CS4641B Machine Learning

Lecture 01: Course overview

Rodrigo Borela ▸ rborelav@gatech.edu

Where can I find information on the…

Schedule
Class policies
Inclusion
Accessibility
Office hours
Grading
Assignments
Project
Quizzes
COVID-19 impacts
Additional resources

→ https://rborelav.github.io/cs4641b-fall20/

Instruction Team

Rodrigo Borela
[he/him/his]

Nimisha Roy

Danrong Zhao

Rafael Hanashiro

Prithvi Alva Suresh

Vidisha Goyal

Tanvi Bhagwat

Gnanaguruparan Aishvaryaadevi

Bo Zhao

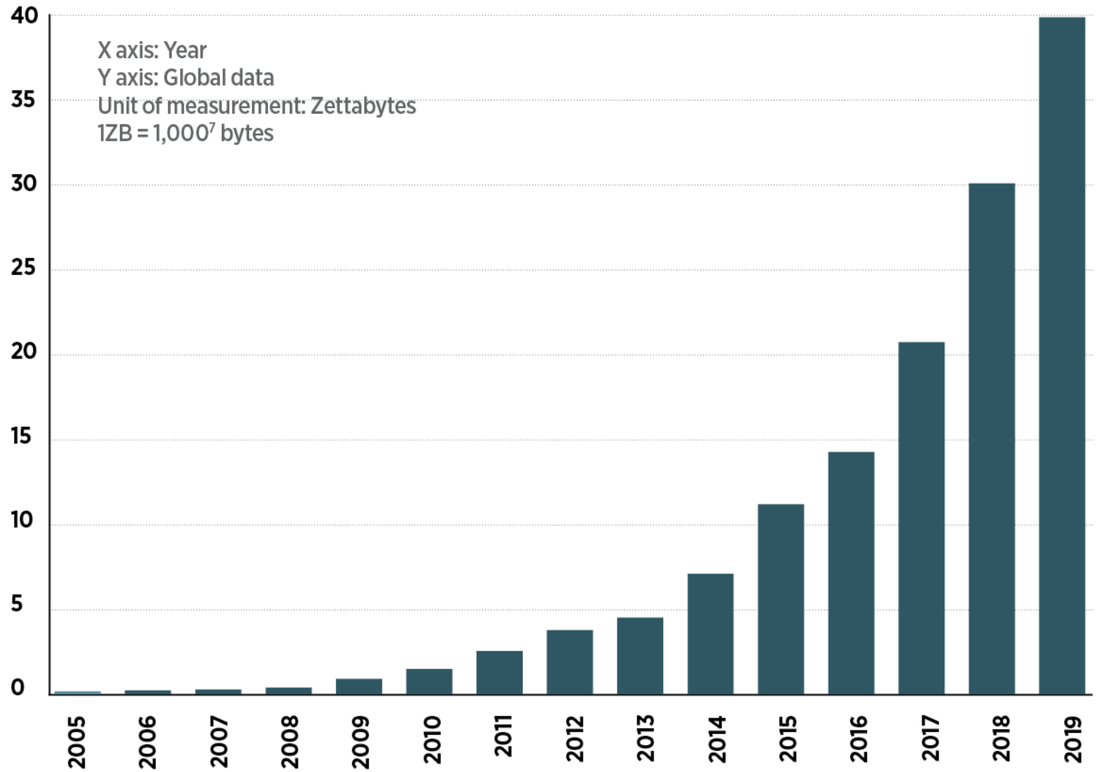
Xueyu Wang

Yuening Tang

Why Machine Learning?

