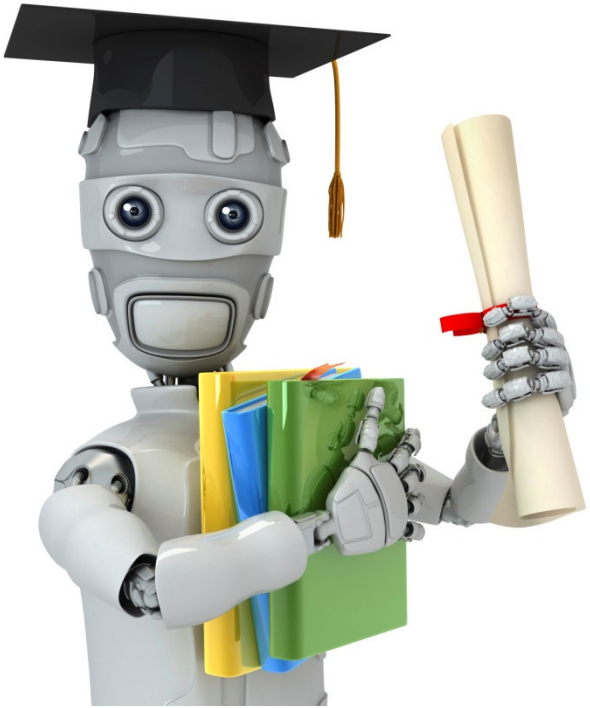


Machine Learning

Introduction

Welcome


Apple - iPhoto - New full-
www.apple.com/ilife/iphoto/
Store
Mac
iPod
iPhone
iPad
iTunes
Support
iLife '11
iPhoto
iMovie
GarageBand
Video Showcase
Resources
Upgrade Now
iPhoto '11
From your Facebook Wall to your coffee table to your best friend's inbox (or mailbox). Do more with your photos than you ever thought possible. And do it all in one place. iPhoto.
Watch the iPhoto video
MacBook Pro
What's New in iPhoto
What is iPhoto?

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

- I

Exam

- I

ıg

- A

ıost of

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

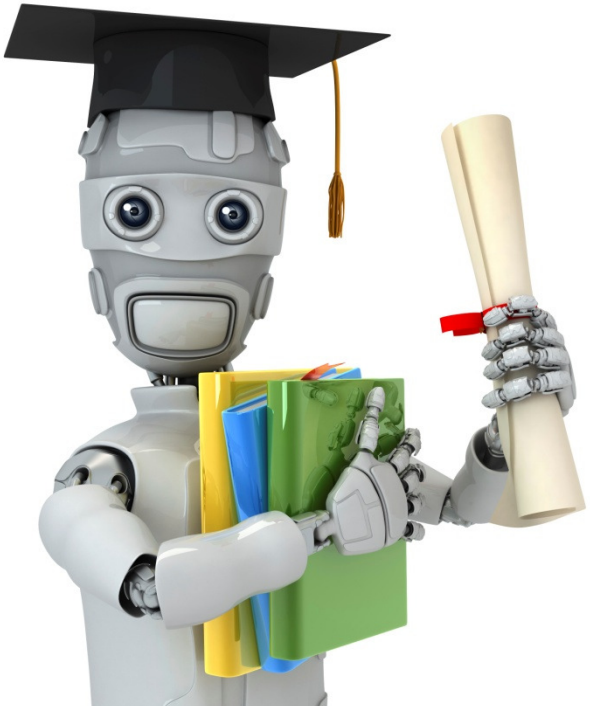
- Applications can't program by hand.

  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).



