

INTRODUCTION TO  
MACHINE LEARNING

David Kauchak  
CS 451 – Fall 2013

Why are you here?


What is Machine Learning?

Why are you taking this course?

What topics would you like to see covered?

Machine Learning is...

Machine learning, a branch of artificial intelligence, concerns the construction and study of systems that can learn from data.



WIKIPEDIA  
The Free Encyclopedia

Machine Learning is...

Machine learning is programming computers to optimize a performance criterion using example data or past experience.  
-- Ethem Alpaydin

The goal of machine learning is to develop methods that can automatically detect patterns in data, and then to use the uncovered patterns to predict future data or other outcomes of interest.  
-- Kevin P. Murphy

The field of pattern recognition is concerned with the automatic discovery of regularities in data through the use of computer algorithms and with the use of these regularities to take actions.  
-- Christopher M. Bishop

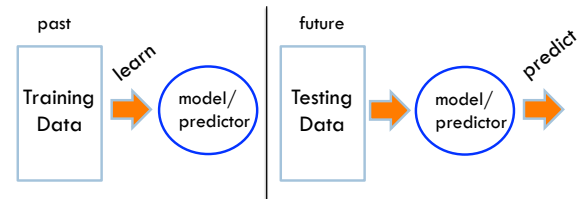
## Machine Learning is...

Machine learning is about predicting the future based on the past.  
-- Hal Daume III



## Machine Learning is...

Machine learning is about predicting the future based on the past.  
-- Hal Daume III



## Machine Learning, aka

*data mining*: machine learning applied to “databases”, i.e. collections of data

*inference and/or estimation* in statistics

*pattern recognition* in engineering

*signal processing* in electrical engineering

*induction*

*optimization*

## Goals of the course: Learn about...

Different machine learning problems

Common techniques/tools used

- ▣ theoretical understanding
- ▣ practical implementation

Proper experimentation and evaluation

Dealing with large (huge) data sets

- ▣ Parallelization frameworks
- ▣ Programming tools

## Goals of the course



Be able to laugh at these signs  
(or at least know why one might...)

## Administrative

### Course page:

- ▣ <http://www.cs.middlebury.edu/~dkauchak/classes/cs451/>
- ▣ [go/cs451](#)

### Assignments

- ▣ Weekly
- ▣ Mostly programming (Java, mostly)
- ▣ Some written/write-up
- ▣ Generally due Friday evenings

### Two exams

### Late Policy

### Honor code

## Course expectations

400-level course

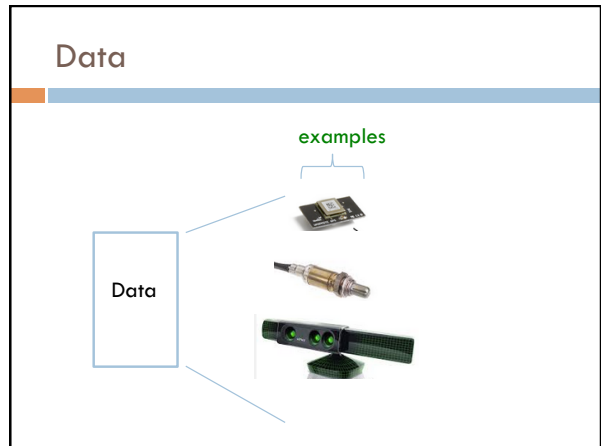
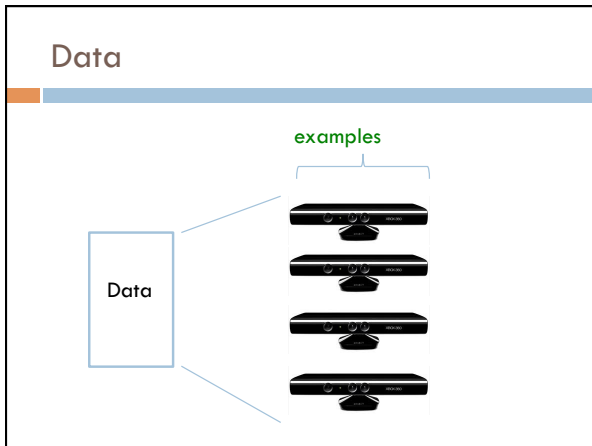
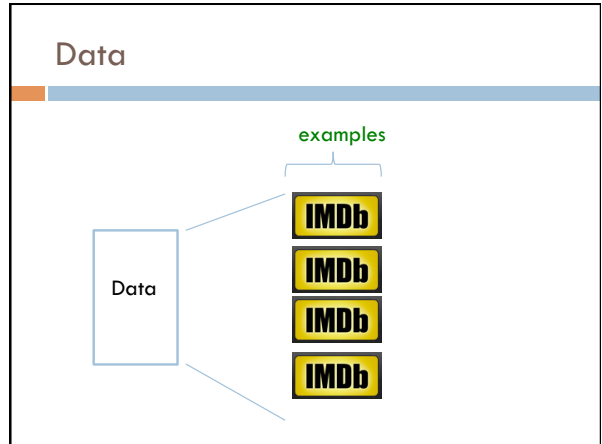
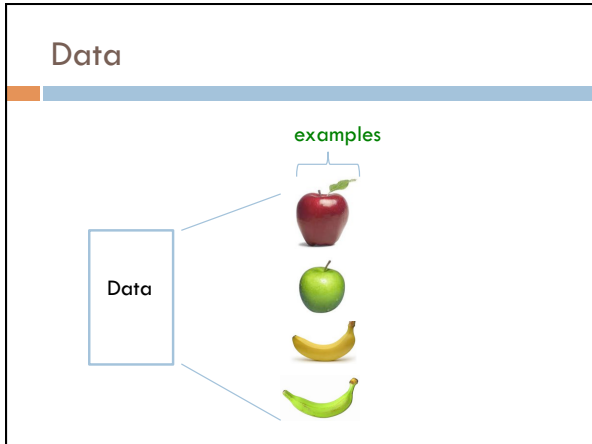
Plan to stay busy!

