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

Assignment	Due Date	Points
Homework 1	21-Jan-14	10
Homework 2	28-Jan-14	20
Project Proposal	6-Feb-14	10
Homework 3	~13-Feb-14	10
Project Status Report	18-Feb-14	10
Homework 4	~6-Mar-14	10
Project Video	21-Mar-14	20
Project website	21-Mar-14	15
	TOTAL POINTS	**105**

A	A-	B+	B	B-	C+	C	C-	Etc…
93+	92-90	89-87	86-83	82-80	79-77	76-73	72-70	69…

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

- T. Mitchell, ***Machine Learning***, McGraw-Hill
- E. Alpaydin, ***Introduction to Machine*** Learning, MIT Press
- (both “required”)
- Papers & Web pages

A Few Quotes

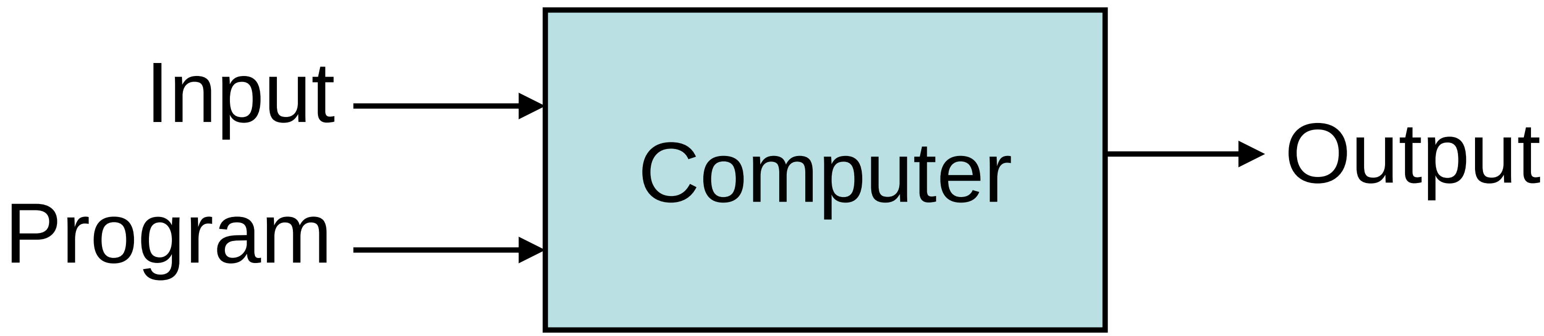
- “A breakthrough in machine learning would be worth ten Microsofts” (Bill Gates, Chairman, Microsoft)
- “Machine learning is the next Internet” (Tony Tether, Director, DARPA)
- “Machine learning is the hot new thing” (John Hennessy, President, Stanford)
- “Web rankings today are mostly a matter of machine 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?

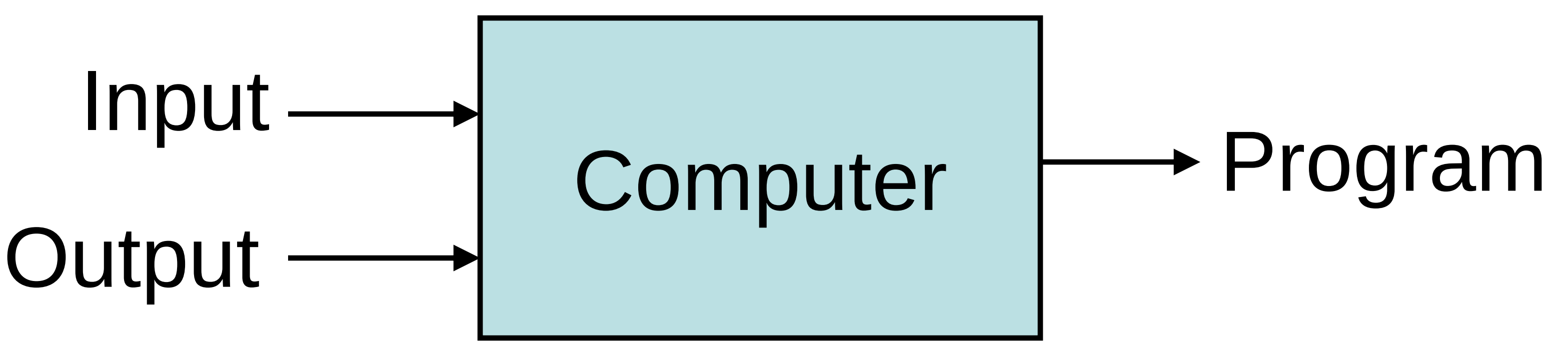
- "The study of computer programs that improve automatically with experience"
  T. Mitchell *Machine Learning*