Machine Learning

Introduction

What is machine learning

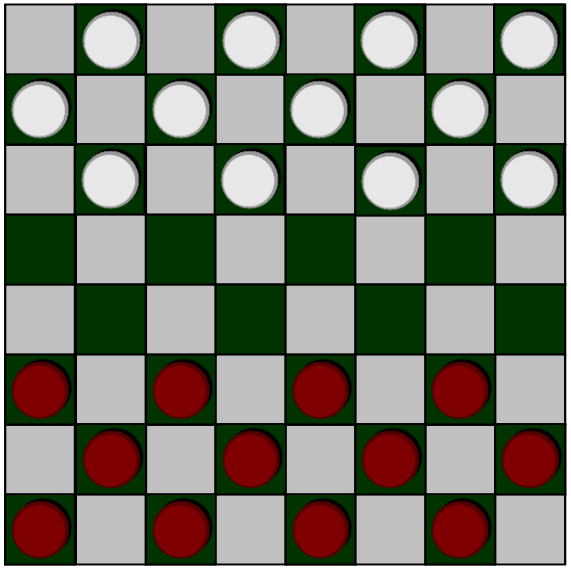
Machine Learning definition

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

old

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

new

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

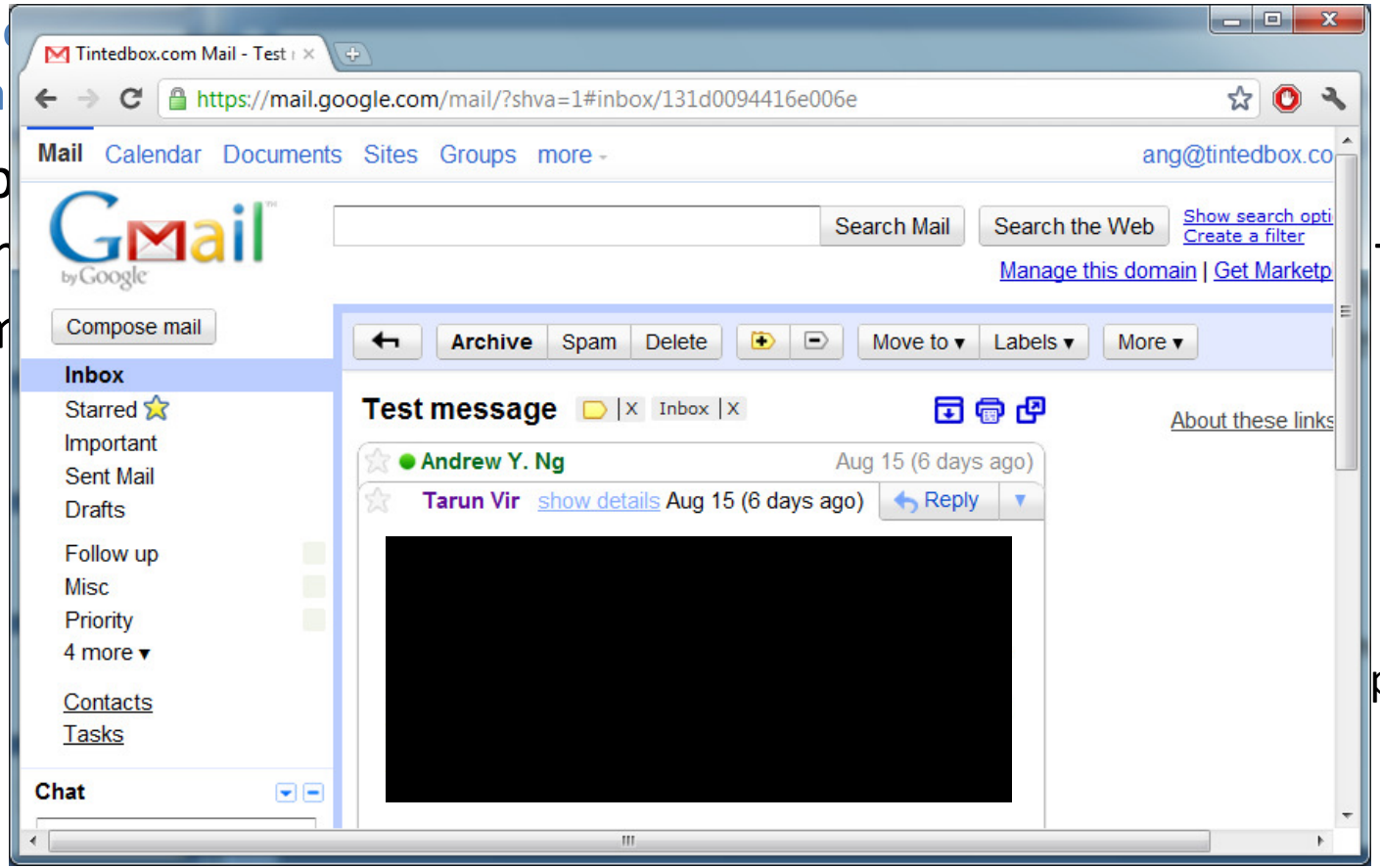
T ←

E ←

P ←

"A computer program is said to *learn* from experience E with respect to s on T, a

Supp or do not r filter spam



pam.

"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.
- Watching you label emails as spam or not spam.
- The number (or fraction) of emails correctly classified as spam/not spam.
- None of the above—this is not a machine learning problem.

T ←

E ←

P ←

Machine learning algorithms:

two most popular machine learning algorithms

- Supervised learning
- Unsupervised learning

teach the computer to how to do sth

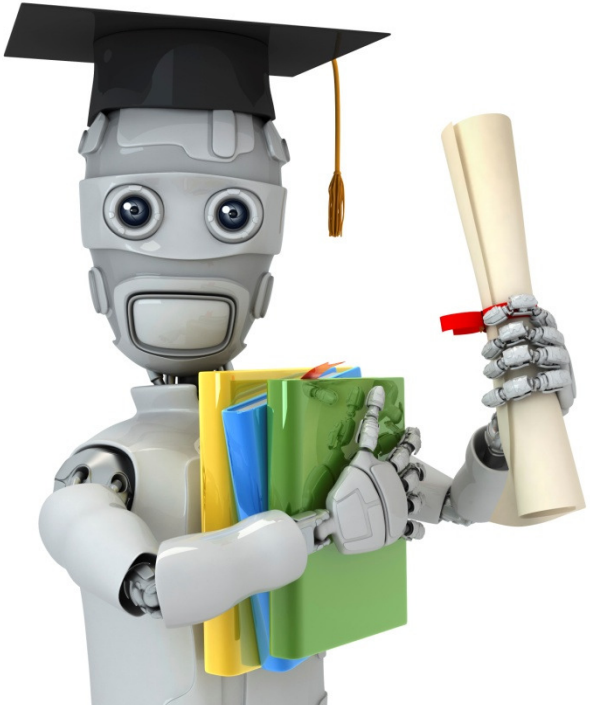
let the computer to learn by itself

other types of machine learning algorithms (later)

Others: Reinforcement learning, recommender systems.

important!

Also talk about: Practical advice for applying learning algorithms.



