Machine Learning

By Ghazal Lalooha

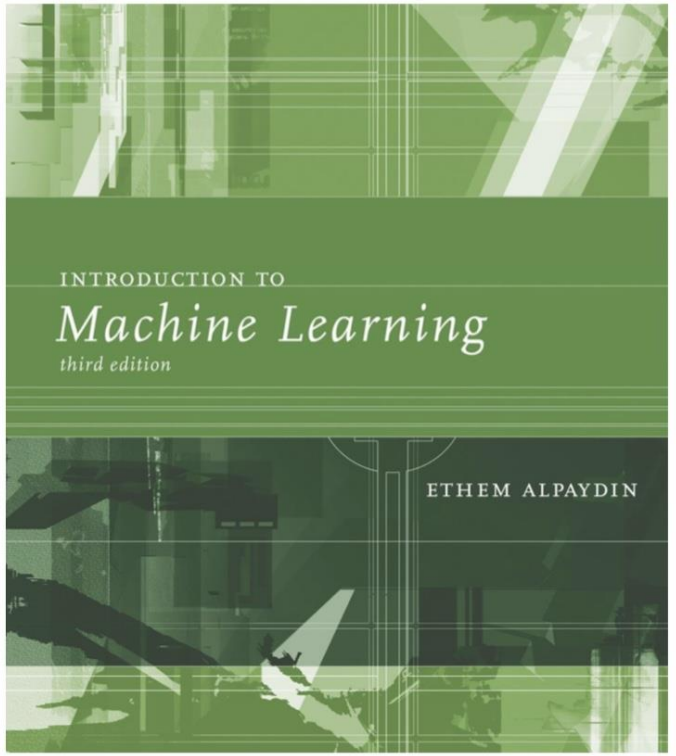
Introduction

Sources and References

- An introduction to machine learning
- (Alpadin)

- Machine learning: a probabilistic perspective
- (Kevin Murphy)

- Pattern recognition and machine learning
- (Christopher Bishop)

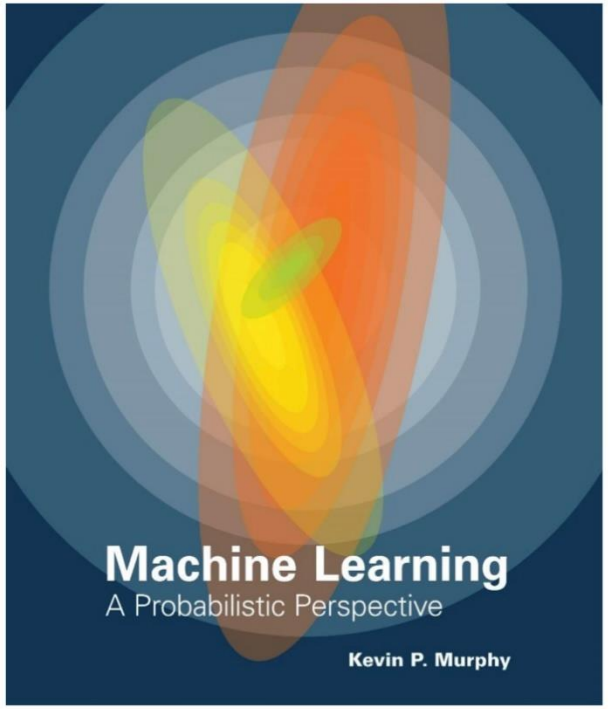


Sources and References

- An introduction to machine learning
- (Alpadin)

- Machine learning: a probabilistic perspective
- (Kevin Murphy)

- Pattern recognition and machine learning
- (Christopher Bishop)

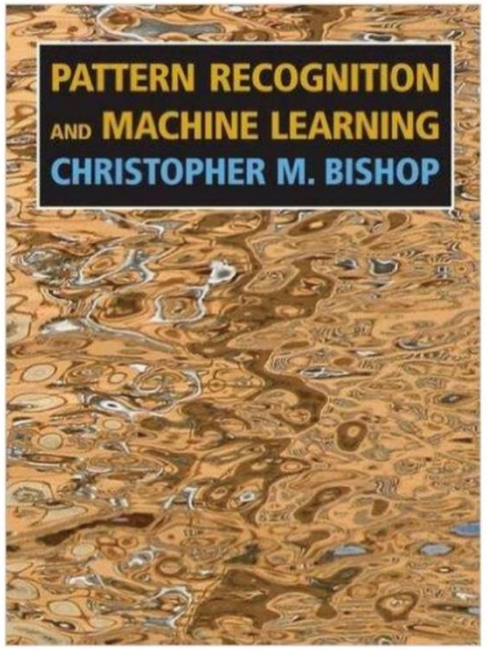


Sources and References

- An introduction to machine learning
- (Alpadin)

- Machine learning: a probabilistic perspective
- (Kevin Murphy)

- Pattern recognition and machine learning
- (Christopher Bishop)



Prerequisites

- Algorithm analysis and design methods
  - Computational complexity analysis of machine learning algorithms
- Linear Algebra
  - Matrices, vectors, matrix operations and linear equations
  - Inverse matrix, eigenvectors, order of matrix, decomposition of singular values
- Multivariate calculus
  - Difrentiel , integral, tangent planes
- Possibilities
  - Random variables, expected value, variance and...

Table of Contents

A few quotes

"Every step forward in machine learning is worth tenfold to Microsoft."
Bill Gates- Microsoft's manager

"The next generation of the Internet is nothing but machine learning."
Tony Tedder - former director of DARPA

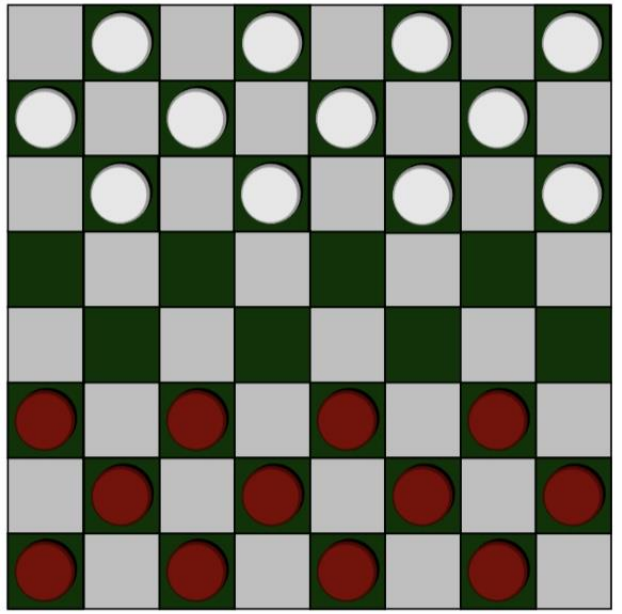
"Machine learning will eventually lead to a true revolution."
Greg Papadopoulos- Former Sun Manager

What is Machine Learning?

