Georgia Tech

Lecture 01. Course Overview

Xin Chen

These slides are based on slides from Mahdi Roozbahani

Outline

- Logistics 
- Syllabus overview
- Introduction to machine learning (ML)
- Machine learning examples
- Take-home messages and next

Logistics

- Registration
  - Please decide soon if you want to take the class
- All resources of this course will be on https://xinchen384.github.io/cs4641B-summer2020/ and Canvas.
- Please sign up in piazza, since we can not see you in person
- Every one-two week, I will post a thread to collect your feedback, including some unclear parts in the lectures or some interesting things you'd like to cover in our future.

Outline

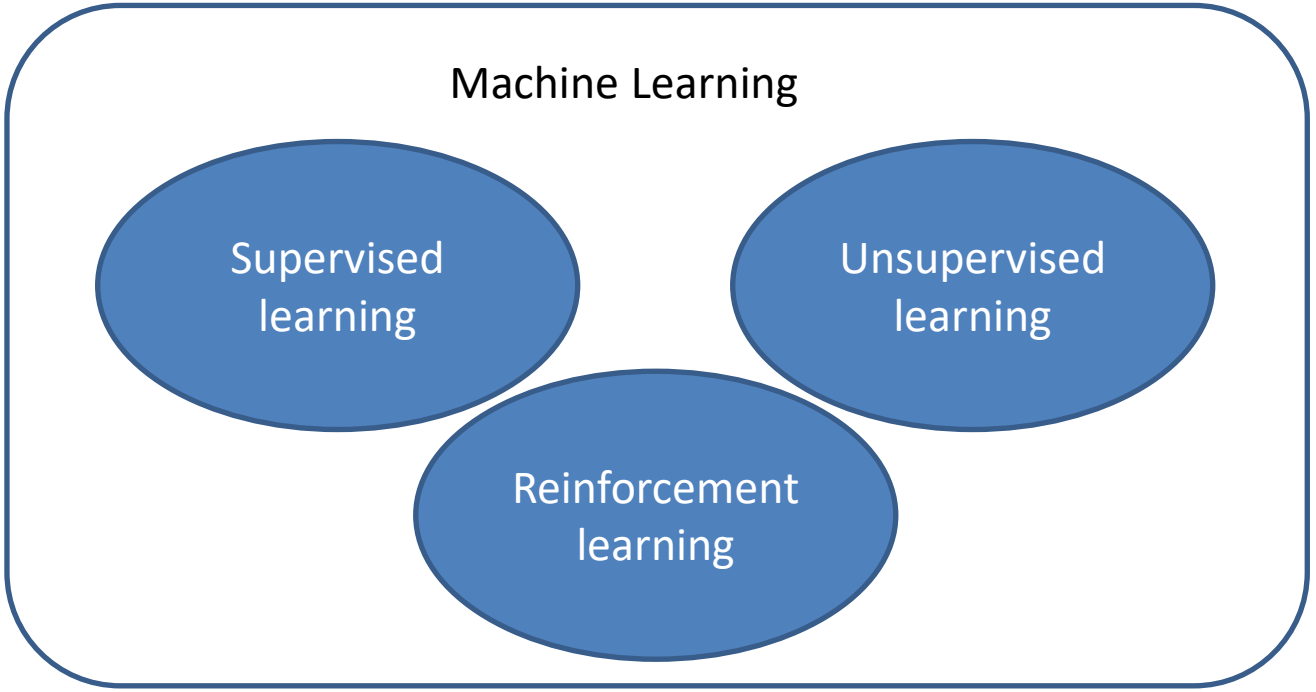
- Logistics
- Syllabus overview 
- Introduction to machine learning (ML)
- Machine learning examples
- Take-home messages and next

Syllabus: Course Objectives

- Introduce to you the **pipeline of Machine Learning**
- Help you understand **major machine learning algorithms**
- Help you learn to **apply tools** for **real data analysis problems**
- Encourage you to **do research** in data science and machine learning

Syllabus: Course Content

Basics of linear algebra and probability



Syllabus: Supervised Learning

- Linear classification/regression models
  - Linear regression
  - Naive Bayes
  - Logistic regression
  - Support vector machine