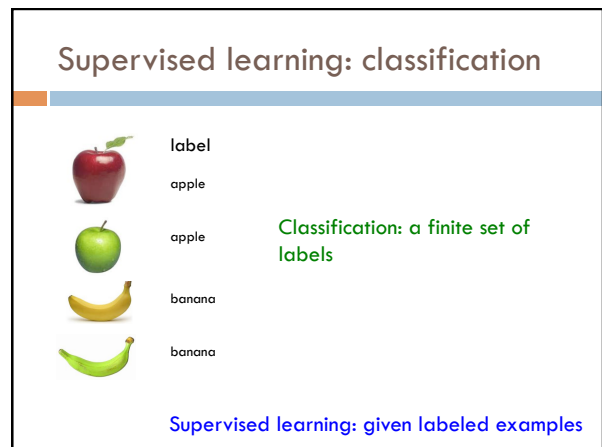
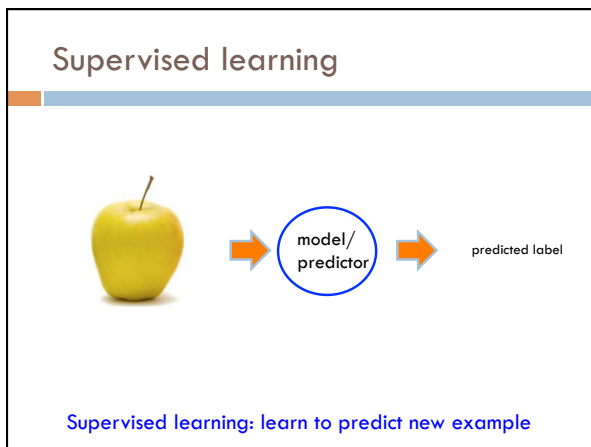
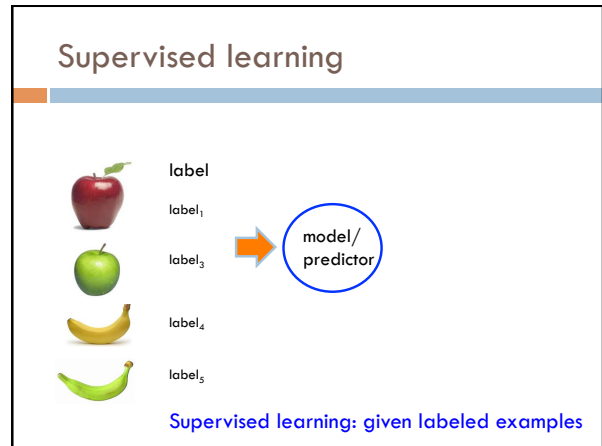
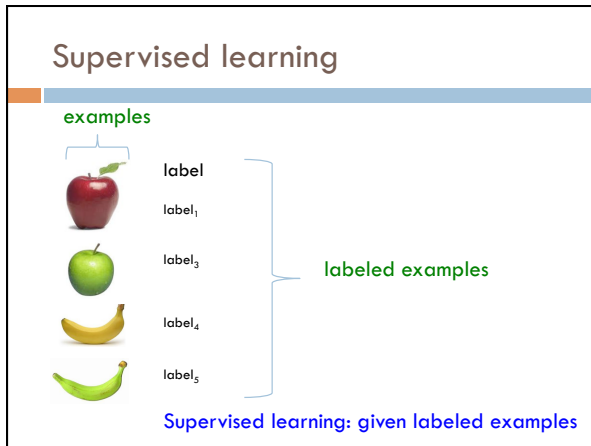
Applied class, so lots of programming

