Course Overview

[Mahdi Roozbahani](#)

Lecturer, College of Computing, CSE,

Georgia Tech

Founder of [Filio](#)

How to succeed in this course?

I will always start the class with “Good Morning” or “Good Afternoon” (of course, if the class starts before noon)

I need a response **full of energy** from **all** students.
It gives me a good feeling and shows me you are here to learn new exciting things. That way I stay motivated.

If I stay motivated → I do my best in teaching → you learn better

→ you will get an A

Let’s practice

Refer to:

Class Website

For anything (updates, lectures, logistics, and so on) related to this class

CIOS - Students

- “The course was great and helpful as a starter to learn machine learning. The project work is really valuable and can make a great addition to resumes if students put in a good effort and learn. The project requires self-learning quite a few things but I think that's a great way to learn in depth. For me the project was very helpful since lots of my interviewers in Intel asked about it and were interested in it. The cherry on the cake was that I landed the job at Intel.”
- “I went back and watched lectures before I took the weekly quizzes. The homework was tough, and they werent kidding when they said it would take a lot of time.”
- “Metaphorically speaking, I felt like whenever I left lecture, I knew how to put up a fence in my backyard. but when the time came to do the homeworks, it felt like I was being asked to rebuild the great wall of china.”
- “This class expects a lot of effort and I had to expend a lot of effort. The concepts are very advanced and complicated, so it is important to pay attention to lectures and allocate plenty of time to complete the homework. There is a lot of learning as you go, so time management is key.”
- “The fact that I got to learn so much, complete some intense coding, do a group project and accomplish so much in one class in one semester is amazing.”
- “I think that the class inherently requires a significant amount of time and effort since the material is very complicated. For the complexity of what we are learning, the effort required is appropriate”

I verbally ask questions in the class

Sometimes I ask questions about previous lectures or something that you have already learnt.

Answer the question even if you think you are wrong (nobody loses point answering the question wrong). It will help you and other students to understand concepts much better.

Ed (the best source for Q/A)

Ask your questions in ed (make it public to other students), and also please see other questions on Ed, it might answer your question. (Please do not send me or TAs Emails regarding hw questions, exams, and other logistics – you can also ask “private question” on ed) =>Class participation

Bonus points: Undergrad and grad

Some important notes:

- No Email please, just Ed (Chat and Threads).
- Please read the website carefully. Website will be the first thing that we update if there any changes.
- Add HWs dues and Quizzes to your desired calendar.
- Please come to the class; it will help a lot ☺
- Lectures will be Math Heavy and HWs will be mostly programming. ML is all about Linear Algebra, Probability, Statistics and Optimization. You need to have both the mathematical and programming skills to be successful in this area.

Office hours:

Notes:
1) Each student may fill in their name and create a BlueJeans Meeting for the office hour
2) The signup policy is first come, first serve. **Please, DO NOT override others' sign-ups without their permissions.**
3) Because of the time limit, each student will have a maximum time of 10 minutes to work with the TA
4) Please come prepared with specific questions so we can help you more effectively

The slots were cleared on Nov 15th. Please add your name again if you had filled it for this week. Apologies for the inconvenience.

