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 2:00-3:00 (or by appt), Ford 3-345
- **TAs:** Mohammed Alam (Rony), Yanran Wang (Joyce), Zack Witten
- **Web:** (linked from prof. homepage) http://www.cs.northwestern.edu/~downey/courses/349_Spring2015/

Grading and Assignments (1 of 2)

Assignment	Due Date	Points
Homework 1	14-Apr-15	10
Homework 2	TBD	15
Project Proposal	9-Apr-15	5+5
Homework 3	TBD	5
Project Status Report	TBD	5+5
Homework 4	TBD	10
Project Video	5-Jun-15	10
Project Website	5-Jun-15	20+5
Quizzes	Every Wednesday	8
	TOTAL POINTS	**103**

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…

Grading and Assignments (2 of 2)

- Four homeworks (40 pts)
  - Submitted via e-mail according to hmwk instructions
    - Late penalty 5% per day – must be within 1 week of original deadline
  - Significant programming, some exercises
    - Any programming language
- Quizzes (8 pts) – Each Wednesday weeks 2-9
  - Bring a device to access Canvas. *Practice* quiz this week
- Project (40 pts + 15 peer review)
  - Teams of *k*
  - Define a task, create/acquire data for the task, train ML algorithm(s), evaluate & report

Prerequisites

- Significant Programming Experience
  - EECS 214, 325 or the equivalent
  - Example: implement decision trees (covered starting Wednesday)
- Basics of probability
  - E.g. independence
- Basics of logic
  - E.g. DeMorgan's laws

Advice

Look at Winter 2014 EECS 349 Homework #2 today

Source Materials

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

Think/Pair/Share

Why study Machine Learning?

Think

Start

End

Think/Pair/Share

Why study Machine Learning?

Think

Start

End

Think/Pair/Share

Why study Machine Learning?

Pair

Start

End

Think/Pair/Share

Why study Machine Learning?

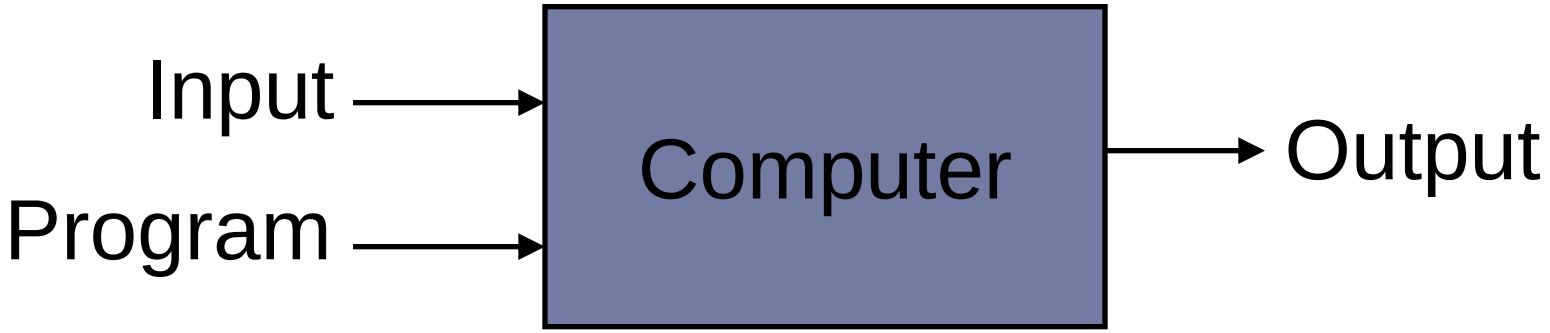
Share

What is Machine Learning?

