

Machine Learning: CS 6375

Introduction

The University of Texas at Dallas

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

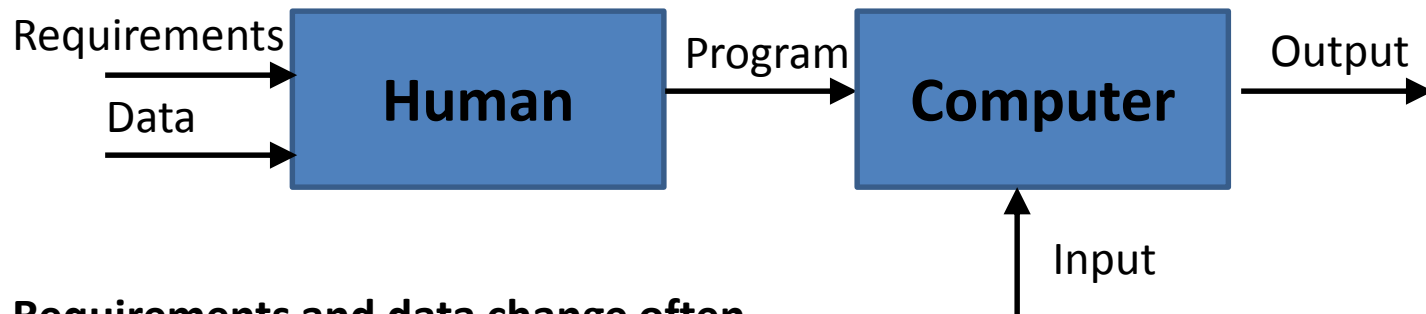
Why Study Machine Learning:

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)

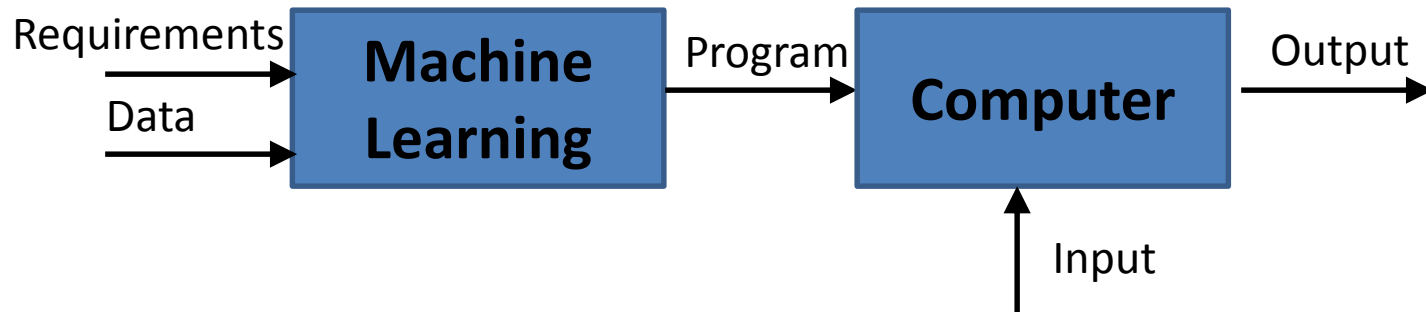
- Getting computers to program themselves
- Writing software is the bottleneck, let data do the work