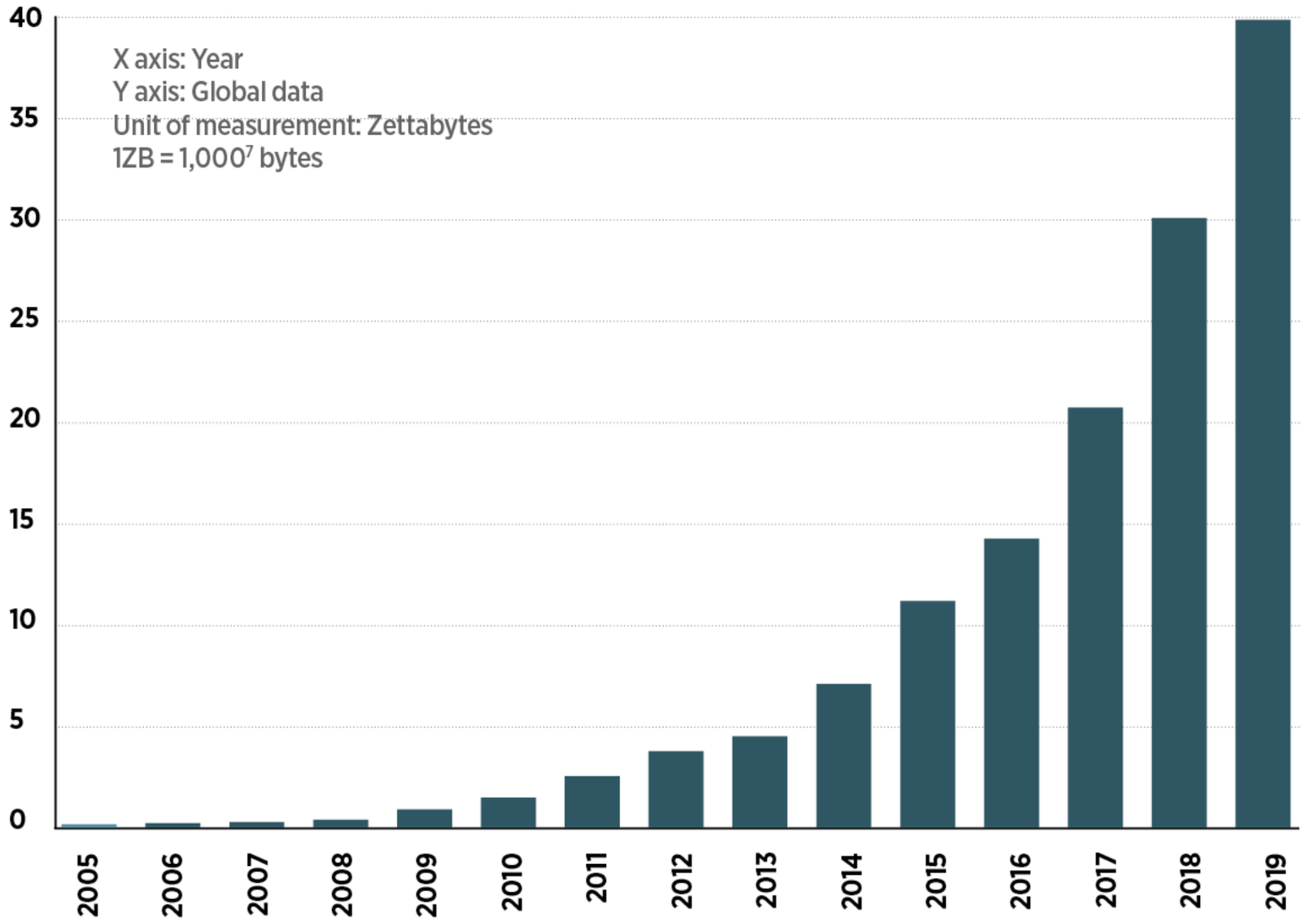
Time Slots	Student Name	Question of Interest (Please be more specific about the question you want to ask)	BlueJeans Address	Label
Jayanta's OH				
12:00--12:10				
12:10--12:20				
12:20--12:30				
12:30--12:40				
12:40--12:50				
12:50--13:00				
Waitlist				
1				
2				
3				
Huili's OH				
14:30--14:40				
14:40--14:50				
14:50--15:00				
15:00--15:10				
15:10--15:20				
15:20--15:30				
Waitlist				
1				
2				

- You have just 10 minutes for each slot.
- Please mindful of other students (you can’t block multiple office hours, just one and we have the waitlist!
- Make sure you pinpoint your problem exactly before joining the office hour.
- Office hour is not for general code debugging. You need to be specific about your question.

Machine Learning

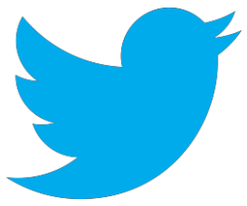
"***We are drowning in information but starved for knowledge.***"