Machine Learning: Definitions

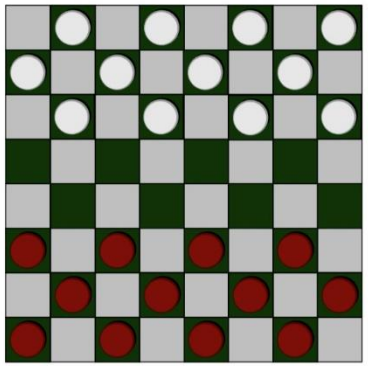
- Arthur Samuel (1955)
  - "A field of study that gives machines the ability to learn, without the need for these machines to be explicitly programmed."

- Checkers game. (Samuel. 50s)



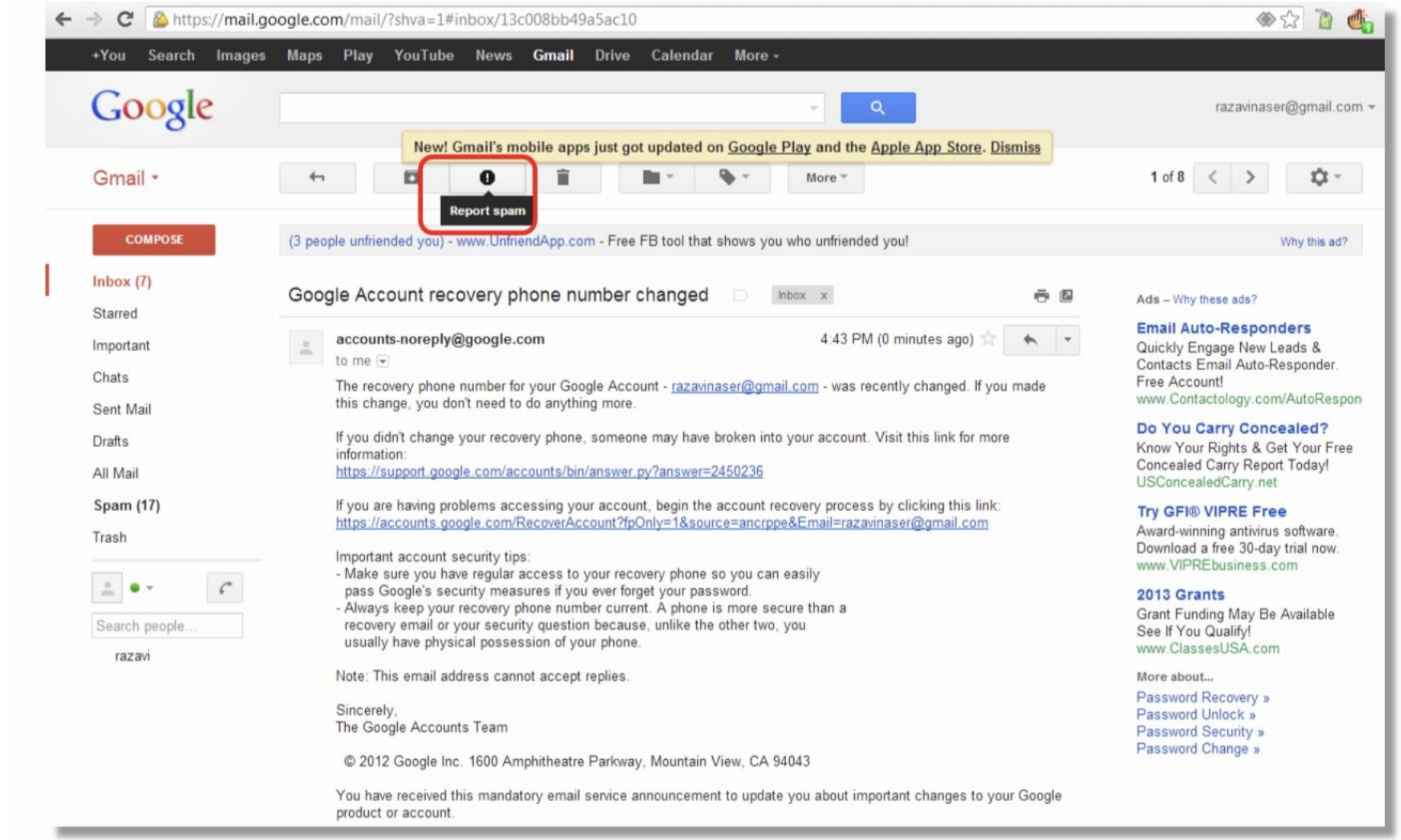
Machine Learning: Definitions

- Tom Mitchell. (1998)
  - "Given a task such as T and a performance measure such as P, we say that a computer program learns from experience E if the program's performance measure P for performing task T is improved by using experience E."



- Example. Checkers game
  - Task: playing checkers
  - Experience: Play against yourself thousands of times
  - Performance measure: Number of wins against new competitors

Example: Spam Detection



Example: Spam Detection

- Example: spam detection
  - Let's say your email program allows you to mark incoming emails as spam, and you learn how to filter spam better.

  - Task: Categorize emails as spam or email
  - Experience: Monitor which emails you mark as spam.
  - Performance measure: the number of emails that are correctly categorized.

Types of machine learning methods

- Machine learning: improving a machine's performance on a task with experience
- Question: How can a car know that its performance has improved?
  - We can give the machine the correct answer for a few 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 the inputs that have a common existence. - Unsupervised learning

Supervised Learning

Supervised Learning

- Input: A training set in which a correct answer is given for each input.