Traditional Programming



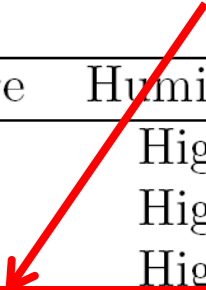
Requirements and data change often

Machine Learning



Training Data

Training Example

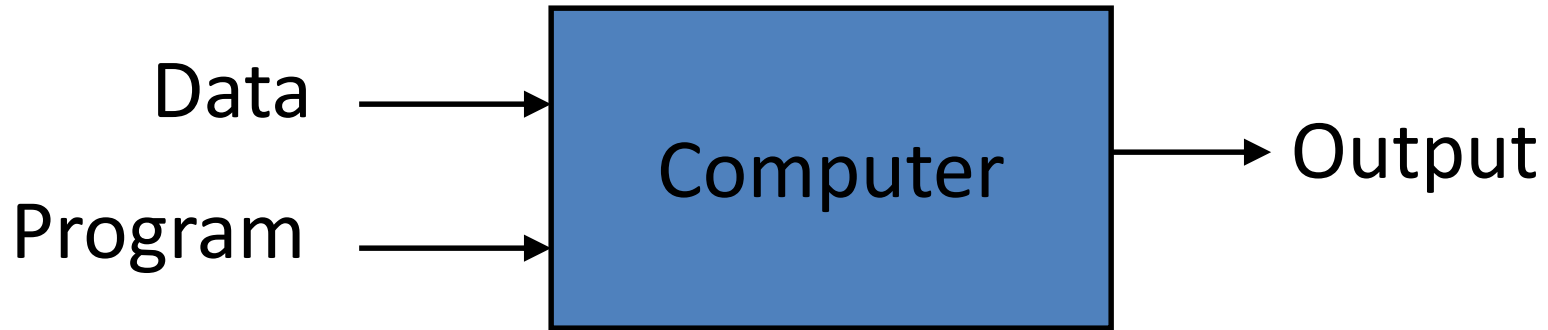


Day	Outlook	Temperature	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

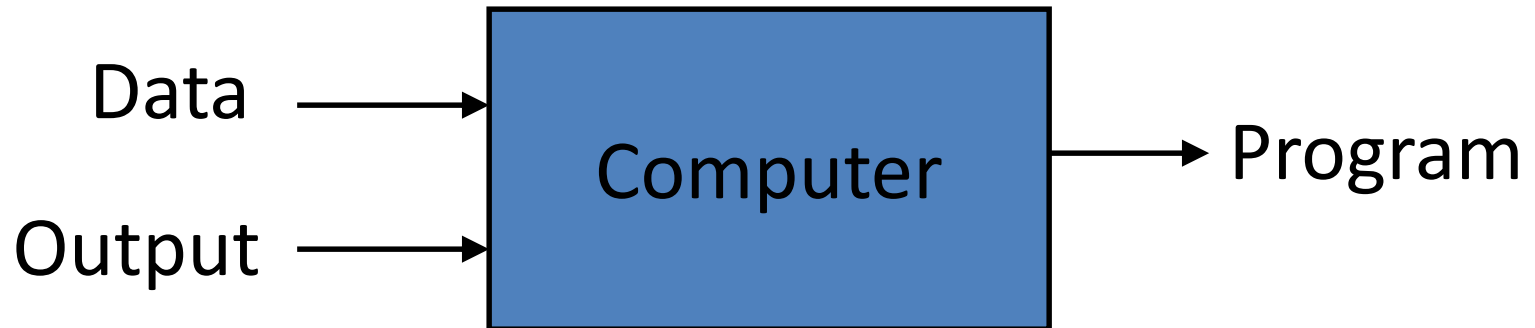
**Two
Classes:
{Yes,No}**

- 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**)

Definition: Machine Learning!

- Pioneer machine learning researcher Arthur Samuel defined machine learning as: “***the field of study that gives computers the ability to learn without being explicitly programmed***”.

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

When to use Machine Learning?