- Tree-based models
  - Linear regression Decision tree
  - Ensemble learning/Random forest

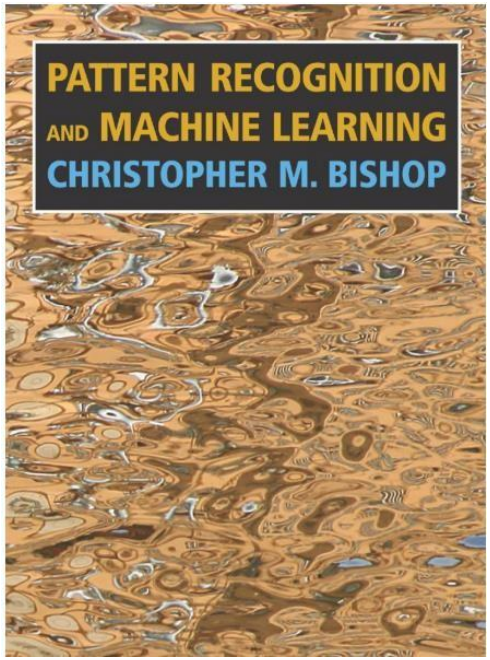
Syllabus: unsupervised Learning

- Clustering Analysis
  - K-means Gaussian mixture model
  - Hierarchical clustering
  - Density-based clustering
- Dimension Reduction
  - Principal component analysis
- Kernel Density Estimation
  - Parametric density estimation
  - Non-parametric density estimation

Syllabus: Reinforcement Learning

- Markov Decision Process
- Reinforcement learning
- Neural network
  - Deep neural network
  - Deep reinforcement learning
  - Feedforward neural networks and backpropagation analysis

Text Books



Pattern Recognition and Machine Learning, by Chris Bishop

Other recommended books:

Learning from data, by Yaser S. Abu-Mostafa

Machine learning, by Tom Mitchell

Deep Learning, by Ian Goodfellow, Yoshua Bengio, and Aaron Courville

Pre-requisite

- Linear algebra
  - Matrix multiplication
  - Eigenvector
- Probability
  - Distribution, random variable, expectation, conditional probability, variance, density
- Basic programming
  - Python

