

Machine Learning

# Introduction

---

## Welcome

Andrew Ng



Andrew Ng



## Machine Learning

- Grew out of work in AI
- New capability for computers

### Examples:

- Database mining  
Large datasets from growth of automation/web.  
E.g., Web click data, medical records, biology, engineering
- Applications can't program by hand.  
E.g., Autonomous helicopter, handwriting recognition, most of  
Natural Language Processing (NLP), Computer Vision.

## Machine Learning

- Grew out of work in AI

## Exam

- |

- |



lost of  
ig

## Machine Learning

- Grew out of work in AI
- New capability for computers

### Examples:

- Database mining  
Large datasets from growth of automation/web.  
E.g., Web click data, medical records, biology, engineering
- Applications can't program by hand.  
E.g., Autonomous helicopter, handwriting recognition, most of  
Natural Language Processing (NLP), Computer Vision.

## Machine Learning

- Grew out of work in AI
- New capability for computers

### Examples:

- Database mining  
Large datasets from growth of automation/web.  
E.g., Web click data, medical records, biology, engineering
- Applications can't program by hand.  
E.g., Autonomous helicopter, handwriting recognition, most of  
Natural Language Processing (NLP), Computer Vision.
- Self-customizing programs  
E.g., Amazon, Netflix product recommendations

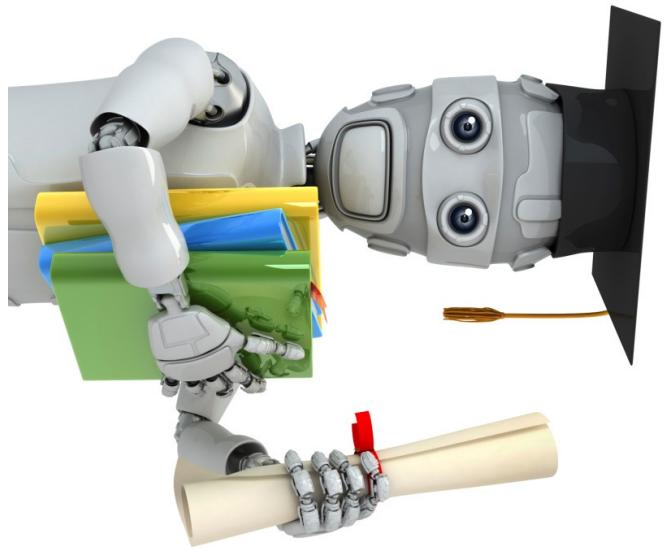
## Machine Learning

- Grew out of work in AI
- New capability for computers

### Examples:

- Database mining  
Large datasets from growth of automation/web.  
E.g., Web click data, medical records, biology, engineering
- Applications can't program by hand.  
E.g., Autonomous helicopter, handwriting recognition, most of  
Natural Language Processing (NLP), Computer Vision.
- Self-customizing programs  
E.g., Amazon, Netflix product recommendations
- Understanding human learning (brain, real AI).

Andrew Ng



Machine Learning

# Introduction

---

## What is machine learning

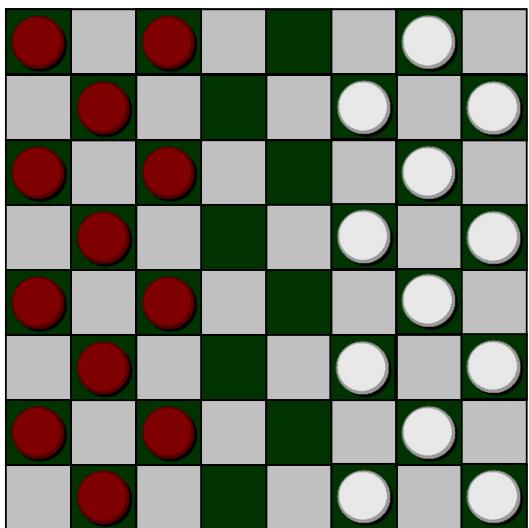
# Machine Learning definition

# Machine Learning definition

- Arthur Samuel (1959). Machine Learning: Field of study that gives computers the ability to learn without being explicitly programmed.

# Machine Learning definition

- Arthur Samuel (1959). Machine Learning: Field of study that gives computers the ability to learn without being explicitly programmed.



# Machine Learning definition

- Arthur Samuel (1959). Machine Learning: Field of study that gives computers the ability to learn without being explicitly programmed.
- Tom Mitchell (1998) Well-posed Learning  
Problem: A computer program is said to *learn* from experience  $E$  with respect to some task  $T$  and some performance measure  $P$ , if its performance on  $T$ , as measured by  $P$ , improves with experience  $E$ .

"A computer program is said to *learn from experience*  $E$  with respect to some task  $T$  and some performance measure  $P$ , if its performance on  $T$ , as measured by  $P$ , improves with experience  $E$ ."