— John Naisbitt



The Booming Age of Data

30 trillion Web pages



500 million tweets per day

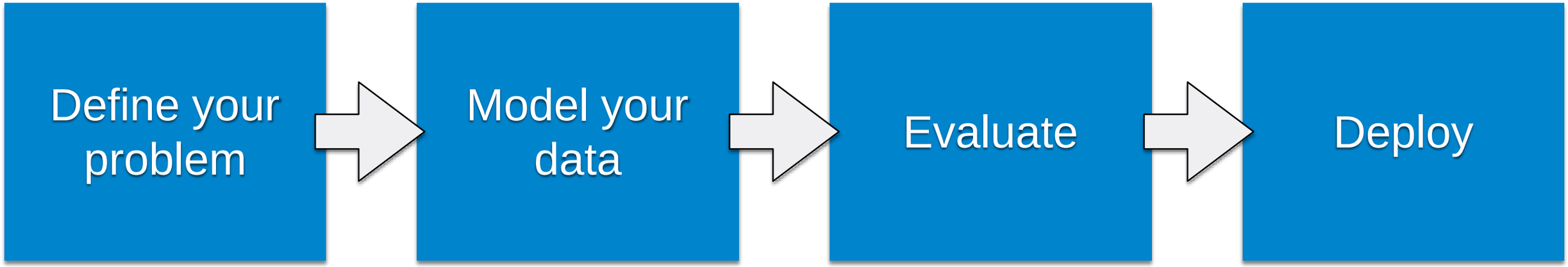
2.27 billion monthly active users

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**.



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

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

Syllabus

Part I: Basic math for computational data analysis

- Probability, statistics, linear algebra

Part II: Unsupervised learning for data exploration

- Clustering analysis, dimensionality reduction, kernel density estimation

Part III: Supervised learning for predictive analysis

- Tree-based models, linear classification/regression, neural networks

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 focus on $X_{n\times d}$ and $Y_{n\times 1}$

We can do better than Cat and Dog

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

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

Syllabus: Unsupervised Learning

Clustering Analysis

- K-means
- Gaussian mixture model
- Hierarchical clustering
- Density-based clustering
- Evaluation of clustering algorithms

Dimension Reduction

- Principal component analysis

Kernel Density Estimation

- Parametric density estimation
- Non-parametric density estimation

Community Detection in Social Networks

What are the inputs and how to represent them?

What are the desired outputs?

What learning algorithms to choose?