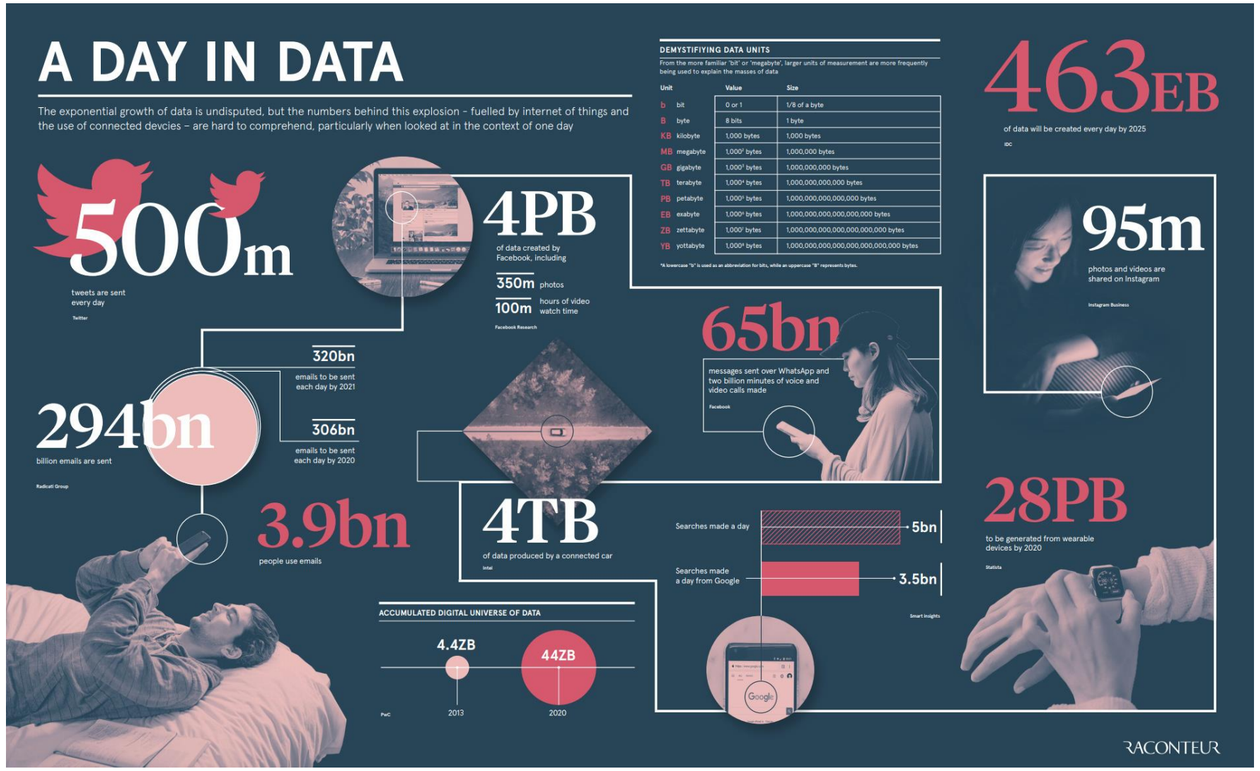
"We are drowning in information but starved for knowledge."

— John Naisbitt



The Booming Age of Data

- 30 trillion web pages
- 500 million tweets per day
- 2.7 billion monthly active users on Facebook
- 1.8 billion images uploaded per day
- 2.9 billion base pairs in human genome



Infographic by Raconteur. Click on the figure to access full size pdf

Has machine learning touched your life today?

- The toothbrush you used this morning
- What you ate for breakfast
- The device you used to track your physical activity
- When you asked your phone to set a timer for your food in the oven
- The message you received from a friend in social media
- Vaccine development and advances in the treatment of COVID19