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

Traditional Programming



Machine Learning

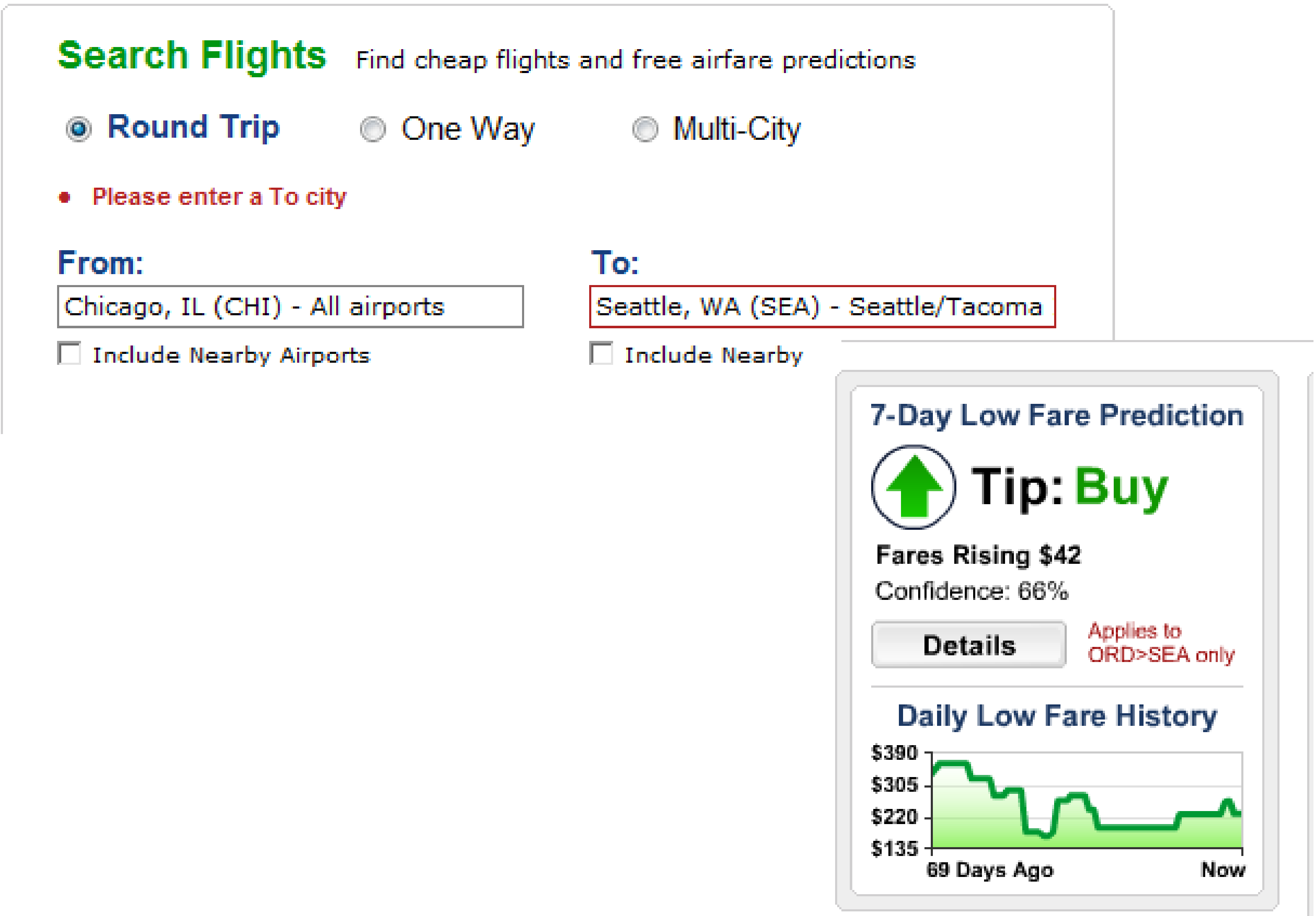


Magic?

No, more like gardening

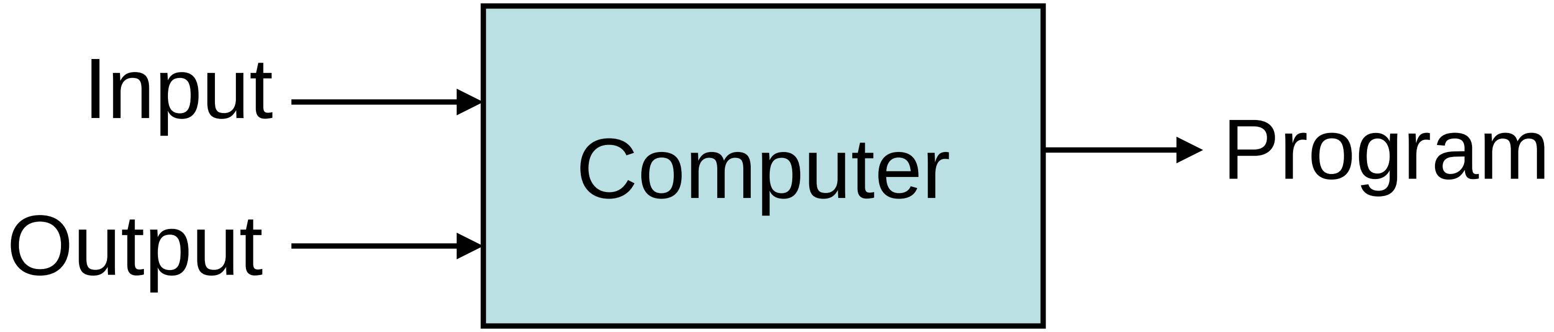
- **Seeds** = Algorithms
- **Nutrients** = Data
- **Gardener** = You
- **Plants** = Programs

Case Study: Farecast



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)* = 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