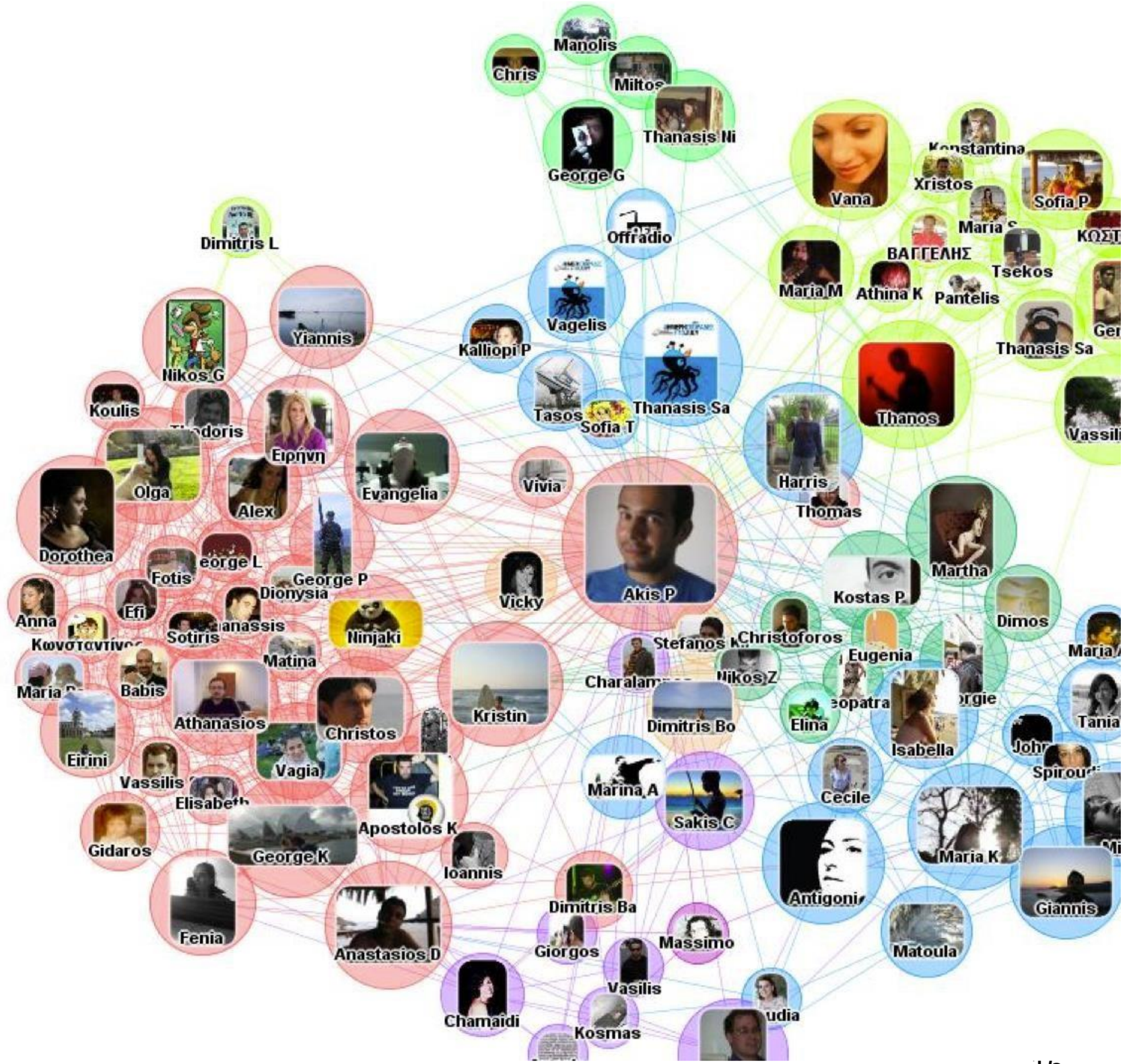


Image credit: Symeon Papadopoulos

Examples on our class website

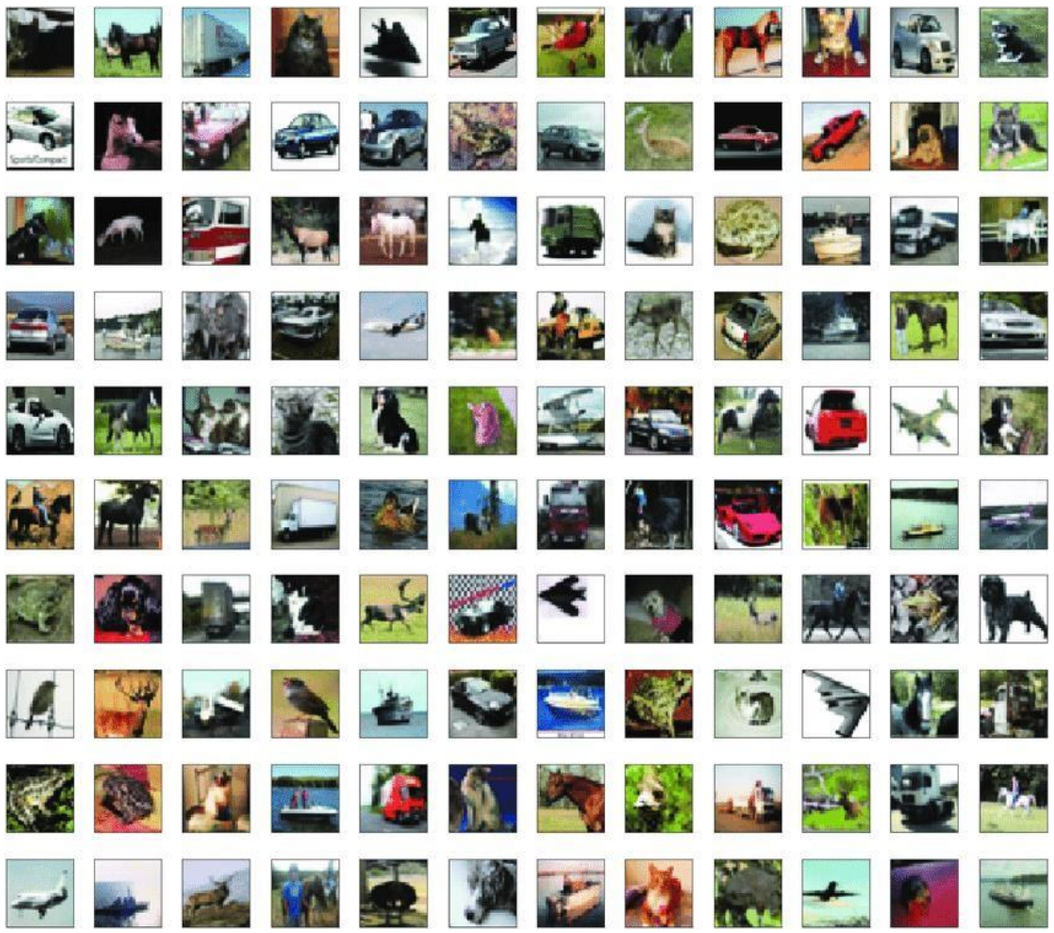
- **Sample Projects**
  - Sample Project from previous semester [Undergrad Canvas Access for previous ML projects]; [Grad Canvas Access for previous ML projects]; Stanford Project Examples;