Syllabus: grading policy

- The syllabus is posted on canvas, including the grading policy

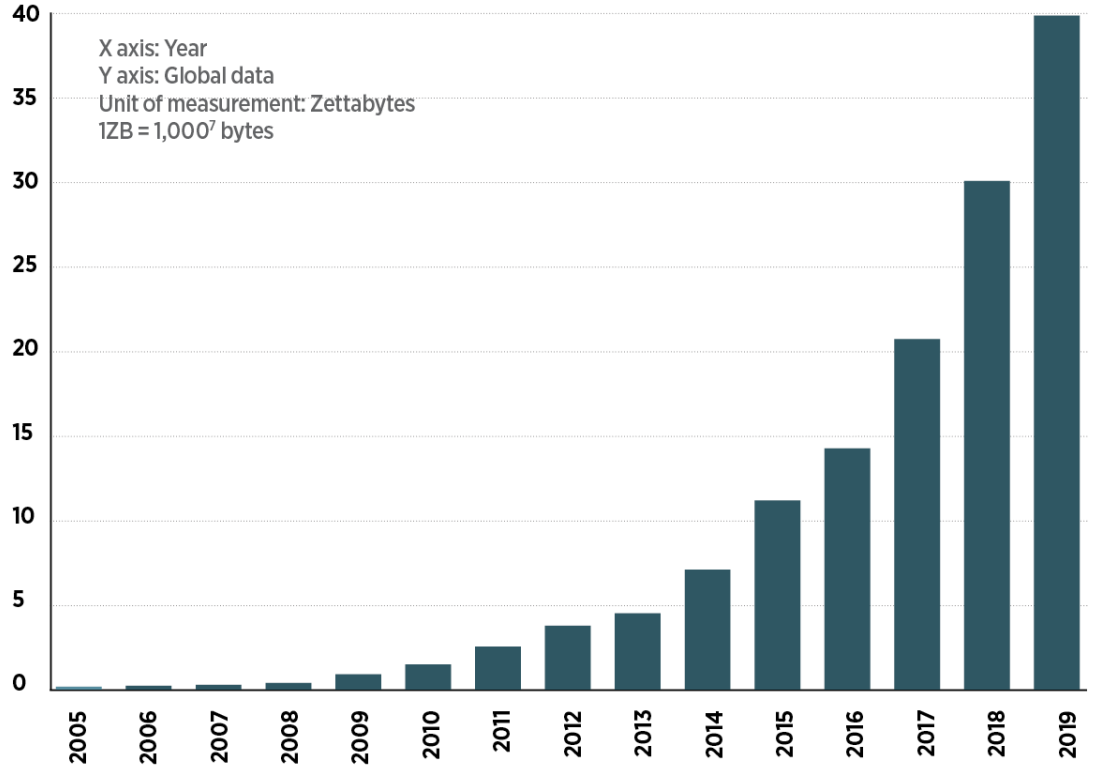
Outline

- Logistics
- Syllabus overview
- Introduction to machine learning (ML) 
- Machine learning examples
- Take-home messages and next

Machine Learning

"***We are drowning in information but starved for knowledge.***"

— John Naisbitt



The Booming Age of Data

30 trillion Web pages

500 million tweets per day

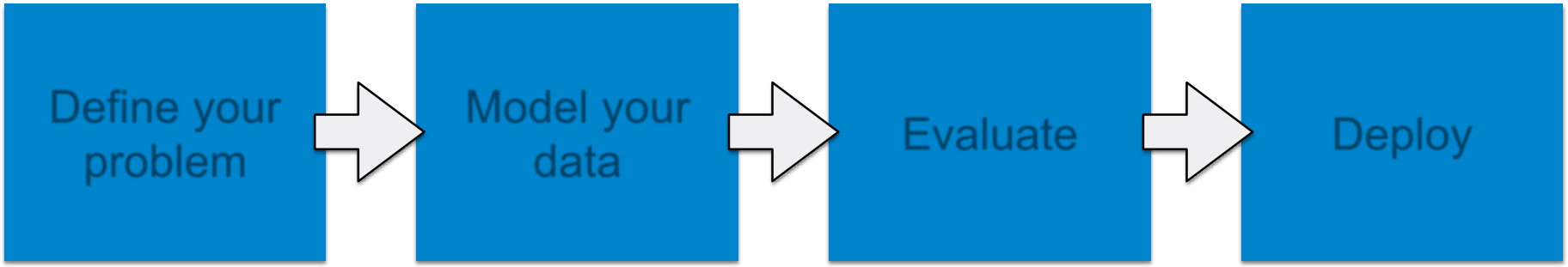
2.27 billion monthly active users

1.8 billion images uploaded to Internet per day

2.9 billion base pairs in human genome

Machine Learning

Machine Learning is the process of **turning data into actionable knowledge** for **task support** and **decision making**.



Brief History of Machine Learning

1950s

Samuel's checker player

Selfridge's Pandemonium

1960s:

Neural networks: Perceptron

Pattern recognition

Learning in the limit theory

Minsky and Papert prove limitations of Perceptron

1970s:

Symbolic concept induction

Winston's arch learner

Expert systems and the knowledge acquisition bottleneck

Quinlan's ID3

Michalski's AQ and soybean diagnosis

Scientific discovery with BACON

Mathematical discovery with AM (Automated Mathematician)

Slide Credit: Ray Mooney

Brief History of Machine Learning

1980s:

Advanced decision tree and rule learning
Explanation-based Learning (EBL)
Learning and planning and problem solving
Utility problem
Analogy
Cognitive architectures
Resurgence of neural networks (connectionism, backpropagation)
Valiant's PAC Learning Theory
Focus on experimental methodology

1990s

Data mining
Adaptive software agents and web applications
Text learning
Reinforcement learning (RL)
Inductive Logic Programming (ILP)
Ensembles: Bagging, Boosting, and Stacking
Bayes Net learning

Slide Credit: Ray Mooney

Brief History of Machine Learning

2000s:

Support vector machines
Kernel methods
Graphical models
Statistical relational learning
Transfer learning
Sequence labeling
Collective classification and structured outputs
Computer Systems Applications
Learning in robotics and vision

2010s:

Deep learning
Reinforcement learning
Generative models
Adversarial learning
Muti-task learning
Learning in NLP, CV, Robotics, …

Based on Ray Mooney's slide

Definition of Machine learning

- Arthur Samuel (1959): Machine learning is the field of study that gives the computer the ability to learn without being explicitly programmed.