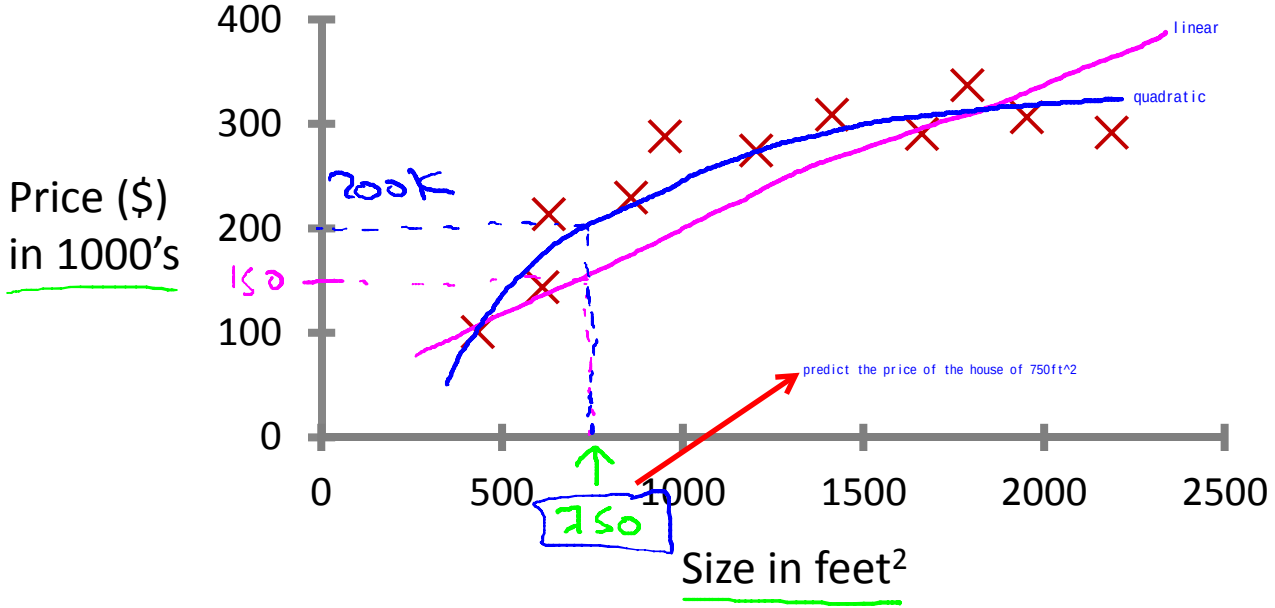
Machine Learning

Introduction

Probably the most popular ML

Supervised Learning

Housing price prediction.



For each house, the exact price (right answer)
The task to predict the price for a new house.