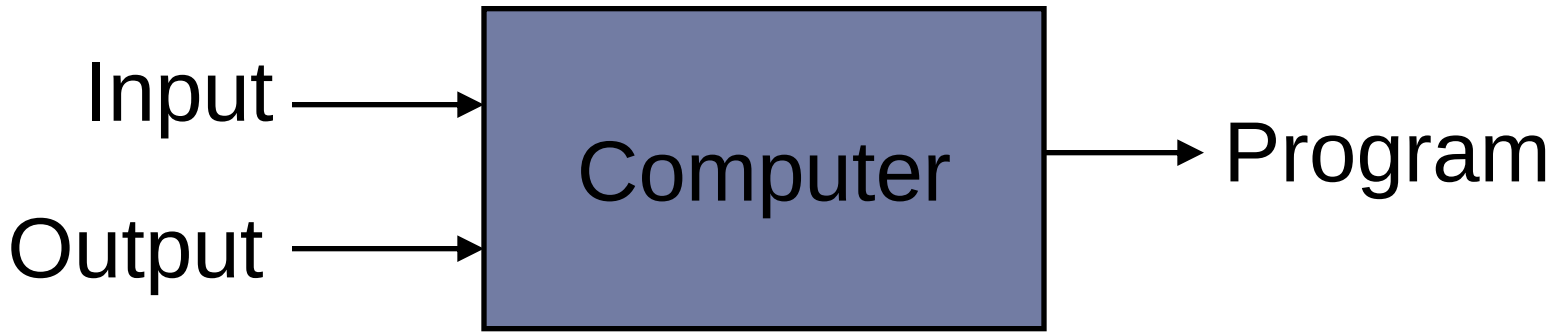
- "The study of computer programs that improve automatically with experience"
  T. Mitchell *Machine Learning*

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

Traditional Programming



Machine Learning

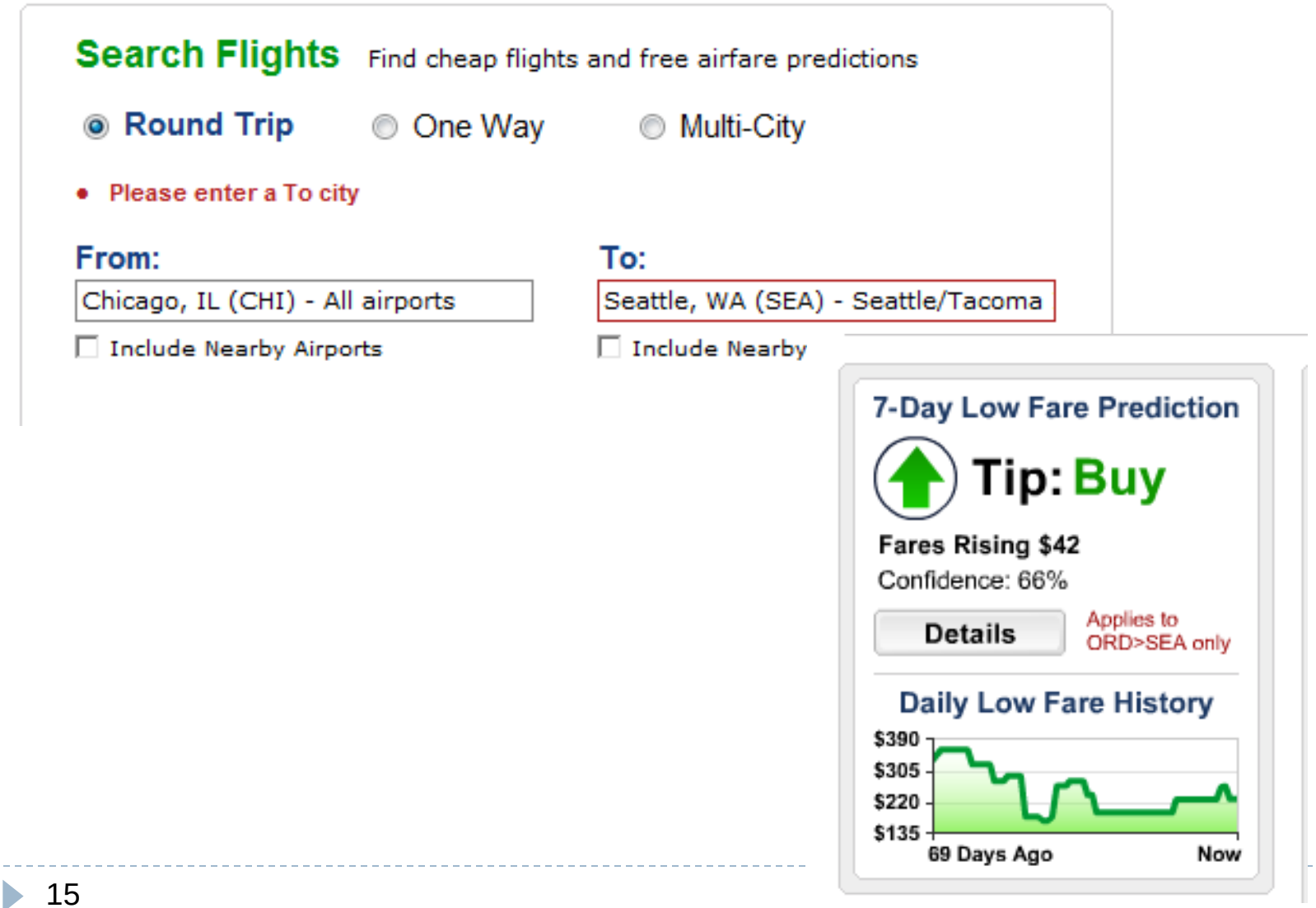


Magic?