- Tom Mitchell (1998): a computer program is said to learn from experience E with respect to some class of task T and performance measure P, if its performance at task in T, as measured by P, improves with experience E.

Outline

- Logistics
- Syllabus overview
- Introduction to machine learning (ML)
- Machine learning examples 
- Take-home messages and next

Unsupervised and Supervised learning

	Weight(lb)	Height(cm)	Fur color	Eye color	Label
Point 1	10	20	w	g	cat
Point 2	50	100	br	bl	dog
Point 3	8	15	bl	bl	dog
Point 4	12	25	w	bl	cat
Point 5	14	10	bl	g	dog

$X_{n\times d}$ $Y_{n\times 1}$

Unsupervised just focuses on $X_{n\times d}$

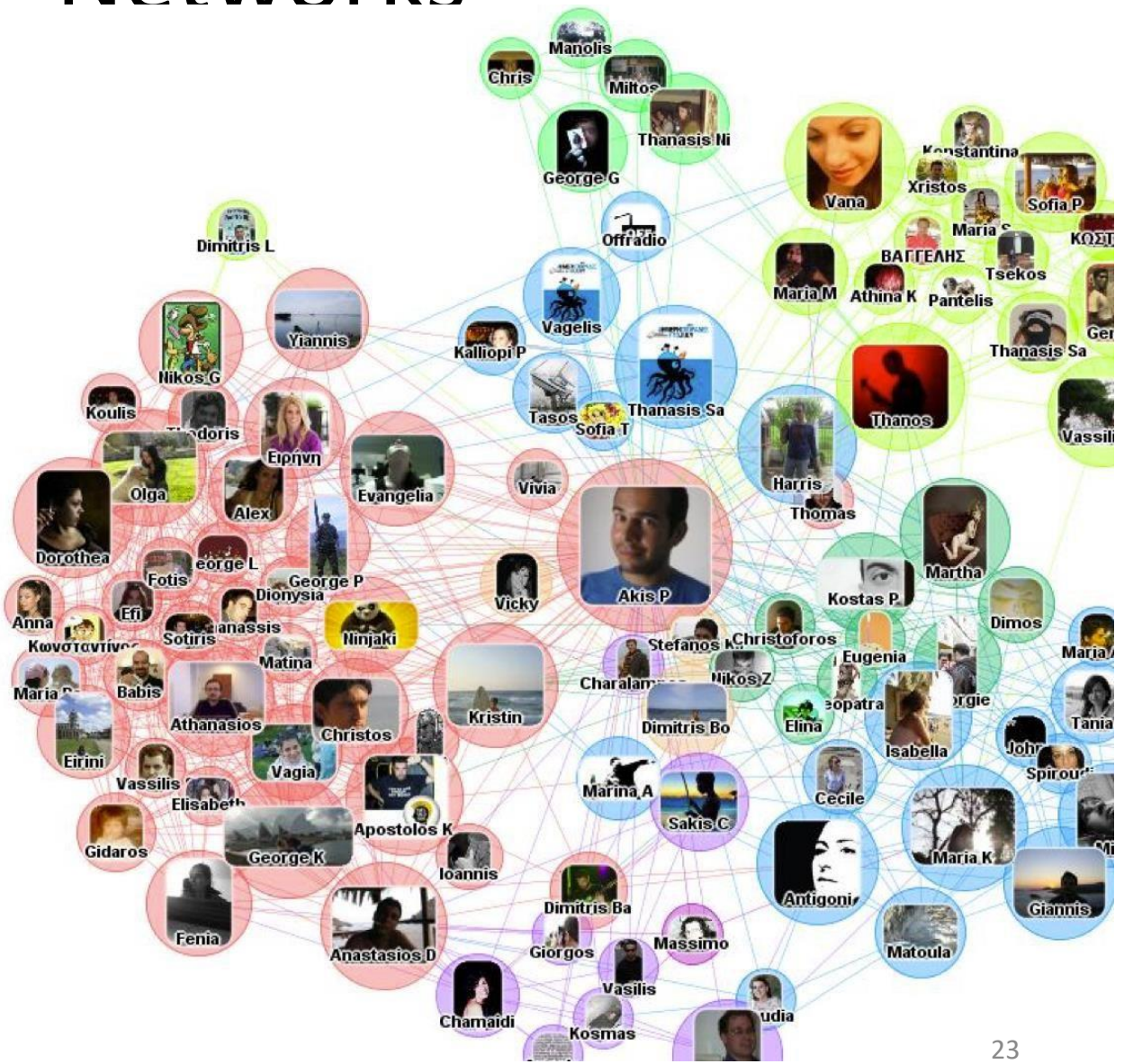
Supervised just focuses on $X_{n\times d}$ and $Y_{n\times 1}$