Also called

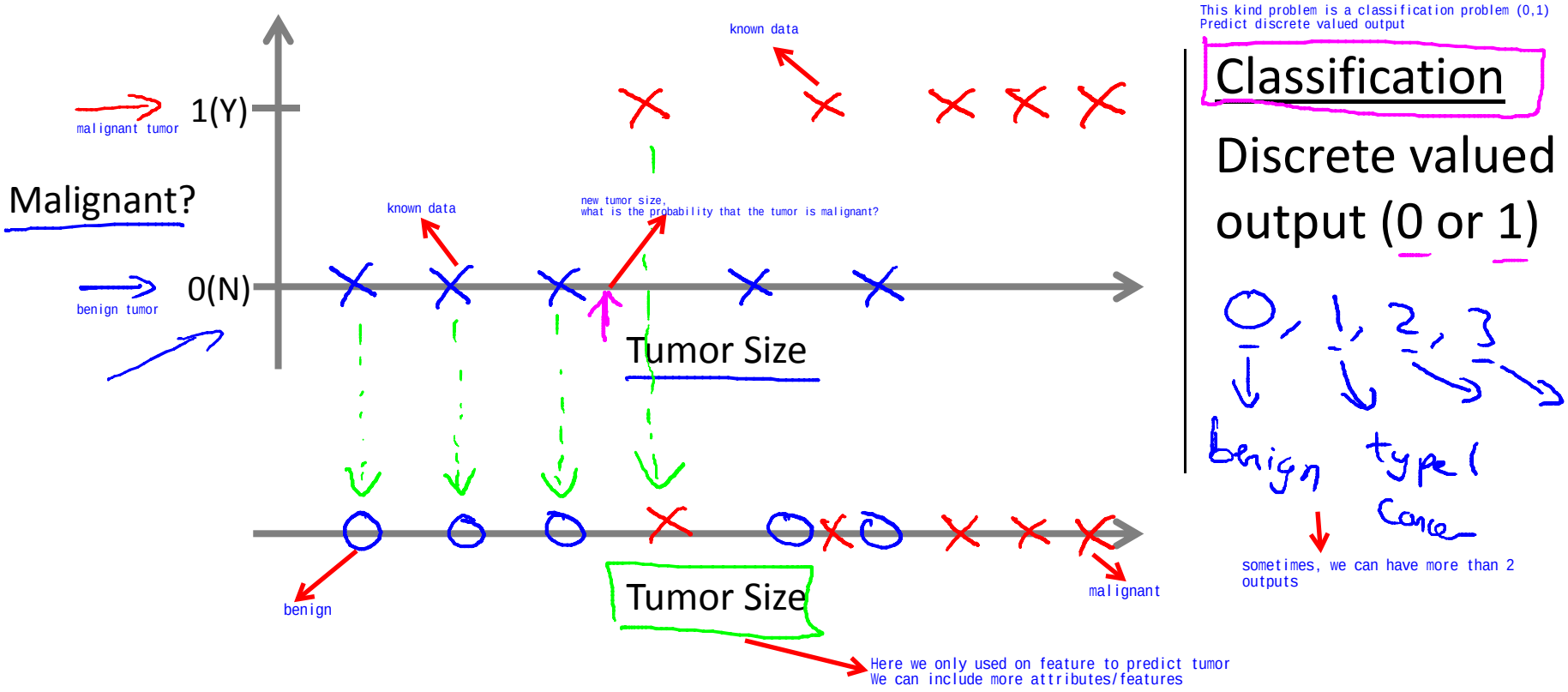
Predict continuous output

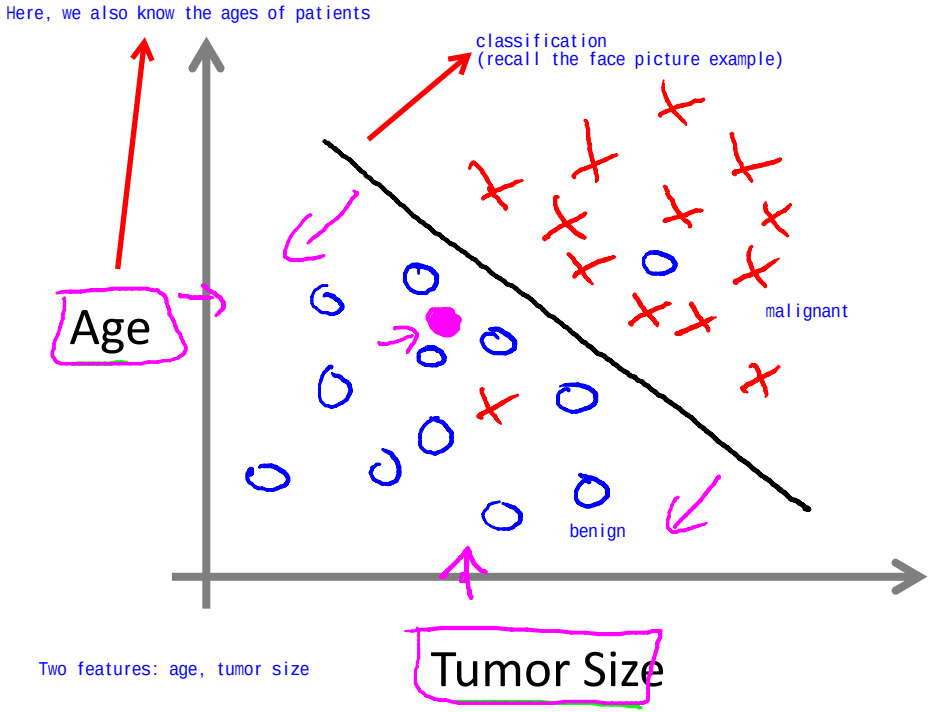
Supervised Learning

"right answers" given

Regression: Predict continuous valued output (price)

Breast cancer (malignant, benign)





We can also include many other features.
Some learning algorithms use infinite number of features (computer limited memory).
(support vector machine allows the computer to deal with infinite features)

- 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.

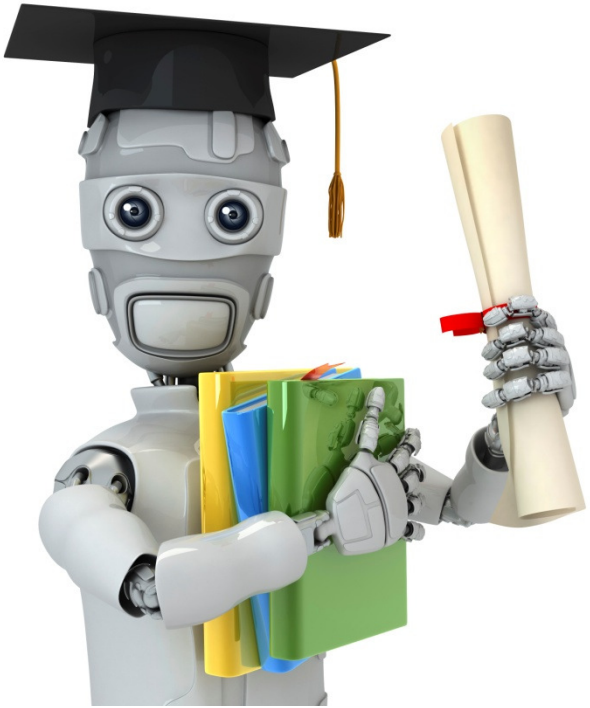
compromised account (受波及的账户)

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.
  - just treat as continuous value
  - (0,1) discrete output values
- Treat both as regression problems.

