
    - How many points do you need to fit a degree 2 polynomial?
    - How many points do you need to fit a degree 100 polynomial?
- Ockham's razor is embodied in a wide range of methods

How to Succeed with Machine Learning

- Theoretical/algorithmic success
  - Maneuver through space of hypotheses efficiently
  - Efficiency
    - Make good use of data
    - Make good use of time
- Practical Success
  - Getting something to learn can be hard (my job!)
  - Know your problem!
    - Pick training data carefully
    - Craft hypothesis space