Community Detection in Social Networks

What are the inputs and how to represent them?

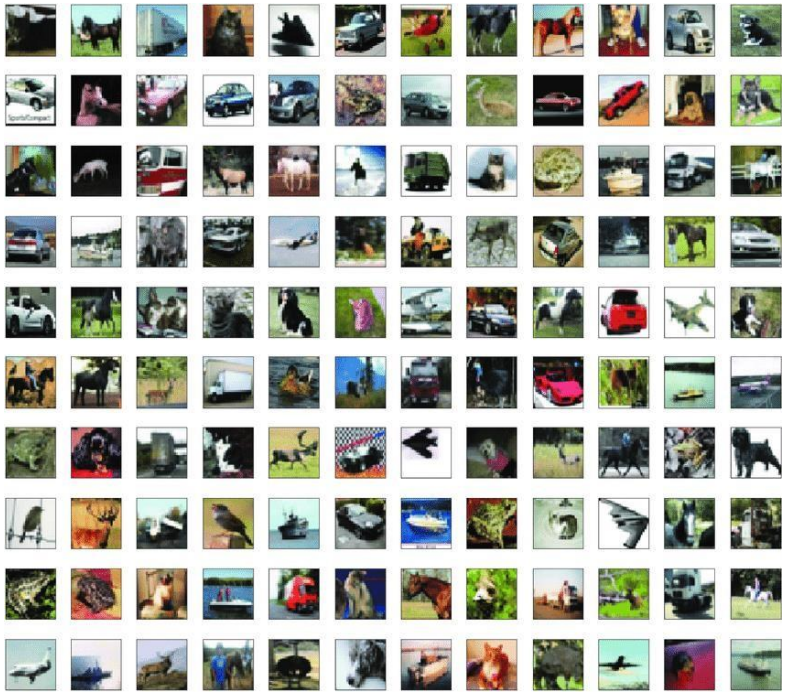
What are the desired outputs?

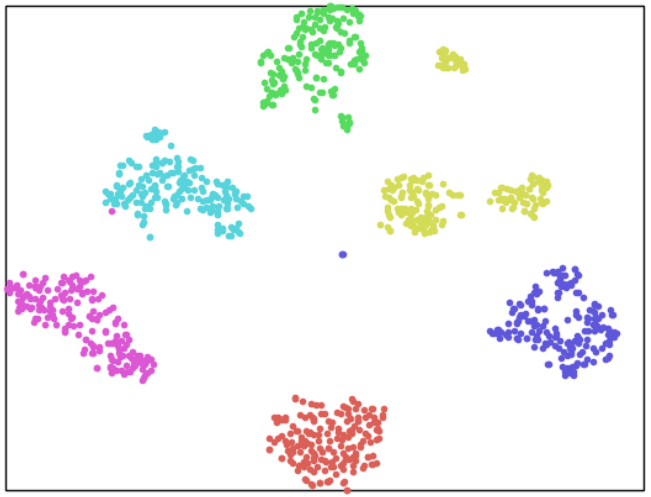
What learning algorithms to choose?