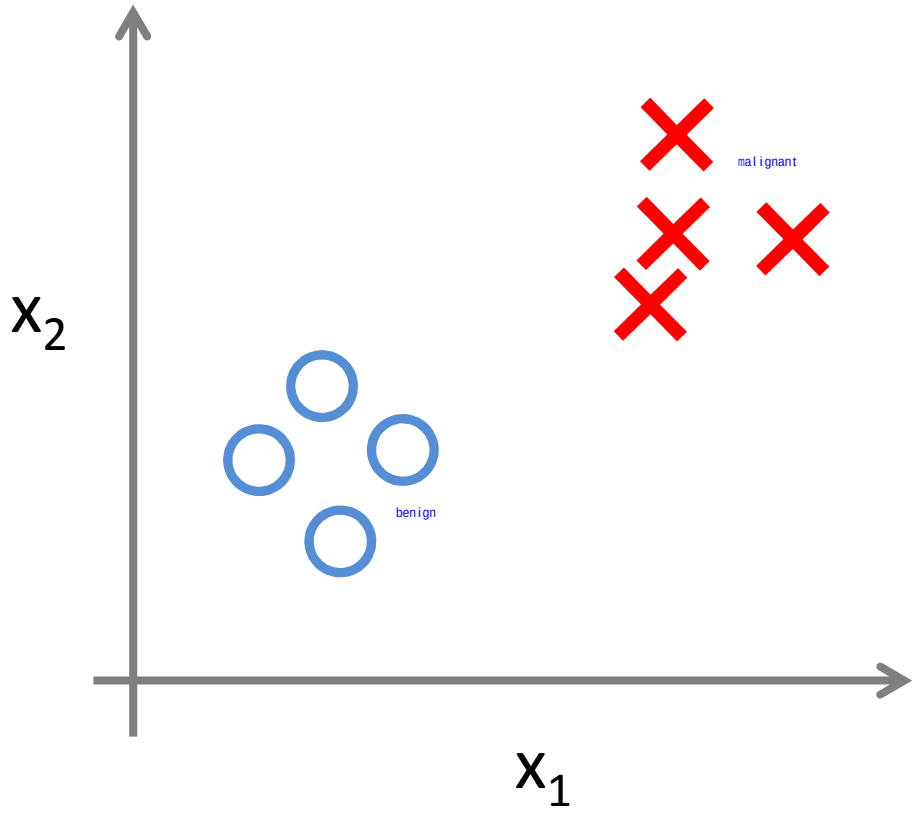


Machine Learning

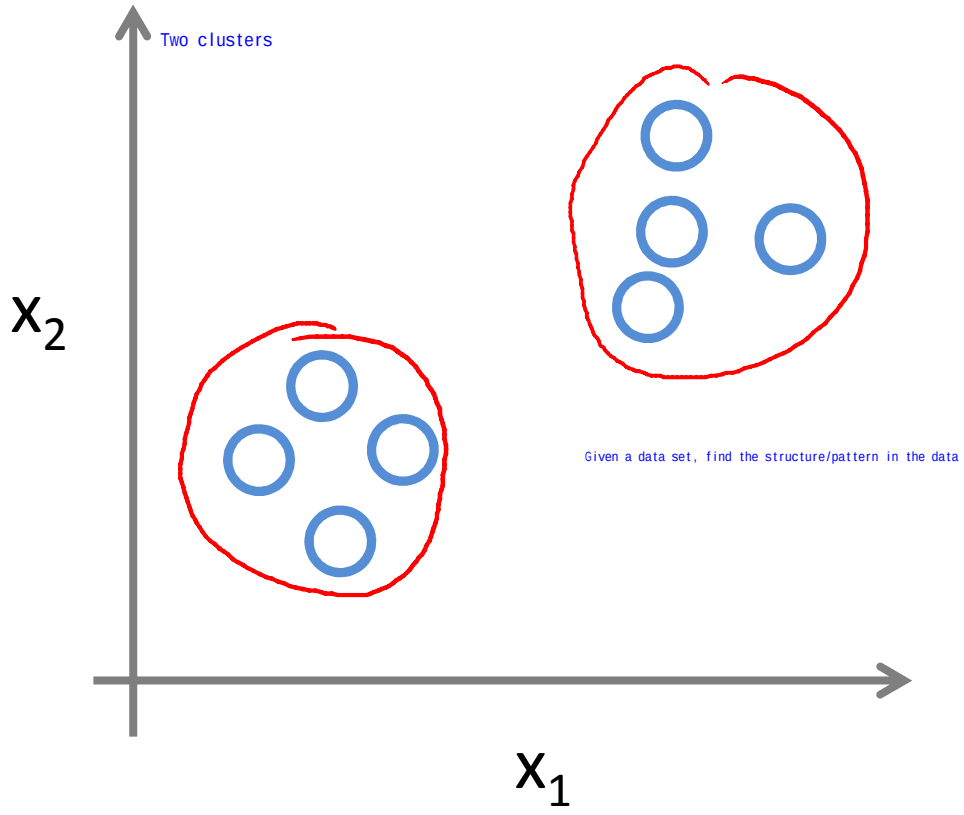
Introduction

Unsupervised Learning

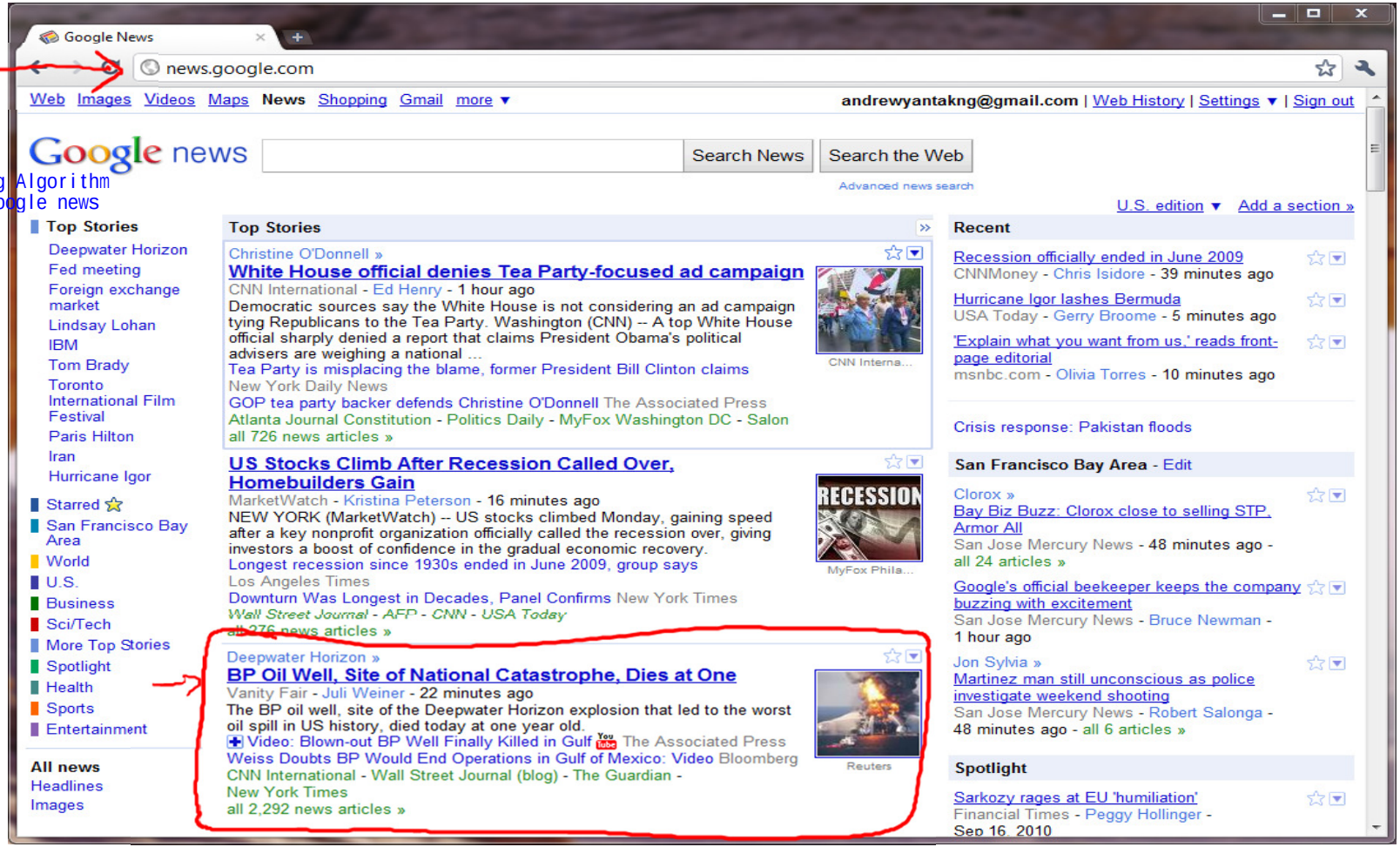
Supervised Learning

