

Machine Learning



Amin Golzari Oskouei

a.golzari@azaruniv.ac.ir

a.golzari@tabrizu.ac.ir

<https://github.com/Amin-Golzari-Oskouei>

Azərbaycan Şahid Mədani Universiteti
2023

Machine Learning : Introduction



References (1)

3

- Introduction To Machine Learning

[Ethem Alpaydin, 2014]

- Machine Learning: A Probabilistic Perspective

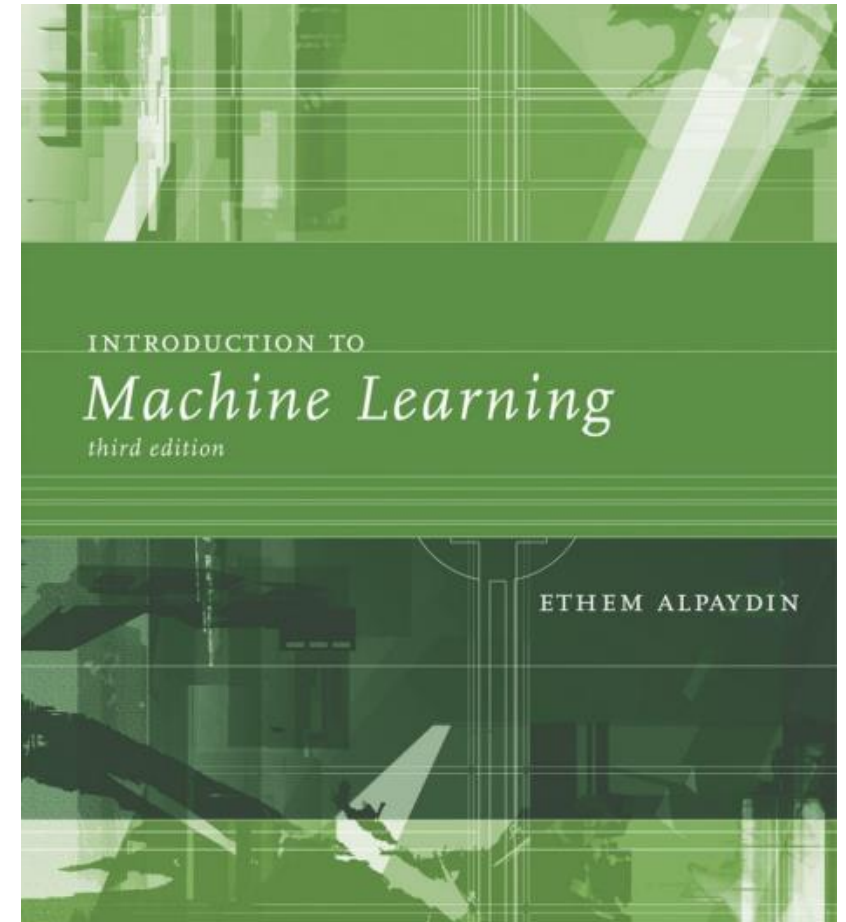
[Kevin P. Murphy, 2012]

- Pattern Recognition and Machine Learning

[Christopher M. Bishop, 2006]

- Machine Learning

[Razavi S. Naser, 2018]



References (2)

4

□ Introduction To Machine Learning

[Ethem Alpaydin, 2014]

□ Machine Learning: A Probabilistic Perspective

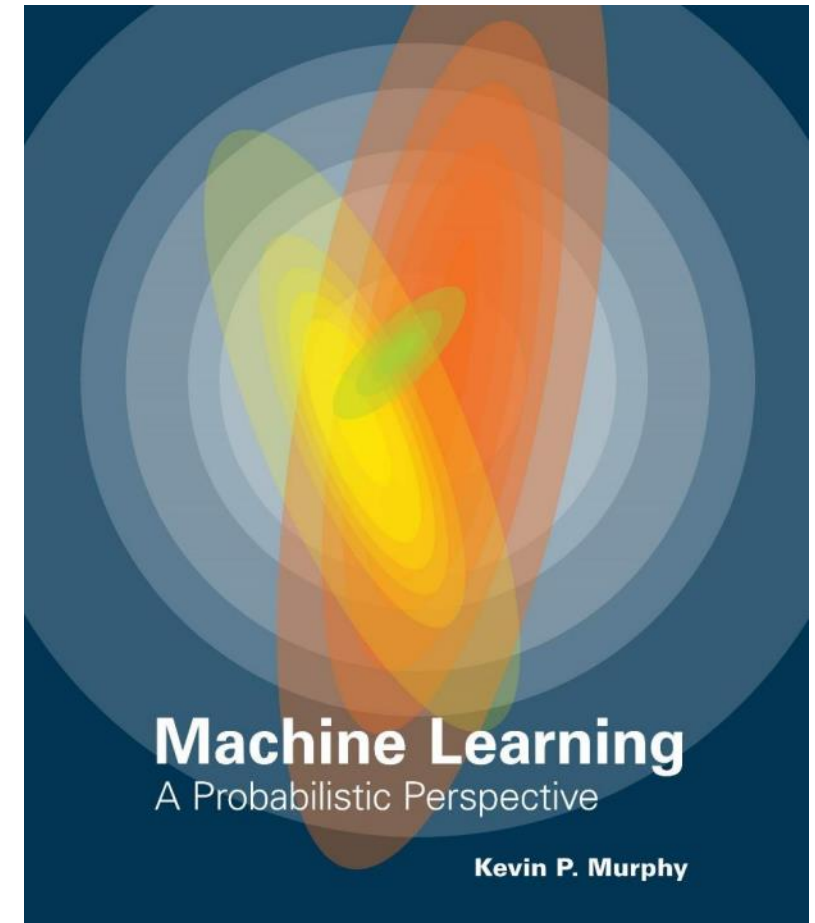
[Kevin P. Murphy, 2012]

□ Pattern Recognition and Machine Learning

[Christopher M. Bishop, 2006]

□ Machine Learning

[Razavi S. Naser, 2018]



References (4)

5

- Introduction To Machine Learning

[Ethem Alpaydin, 2014]

- Machine Learning: A Probabilistic Perspective

[Kevin P. Murphy, 2012]

- Pattern Recognition and Machine Learning

[Christopher M. Bishop, 2006]

- Machine Learning

[Razavi S. Naser, 2018]

References (5)

6

□ These slides have been translated and edited based on Dr. Seyed Naser Razavi's machine learning course.