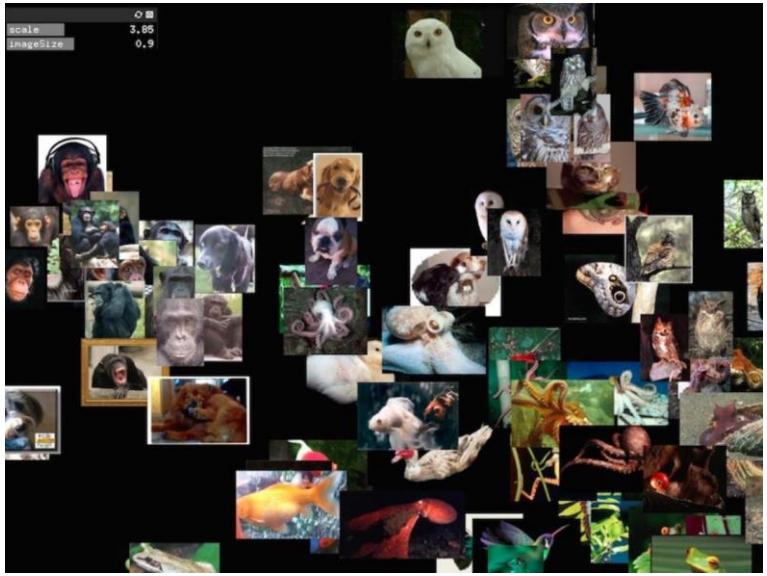


mage credit: Symeon Papadopoulos

Dimensionality Reduction







What are the inputs and how to represent them?

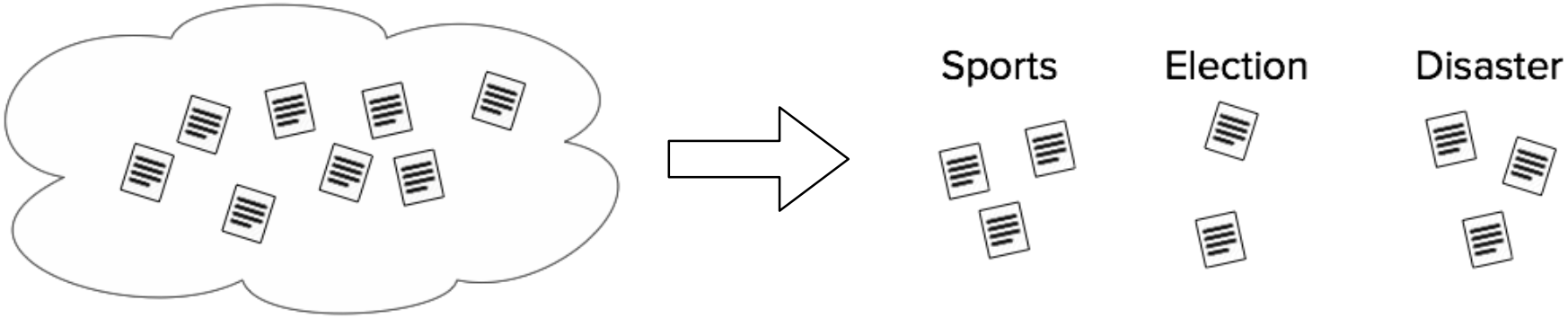
What are the desired outputs?

What learning algorithms to choose?