…machine learning was involved in all of these activities directly or indirectly

Machine learning as a discipline

Study of algorithms that

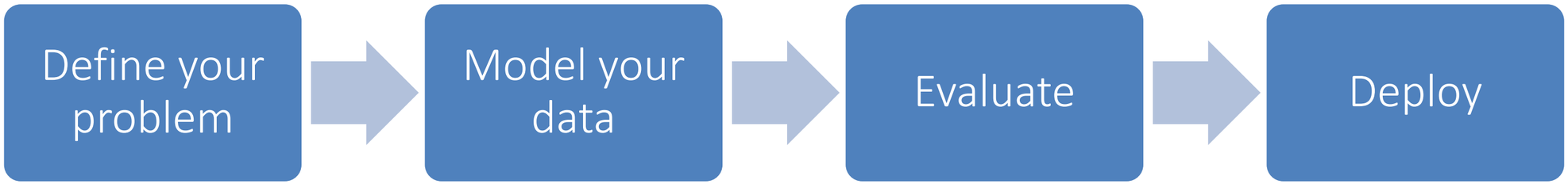
- improve their **performance** P
- at some **task** T
- with **experience** E

well-defined learning task: <P, T, E>

— Tom Mitchell

Machine learning in practice

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



Course objectives

- Introduce you to the machine learning **workflow**
- Develop deep understanding of major machine learning **algorithms**
- Learn how to **apply tools for real-data** analysis problems
- Create **effective visualizations** for data modeling
- Improve your written and verbal **communication skills**
- Experience **teamwork in a remote environment**
- Motivate you to do **research in data science** and machine learning

Our approach to this course

Lectures

Introduce the key concepts that will enable you to develop a basic understanding of how different algorithms operate and how to measure their performance

Assignments

Reinforce concepts, deepen your understanding of topics covered in class and expand your knowledge on related techniques, enhance your debugging skills

Project

Give you an opportunity to expand beyond topics covered in class and gain hands-on-experience applying machine learning to a real-world problem

Quizzes

Assess your understanding of the materials to improve on course delivery

Prerequisites

- Basic knowledge in probability, statistics, linear algebra and optimization
- Basic programming skills in Python, HTML and LaTeX/Markdown
- *No prior background in machine learning is required*

What have students said about other editions of this course?

- It is like drinking out of a firehose
- There is a lot of math
- Homework is extremely hard
- Put more effort into the course than any other

But wait, before you drop the call, they have also said…

- Helped them get jobs
- Learned an enormous amount
- Improved several skills (coding, etc.)

How to succeed in this course

Sit back, relax and enjoy the class

Of all courses you could take this Fall, you picked this one, and we appreciate your being here! It's time to put your phone away, grab a cup of tea, or coffee, or water, sit back and focus on the class!

Closed mouths do not get fed

The instruction team is here to support you, but we need your help too! If you get stuck on assignments, are struggling with the project or didn't grasp something in class, ask for help!

The early bird gets the worm

Time management is a key aspect of this course, make sure to start early and work consistently on the assignments and the project.

Class participation

Live-lectures

There will be poll questions during class, these are not graded, so please (please!) answer them even if you are not sure whether you are wrong or right

No such thing as a stupid question

I truly mean it. Many times, questions prefaced by “this might be silly […]” reveal important aspects of a problem that we might not have thought or talked about

It takes a village

We might all be in different places but you are not taking this course alone! The best place to ask questions and have discussions about the topics covered in class is the Piazza page. You can even get bonus points for answering your classmates’ questions **(!)**
Finally, **do not email your questions to the instruction team, ask them on Piazza so all students can benefit from the responses**

Office hours

One-on-one with the TAs

You can sign up to 10-minute slots where you can share your screen to discuss your assignments and project more effectively. You are allowed to share your screen. **You may not join other student's appointment***. The schedule is available here.

Open office hours with the instructor

Every Thursday evening from 7:00pm to 8:00pm for questions about topics covered in class, no screen sharing is allowed.

One-on-one with the instructor

If you want to have a one-on-one meeting with the instructor, please make a private note on Piazza and we will schedule an appointment

Grading

- Assignments (50%)
- Project (35%)
  - Proposal (10%)
  - Midterm report (10%)
  - Final report (15%)
- Quizzes (10%)
- Class participation (5%)
- Bonus points (up to 7%)

Assignments

Format

Programming and written portions. The former involve implementation of algorithms covered in class from scratch. You will not be allowed to use machine learning libraries available in Python.

