EECS 349
Machine Learning

Instructor: Doug Downey

(some slides from Pedro Domingos, University of Washington)
Logistics

• **Instructor**: Doug Downey
  – Email: ddowney@eecs.northwestern.edu
  – Office hours: Mondays 1:00-2:00 (or by appt), Ford 3-345

• **TAs**: Chandra Sekhar Bhagavatula and Kathy Lee

• **Web**: (linked from prof. homepage)
  www.cs.northwestern.edu/~downey/courses/349_Winter2014/
# Grading and Assignments

<table>
<thead>
<tr>
<th>Assignment</th>
<th>Due Date</th>
<th>Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>Homework 1</td>
<td>21-Jan-14</td>
<td>10</td>
</tr>
<tr>
<td>Homework 2</td>
<td>28-Jan-14</td>
<td>20</td>
</tr>
<tr>
<td>Project Proposal</td>
<td>6-Feb-14</td>
<td>10</td>
</tr>
<tr>
<td>Homework 3</td>
<td>~13-Feb-14</td>
<td>10</td>
</tr>
<tr>
<td>Project Status Report</td>
<td>18-Feb-14</td>
<td>10</td>
</tr>
<tr>
<td>Homework 4</td>
<td>~6-Mar-14</td>
<td>10</td>
</tr>
<tr>
<td>Project Video</td>
<td>21-Mar-14</td>
<td>20</td>
</tr>
<tr>
<td>Project website</td>
<td>21-Mar-14</td>
<td>15</td>
</tr>
<tr>
<td><strong>TOTAL POINTS</strong></td>
<td></td>
<td><strong>105</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Grade</th>
<th>A-</th>
<th>B+</th>
<th>B</th>
<th>B-</th>
<th>C+</th>
<th>C</th>
<th>C-</th>
<th>Etc…</th>
</tr>
</thead>
<tbody>
<tr>
<td>93+</td>
<td>92-90</td>
<td>89-87</td>
<td>86-83</td>
<td>82-80</td>
<td>79-77</td>
<td>76-73</td>
<td>72-70</td>
<td>69…</td>
</tr>
</tbody>
</table>
Homework

• Four homeworks (50 pts)
  – Submitted via e-mail according to hmwk instructions
    • Late assignments penalized 5% per day – must be within 1 week of original deadline
  – Significant programming, some exercises
  – Any programming language
    • Program grading based on sample output, code, explanation

• Final Project (55 pts)
  – Teams of $k$
  – Define a task, create/acquire data for the task, train ML algorithm(s), evaluate & report
Prerequisites

• Significant Programming Experience
  – EECS 214/311, 325 or the equivalent
  – Example: implement decision trees
    (covered next Monday)

• Basics of probability
  – E.g. independence

• Basics of logic
  – E.g. DeMorgan’s laws
Advice

Look at Homework #2 today
Source Materials

• (both “required”)
• Papers & Web pages
A Few Quotes

• “A breakthrough in machine learning would be worth ten Microsofts” (Bill Gates, Chairman, Microsoft)

• “Machine learning is the next Internet” (Tony Tether, 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)

• “Machine learning is today’s discontinuity” (Jerry Yang, CEO, Yahoo)
So What Is Machine Learning?

• “The study of computer programs that improve automatically with experience”
  T. Mitchell *Machine Learning*

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

Input → Computer → Output

Program → Computer

Machine Learning

Input → Computer → Program

Output → Computer
Magic?

No, more like gardening

- **Seeds** = Algorithms
- **Nutrients** = Data
- **Gardener** = You
- **Plants** = Programs
Case Study: Farecast

Search Flights
Find cheap flights and free airfare predictions

- Round Trip
- One Way
- Multi-City

Please enter a To city

From: Chicago, IL (CHI) - All airports

To: Seattle, WA (SEA) - Seattle/Tacoma

7-Day Low Fare Prediction
Tip: Buy
Fares Rising $42
Confidence: 66%

Daily Low Fare History

- $390
- $305
- $220
- $135

69 Days Ago — Now

Details
Applies to ORD->SEA only
Sample Applications

- Web search
- Computational biology
- Finance
- E-commerce
- Space exploration

- Robotics
- Information extraction
- Social networks
- Finance
- Debugging
- [Your favorite area]
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

• How do we represent the function from input to output?
  – Decision trees
  – Sets of rules / Logic programs
  – Instances
  – Graphical models (Bayes/Markov nets)
  – Neural networks
  – Support vector machines
  – Model ensembles
  – Etc.
Evaluation

• Given some data, how can we tell if a function is “good”?
  – Accuracy
  – Precision and recall
  – Squared error
  – Likelihood
  – Posterior probability
  – Cost / Utility
  – Margin
  – Entropy
  – K-L divergence
  – Etc.
Optimization

• *Given some data, how do we *find the* “best” function?*
  – 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

• Semi-supervised learning
  – Training data includes a few desired outputs

• Reinforcement learning
  – Rewards from sequence of actions
Inductive Learning

• **Given** examples of a function \((X, F(X))\)
• **Predict** function \(F(X)\) for new examples \(X\)
  – Discrete \(F(X)\): Classification
  – Continuous \(F(X)\): Regression
  – \(F(X) = \text{Probability}(X)\): Probability estimation
What We’LL Cover

• Inductive learning
  – Decision tree induction
  – Instance-based learning
  – Neural networks
  – Genetic Algorithms
  – Support vector machines
  – Bayesian Learning
  – Hidden Markov Models
  – Learning theory
  – Reinforcement Learning

• Unsupervised learning
  – Clustering
  – Dimensionality reduction
What You’ll Learn

• Where can I use ML?
• For a given problem, how do I:
  – Express it as an ML task
  – Choose the right ML algorithm
  – Evaluate the results
• What are the unsolved problems/new frontiers?
ML in Practice

- Understanding domain, prior knowledge, and goals
- Data integration, selection, cleaning, pre-processing, etc.
- Learning models
- Interpreting results
- Consolidating and deploying discovered knowledge
- Loop
Reading for This Week

• Wired data mining article, Forbes article (linked on course Web page)

• Recommended:
  – Mitchell, Chapters 1 & 2
  – Alpaydin, Ch 1 & 2