- Human expertise is absent
 - Example: navigating on mars
- Humans are unable to explain their expertise
 - Example: vision, speech, language
- Requirements and data change over time
 - Example: Tracking, Biometrics, Personalized fingerprint recognition
- The problem or the data size is just too large
 - Example: Web Search



Types of Learning

- **Supervised (inductive) learning**
 - Training data includes desired outputs
- **Unsupervised learning**
 - Training data does not include desired outputs
 - Find hidden/interesting 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

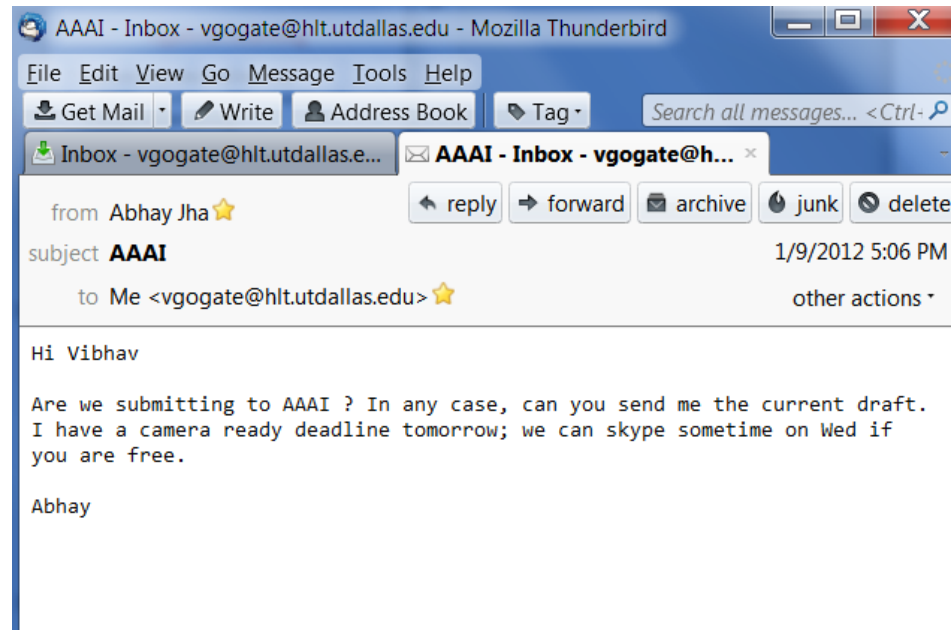
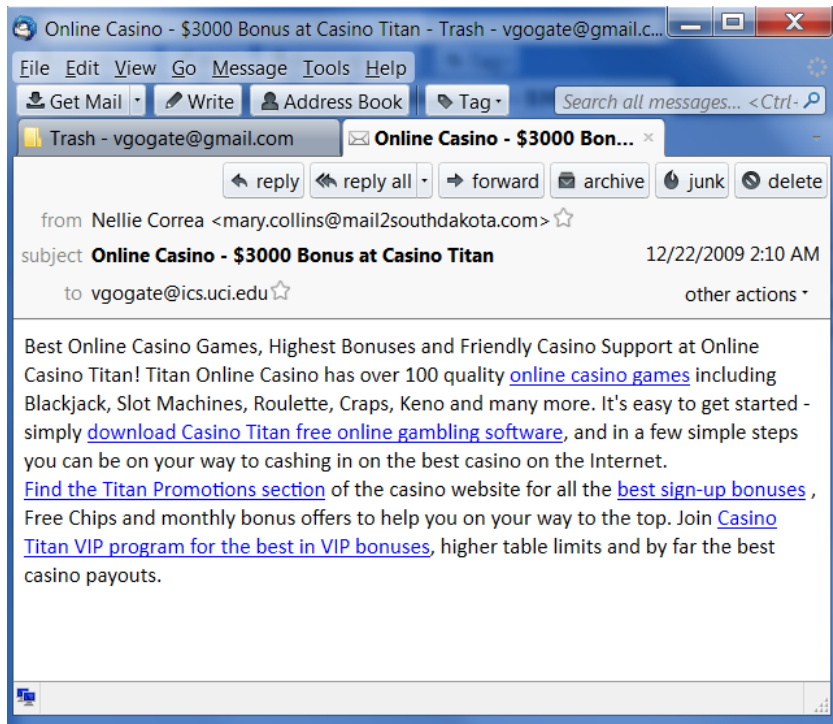
Examples/Types of Machine Learning Tasks

- Forecasting or Prediction
 - Stock price of Google tomorrow?
- Classification and Regression
 - Is Ana credit-worthy?
 - What is Ana's credit score?
- Ranking
 - How to rank images that contain "An awesome machine learning model"?
- Outlier/Anomaly/Fraud detection
 - Is it Ana" using the credit card in Mexico or is it someone else?
- Finding patterns
 - Almost 60% of shoppers buy Diapers and Milk together!