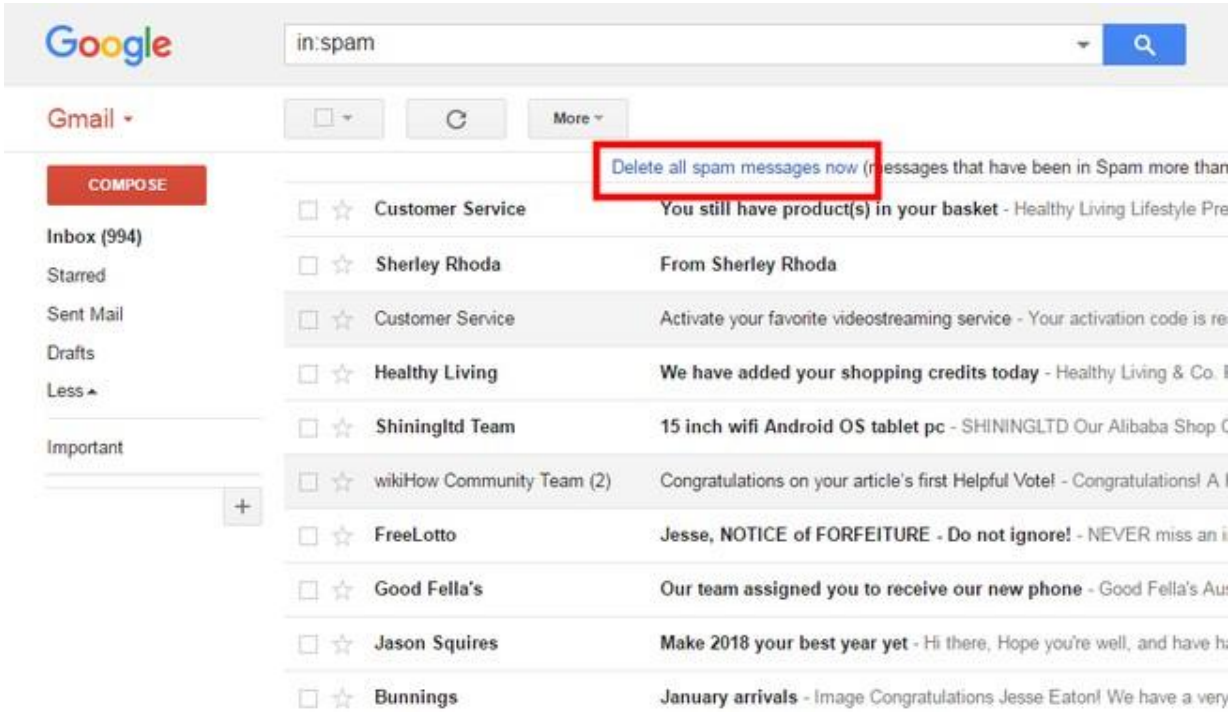
mage credit: Gene Kogan

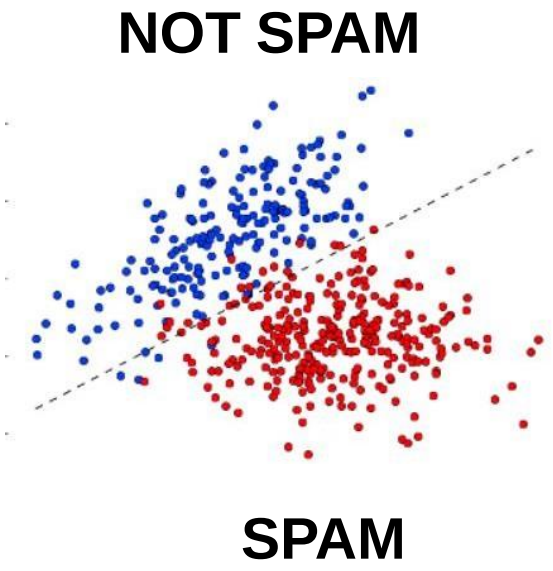
News Classification

- What are the inputs and how to represent them?
- What are the desired outputs?
- What learning algorithms to choose?



Spam Detection





What are the inputs and how to represent them?

What are the desired outputs?

What learning algorithms to choose?

