

# Machine Learning: CS 6375

## **Introduction**

Instructor: Vibhav Gogate  
The University of Texas at Dallas

# Logistics

- **Instructor:** Vibhav Gogate
  - Email: [vgogate@hlt.utdallas.edu](mailto:vgogate@hlt.utdallas.edu)
  - Office: ECSS 3.406
  - Office hours: M/W – 12:45 p.m. to 1:45 p.m.
- **TA:** To be announced
- **Web:**  
<http://www.hlt.utdallas.edu/~vgogate/ml/2013f/index.html>
- **Discussion Board**
  - Discussion board on **Piazza**. This will be the main on-line forum for discussing assignments and course material, and interacting with other students, TA and me. We will also post course-wide announcements on **Piazza**.

# Evaluation

- Five homeworks (35%)
  - 7% each
  - Due two weeks later
  - Some programming, some exercises
  - Assigned via ELearning.
- One Project (15%)
- One Midterm (20%), One Final (30%)
  - Exams are closed book. You will be allowed a cheat sheet, a double-sided 8.5 x 11 page.
- Attendance is mandatory
- Fast paced and rigorous. But you are going to have fun.

# Source Materials

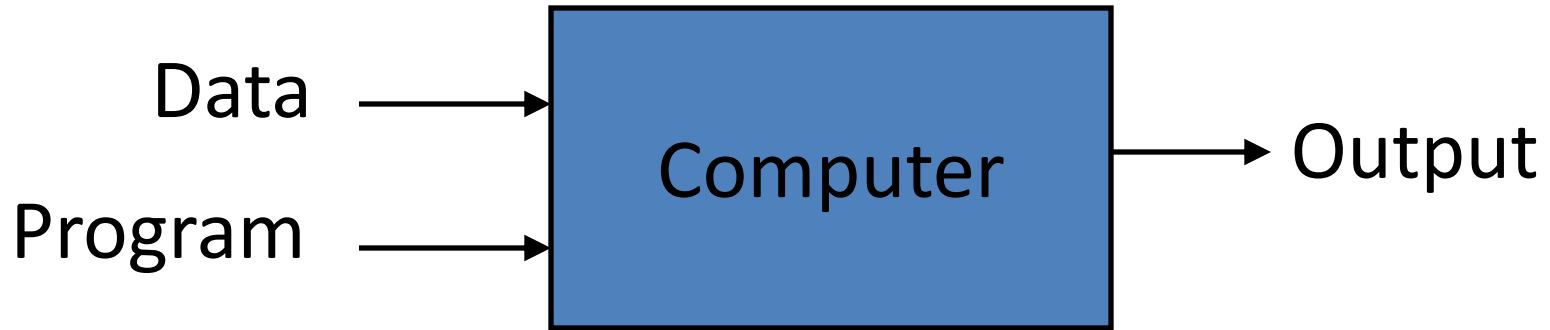
- T. Mitchell, ***Machine Learning***, McGraw-Hill
- C. Bishop, **Pattern Recognition and Machine Learning**, Springer
- Kevin Murphy, **Machine Learning: A probabilistic perspective**
- Class Notes/Slides