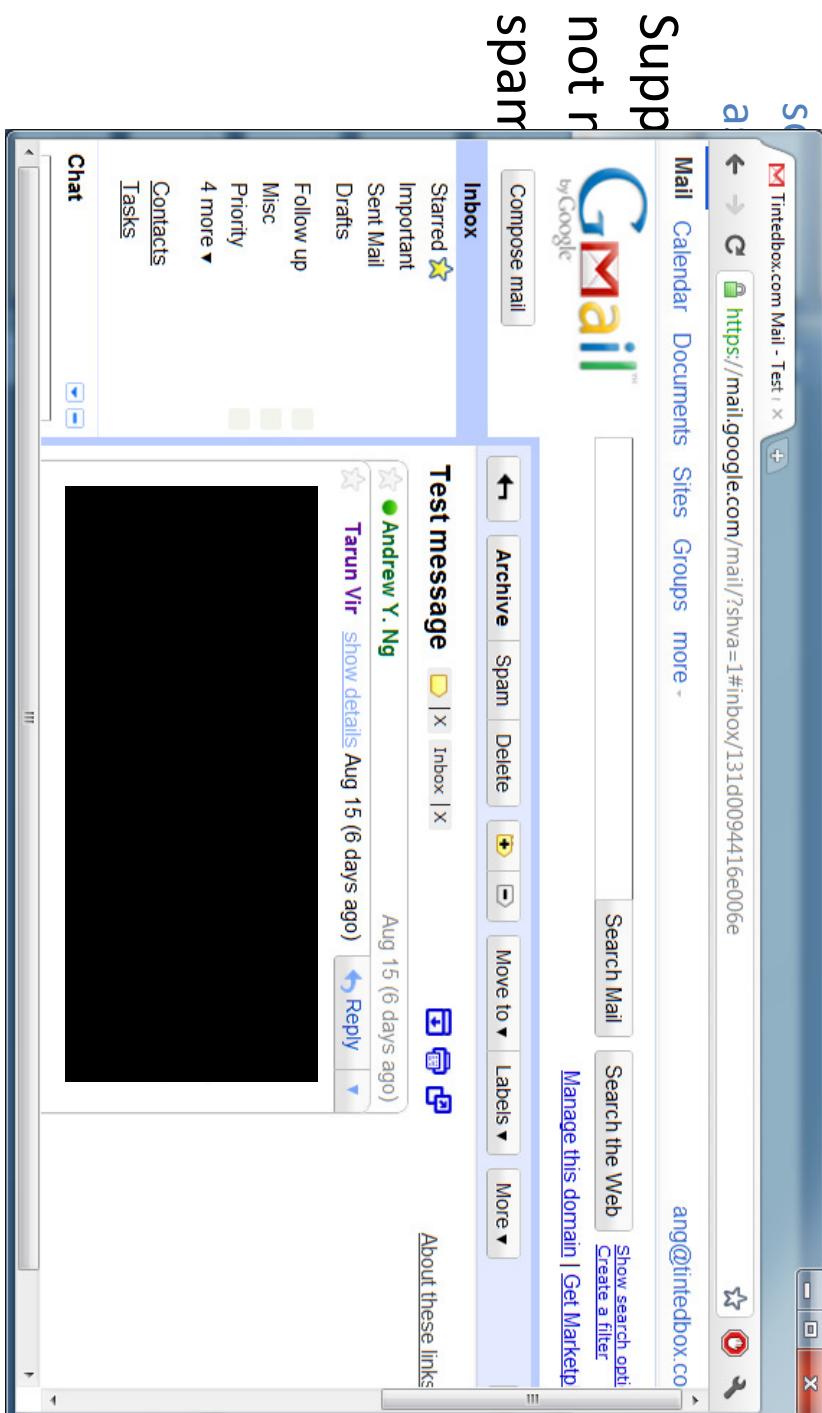
Suppose your email program watches which emails you do or do not mark as spam, and based on that learns how to better filter spam. What is the task  $T$  in this setting?

- Classifying emails as spam or not spam.  $T \leftarrow$
- Watching you label emails as spam or not spam.  $E \leftarrow$
- The number (or fraction) of emails correctly classified as spam/not spam.
- None of the above—this is not a machine learning problem.  $P \leftarrow$

"A computer program is said to *learn from experience E* with respect to

on T,

Support not on T,  
spam or do  
filter



spam.

*"A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E."*

Suppose your email program watches which emails you do or do not mark as spam, and based on that learns how to better filter spam. What is the task T in this setting?

- Classifying emails as spam or not spam.       $T \leftarrow$
- Watching you label emails as spam or not spam.       $E \leftarrow$
- The number (or fraction) of emails correctly classified as spam/not spam.
- None of the above—this is not a machine learning problem.       $P \leftarrow$

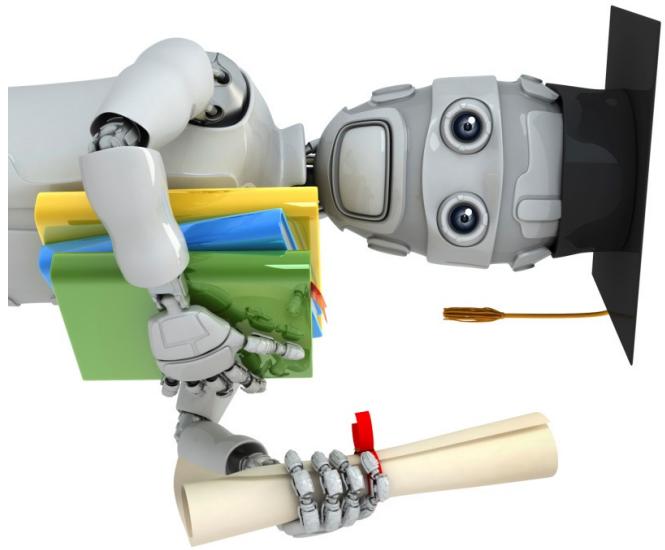
Machine learning algorithms:

- Supervised learning
- Unsupervised learning

Others: Reinforcement learning, recommender systems.