Requirements

7

- Algorithm analysis and design methods
 - Analysis of computational complexity of learning algorithms
- Linear algebra
 - Matrices, vectors, matrix operations and linear equations
 - Inverse matrix, eigenvectors, order of matrix, decomposition of singular values
- Multivariate calculus
 - Derivative, integral, tangent planes
- Probabilities
 - Random variables, expected value, variance and ...

Evaluation

8

- **Tasks** [45%]

- Theoretical topics

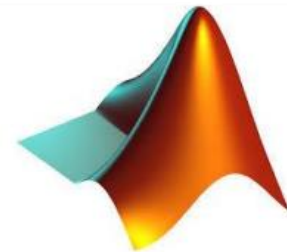
- Programming

- **Final Exam** [50%]

- **Effective attendance in class** [5%]



OCTAVE



MATLAB

Table of Contents

9

- Supervised learning.
 - Regression - univariate and multivariate linear regression
 - Classification – logistic regression, neural networks, support vector machines

- Unsupervised learning.
 - Clustering

- Reinforcement learning.

- Programming using Python (or Octave).

- Practical recommendations in using machine learning algorithms.

Some Quotes

10



“If you invent a breakthrough in artificial intelligence, so machines can learn, that is worth 10 Microsofts.”

-Bill Gates (Former CEO of Microsoft)



“Machine Learning is the next internet.”

-Anthony Tether (Director of DARPA)



“Machine learning is going to result in a real revolution.”

-Greg Papadopoulos (Creator of Redshift)

What is Machine Learning?

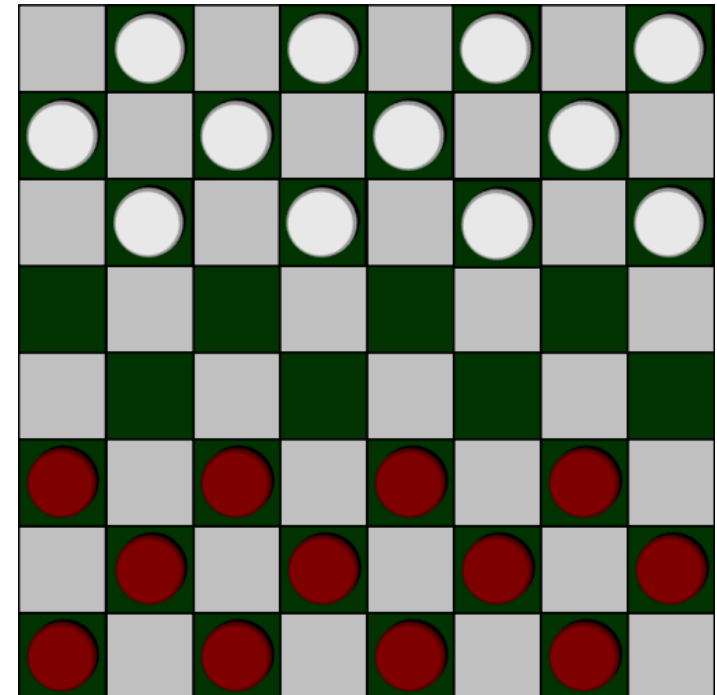
Machine Learning: Definitions

12

- Arthur Samuel [1959]

“Field of study that gives computers the ability to learn without being explicitly programmed.”

- Samuel’s Checkers Player [Samuel, 1959, 1967]



Machine Learning: Definitions

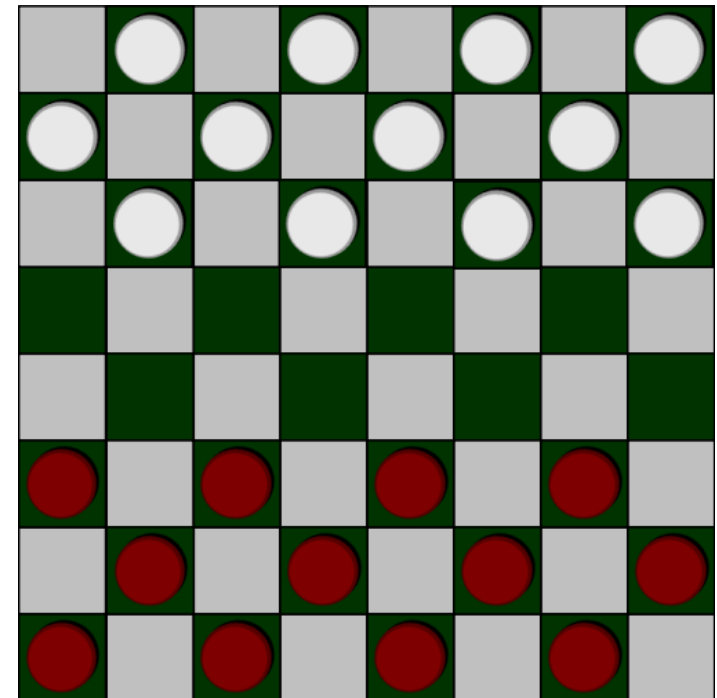
13

□ Tom Mitchell [1998]

“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 E .”