# Why Study Machine Learning: A Few Quotes

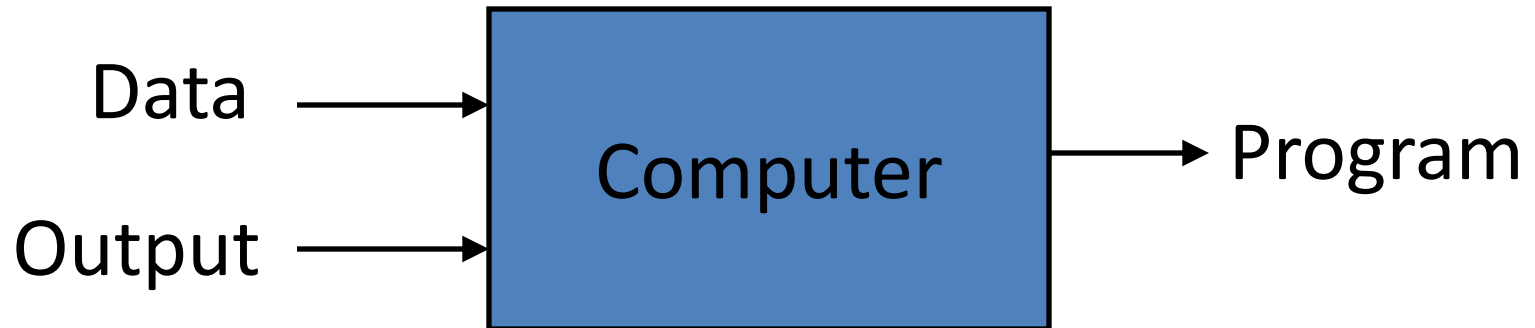
- “A breakthrough in machine learning would be worth ten Microsofts” (Bill Gates, Microsoft)
- “Machine learning is the next Internet” (Tony Tether, Former Director, DARPA)
- Machine learning is the hot new thing” (John Hennessy, President, Stanford)
- “Web rankings today are mostly a matter of machine learning” (Prabhakar Raghavan, Former Dir. Research, Yahoo)
- “Machine learning is going to result in a real revolution” (Greg Papadopoulos, CTO, Sun)

- Getting computers to program themselves
- Writing software is the bottleneck, let data do the work

## Traditional Programming



## Machine Learning



# Magic?

No, more like gardening

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



# Definition: Machine Learning!

- T. Mitchell: Well posed machine learning
  - Improving performance via experience
  - Formally, 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.
- H. Simon
  - Learning denotes changes in the system that are adaptive in the sense that they enable the system to do the task or tasks drawn from the same population more efficiently and more effectively the next time.

The ability to perform a task in a situation which has never been encountered before (**Learning = Generalization**)



# 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

# 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
- Sets of rules / Logic programs
- Instances
- Graphical models (Bayes/Markov nets)
- Neural networks
- Support vector machines
- Model ensembles
- 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