Also talk about: Practical advice for applying learning algorithms.

Andrew Ng



Machine Learning

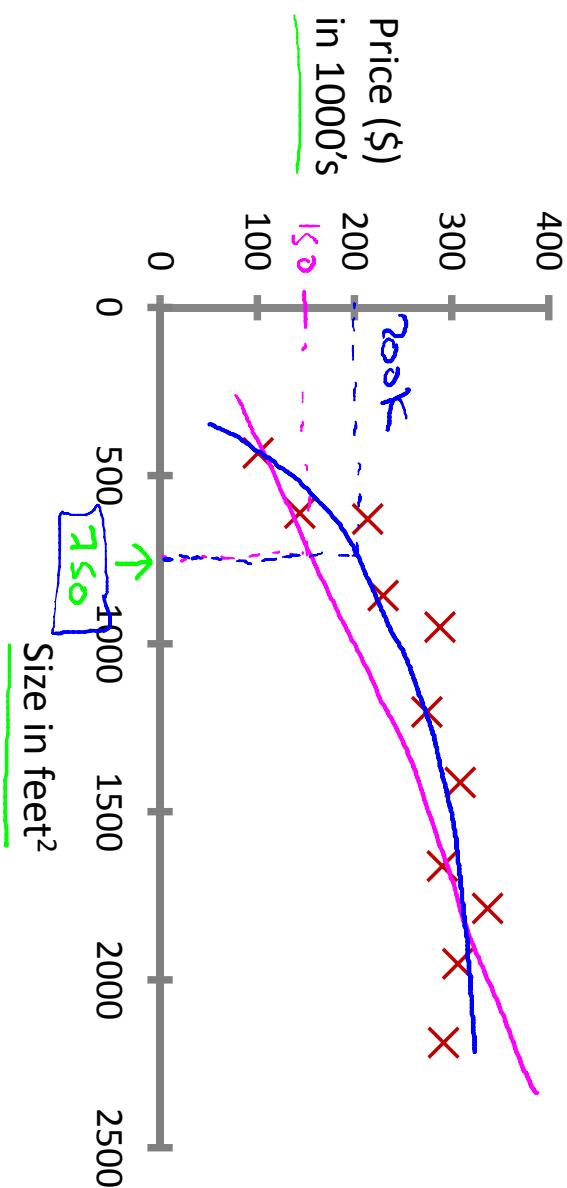
# Introduction

---

# Supervised

# Learning

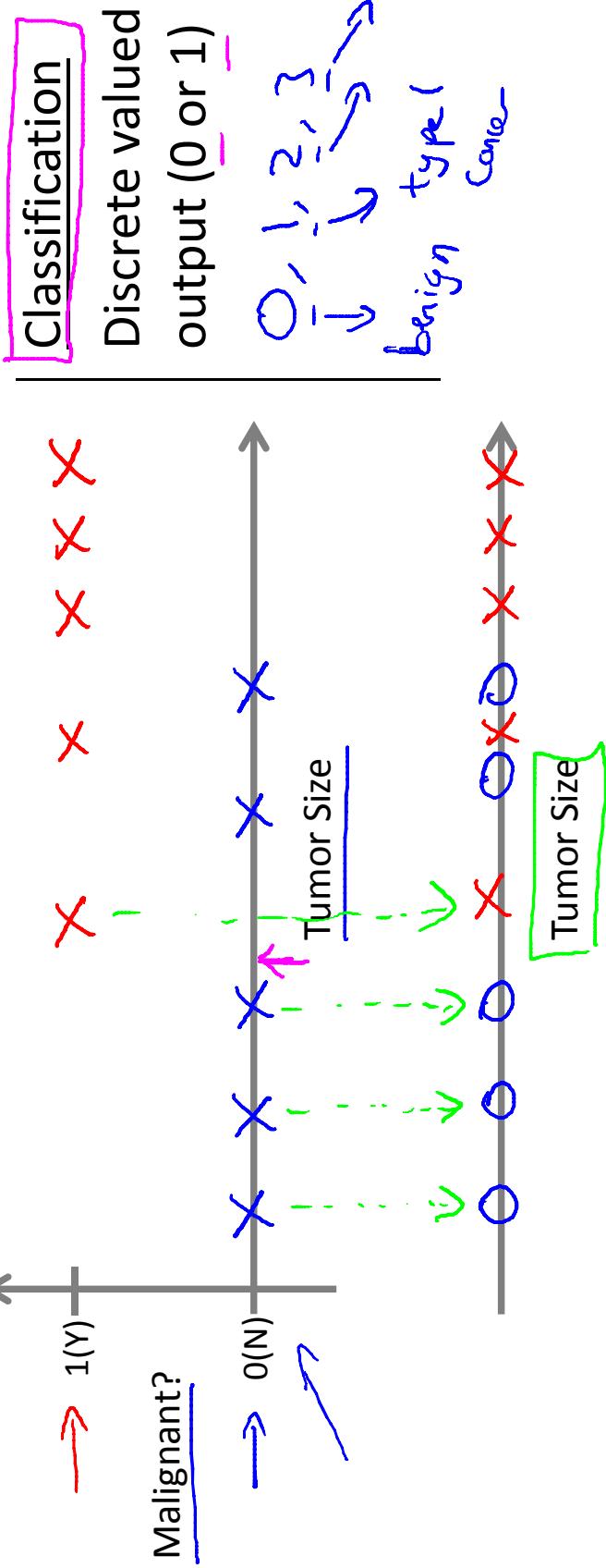
## Housing price prediction.