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



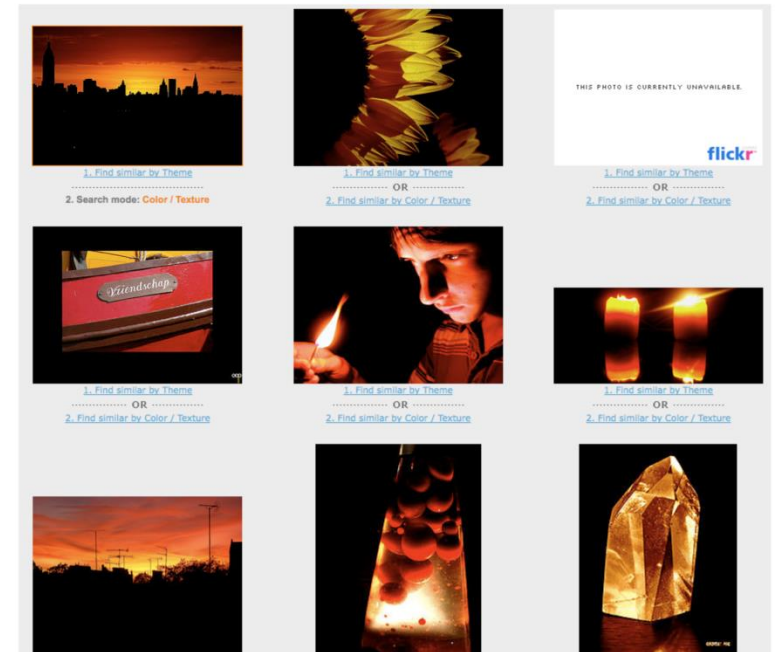
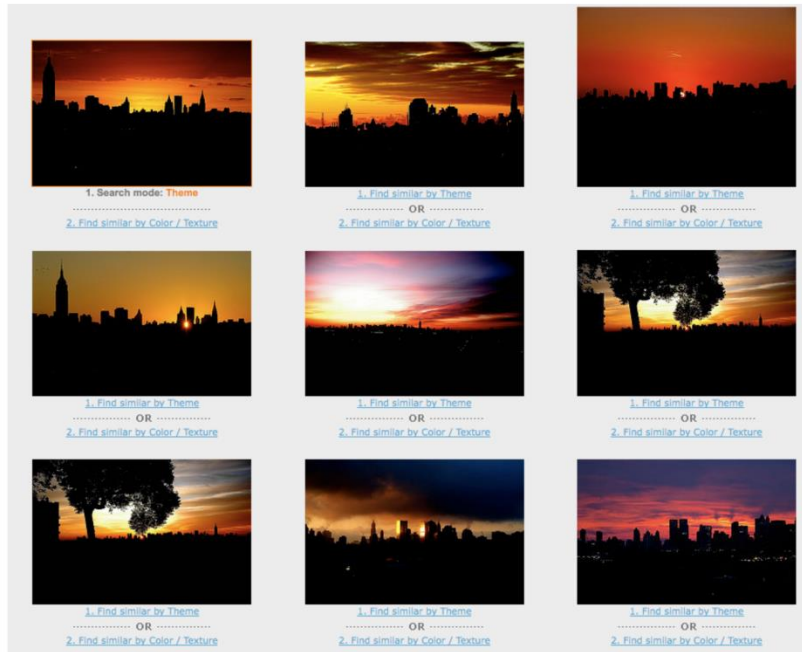
Regression example: Predicting Gold/Stock prices