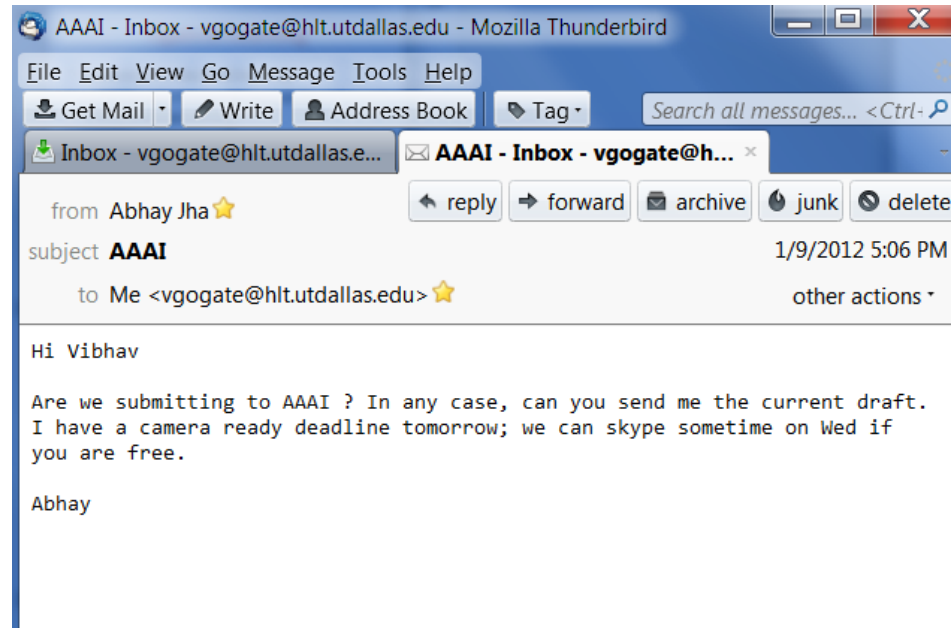
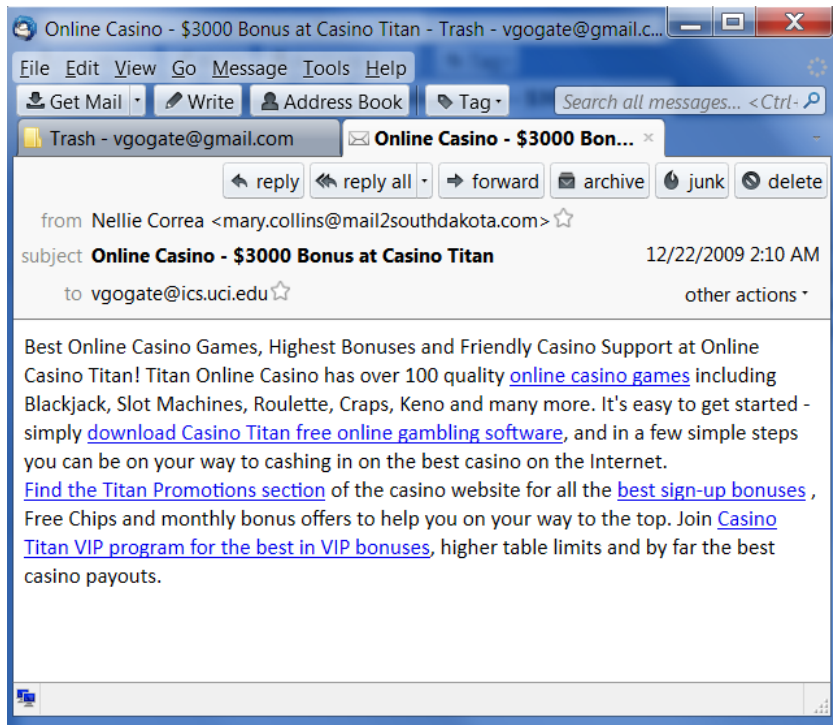
- Classification: learning to predict a discrete value from a predefined set of values
- Regression: learning to predict a continuous/real value



# Machine Learning: Applications

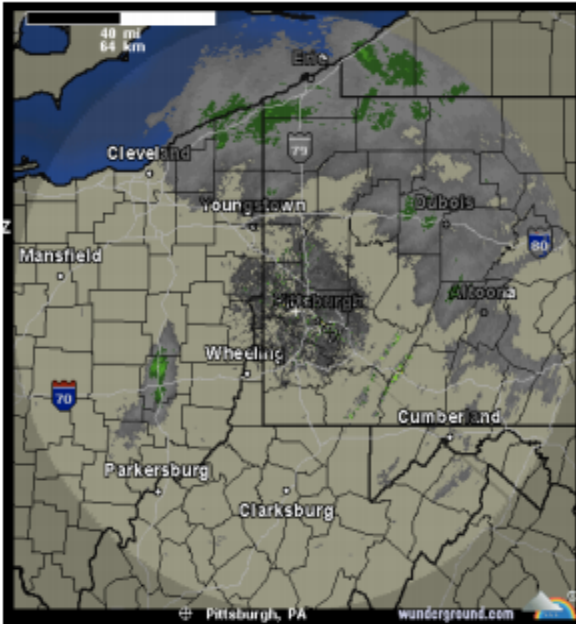
Examples of what you will study in  
class in action!