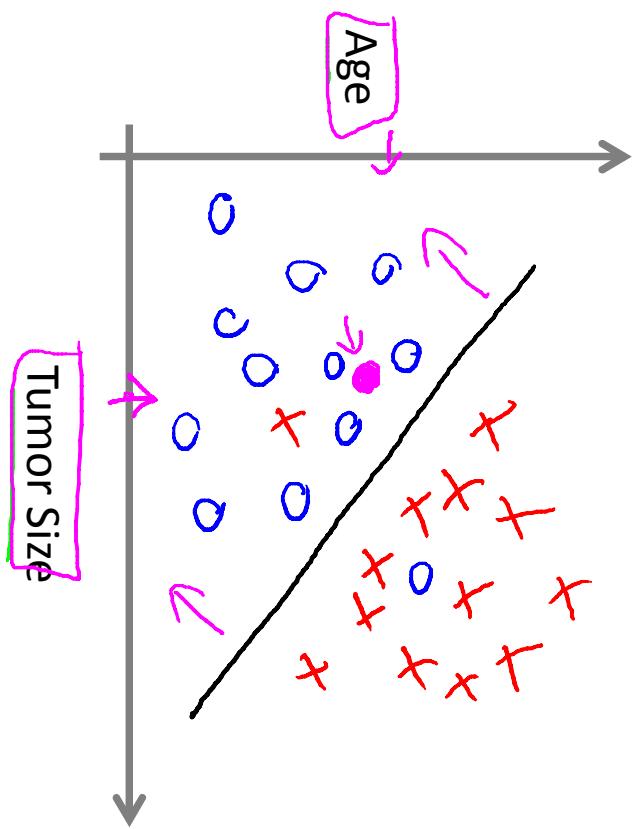


Supervised Learning  
"right answers" given

Regression: Predict continuous valued output (price)

## Breast cancer (malignant, benign)





- Clump Thickness
- Uniformity of Cell Size
- Uniformity of Cell Shape
- ...

You're running a company, and you want to develop learning algorithms to address each of two problems.

1000's

→ Problem 1: You have a large inventory of identical items. You want to predict how many of these items will sell over the next 3 months.

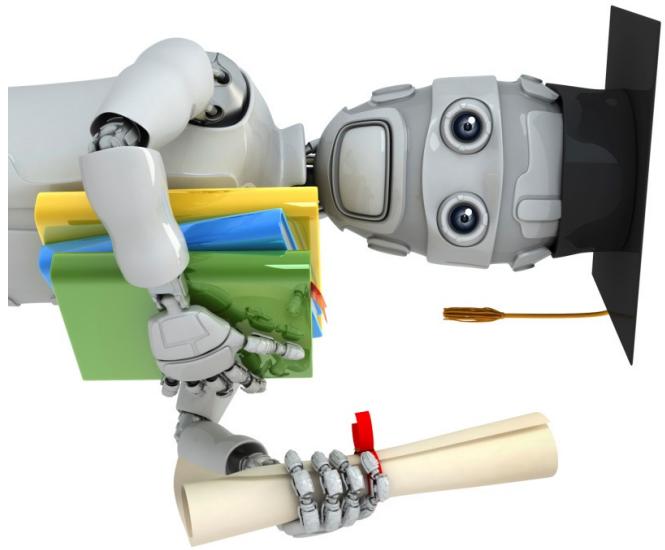
→ Problem 2: You'd like software to examine individual customer accounts, and for each account decide if it has been hacked/compromised.

→ 0 — not hacked  
→ 1 — hacked

Should you treat these as classification or as regression problems?

- Treat both as classification problems.
- Treat problem 1 as a classification problem, problem 2 as a regression problem.
- Treat problem 1 as a regression problem, problem 2 as a classification problem.
- Treat both as regression problems.