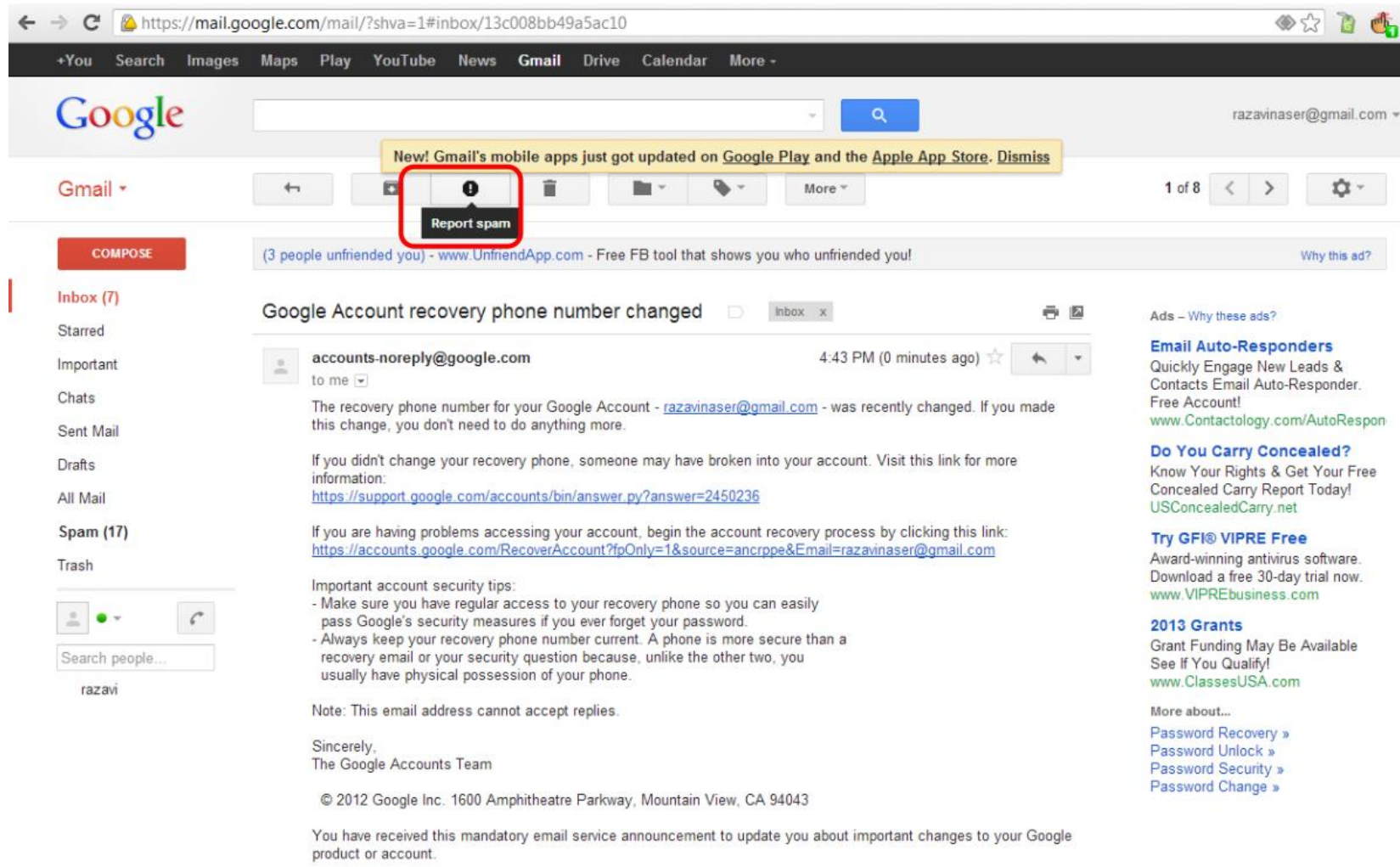
□ Example: Checkers Game

- **Task:** Playing checkers
- **Experience:** Playing thousands of time against yourself
- **Efficiency Measure:** Number of wins against new opponents



Example: Detecting Spam Mail

14



Example: Detecting Spam Mail

15

□ Example: Detecting Spam Mail

Let's say your email program allows you to mark incoming emails as spam and learns how to detect spam mails accordingly.

- **Task:** Categorizing emails as spam or email
- **Experience:** Monitoring which emails you mark as spam
- **Efficiency Measure:** The number of emails that are correctly categorized

Types of Machine Learning Methods

16

- **Machine learning** Improving the machine's performance in performing a task with experience.

- **Q.** How can a machine know that its performance has improved?
 - We can give the machine the correct answer for some limited examples of input in the hope that it can generalize it to other examples - **supervised learning**

 - We can tell the machine to what extent its answer was correct (for example by giving a point) and the machine itself is responsible for finding the correct answers - **reinforcement learning**

 - We may not give the machine any information about the correct answer and only ask the machine to find inputs that have common features - **unsupervised learning**



Supervised Learning

Supervised Learning

18

- **Input:** A training set where the correct answer is given for each input.

$$\{(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \dots, (x^{(m)}, y^{(m)})\}$$

← Training Set

- **Goal:** Find a suitable approximation for the following mapping:

$$f: X \rightarrow Y$$

- **Example:**

- **Spam Detection:** Mapping emails to the {Spam, Non-Spam} collection
- **Digit Recognition:** Mapping a set of pixels to the set {0, 1, 2, ..., 9}
- **Cancer Diagnosis:** Mapping medical data to the set {Malignant, Benign}

Example: Detecting Spam Mail

19

- **Input:** Email
- **Output:** Spam, Non-Spam

Dear Sir. First, I must solicit your confidence in this transaction, this is by virtue of its nature as being utterly confidential and top secret. ...



To be removed from future mailings, simply reply to this message and put "remove" in the subject. 99 million email addresses for only \$99



Ok, I know this is blatantly OT but I'm beginning to go insane. Had an old Dell Dimension XPS sitting in the corner and decided to put it to use, I know it was working pre being stuck in the corner, but when I plugged it in, hit the power nothing happened.