# Classification Example: Spam Filtering



Classify as “Spam” or “Not Spam”

# Classification Example: Weather Prediction



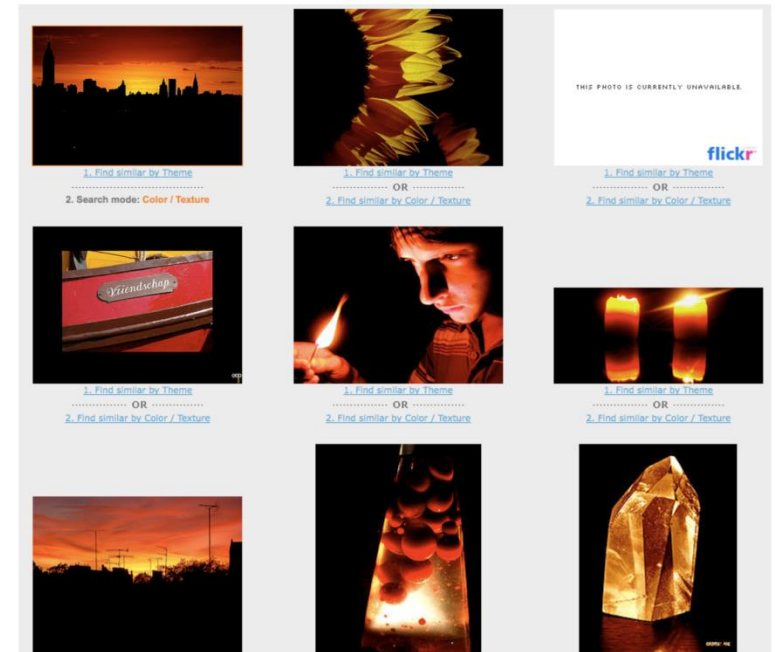
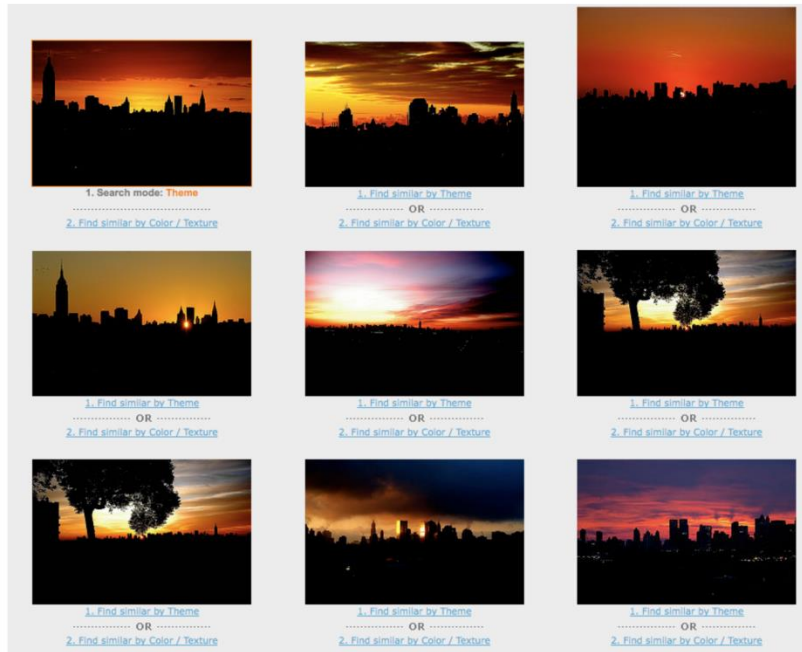
# Regression example: Predicting Gold/Stock prices



Good ML can make you rich (but there is still some risk involved).

**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


Vibhav Gogate | Your Account & Help

Watch InstantlyJust for KidsBrowse DVDsYour Queue★ Suggestions For You

Suggestions (4,663)Rate Shows & MoviesTaste PreferencesWhat You've Rated (316)

Suggestions In: All GenresRATINGS 316

Suggestions to Watch Instantly



**Bob the Builder: Three Musketrucks**


Because you enjoyed:

- Caillou: Caillou's World of Wonder
- Super Why!: Jack and the Beanstalk
- Care Bears: To the Rescue: The Movie