Andrew Ng



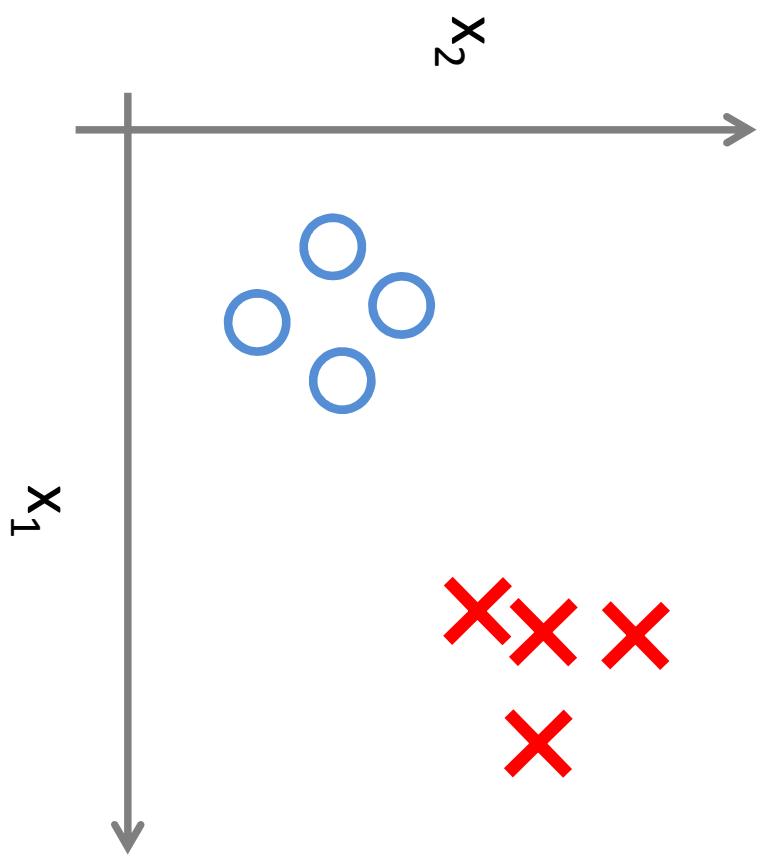
Machine Learning

# Introduction

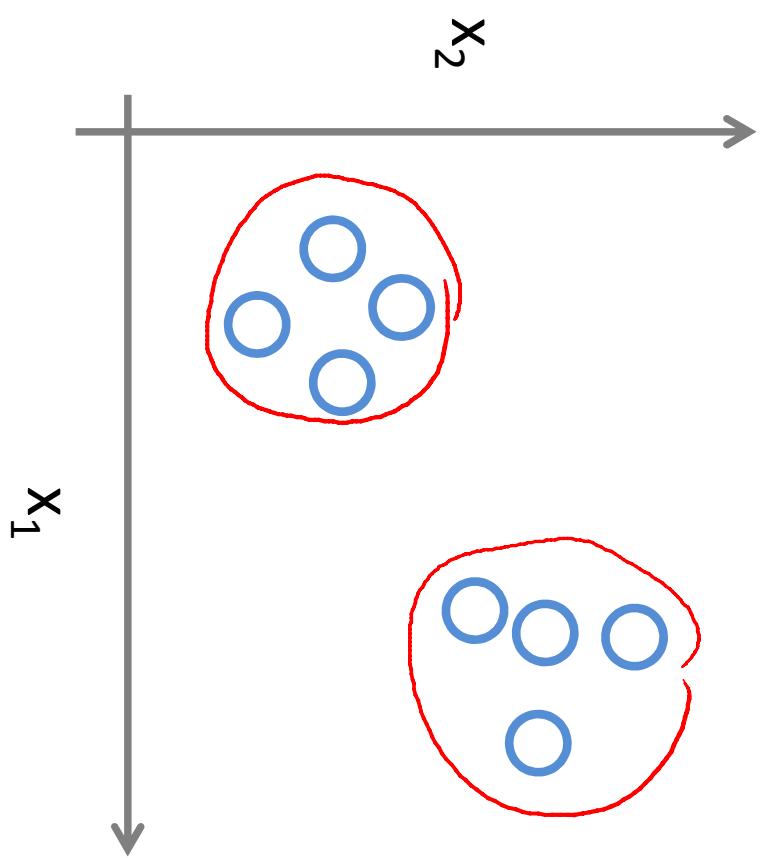
---

## Unsupervised Learning

# Supervised Learning



# Unsupervised Learning



Google News | news.google.com | andrewyantaking@gmail.com | Web History | Settings ▾ | Sign out

**Top Stories**

- Deepwater Horizon
- Fed meeting
- Foreign exchange market
- Lindsay Lohan
- IBM
- Tom Brady
- Toronto International Film Festival
- Paris Hilton
- Iran
- Hurricane Igor
- Starred
- San Francisco Bay Area
- World
- U.S.
- Business
- Sci/Tech
- More Top Stories
- Spotlight
- Health
- Sports
- Entertainment

**Top Stories**

- Christine O'Donnell »
- White House official denies Tea Party-focused ad campaign**
- Democratic sources say the White House is not considering an ad campaign tying Republicans to the Tea Party. Washington (CNN) - A top White House official sharply denied a report that claims President Obama's political advisers are weighing a national ...
- Tea Party is misplacing the blame, former President Bill Clinton claims
- New York Daily News
- GOP tea party backer defends Christine O'Donnell
- The Associated Press
- Atlanta Journal Constitution - Politics Daily - MyFox Washington DC - Salon
- all 726 news articles »

**US Stocks Climb After Recession Called Over.**

**Homebuilders Gain**

MarketWatch - Kristina Peterson - 16 minutes ago

**NEW YORK** (MarketWatch) - US stocks climbed Monday, gaining speed after a key nonprofit organization officially called the recession over, giving investors a boost of confidence in the gradual economic recovery.