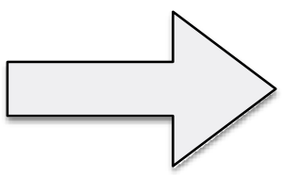
- **General project guidance**
  - Your project will be graded based on the following criteria:

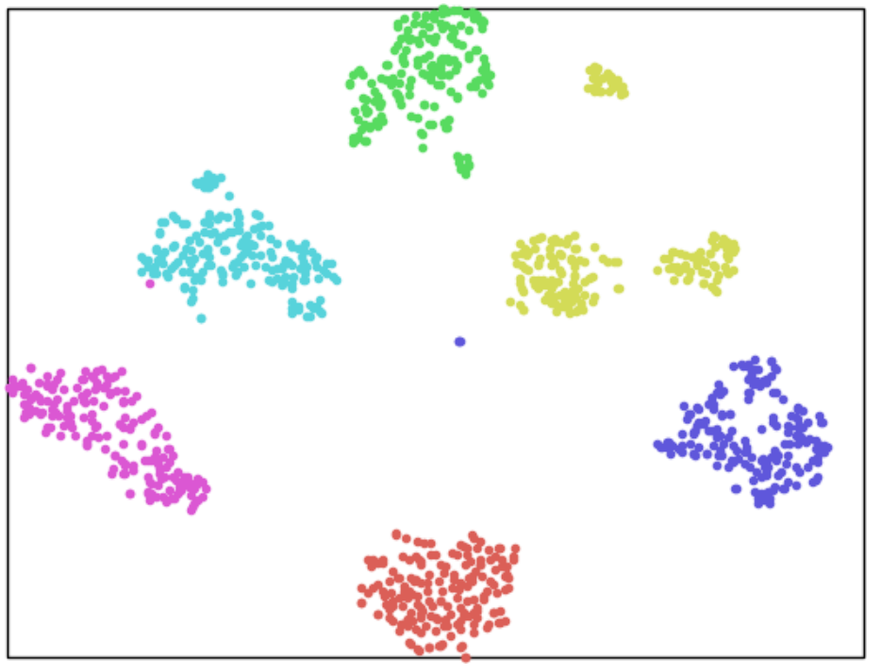
    Was the motivation clear?
    - What is the problem?
    - Why is it important and why we should care?

    Were the dataset and approach used effectively?