Example: Recognizing Handwritten Numbers

20

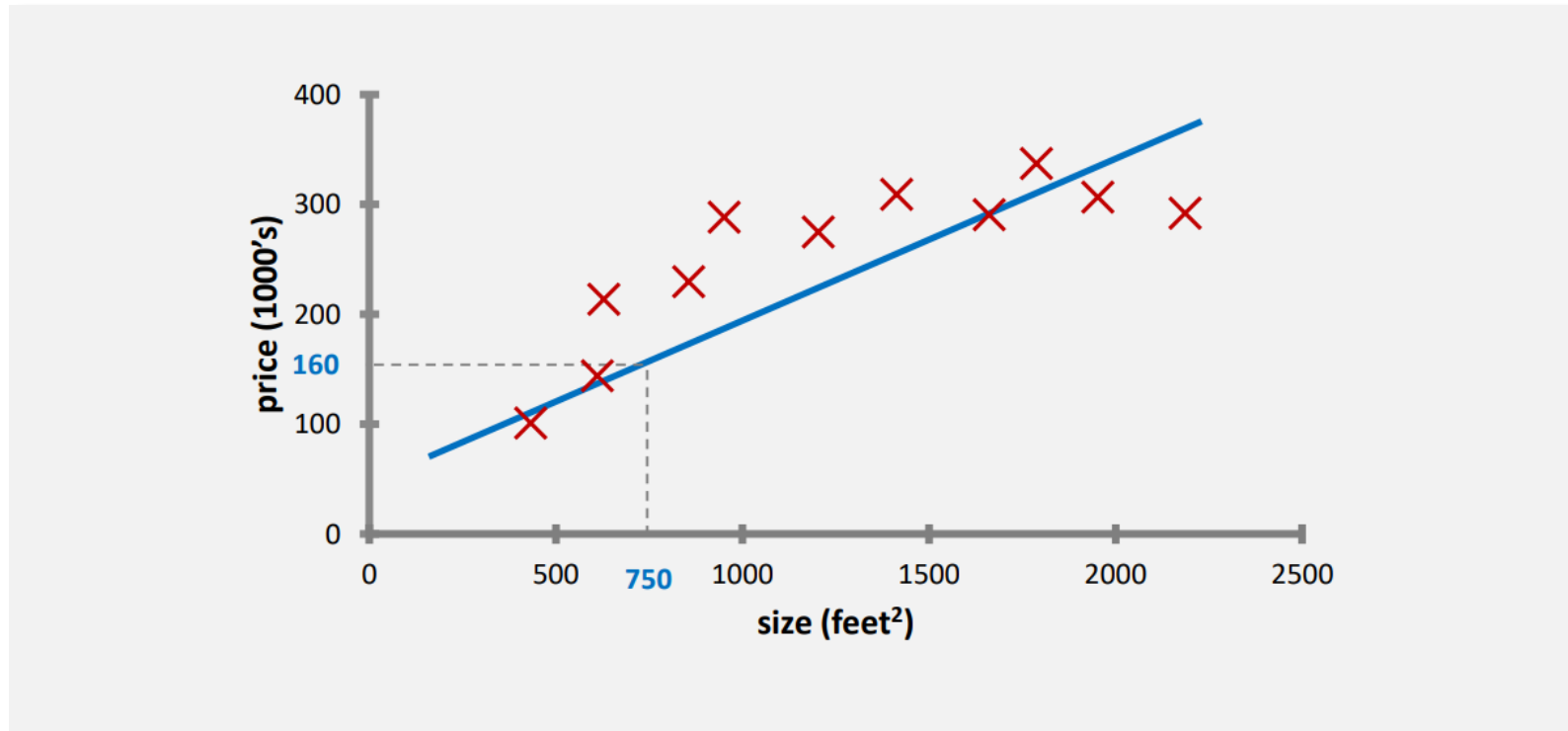
- **Input:** Image of A Number
- **Output:** A Number

8	3	9	3	8	5	8	5	6	5
9	4	9	5	7	1	7	6	1	1
6	8	3	6	8	8	8	1	1	4
4	9	5	0	1	2	1	4	5	3
7	2	7	2	6	3	1	1	2	1
3	2	7	0	4	6	0	8	1	8
6	0	7	4	1	1	7	4	2	1
2	9	5	3	7	4	1	0	5	8
3	5	5	7	6	5	9	9	9	3
1	9	9	6	1	2	1	3	6	7

