

Learning Intro

CPS 270
Ron Parr

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

Formalizing 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

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
- Need to make some assumptions:
 - Training examples must be representative
 - Algorithm needn't always work, but should scale well
 - Number of samples or run time needed should have polynomial relationship to inverse probability of error and percentage of misclassifications
- Probably Approximately Correct (PAC) learning
- Goal: Algorithms that have a low error rate with high probability

COLT

- Learning theory is elegant and mathematically rich. However,
 - It sometimes isn't constructive
 - It sometimes tells us how many data are needed, but not how to manipulate the data efficiently
- Until recently, learning theory has been disconnected from practical learning algorithms
- New advances are leading to a rapprochement
 - Support vector machines
 - Boosting

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 even when confronted with novel objects
 - This is what we would expect from people
 - A blue Broccolisaurus is still blue



Examples

- Distinguish healthy/diseased tissue
- Predict good/bad credit risks
- Drilling sites likely to have oil
- Document categorization
- Predict failures in physical systems
- Predict if mushrooms are safe to eat
- Good/bad moves in a game
- Public database of learning problems:
 - <http://www.ics.uci.edu/~mllearn/MLSummary.html>

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 procedure through a space of concepts
- Space of concepts determines
 - Difficulty of task
 - Appropriate algorithm
 - Restricting too aggressively can trivialize problem
- Space: Conjunctions of colors and shapes
 - Eliminates primes and (possibly) cross sections