Machine Learning: CIS 472/572

Introduction

Instructor: Daniel Lowd
Based on slides by Vibhav Gogate, Pedro Domingos, and others.

Evaluation

• 3-5 homeworks (40%)
  – Some programming, some exercises
• One Midterm (25%)
  – 2/3rds of the way through
• One Project (30%)
  – Apply machine learning to a real problem of your choice
  – Groups allowed
  – Written report
  – Presentations during final exam time
  – More details will be announced soon
• Participation (5%)

Source Materials

• K. Murphy, Machine Learning: A Probabilistic Perspective, MIT Press, 2012. (Recommended, in bookstore)
• T. Mitchell, Machine Learning, McGraw-Hill, 1997. (Recommended, but old and expensive)
• C. Bishop, Pattern Recognition and Machine Learning, Springer, 2006
• R. Duda, P. Hart & D. Stork, Pattern Classification (2nd ed.), Wiley, 2000
• D. Barber, Bayesian Reasoning and Machine Learning, Cambridge University Press, 2012. (Free online!)
• T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning, Springer, 2009. (Free online!)

Why Study Machine Learning:
A Few Quotes

“A breakthrough in machine learning would be worth ten Microsofts.”
-Bill Gates, Microsoft

“Machine learning is the next Internet.”
-Tony Tether, Former Director, DARPA

“Machine learning is the hot new thing.”
-John Hennessy, President, Stanford

“Web rankings today are mostly a matter of machine learning.”
-Prabhakar Raghavan, Dir. Research, Yahoo

“Machine learning is going to result in a real revolution.”
-Greg Papadopoulos, CTO, Sun

So What Is Machine Learning?

• Automating automation
• Getting computers to program themselves
• Writing software is the bottleneck
• Let the data do the work instead!
Traditional Programming

- Automating automation
- Getting computers to program themselves

Machine Learning

Related Fields

- Fields that use machine learning:
  - Artificial intelligence
  - Computer vision
  - Natural language processing
  - Computational biology
  - Robotics
  - ...many more...
- Fields with similar goals to machine learning:
  - Statistics
  - Data mining
  - Data science
  - Psychology (developmental, cognitive)
- Fields used by machine learning:
  - Information theory
  - Numerical optimization
  - Computational complexity

Definition: Machine Learning!

- T. Mitchell: Improving performance via experience
  - A computer program is said to learn from experience $E$ with respect to some class of tasks $T$ and performance measure $P$, if its performance at tasks in $T$ as measured by $P$, improves with experience.

Example 1: A Chess learning problem

- Task $T$: playing chess
- Performance measure $P$: percent of games won against opponents
- Training Experience $E$: playing practice games against itself

Example 2: Autonomous Vehicle Problem

- Task $T$: driving on a public highway/roads using vision sensors
- Performance Measure $P$: percentage of time the vehicle is involved in an accident
- Training Experience $E$: a sequence of images and steering commands recorded while observing a human driver

Magic?

No, more like gardening

- Seeds = Algorithms
- Nutrients = Data
- Gardener = You
- Plants = Programs
### ML in a Nutshell

- Tens of thousands of machine learning algorithms
- Hundreds new every year
- Every machine learning algorithm has three components:
  - Representation
  - Evaluation
  - Optimization

### Representation

- Decision trees
- Instances
- Linear function (hyperplane)
- Neural networks
- Support vector machines
- Model ensembles
  - (Sets of rules / Logic programs)
  - (Graphical models (Bayes/Markov nets))
- Etc.

### Evaluation

- Accuracy
- Precision and recall
- Squared error
- Likelihood
- Posterior probability
- Cost / Utility
- Margin
- Entropy
- K-L divergence
- Etc.

### Optimization

- Combinatorial optimization
  - E.g.: Greedy search
- Convex optimization
  - E.g.: Gradient descent
- Constrained optimization
  - E.g.: Linear programming

### Types of Learning

- **Supervised (inductive) learning**
  - Training data includes desired outputs
- **Unsupervised learning**
  - Training data does not include desired outputs
  - Find hidden structure in data
- **Semi-supervised learning**
  - Training data includes a few desired outputs
- **Reinforcement learning**
  - the learner interacts with the world via “actions” and tries to find an optimal policy of behavior with respect to “rewards” it receives from the environment

### Types of Supervised Learning Problems

- Classification: learning to predict a discrete value from a predefined set of values
- Regression: learning to predict a continuous/real value
- Structured prediction: learning to predict a complex output, such as a sequence or tree
Machine Learning: Applications

Examples of what you will study in class in action!

Classification Example: Spam Filtering

Classify as “Spam” or “Not Spam”

Classification Example: Weather Prediction

Regression example: Predicting Gold/Stock prices

Given historical data on gold prices, predict tomorrow’s price!

Similarity Determination

Collaborative Filtering

- The problem of collaborative filtering is to predict how well a user will like an item that he has not rated given a set of historical preference judgments for a community of users.
Collaborative Filtering

Clustering: Discover Structure in data

Clustering images

Machine learning has grown in leaps and bounds

- The main approach for
  - Speech Recognition
  - Robotics
  - Natural Language Processing
  - Computational Biology
  - Sensor networks
  - Computer Vision
  - Web
  - ...and many more each year...

What We’ll Cover

- **Supervised learning**: Decision tree induction, Instance-based learning, Bayesian learning, Neural networks, Support vector machines, Linear Regression, Model ensembles, Learning theory, etc.
- **Unsupervised learning**: Clustering, Dimensionality reduction
- **General machine learning concepts and techniques**: Feature selection, cross-validation, maximum likelihood estimation, gradient descent, expectation-maximization

Not covering:

- Reinforcement learning (471/571)
- Graphical models (471/571, 410/510pm)
- Topic models ⊕