Longest recession since 1930s ended in June 2009, group says

Los Angeles Times

Downturn Was Longest in Decades, Panel Confirms

New York Times

Wall Street Journal - AFP - CNN - USA Today

all 276 news articles »

**RECESsion**

CNN Money - Chris Isidore - 39 minutes ago

Hurricane Igor lashes Bermuda

USA Today - Gerry Broome - 5 minutes ago

Explain what you want from us,' reads front-page editorial

msnbc.com - Olivia Torres - 10 minutes ago

CNN Interna...

Crisis response: Pakistan floods

**San Francisco Bay Area - Edit**

Clorox »

Bay Buzz Buzz. Clorox close to selling S.T.P.

Bay All

San Jose Mercury News - 48 minutes ago - all 28 articles »

Google's official beekeeper keeps the company buzzing with excitement

San Jose Mercury News - Bruce Newman - 1 hour ago

Jon Sylva »

Martinez man still unconscious as police investigate weekend shooting

San Jose Mercury News - Robert Salonga - 48 minutes ago - all 6 articles »

Weiss Doubts BP Would End Operations in Gulf of Mexico. Video Bloomberg

Reuters

**BP Oil Well, Site of National Catastrophe, Dies at One**

Vanity Fair - Juli Werner - 22 minutes ago

The BP oil well site of the Deepwater Horizon explosion that led to the worst oil spill in US history, died today at one year old.

Video: Blow-out BP Well Finally Killed in Gulf  The Associated Press

Weiss Doubts BP Would End Operations in Gulf of Mexico. Video Bloomberg

CNN International - Wall Street Journal (blog) - The Guardian - New York Times

all 2,292 news articles »

**Spotlight**

Sarkozy rages at EU 'humiliation'

Financial Times - Peggy Hollinger -

Seo 16, 2010

Google News

Top Stories

White House official denies Tea Party-focused ad campaign

U.S.