No, more like gardening

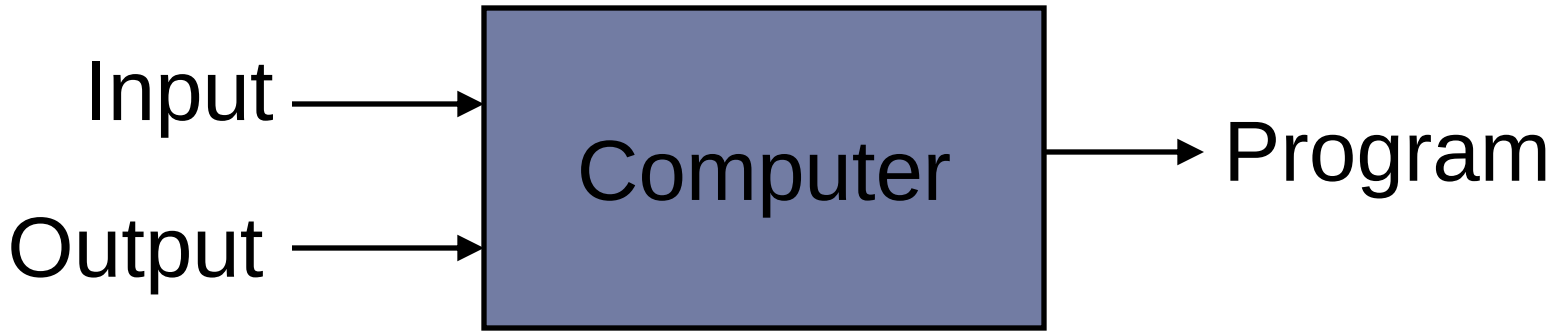
- **Seeds** = Algorithms
- **Nutrients** = Data
- **Gardener** = You
- **Plants** = Programs

Case Study: Farecast



Sample Applications

- Web search
- Computational biology
- Finance
- E-commerce
- Space exploration
- Robotics
- Information extraction
- Social networks
- Finance
- Debugging
- [Your favorite area]



Relationship of Machine Learning to…

- **Statistics**
- **Analytics**
- **Data Mining**
- **Artificial Intelligence**

Why study Machine Learning? (1 of 3)

- "A breakthrough in machine learning would be worth ten Microsofts" (Bill Gates, Chairman, 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)
- "Machine learning is today's discontinuity" (Jerry Yang, CEO, Yahoo)

Why study Machine Learning? (2 of 3)

http://www.emc.com/leadership/digital-universe/2014iview/executive-summary.htm

Why study Machine Learning? (3 of 3)

- **One example, proportion of physicians using EMRs**
  - **2001: 18%**
  - **2011: 57%**
  - **2013: 78%**

- **…what will be able to learn from these?**

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

What You'll Learn in this Class

- How do ML algorithms work?
  - Learn by **implementing**, using
- For a **real** problem, how do I:
  - Express my problem as an ML task
  - Choose the right ML algorithm
  - Evaluate the results

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 instances **x**
  - Discrete *f*(**x**): Classification
  - Continuous *f*(**x**): Regression
  - *f*(**x**) = Probability(**x**): Probability estimation

- Example:
  - **x** = <Flight=United 102, FlightDate=May 26, Today=May 7>
  - *f*(**x**) = +1 if flight price will increase in the next week, or
    -1 otherwise

What We'll Cover

- **Inductive learning**
  - Decision tree induction
  - Instance-based learning
  - Linear Regression and Classification
  - Neural networks
  - Genetic Algorithms
  - Support vector machines
  - Bayesian Learning
  - Learning theory
  - Reinforcement Learning
- **Unsupervised learning**
  - Clustering
  - Dimensionality reduction

Parting Notes

- Bring a device to access Canvas for quiz on Wednesday
- Take a look at Homework #2 from EECS 349 Winter 2014 (see my Web page)
- Reading:
  - Skim: Forbes article

    (linked on course Web page)
  - Recommended:
    - Mitchell, Chapters 1 & 2

      Alpaydin, Ch 1 & 2