Logistics

You will submit your assignments via Gradescope, following the instructions that accompany each one of them. The programming portions are autograded and you can resubmit as many times as you like

Late submission and collaboration policies

Assignments submitted **after the deadline will receive zero credit**. You are **not allowed to collaborate on the assignments** and the work you submit should be yours only. Use Piazza for questions!

Bonus points

Every assignment will have bonus questions that count toward the overall bonus points grade. You can also obtain bonus points from the “early bird specials”

Project

Expectations

You will perform an independent study on a machine learning task of your choosing as a team (exactly 4 people). You should not be tied uniquely to the topics covered in class and aim at expanding your knowledge in ML. You are allowed to use any library and software you would like

Guidance

You will be assigned a mentor who will be your point of contact for most questions related to the project. They will lead the touchpoint discussions and help you should you find yourself unable to make progress independently

Seminars

The instruction team will offer seminars at the beginning of the semester to talk about projects they have worked on and inspire you in deciding what topics to work on

Project

Deliverables

All your deliverables will be submitted as a link to your project webpage via Gradescope. This includes your proposal, midterm progress report and final project. Request your Github education account as soon as possible following this link.

Touchpoints

There will be three touchpoints during the semester, each of them before the deadline of your deliverables so you can incorporate feedback into your submission. You will meet via BlueJeans with your mentor and other teams working on similar topics to discuss your progress and learn from your peers. In addition to these remote touchpoints the instructor will hold an in-person version of the event to which you will register ahead of time. Instructions on the submissions before the touchpoints are will be available on the course webpage this coming week.

Quizzes

Format

Five multiple choice questions about the content covered in class, almost every Friday. The quiz will be available from 6:00am EST until the following day at 6:00am EST. Once you start the quiz you will have 7 minutes to complete it. Sets of questions are randomly assigned to each student.

Grading

We will count only your top 10 quizzes towards your grade

Syllabus

Part I: Basic math for computational data analysis

Probability and statistics, linear algebra, optimization

Part II: Unsupervised learning for data exploration

Clustering analysis, dimensionality reduction, density estimation

Part III: Supervised learning for predictive analysis

Linear classification/regression, tree-based models, sparse kernel methods, neural networks

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)

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

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
Multi-task learning
Transfer learning
Learning in NLP, CV, Robotics, …

2020s

What will your contribution be?

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 $Y_{n\times 1}$

Unsupervised Learning

Clustering Analysis

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

Dimensionality reduction

- Principal component analysis
- Manifold learning

Probability distributions

- 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?