Recession

San Francisco Bay Area

BP oil well blowout

BP oil well blowout

BP kills Macondo, But Its Legacy Lives On

The Wall Street Journal

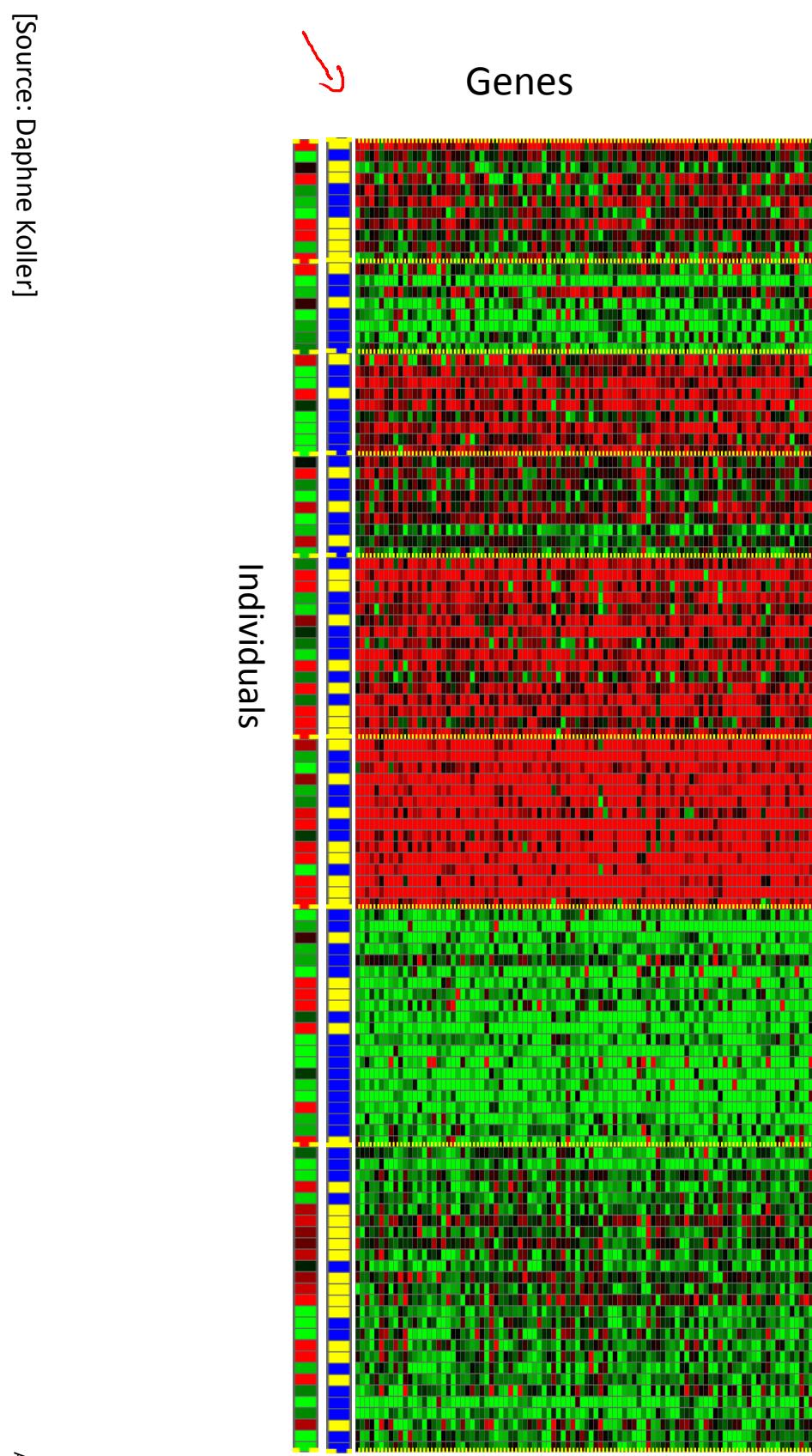
THE SOURCE

Allen: Well is dead, but much Gulf Coast work remains

BP oil spill cost hits nearly \$10bn

BP's costs for the Deepwater Horizon disaster have hit \$10bn. Photo: AP

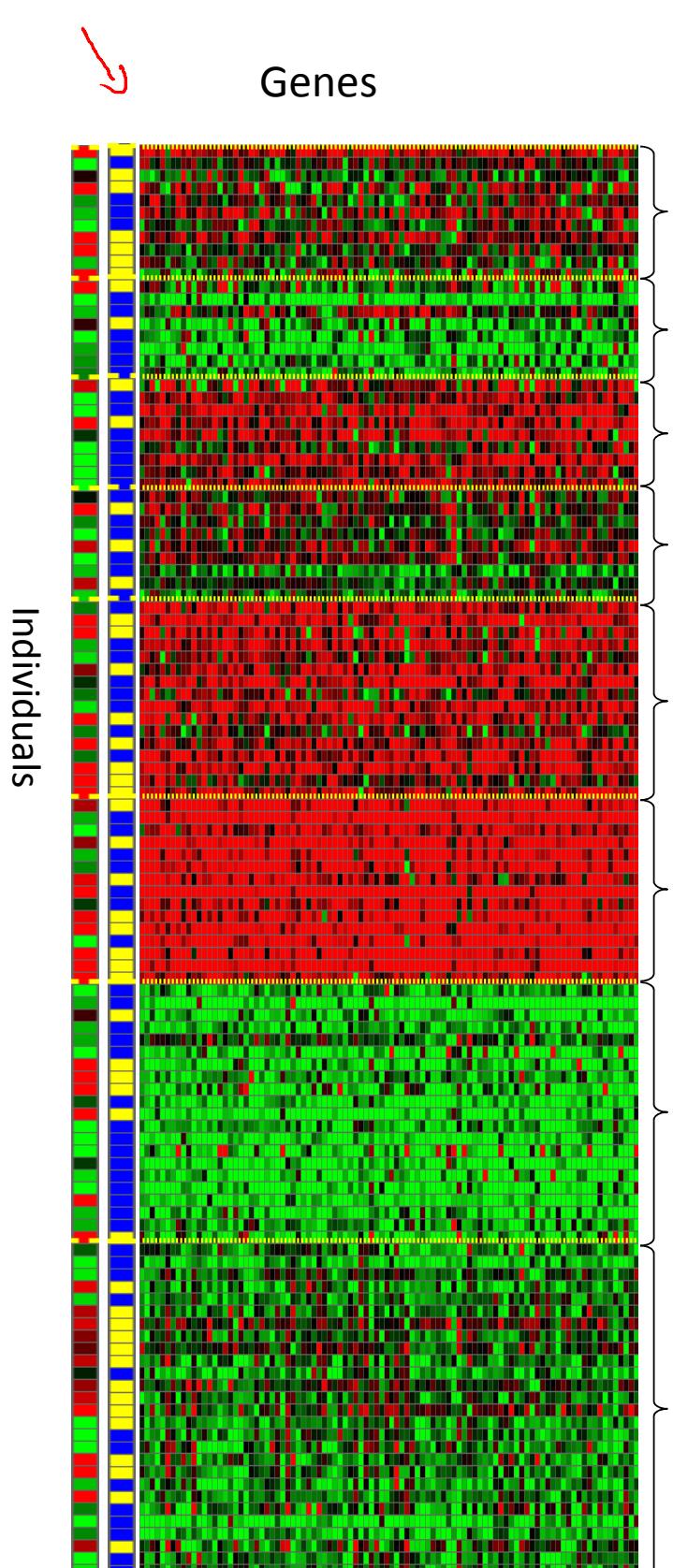
Andrew Ng



Andrew Ng

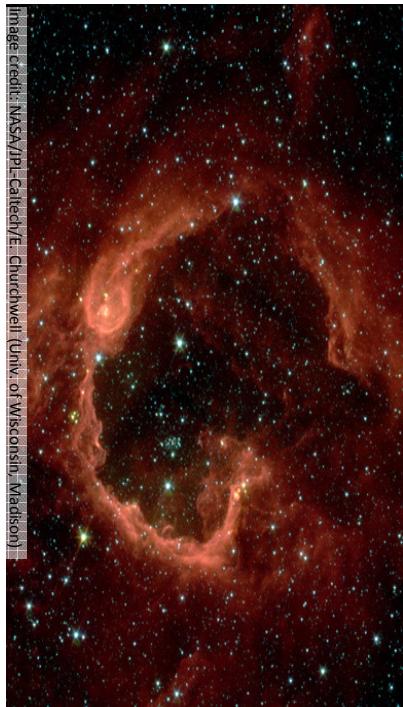
[Source: Daphne Koller]

[Source: Daphne Koller]