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

Training set

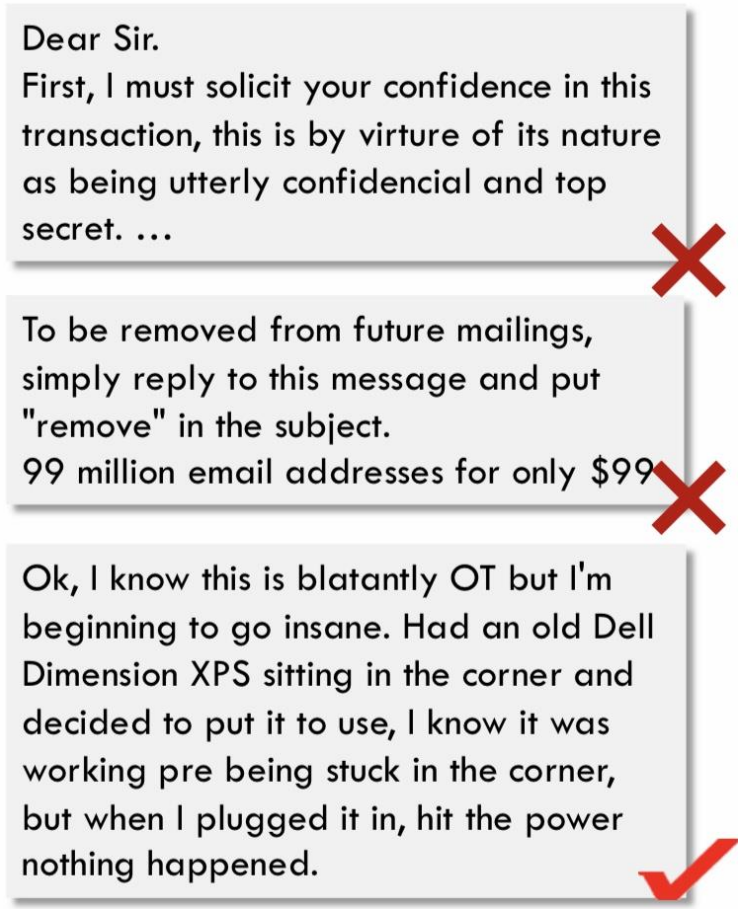
Objective: to find a suitable approximation for the following mapping:

$$f: X \to Y$$

- Example:
  - Spam detection: mapping emails to {spam, non-spam} sets
  - Digit recognition: mapping a set of pixels to the set {0, 1, 2, ..., 9}
  - Cancer diagnosis: mapping medical data to {malignant, benign} sets