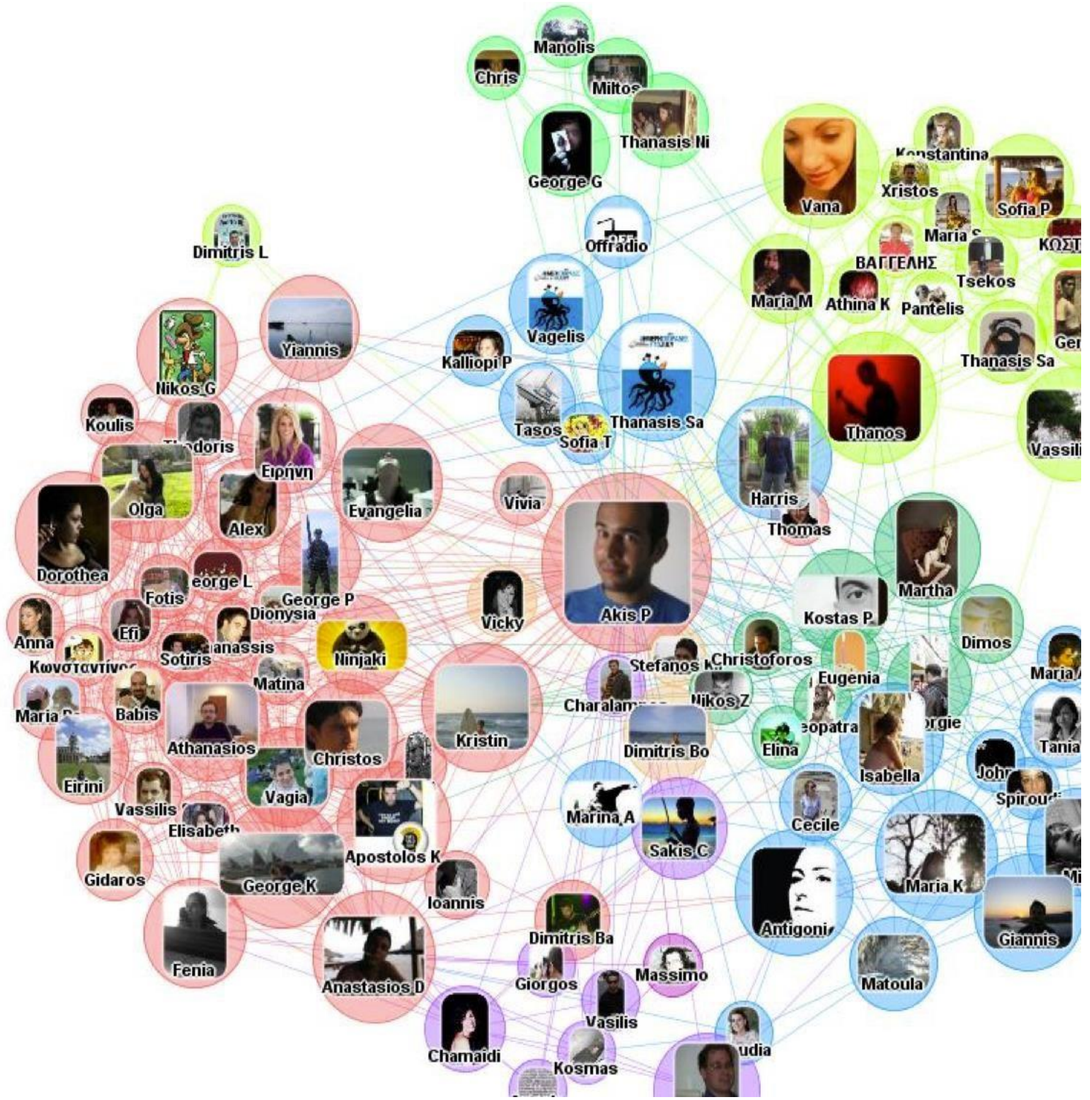
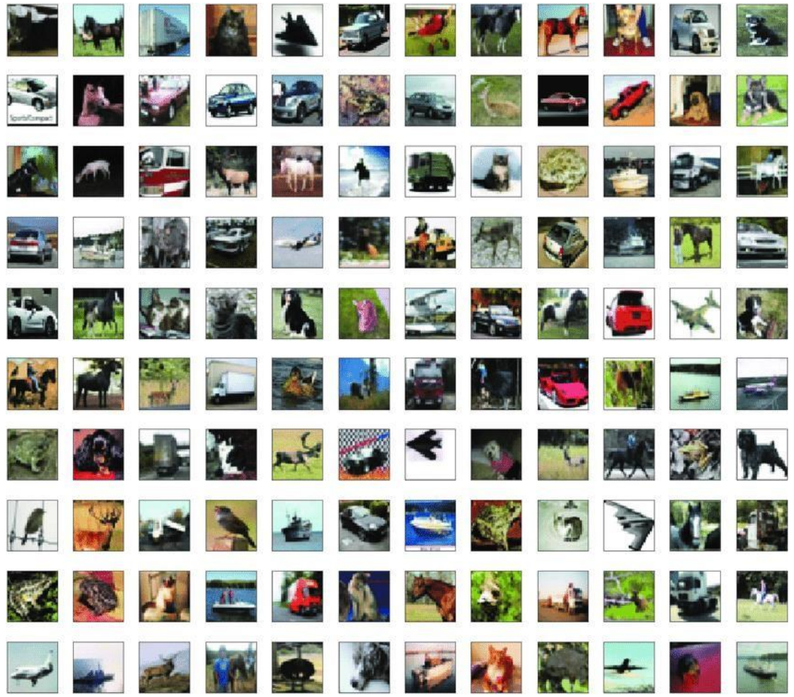