Example: Pricing A House

21

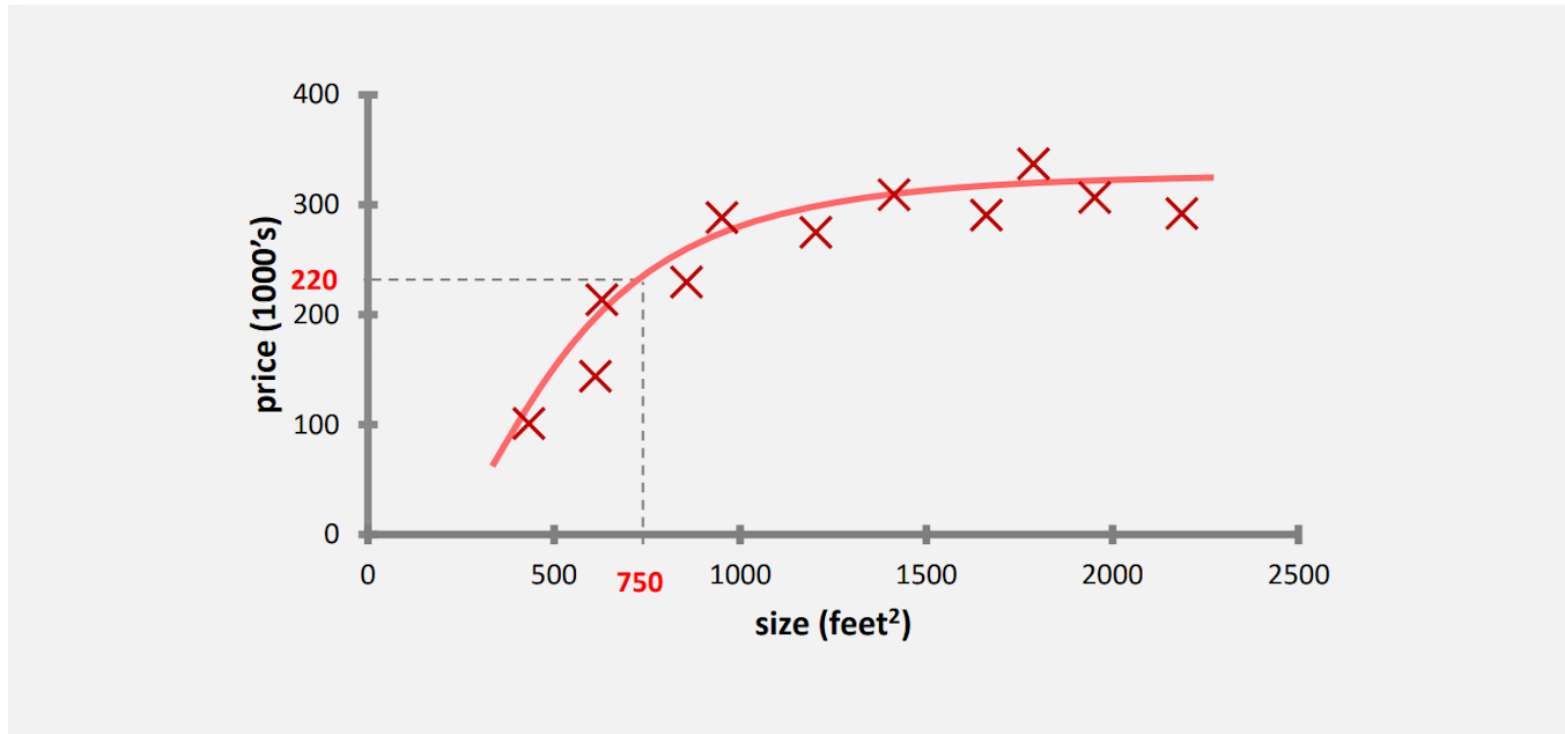
- **Input:** The Size of The House [In Feet²]
- **Output:** Estimated Price



Example: Pricing A House

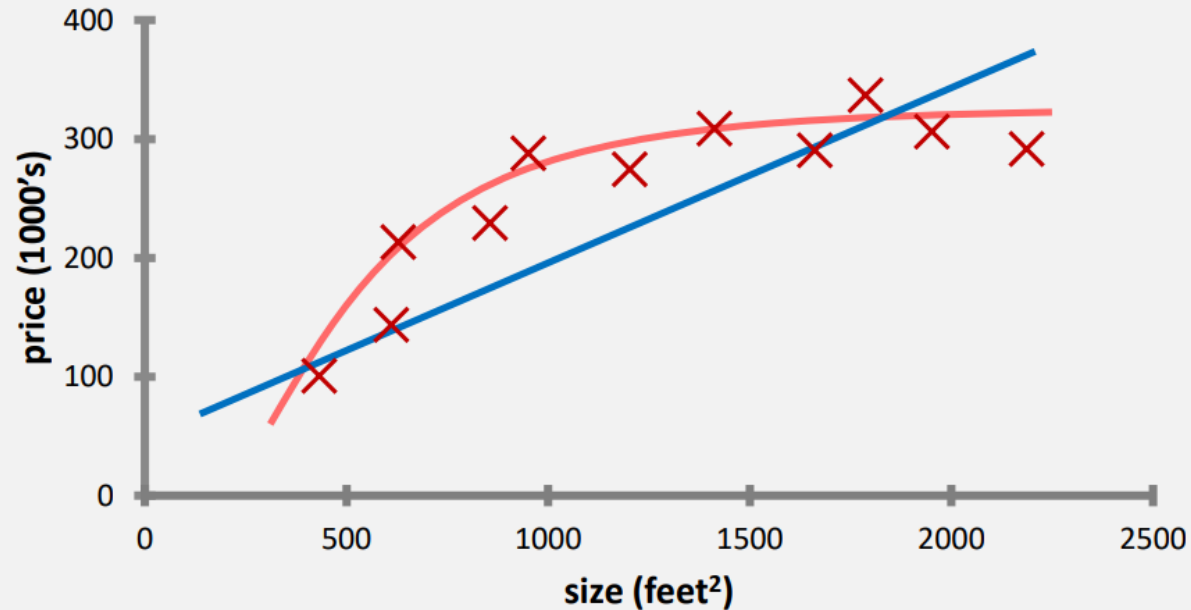
22

□ Q. Which one is better? A linear function or a quadratic function?