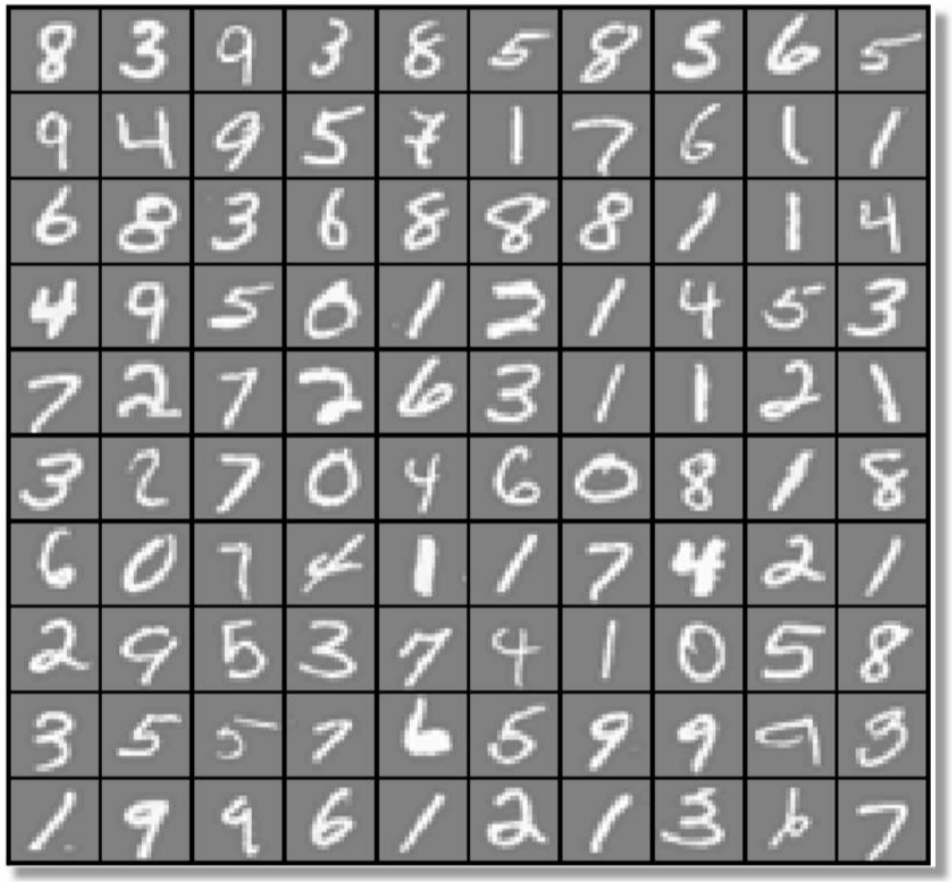
Example: Spam Detection

Input: email
Output: spam, non-spam



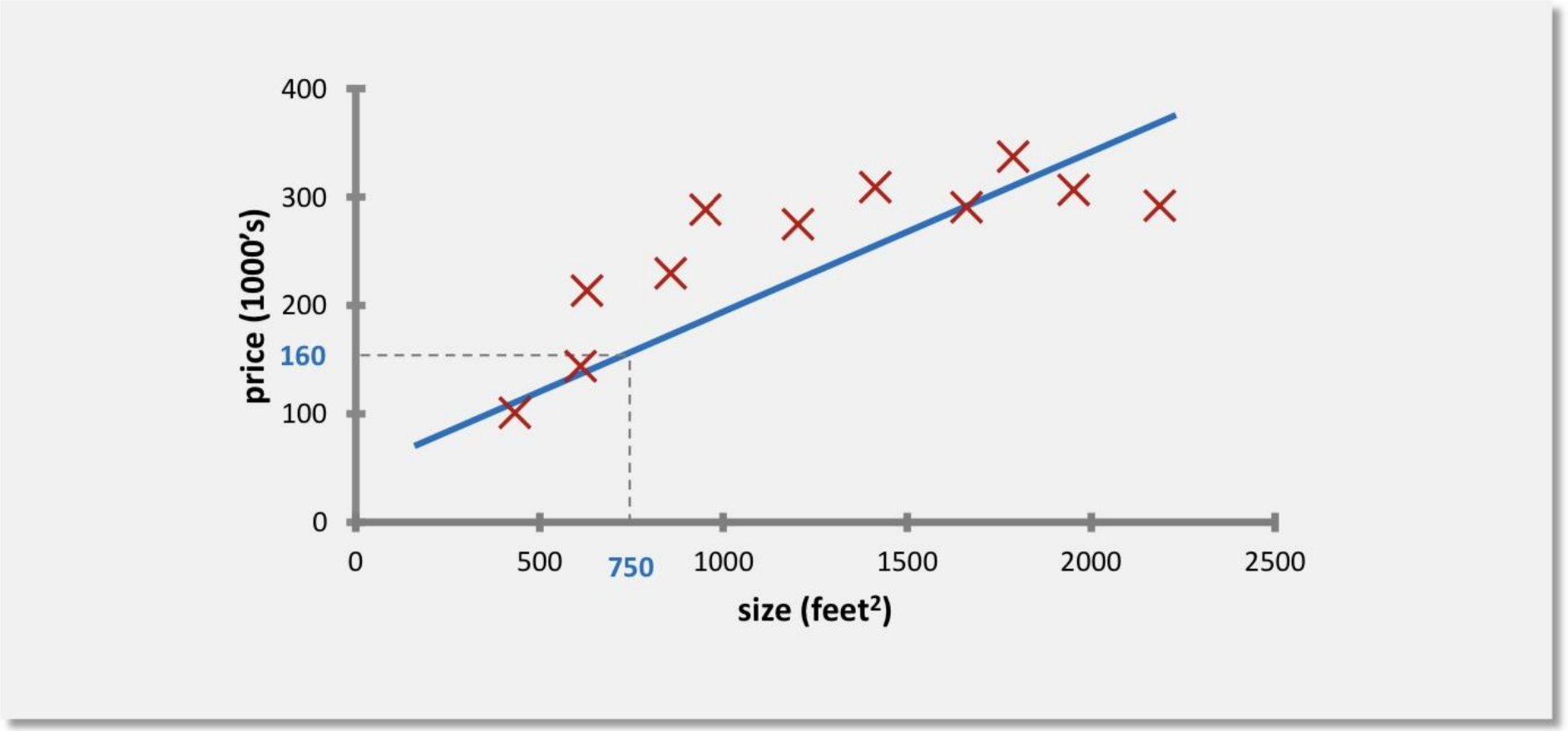
Example: Handwritten-digits Detection

Input: a digit's image
Output: a digit



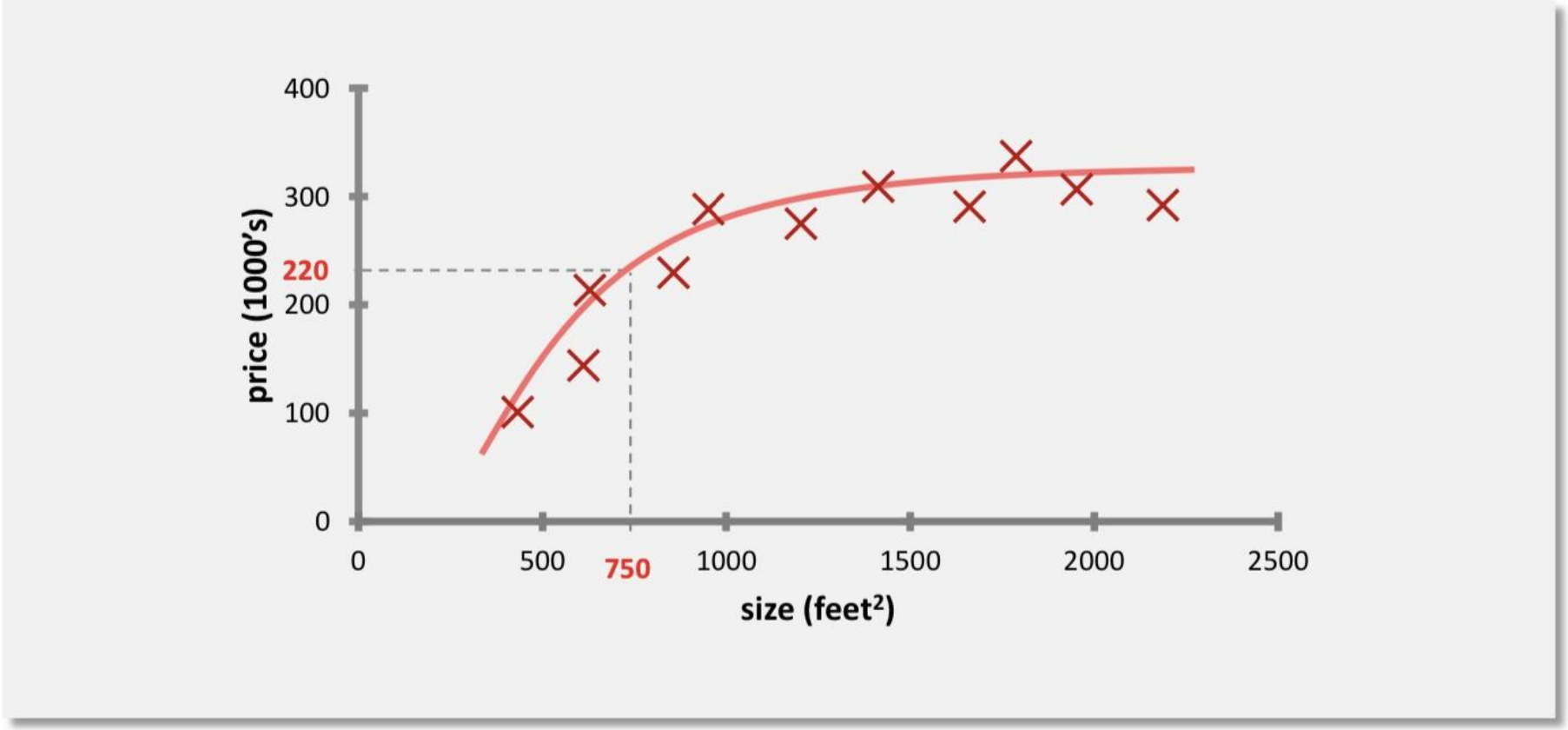
Example: Pricing a House

Input: house's size (square foot)
Output: Estimated price

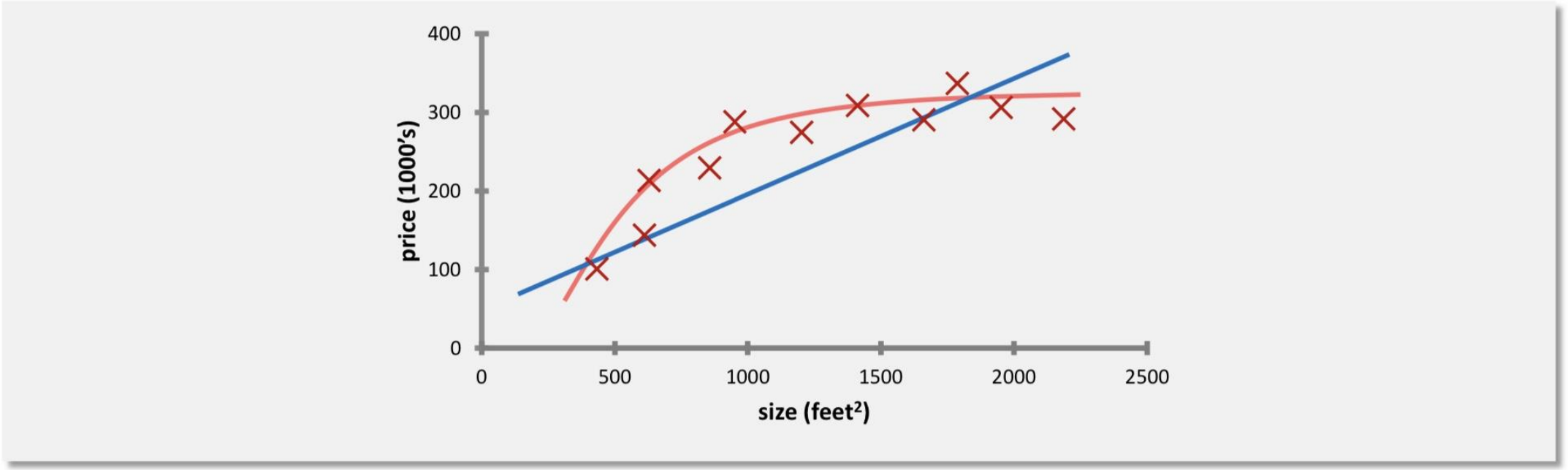


Example: Pricing a House