Image credit: WikiHow

Examples

FONT TELLER

Root Inspired Anchor Model Project

Geographic Classification of Cuisine Recipes

Let us try a simple ML problem

Problem 1:

Problem 2:

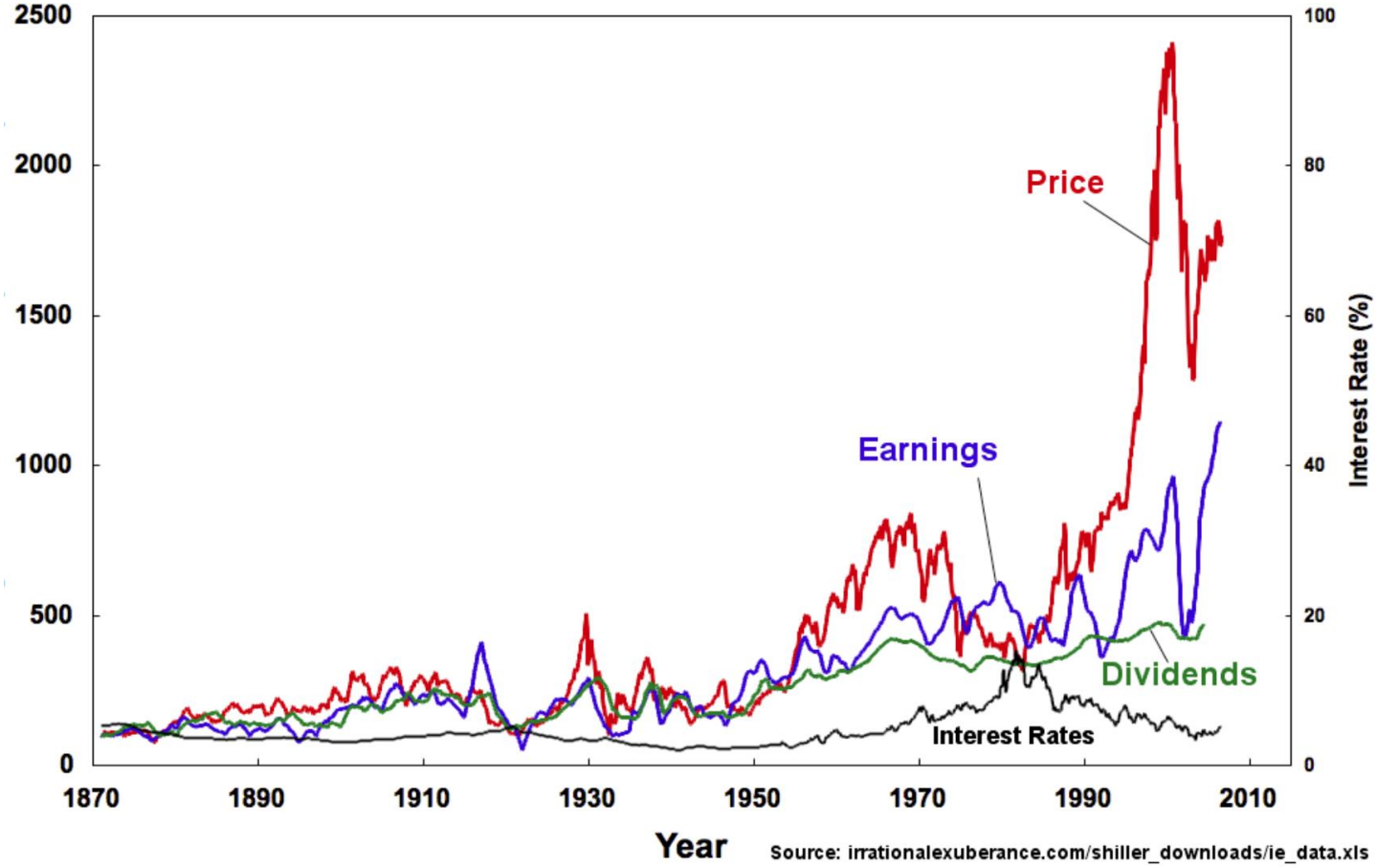
What can we learn from this ML problem?

Problem 1:

Problem 2:

Are we part of a ML problem?

Can we apply ML on stock prices?



Outline

- Logistics
- Syllabus overview
- Introduction to machine learning (ML)
- Machine learning examples
- Take-home messages and next 

Take-home messages

- Machine learning is about learning a model that helps people explain the history and predict the next.
- We are going to cover three main topics in ML: supervised learning, unsupervised learning, and reinforcement learning.
- For each topic (algorithm), we need to focus two aspects: the underlying math and its application.
- Next, we will talk about some math.