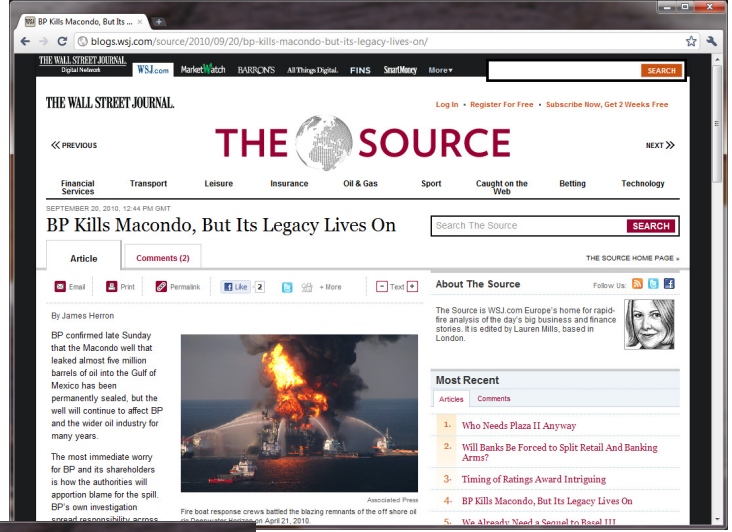


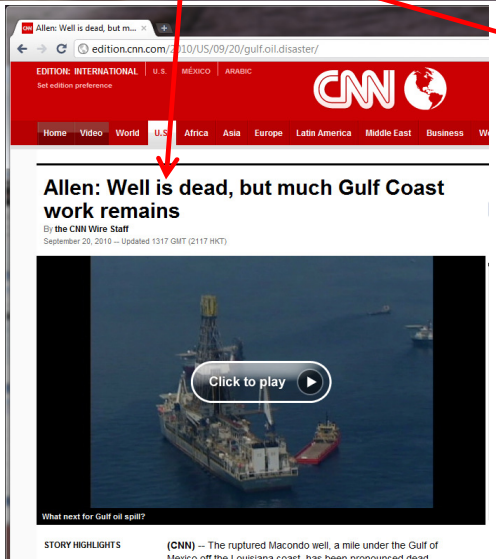
Unsupervised Learning



Clustering Algorithm
used by Google news



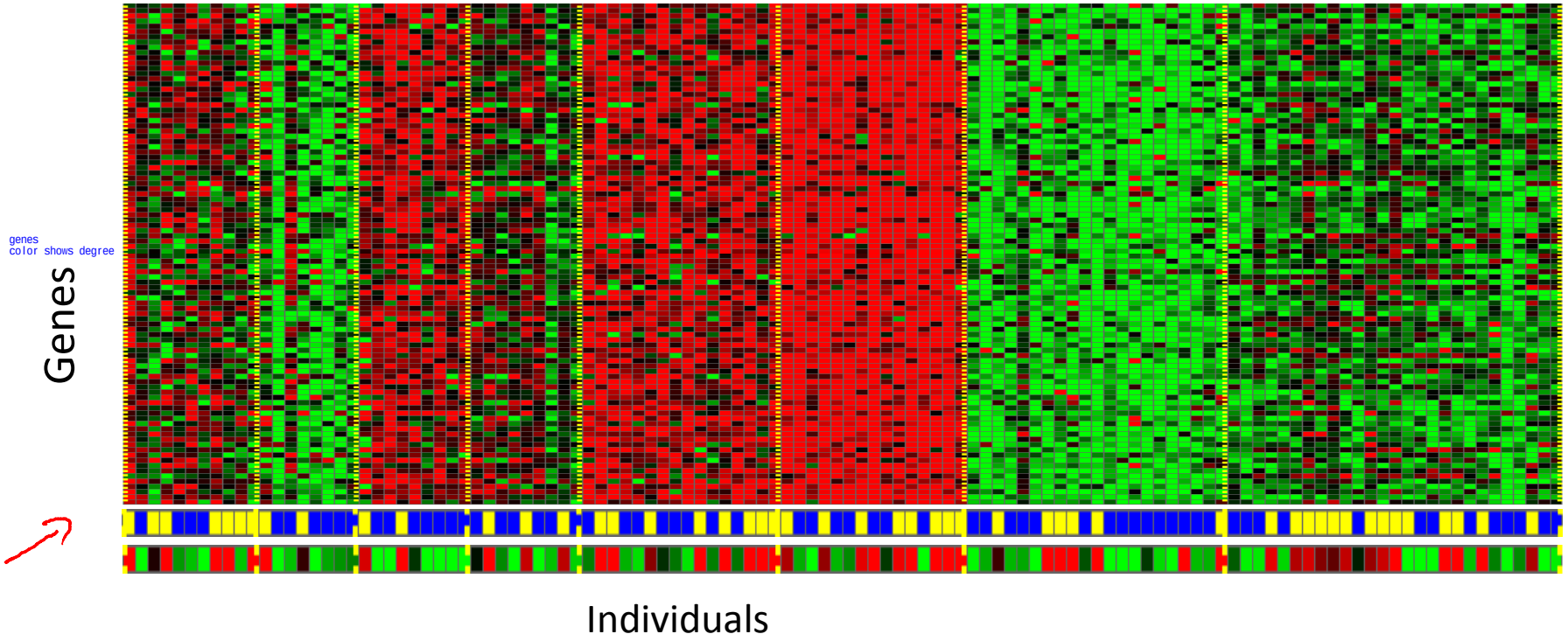