Question: Which one is better? A linear function or a quadratic function?



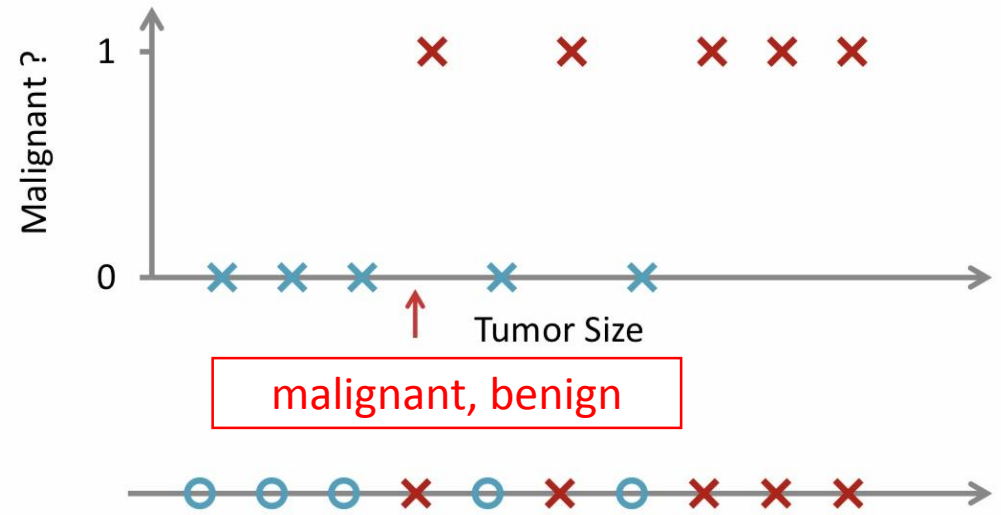
Example: Pricing a House



Supervised learning:
For each training example,
a "correct answer" is given.

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

Example: diagnosis of the type of cancer (malignant, benign)



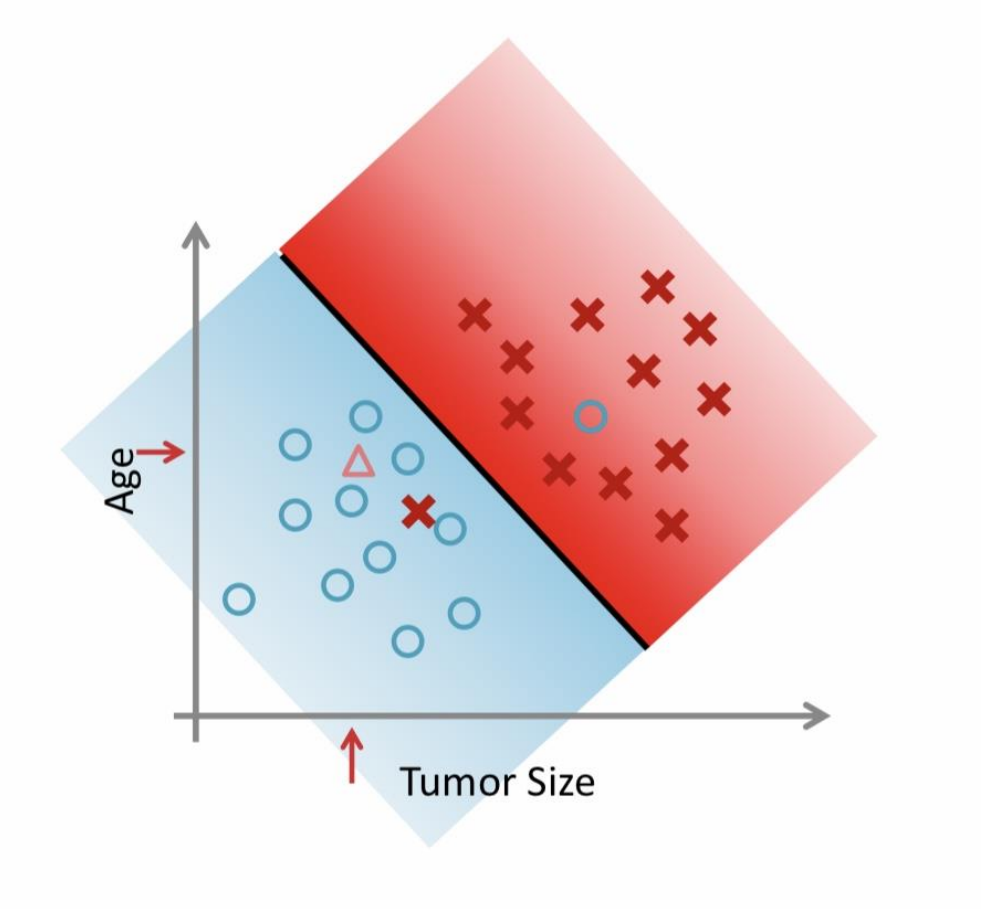
Supervised learning:
For each training example,
a "correct answer" is given.