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

NETFLIX

Vibhav Gogate | Your Account & Help

Watch Instantly

Just for Kids

Browse DVDs

Your Queue

★ Suggestions For You

Movies, TV shows, actors, directors, genres

Suggestions (4,663)

Rate Shows & Movies

Taste Preferences


What You've Rated (316)

Suggestions In: All Genres

RATINGS 316

Suggestions to Watch Instantly

See all >



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



Ben 10: Alien Force


Because you enjoyed:

- Astro Boy

Choose Discs

★★★★★

Not Interested



Handy Manny: Manny's Motorcycle Adventure

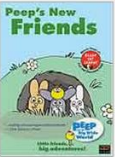
Because you enjoyed:

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

Add

★★★★★

Not Interested



Peep's New Friends

Because you enjoyed:

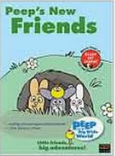
- Caillou: Caillou's World of Wonder
- Clifford's Really Big Movie
- Dragon Tales: Easy as 1-2-3

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 Forward Stop Home Search Favorites RSS Print Mail News Groups

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

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
[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](#)


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

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 X|★★★★★ ☐ 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 X|★★★★★ ☐ 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 X|★★★★★ ☐ I own it ☐ Not interested

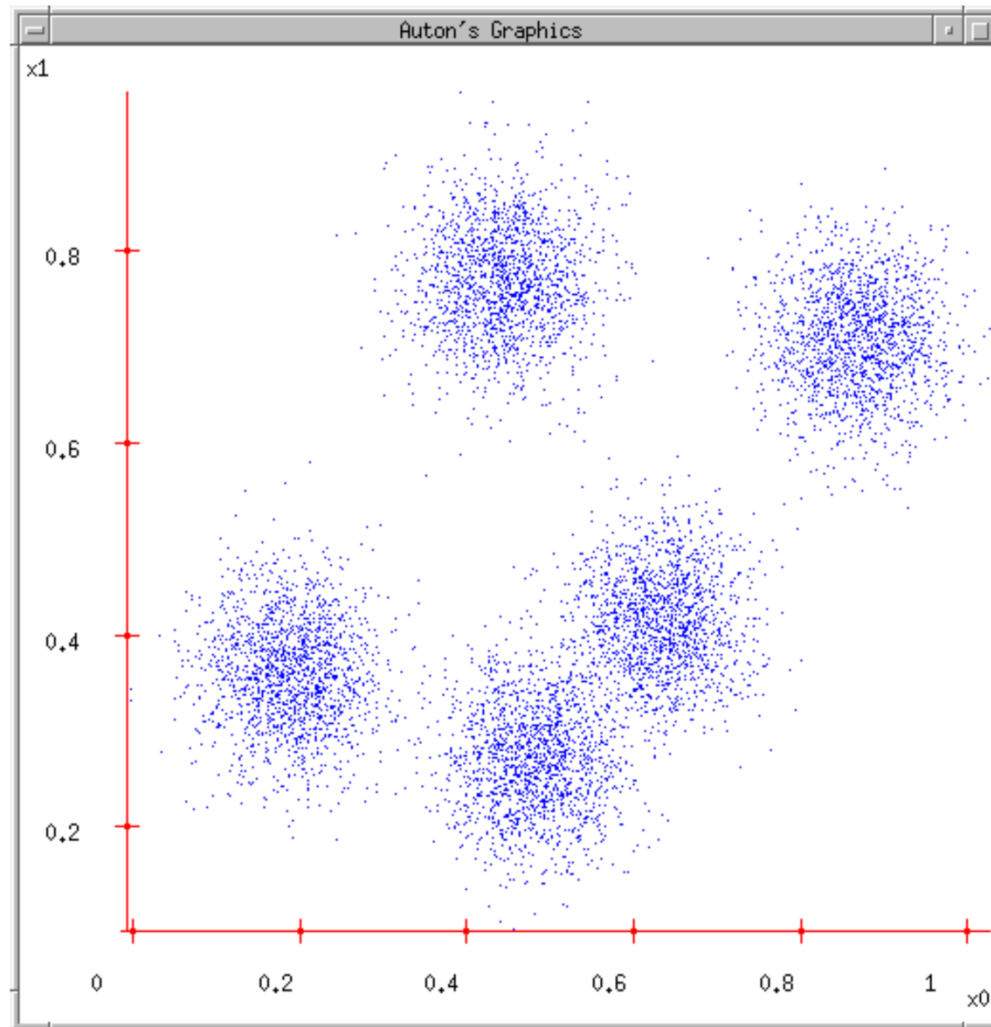
Recommended because you added [Programming Perl \(2nd Edition\)](#) to your Wish List and more ([edit](#))
- 