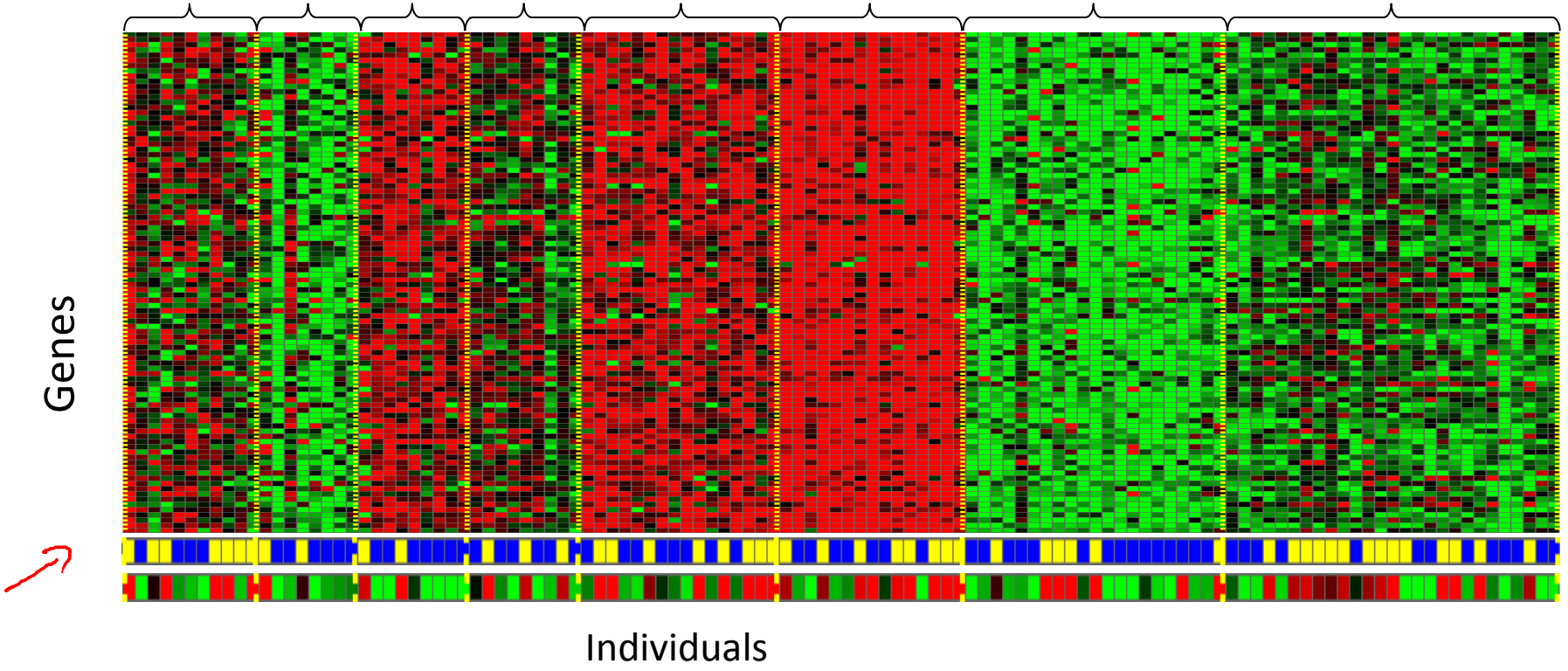
Different stories on the same topic been clustered together

Andrew Ng

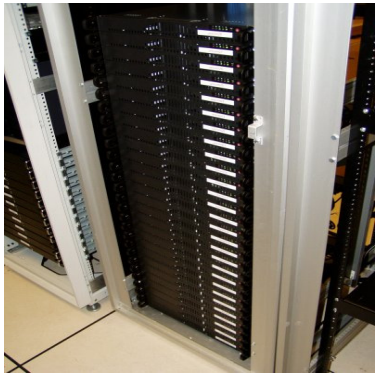
Understanding Genomics



[Source: Daphne Koller]