classification:
Predicting quantities with discrete values (such as 0 and 1)

Example: diagnosis of the type of cancer (malignant, benign)

Other features:

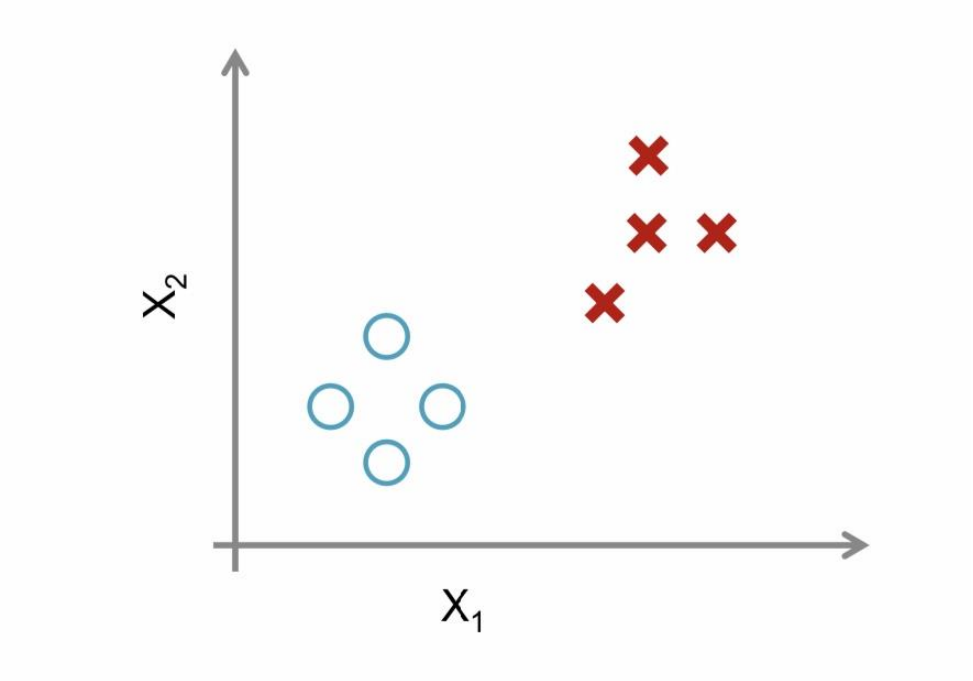
- Uniformity of cell size
- Uniformity of cell shape
- …



Unsupervised Learning

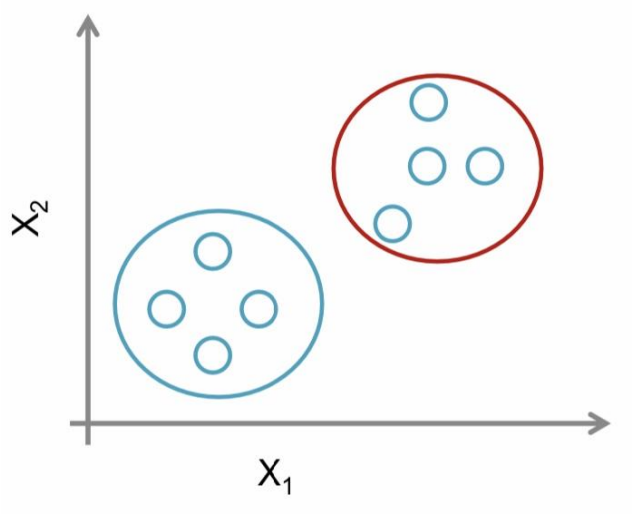
Supervised Learning

For each training example, a correct answer is given.