Dimensionality Reduction



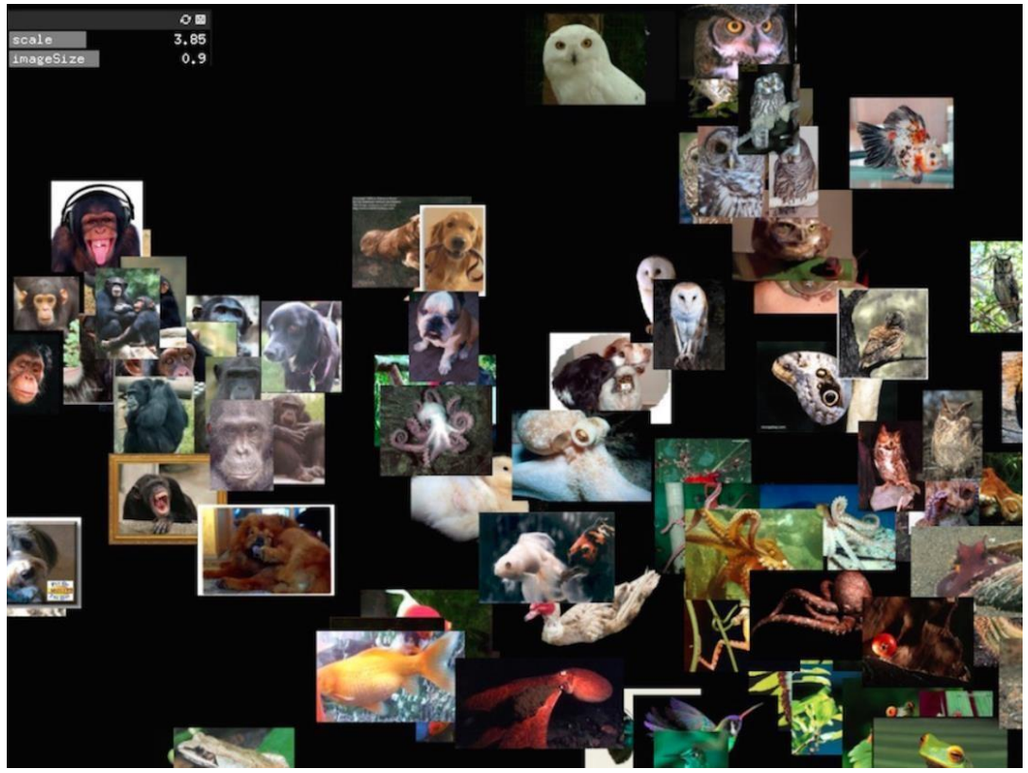




What are the inputs and how to represent them?

What are the desired outputs?

What learning algorithms to choose?



mage credit: Gene Kogan

Syllabus: Supervised Learning

Tree-based models

- Decision tree
- Ensemble learning/Random forest

Linear classification/regression models

- Linear regression
- Naive Bayes
- Logistic regression
- Support vector machine

Neural networks

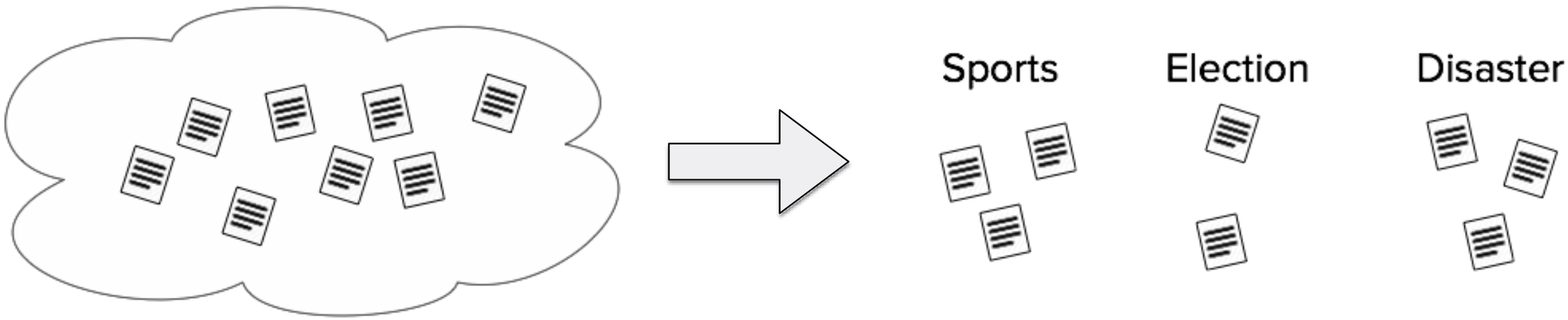
- Feedforward neural networks and backpropagation analysis
- CNN