Example: Pricing A House

23



□ Supervised Learning:

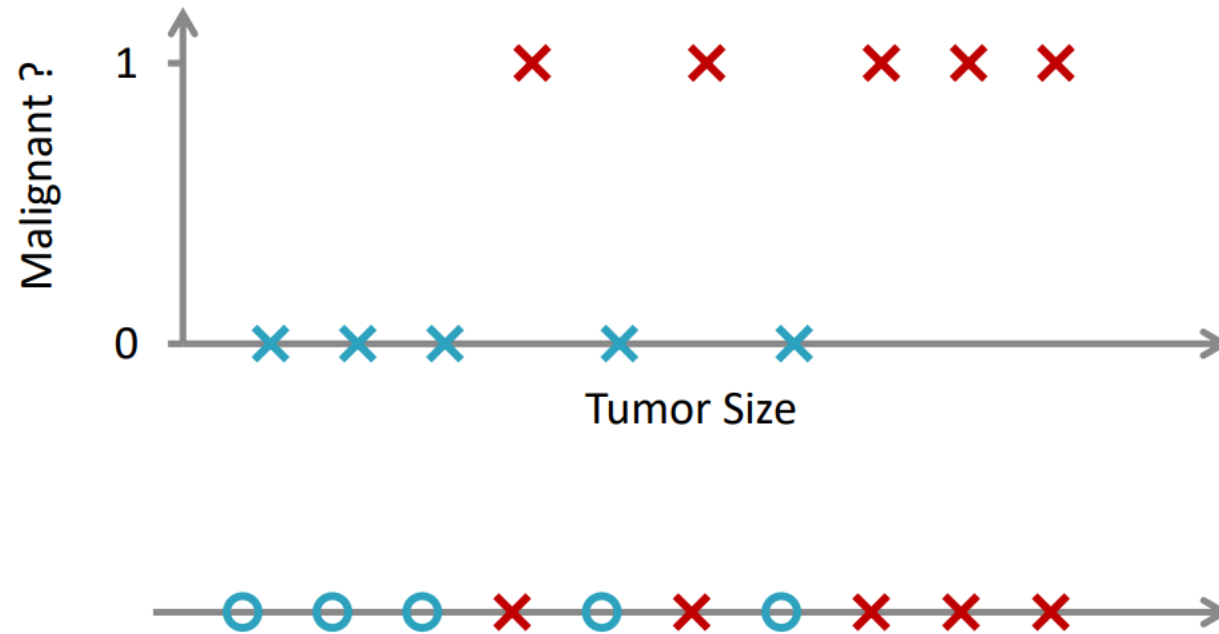
For each training sample, a "correct answer" is given.

□ Regression:

Predicting quantities with continuous values (such as the price of a house)

Example: Cancer Diagnosis (Malignant, Benign)

24



□ Supervised Learning:

For each educational example, a "correct answer" is given.

□ Classification:

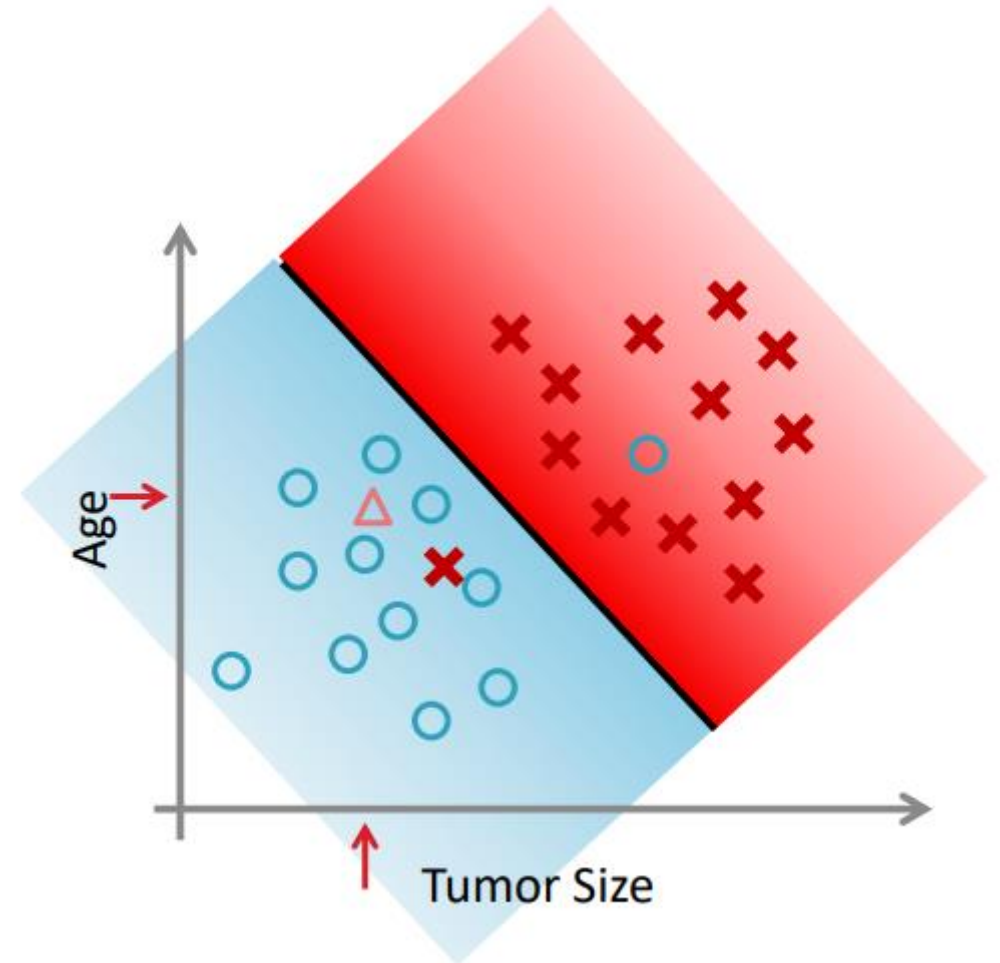
Predicting quantities with discrete values (such as zero and one).

Example: Cancer Diagnosis (Malignant, Benign)

25

□ Other Features:

- Uniformity of cell shape
- Uniformity of cell size
- ...



