Machine Learning Intro

CPS 570
Ron Parr

Why Study Learning?

• Considered a hallmark of intelligence

• Viewed as way to reduce programming burden
  – Not enough programmers in the world to produce custom solutions to all problems – even if we knew how
  – Programmers are expensive!

• Many algorithms assume parameters that are difficult to determine exactly a priori
  – What is the right formula to filter spam?
  – When should your smart thermostat turn on the heat?
Examples

- Image analysis/categorization (people, things, products)
- Document categorization
- SPAM classification (spam/ham, news aggregation, spying)
- Computational Biology/medicine
  - Distinguish healthy/diseased tissue (e.g., skin/colon cancer)
  - Find structure in biological data (regulatory pathways)
- Financial events (good/bad credit risks, price changes, marketing response)
- Drilling sites likely to have oil
- Learn to play games chess, go, etc.
- Learn to control systems (robotics, helicopters, operating systems)
- Public database of learning problems:

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**: Pankaj Agarwal, Robert Calderbank, Rong Ge, Raluca Gordan, Alex Hartemink, Kamesh Munagala, Ron Parr, Sudeepa Roy, Cynthia Rudin, Carlo Tomasi, Jun Yang

- **Stats (everybody, but especially)**: David Dunson, Scott Schmidler, Katherine Heller, Sayan Mukherjee, Beka Steorts

- **Engineering**: Larry Carin, Yiran Chen, Helen Li, Guillermo Sapiro

With apologies to those I missed!

Who Hires in Machine Learning?

- **Universities**
- **Apple, MS, Google, FB, Amazon, etc.**
- **Defense contractors and car companies**
- **Some financial institutions (quietly)**
- **Many startups**

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

Look at this cool result! Maybe somebody can explain why it works later?

Let’s model this situation and prove that we converge to a consistent answer!
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
  - Example input: Labeled x-rays
  - Example use: Cancer diagnosis

- **Unsupervised Learning**
  - No external notion of what is correct
  - Example: Unlabeled x-rays
  - Example use: Clustering based on appearance

- **Reinforcement Learning**
  - Indirect indication of effectiveness
  - Example use: PacMan, go, chess

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 (but in practice always does)

- Raises issues for “benchmark” learning problems
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

Credit: http://abduc-tees.com
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 implicit (representation) and explicit restrictions placed on the hypothesis space – inductive bias
• Learning 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
• Example Space: Space of linear or quadratic functions of inputs
  – Eliminates higher order polynomials, etc.
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 (without additional assumptions), but
  - Theoretical results can prove 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

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:
  - Prove low error rate with high probability (WHP)
  - (Loosely) connect training error to test error WHP
  - Good characterization of how performance scales with data
Learning Intro Final Thoughts

• ML = one of the most successful areas of AI
  – Many practical applications
  – Rich theory that inspires, but lags practice
  – Many ways to “succeed” w/o solving entire AI problem
  – Many fields view machine learning as a special sauce that will give them an advantage

• Machine learning conferences are as large as or larger than the general AI conferences