News Classification

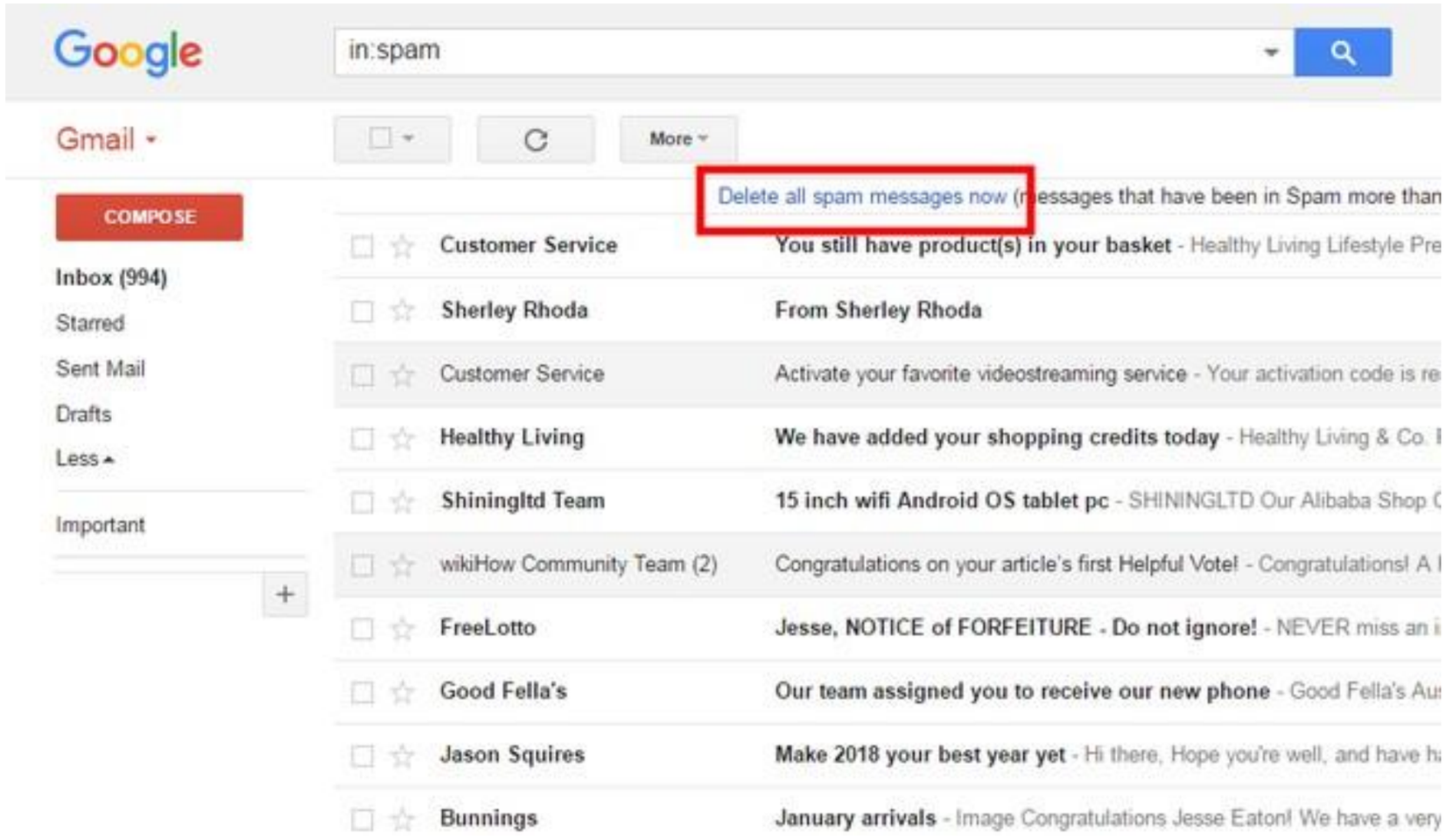
What are the inputs and how to represent them?

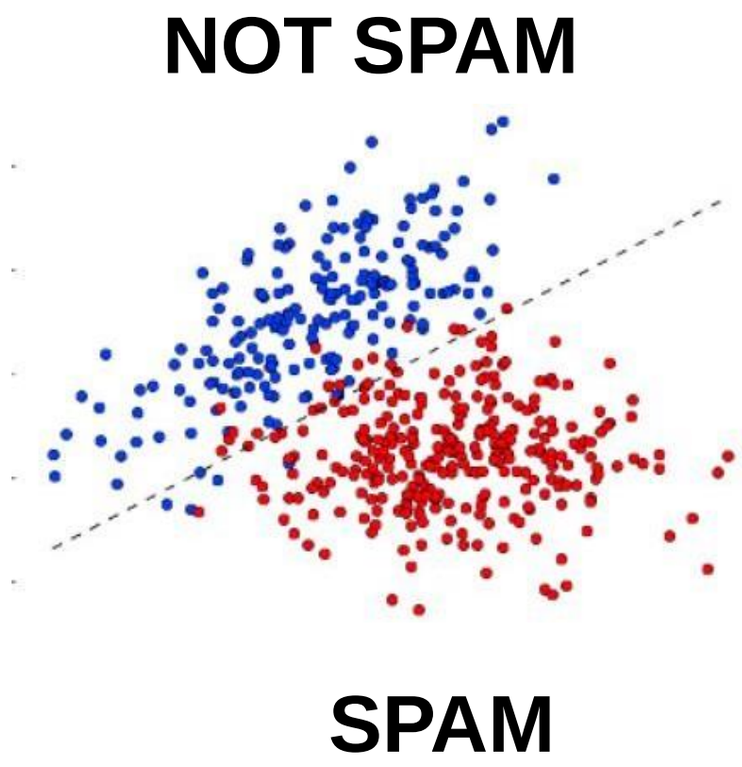
What are the desired outputs?