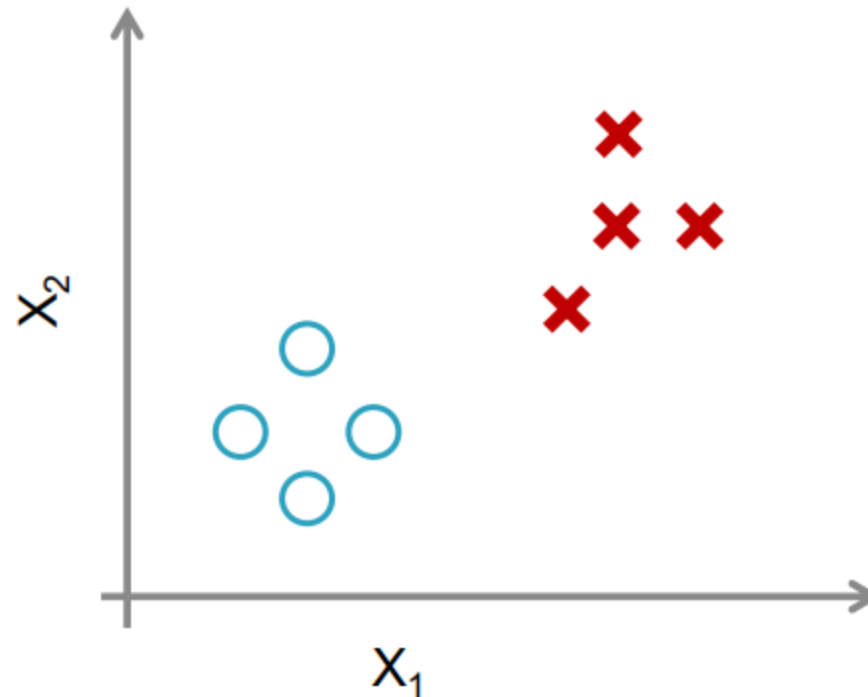


Unsupervised Learning

Supervised Learning

27

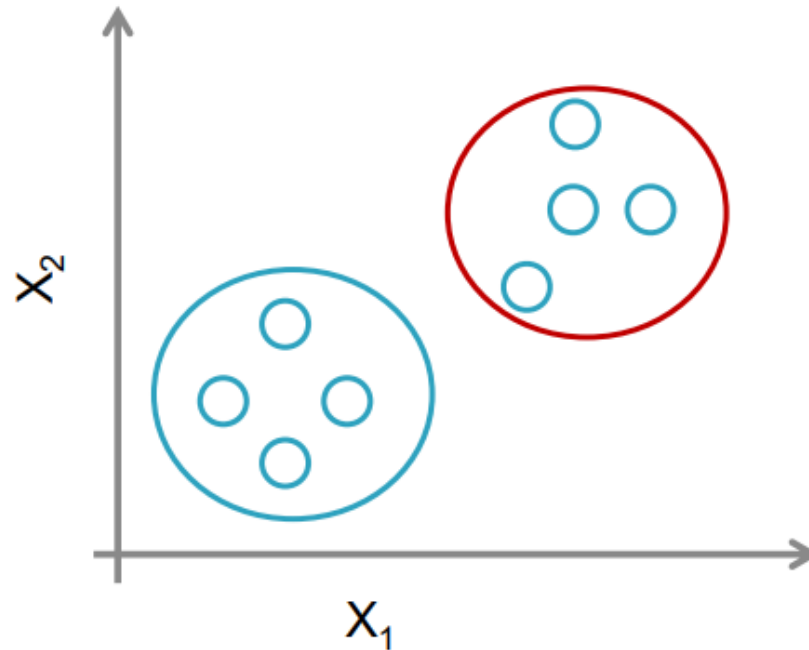
- **Supervised Learning:** For each training example, a "correct answer" is given.



Unsupervised Learning

28

- **Unsupervised Learning:** No information about correct answers is given!



- **Goal:** Detecting structure in input data (grouping similar data).

Application of Clustering: Grouping Related News

29

The screenshot shows the Google News homepage with several news sections. A red box highlights a specific article in the 'Top Stories' section.

Top Stories

- White House official denies Tea Party-focused ad campaign**
CNN International - Ed Henry - 1 hour ago
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 »
- BP Oil Well, Site of National Catastrophe, Dies at One**
Vanity Fair - Juli Weiner - 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: Blown-out BP Well Finally Killed in Gulf
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 »

Recent

- Recession officially ended in June 2009
CNMoney - 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

San Francisco Bay Area - Edit

- Clorox »
Bay Biz Buzz: Clorox close to selling STP, Armor All
San Jose Mercury News - 48 minutes ago - all 24 articles »
- Google's official beekeeper keeps the company buzzing with excitement
San Jose Mercury News - Bruce Newman - 1 hour ago
- Jon Sylvia »
Martinez man still unconscious as police investigate weekend shooting
San Jose Mercury News - Robert Salonga - 48 minutes ago - all 6 articles »

Spotlight

- Sarkozy rages at EU 'humiliation'
Financial Times - Peggy Hollinger - Sep 16, 2010

Application of Clustering: Grouping Related News

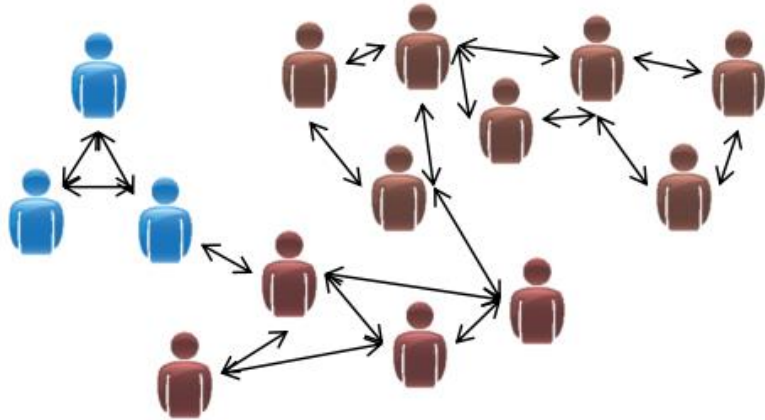
30

The image displays four news articles related to the BP oil spill, connected by red arrows that illustrate a flow of information or a cluster of related news items. The articles are:

- Google News:** A search results page for "BP Oil Well, Site of National Catastrophe, Dies at One". The article is highlighted with a red box, and a red arrow points from it to the CNN article.
- The Wall Street Journal:** An article titled "BP Kills Macondo, But Its Legacy Lives On". A red arrow points from the Google News article to this one.
- CNN:** An article titled "Allen: Well is dead, but much Gulf Coast work remains". A red arrow points from the Google News article to this one.
- The Guardian:** An article titled "BP oil spill cost hits nearly \$10bn". A red arrow points from the CNN article to this one.