Play

★★★★★

Not Interested



**Thomas & Friends: Carnival Capers**

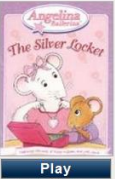
Because you enjoyed:

- Caillou: Caillou's World of Wonder
- Clifford's Really Big Movie
- Dragon Tales: Easy as 1-2-3

Play

★★★★★

Not Interested



**Angelina Ballerina: The Silver Locket**

Because you enjoyed:


- Clifford's Really Big Movie
- Super Why!: Jack and the Beanstalk
- Care Bears: To the Rescue: The Movie

Play

★★★★★

Not Interested

New Suggestions



**Ben 10: Alien Force**


Because you enjoyed:

- Astro Boy

Choose Discs

★★★★★

Not Interested



**Handy Manny: Manny's Motorcycle Adventure**

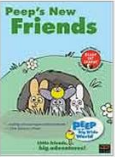
Because you enjoyed:

- Astro Boy
- Caillou: Caillou's World of Wonder
- Toy Story 2

Add

★★★★★

Not Interested



**Peep's New Friends**

Because you enjoyed:


- Caillou: Caillou's World of Wonder
- Clifford's Really Big Movie
- Dragon Tales: Easy as 1-2-3

Add

★★★★★

Not Interested

Action & Adventure



**Gladiator: Extended Edition**

Because you enjoyed:

- The Patriot
- Braveheart
- A Beautiful Mind

*Rate more Action & Adventure*

So we can give you more

# Collaborative Filtering

Amazon.com: Recommended for You - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Search Favorites Go Links

Address <http://www.amazon.com/gp/yourstore/002-8908355-5636015?rGroup=all&>

amazon.com Sue's Store See All 32 Product Categories Your Account | Cart | Wish List | Help |

Improve Your Recommendations | Your Amazon Home | Your Profile | Learn More

Search Amazon.com GO! Web Search GO!

**Recommended for Sue Yeon Syn** (If you're not Sue Yeon Syn, [click here.](#))

**Narrow by Event** Recommendations for you are based on [items you own](#) and more. [More results](#)

[Your Watch List](#) (Beta)


**Narrow by Category**

- [Apparel & Accessories](#)
- [Baby](#)
- [Beauty](#)
- [Books](#)
- [Camera & Photo](#)
- [Computer & Video Games](#)
- [Computers](#)
- [DVD](#)
- [Electronics](#)
- [Health & Personal Care](#)
- [Jewelry & Watches](#)
- [Kitchen & Housewares](#)
- [Magazine Subscriptions](#)
- [Music](#)
- [Outdoor Living](#)
- [Software](#)
- [Sports & Outdoors](#)
- [Tools & Hardware](#)
- [Toys & Games](#)
- [Video](#)

[Select Favorites](#)


**Improve Your Recommendations**  
Update your Amazon history to improve your recommendations  
[Items you own](#)  
[Rated items](#)  
[Not Interested](#)

view: [All](#) | [New Releases](#) | [Coming Soon](#)

- 


**When Things Start to Think**  
by Gershenfeld Neil  
Average Customer Review: ★★★★★  
Publication Date: February 15, 2000  
Our Price: **\$11.20** [Used & new](#) from \$2.00 [Add to cart](#) [Add to Wish List](#)

Rate this item ☐ ★ ☐ ☆ ☐ ☆ ☐ ☆ ☐ ☆ ☐ ☆ ☐ I own it ☐ Not interested

Recommended because you added [The Unfinished Revolution](#) to your Shopping Cart ([edit](#))
- 


**Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web**  
by Tim Berners-Lee  
Average Customer Review: ★★★★★  
Publication Date: November 1, 2000  
Our Price: **\$10.20** [Used & new](#) from \$2.71 [Add to cart](#) [Add to Wish List](#)

Rate this item ☐ ★ ☐ ☆ ☐ ☆ ☐ ☆ ☐ ☆ ☐ ☆ ☐ I own it ☐ Not interested

Recommended because you added [The Unfinished Revolution](#) to your Shopping Cart ([edit](#))
- 