Machine learning involves math

## Machine learning problems

What high-level machine learning problems have you seen or heard of before?





### Classification Example

Differentiate between **low-risk** and **high-risk** customers from their **income** and **savings**

### Classification Applications

- Face recognition
- Character recognition
- Spam detection
- Medical diagnosis: From symptoms to illnesses
- Biometrics: Recognition/authentication using physical and/or behavioral characteristics: Face, iris, signature, etc
- ...

### Supervised learning: regression

	label	
	-4.5	
	10.1	Regression: label is real-valued
	3.2	
	4.3	

Supervised learning: given labeled examples

### Regression Example

Price of a used car





x : car attributes (e.g. mileage)

y : price

## Regression Applications

- Economics/Finance: predict the value of a stock
- Epidemiology
- Car/plane navigation: angle of the steering wheel, acceleration, ...
- Temporal trends: weather over time
- ...

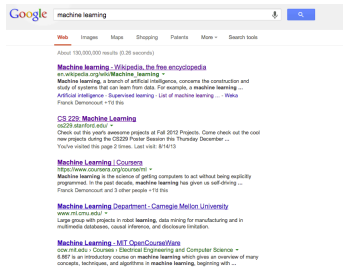
## Supervised learning: ranking

	label 1	Ranking: label is a ranking
	4	
	2	
	3	

Supervised learning: given labeled examples

## Ranking example

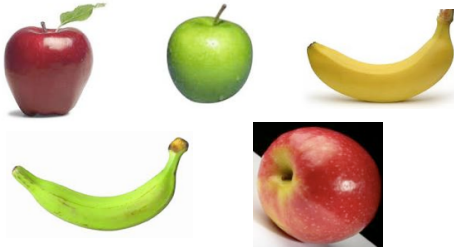
Given a query and a set of web pages, rank them according to relevance



## Ranking Applications

- User preference, e.g. Netflix "My List" -- movie queue ranking
- iTunes
- flight search (search in general)
- reranking N-best output lists
- ...

## Unsupervised learning



Unsupervised learning: given data, i.e. examples, but no labels

## Unsupervised learning applications

- learn clusters/groups without any label
- customer segmentation (i.e. grouping)
- image compression
- bioinformatics: learn motifs
- ...

## Reinforcement learning

left, right, straight, left, left, left, straight **GOOD**

left, straight, straight, left, right, straight, straight **BAD**

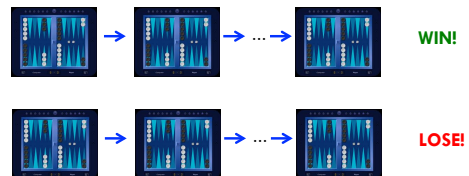
left, right, straight, left, left, left, straight **18.5**

left, straight, straight, left, right, straight, straight **-3**

Given a *sequence* of examples/states and a *reward* after completing that sequence, learn to predict the action to take in for an individual example/state

## Reinforcement learning example

Backgammon



Given sequences of moves and whether or not the player won at the end, learn to make good moves



## Reinforcement learning example



<http://www.youtube.com/watch?v=VCdxqn0fcnE>

## Other learning variations

### What data is available:

- Supervised, unsupervised, reinforcement learning
- semi-supervised, active learning, ...

### How are we getting the data:

- online vs. offline learning

### Type of model:

- generative vs. discriminative
- parametric vs. non-parametric