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](#)

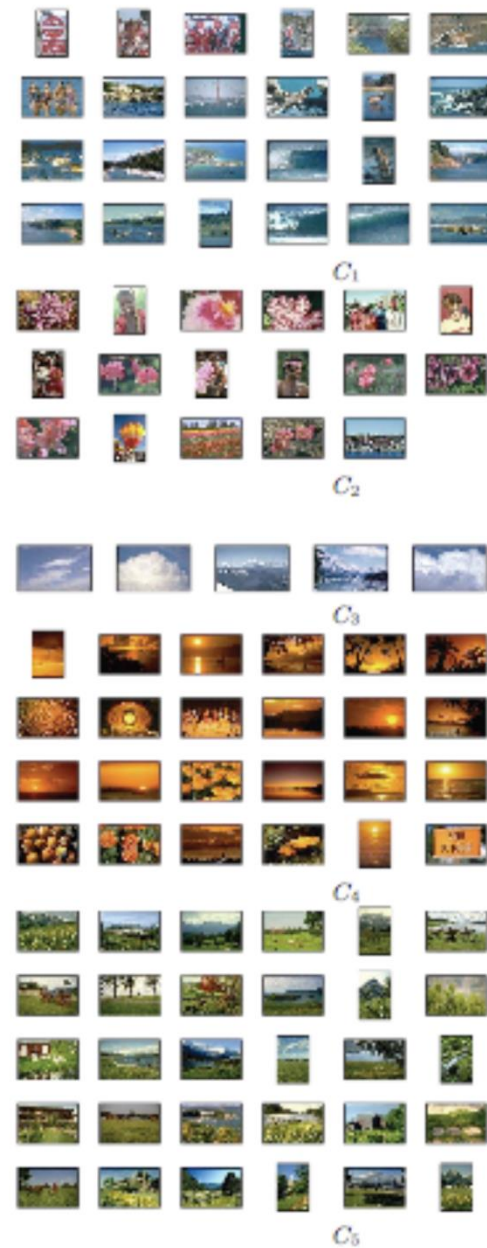
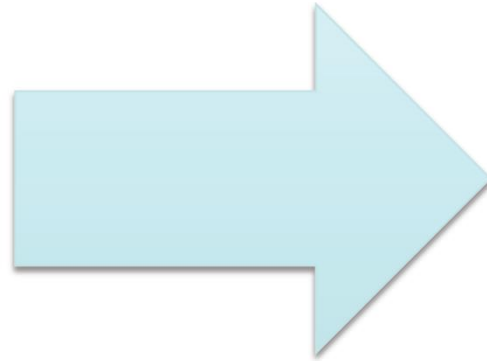
Improve Your Recommendations
Update your Amazon history to improve your recommendations
[Items you own](#)
[Rated items](#)
[Not Interested](#)

Internet

Clustering: Discover Structure in data



Clustering images



[Goldberger et al.]

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

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