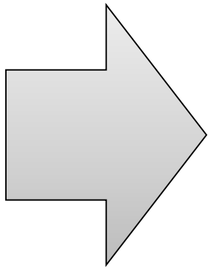
Image credit: Symeon Papadopoulos

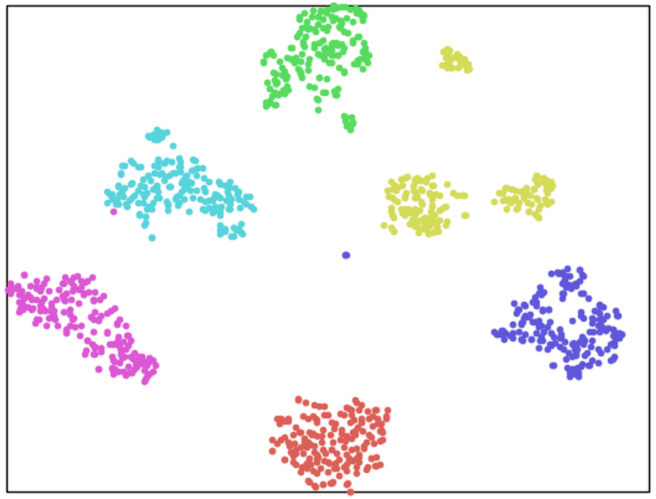
Examples of unsupervised learning

- Clothing clustering
- Wearable Sensors
- The Role of Clustering in Dynamic Functional Network Connectivity for Schizophrenia Diagnosis