**Perl Cookbook, Second Edition**  
by Tom Christiansen, Nathan Torkington  
Average Customer Review: ★★★★★  
Publication Date: August 21, 2003  
Our Price: **\$32.97** [Used & new](#) from \$15.64 [Add to cart](#) [Add to Wish List](#)

Rate this item ☐ ★ ☐ ☆ ☐ ☆ ☐ ☆ ☐ ☆ ☐ ☆ ☐ I own it ☐ Not interested

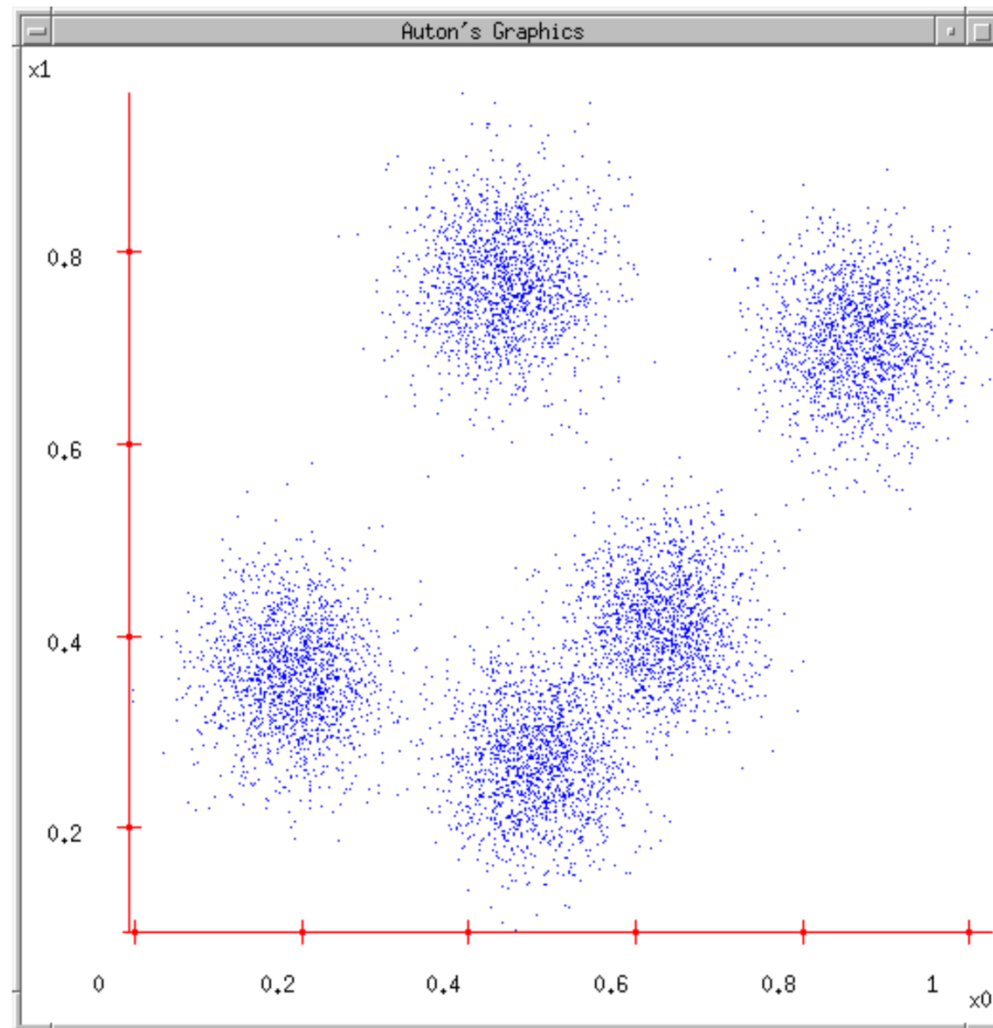
Recommended because you added [Programming Perl \(2nd Edition\)](#) to your Wish List and more ([edit](#))
- 