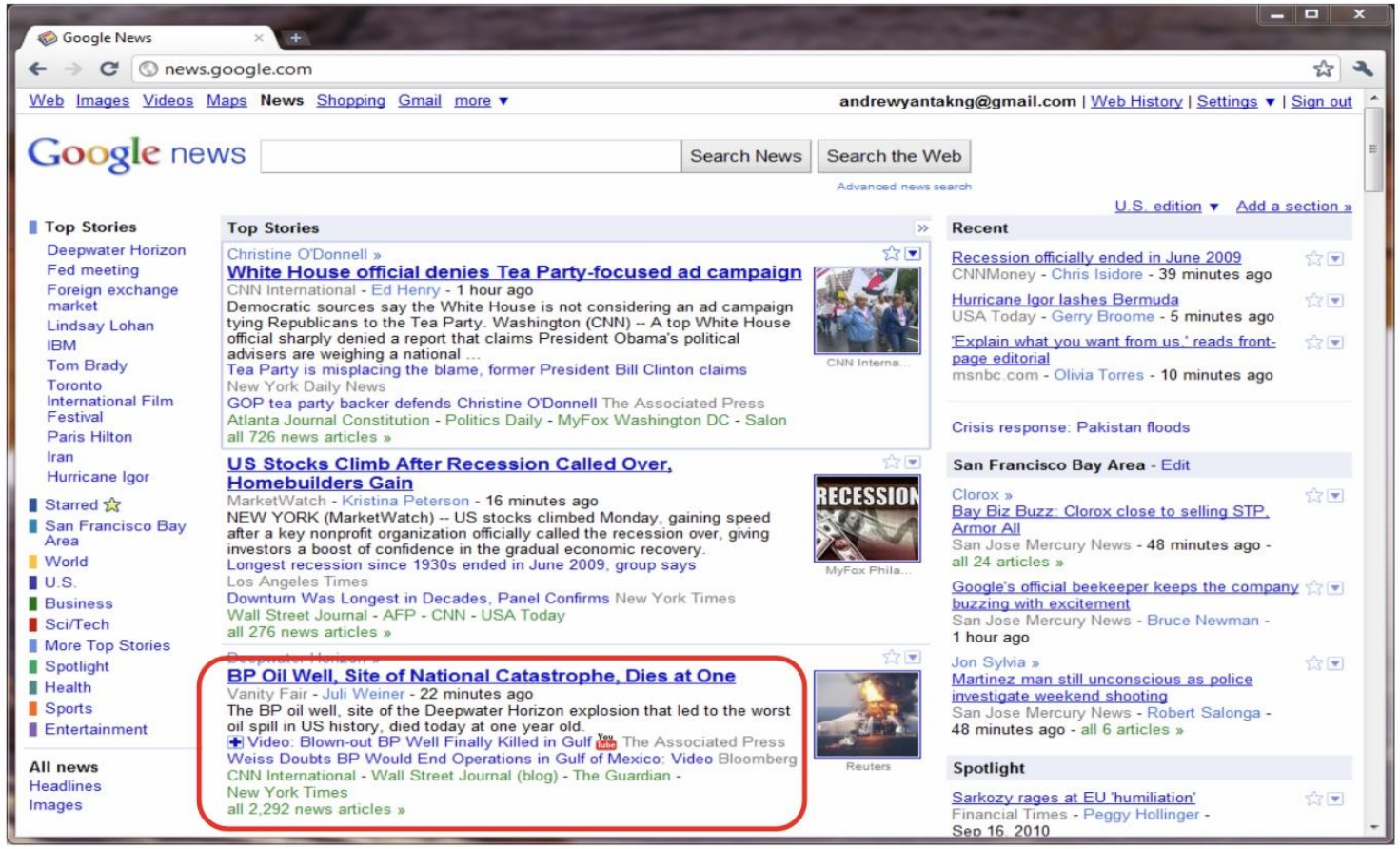
Unsupervised Learning

- Unsupervised learning: no information about correct answers is given!

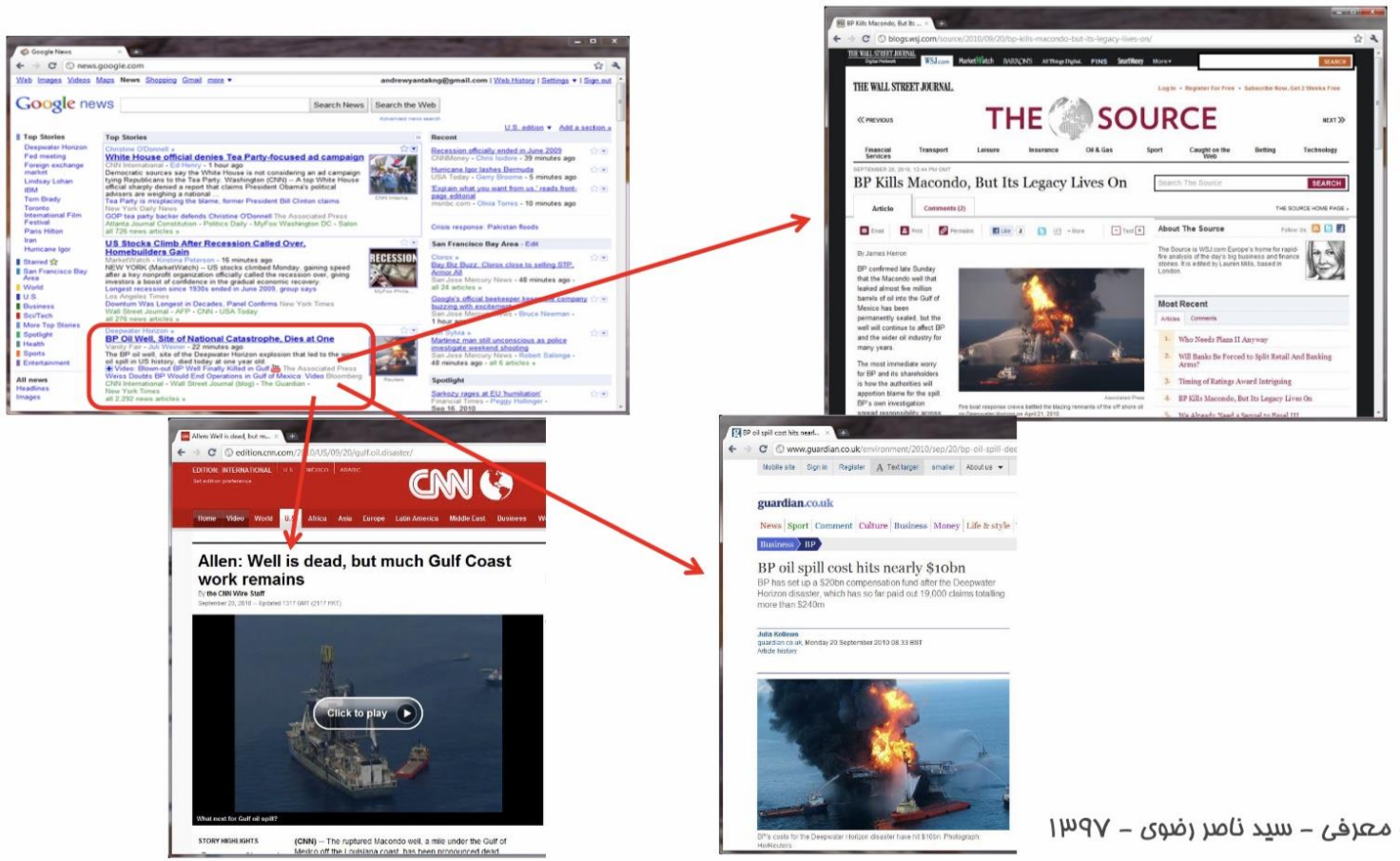


- Goal: recognizing the structure in the input data (grouping similar data)