Dimensionality reduction







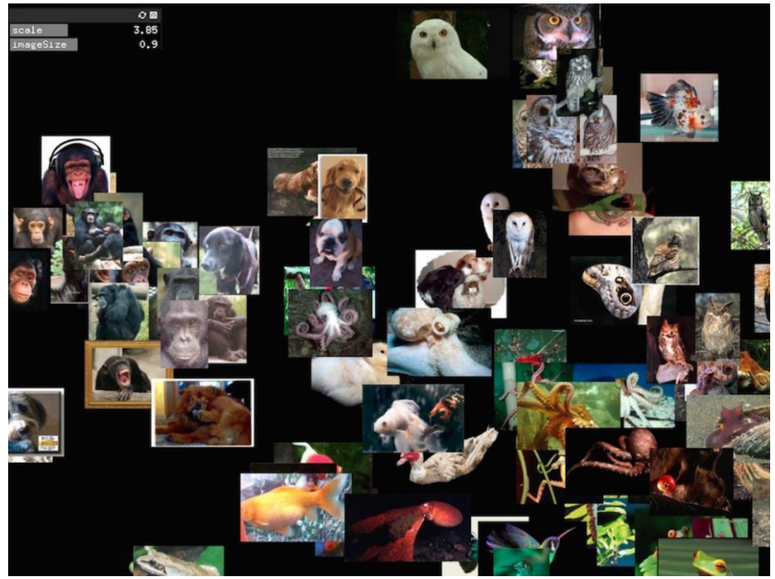


Image credit: Gene Kogan

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

Supervised Learning

Tree-based models

- Decision tree
- Ensemble learning/Random forest

Linear classification/regression models

- Linear regression
- Naive Bayes
- Logistic regression

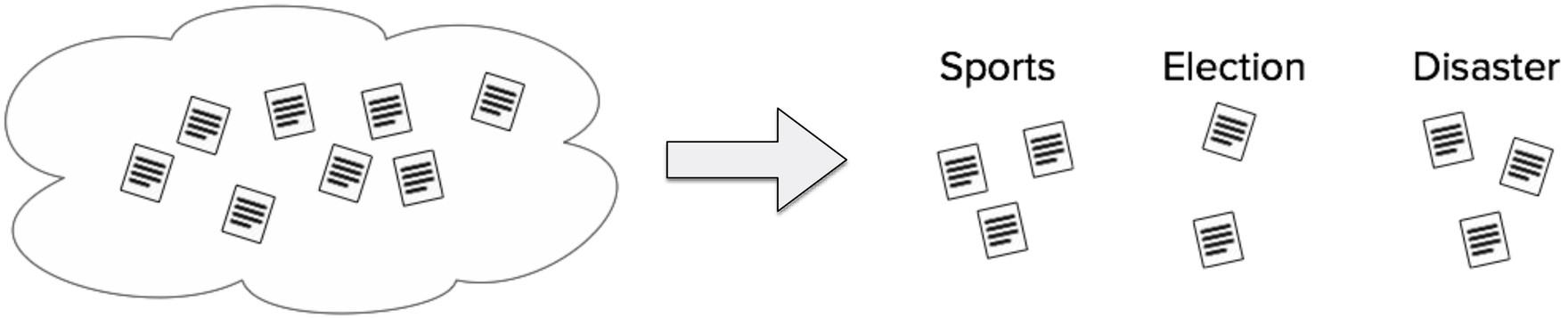
Support vector machine

Neural networks

- Feedforward neural networks and backpropagation analysis
- Convolutional neural networks

News classification

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



Email classification

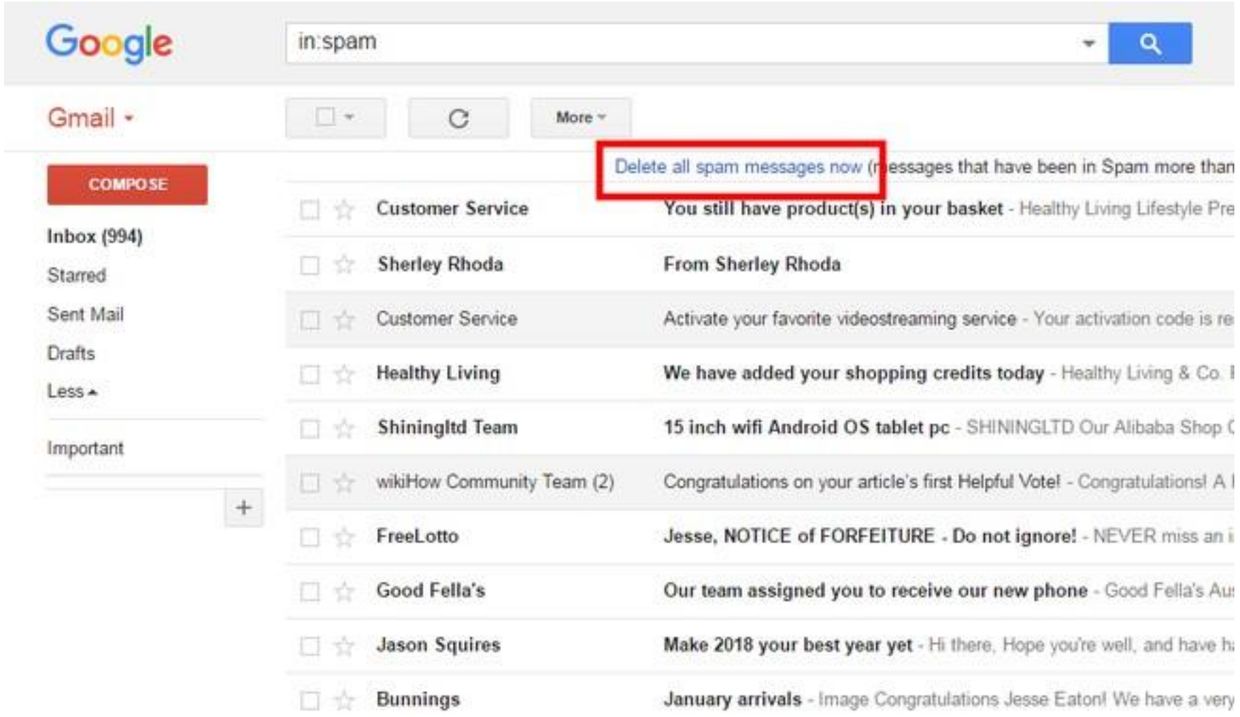
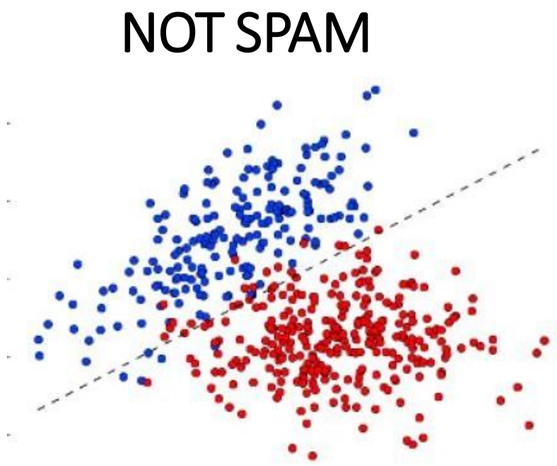