**Network Analysis, Architecture and Design, Second Edition (The Morgan Kaufmann Series in Networking)**  
by James D. McCabe  
Average Customer Review: ★★★★★  
Publication Date: April 1, 2003  
Our Price: **\$58.46** [Used & new](#) from \$46.77 [Add to cart](#) [Add to Wish List](#)

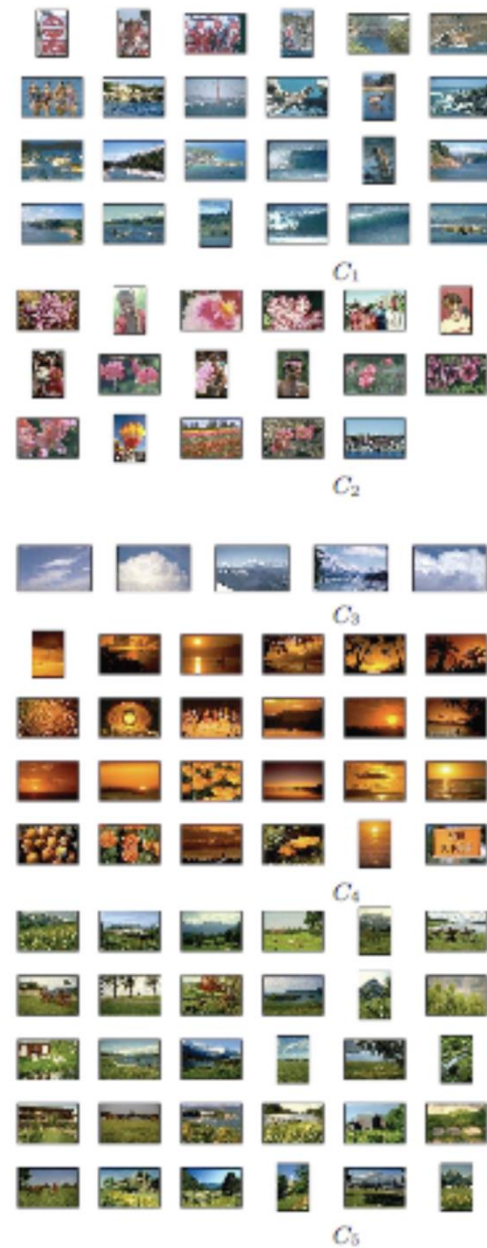
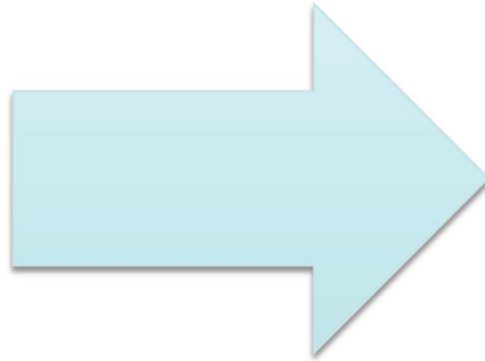
Internet



# Clustering: Discover Structure in data



# Clustering images



[Goldberger et al.]

# 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 so on

**Alice/Bob says: I  
know machine  
learning very well!**

**Potential Employer:  
You are hired!!!**

# What We'll Cover

- **Supervised learning:** Decision tree induction, Rule induction, Instance-based learning, Bayesian learning, Neural networks, Support vector machines, Linear Regression, Model ensembles, Graphical models, Learning theory, etc.
- **Unsupervised learning:** Clustering, Dimensionality reduction
- **Reinforcement learning:** Markov Decision Processes, Q-learning, etc.
- **General machine learning concepts and techniques:** Feature selection, cross-validation, maximum likelihood estimation, gradient descent, expectation-maximization
- **Your responsibility:**
  - Brush up on some important background
  - Linear algebra, Statistics 101, Vectors, Probability theory