Some Other Applications of Unsupervised Learning

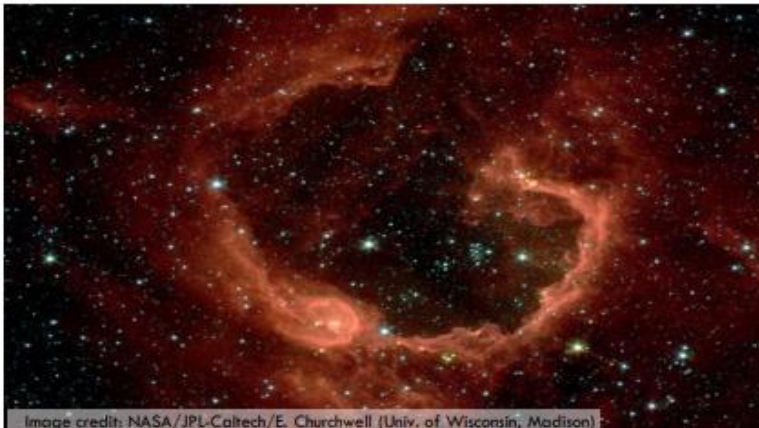
31



Analysis of social networks



Organization of computing clusters (data centers)

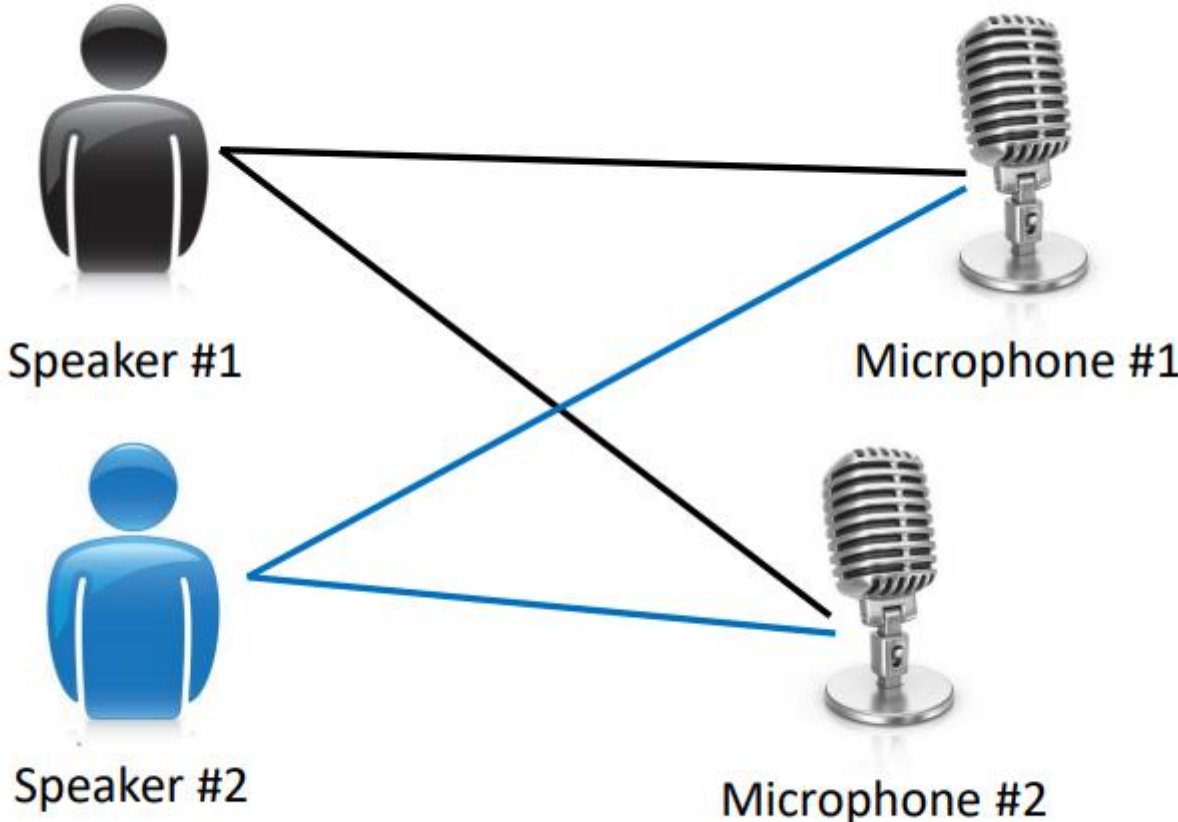


Analysis of astronomical data (how galaxies form)



Market segmentation

Cocktail Party Problem



Cocktail Party Problem

33

Microphone #1: 🗣️

Output #1: 🗣️

Microphone #2: 🗣️

Output #2: 🗣️

Microphone #1: 🗣️

Output #1: 🗣️

Microphone #2: 🗣️

Output #2: 🗣️

Cocktail Party Problem Algorithm

34

□ Octave Code:

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

A Question For Class

35

- For which one of the following situations should an unsupervised learning algorithm be used?
 - Developing a program to filter spam mails by having some normal emails and some spam emails
 - Group a collection of newly found articles on the web by topic
 - Diagnosing diabetes in new patients by having data on a number of healthy and diabetic people
 - Grouping a set of customers into several different market segments by having a database about customers