Application of clustering: grouping related news

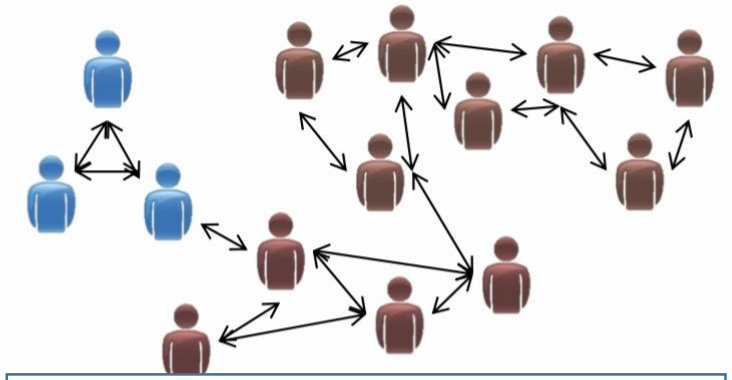


Application of clustering: grouping related news

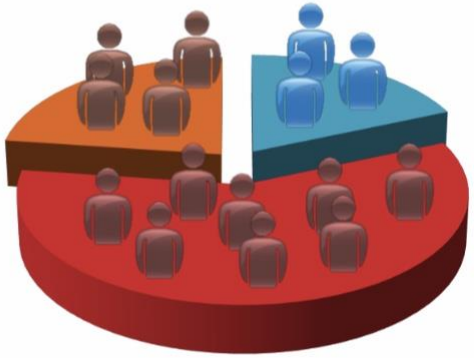


Some other applications of unsupervised learning

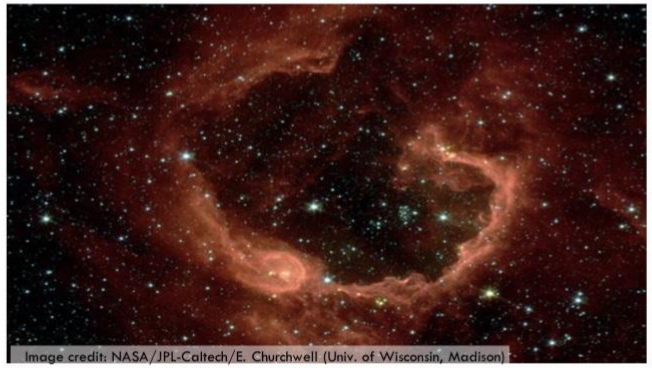
Organization of computing clusters (data center)



Social Networks Analysis

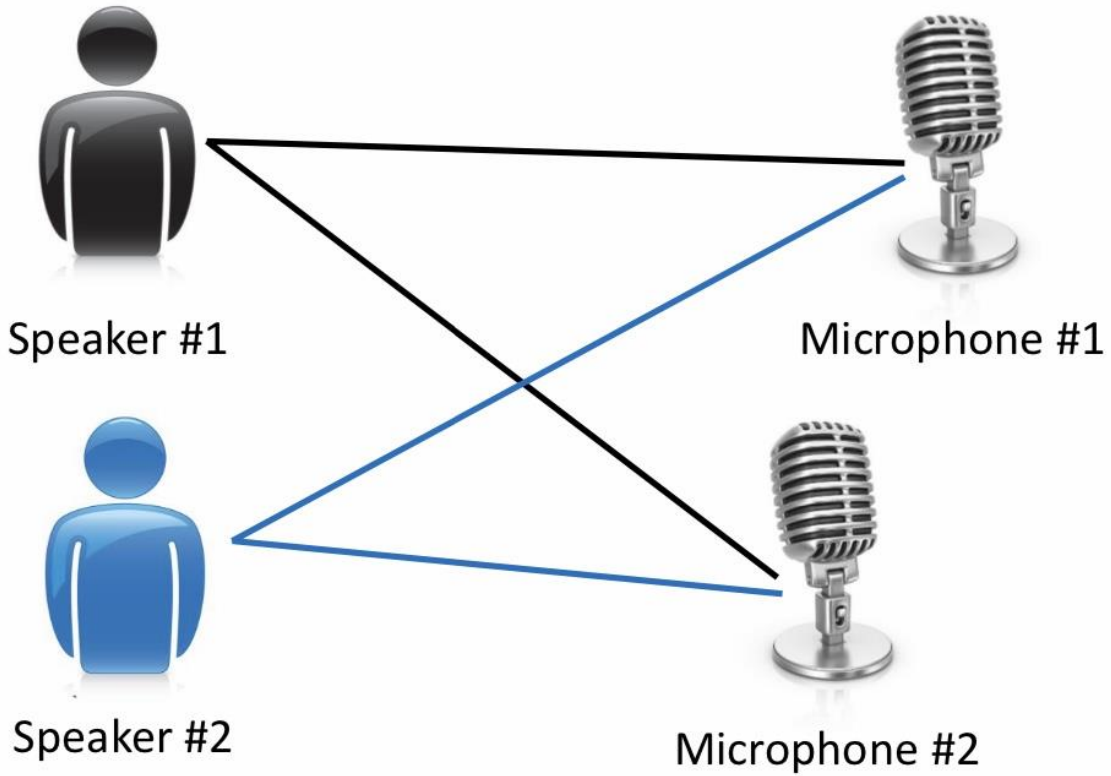


Market Segmentation



Analysis of astronomical data (how galaxies form)

The cocktail party issue



The cocktail party issue

Microphone #1: 🔊 Output #1: 🔊

Microphone #2: 🔊 Output #2: 🔊

Microphone #1: 🔊 Output #1: 🔊

Microphone #2: 🔊 Output #2: 🔊

The algorithm of cocktail party issue

Octave code

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

Class Question

- For which of the following problems should an unsupervised learning algorithm be used?

  - Develop a program to filter spam by having some normal emails and some spam
  - Grouping a collection of newly found articles on the web by topic
  - Grouping a set of customers into several different market segments by having a database of customers
  - Diagnosing diabetes in new patients with data on a number of healthy and diabetic individuals