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 on our class website

- **Sample Projects**
  - Sample Project from previous semester [Undergrad Canvas Access for previous ML projects]; [Grad Canvas Access for previous ML projects]; Stanford Project Examples;

- **General project guidance**
  - Your project will be graded based on the following criteria:

    Was the motivation clear?
    - What is the problem?
    - Why is it important and why we should care?

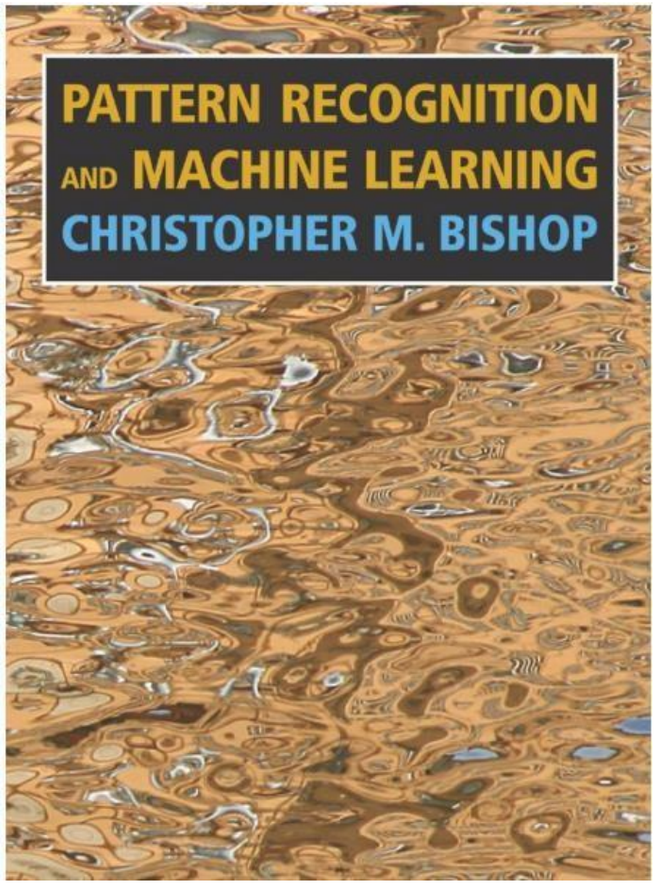
    Were the dataset and approach used effectively?

Prerequisites

Basic knowledge in probability, statistics, and linear algebra

Basic programming skills in Python (Jupyter Notebook)

No background in machine learning is required



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

Assignments

Four assignments (GradeScope)
Each can include written analysis or programming

Start Early as soon as they are out

No-late policy

Assignments received after the due time will receive zero credit

Don’t copy

Because of the large size of our class, if we observe any (even small) similarity\plagiarisms detected by GradeScope or our TAs, WE WILL DIRECTLY REPORT ALL CASES TO OSI, which may unfortunately lead to a very harsh outcome.

Projects

- Work on a real-life Machine Learning problem
  - What is the problem? What is your method? How do you evaluate it?
- Exactly 5 people in a team (Grad and undergrad can’t be mixed in a group)
- GitHub Pages (index.html)
- Start your projects early
- Ask for comments and feedbacks from the teaching staff
- You need to have a group meeting with your team at least once in a week (either virtual or in-person). Otherwise, team conflicts will most probably happen.