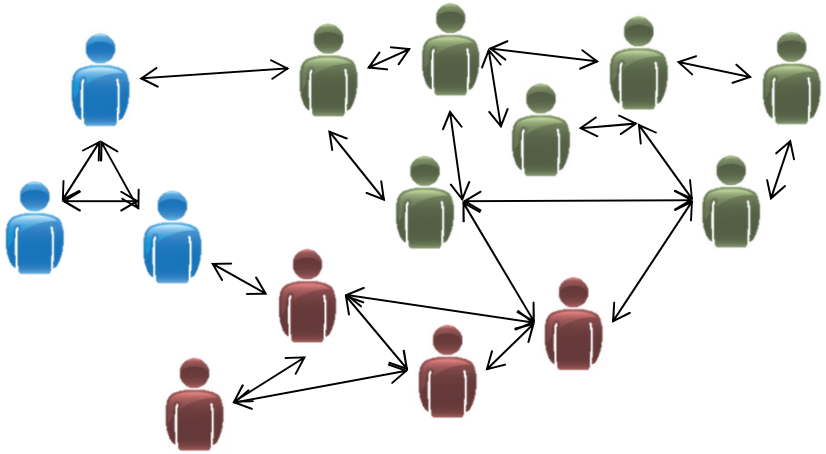


[Source: Daphne Koller]

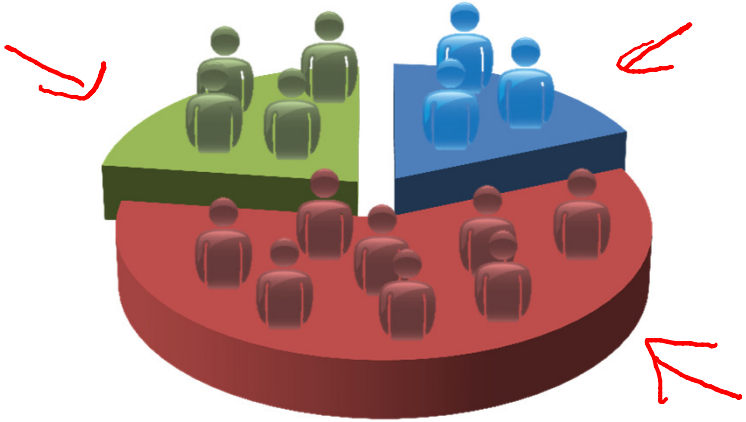


Organize computing clusters

clustering: one type of unsupervised-learning
look at the data set
and automatically clustering



Social network analysis



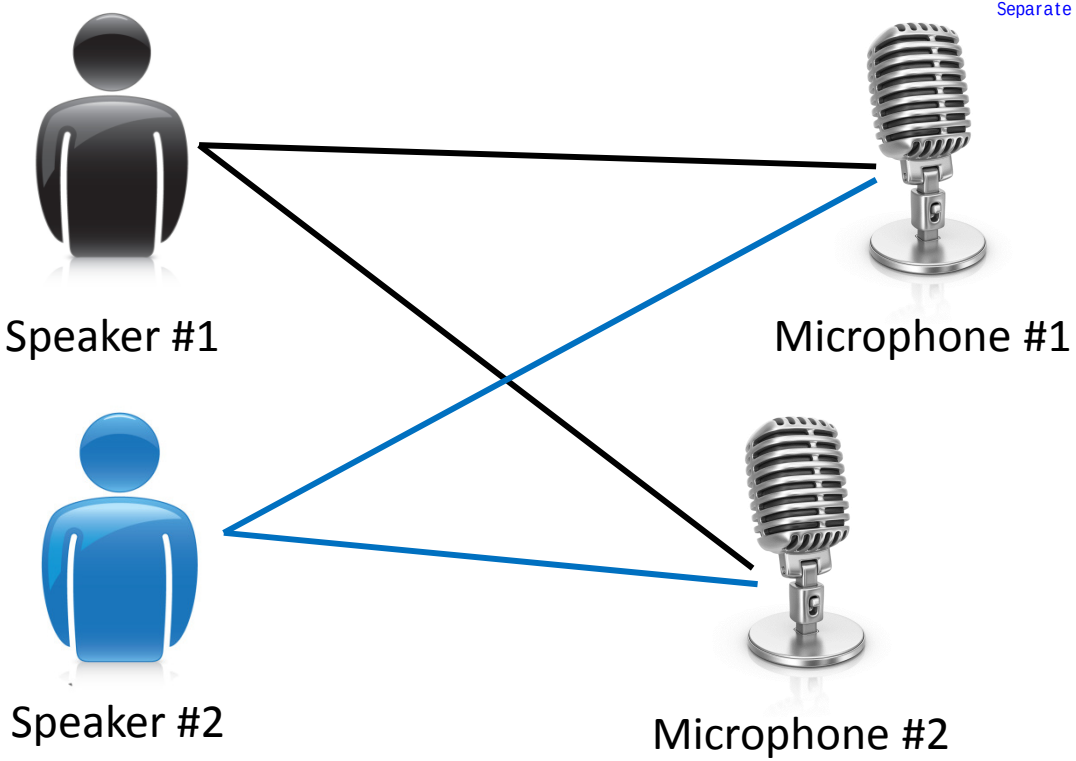
Market segmentation

Astronomical data analysis

Andrew Ng

Cocktail party problem

Another type of unsupervised-learning
Separate the overlapped recording



1st recording

Microphone #1:

Output #1:

separate the overlapped algorithm
(filtering,filter the audio?)

Microphone #2:

Output #2:

2nd recording

Microphone #1:

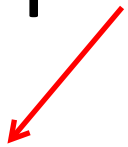
Output #1:

Microphone #2:

Output #2:

[Audio clips courtesy of Te-Won Lee.]

Cocktail party problem algorithm



```
1.Can be solved with one line in MATLAB/Octave
2.C++/Python will be much complicated. e.g. svd
3.Company in silicon valley: prototype in Octave then migrate your code to c++ once you have good results
```

```
[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. supervised
- [x] Given a set of news articles found on the web, group them into set of articles about the same story.
- [x] 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. supervised