Image credit: WikiHow



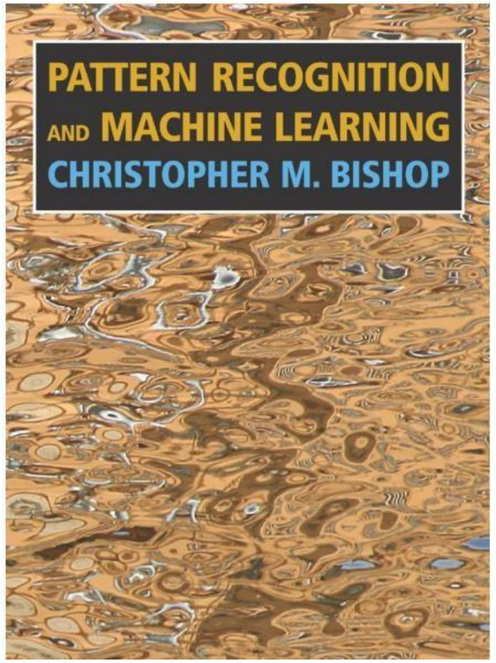
- What are the inputs and how to represent them?
- What are the desired outputs?
- What learning algorithms to choose?

Examples of supervised learning

- Font teller
- Root Inspired Anchor Model Project
- Music Generation using Machine Learning
- Prediction-of-Hard-Drive-Failure

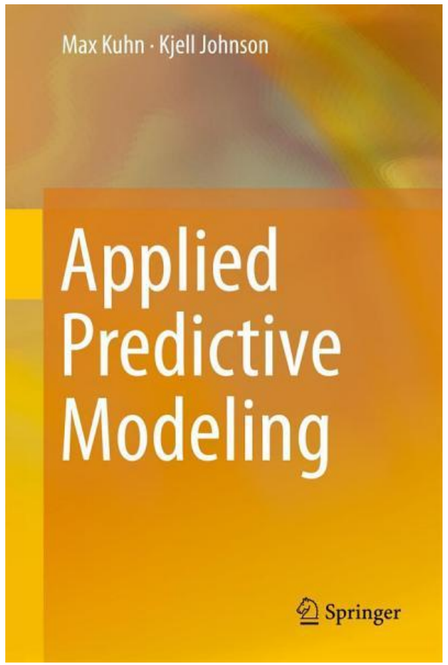
Resources

Lectures and assignments



[Pattern Recognition and Machine Learning, by Chris Bishop]

Project



[Applied Predictive modeling, by Max Kuhn and Kjell Johnson]

Questions?