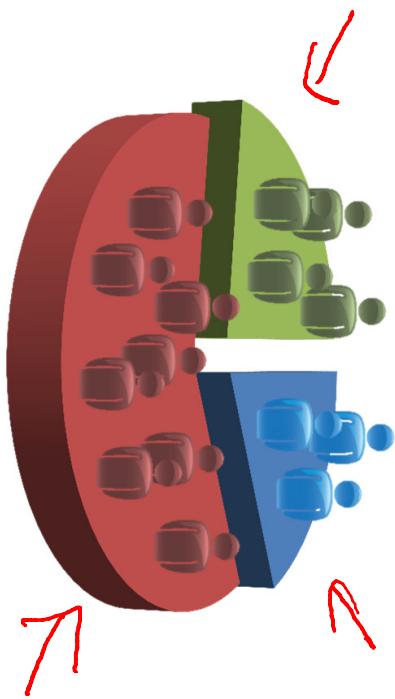


Andrew Ng

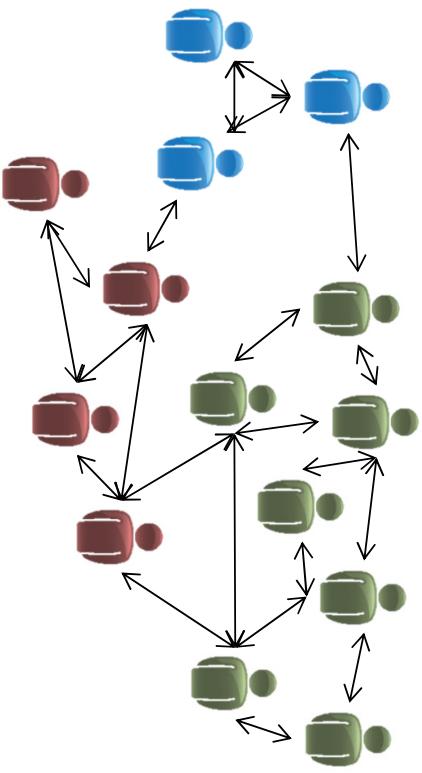
## Astronomical data analysis



## Market segmentation



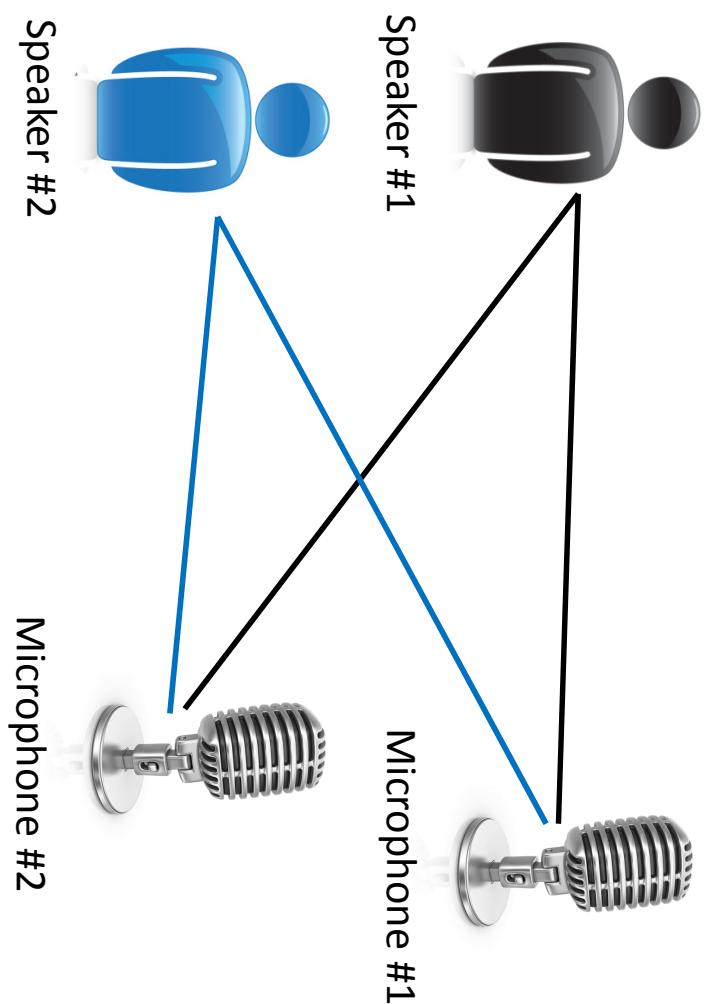
## Social network analysis



## Organize computing clusters



# Cocktail party problem



Microphone #1:



Output #1:



Microphone #2:



Output #2:



Microphone #1:



Output #1:



Microphone #2:



Output #2:



[Audio clips courtesy of Te-Won Lee.]

# Cocktail party problem algorithm

```
[W,S,V] = svd((repmat(sum(x.*x,1),size(x,1),1).*x).*x');
```

[Source: Sam Roweis, Yair Weiss & Eero Simoncelli]

Of the following examples, which would you address using an unsupervised learning algorithm? (Check all that apply.)

- Given email labeled as spam/not spam, learn a spam filter.
- Given a set of news articles found on the web, group them into a set of articles about the same story.
- Given a database of customer data, automatically discover market segments and group customers into different market segments.
- Given a dataset of patients diagnosed as either having diabetes or not, learn to classify new patients as having diabetes or not.

Andrew Ng