What is Machine Learning?

Our goal today

And through the semester

What is (machine) learning?

Let’s play a game

The badges game

Attendees of the 1994 version of the International Conference on Machine Learning (ICML) received conference badges labeled

+ or —

Only one person (Haym Hirsh) knew the function that generated the labels

Depended *only* on the attendee's name

The task for the attendees: Look at as many examples as you want in the conference and find the hidden function

Some example rules

```
If the second letter of the first name is a vowel:
   label = +
 else
   label = -
```

```
If the first name is longer than the last name:
   label = +
 else
   label = -
```

Let’s play

Name	Label
Claire Cardie	+
Peter Bartlett	–
Eric Baum	?
Haym Hirsh	?
Leslie Pack Kaelbling	?
Carla E. Brodley	?

Let’s play

Name	Label
Claire Cardie	+
Peter Bartlett	–
Eric Baum	?
Haym Hirsh	?
Leslie Pack Kaelbling	?
Carla E. Brodley	?

How were the labels generated?

What is the label for *Indiana Jones*?

Let’s play

Name	Label
Claire Cardie	+
Peter Bartlett	−
Eric Baum	+
Haym Hirsh	−
Leslie Pack Kaelbling	−
Carla E. Brodley	+

How were the labels generated?

What is the label for *Indiana Jones*?

Can you guess the label for my name? Yours?

Let's play

Name	Label
Claire Cardie	+
Peter Bartlett	−
Eric Baum	+
Haym Hirsh	−
Leslie Pack Kaelbling	−
Carla E. Brodley	+

Full data on the class website.

Take a look at it to guess how the names were labeled

How were the labels generated?

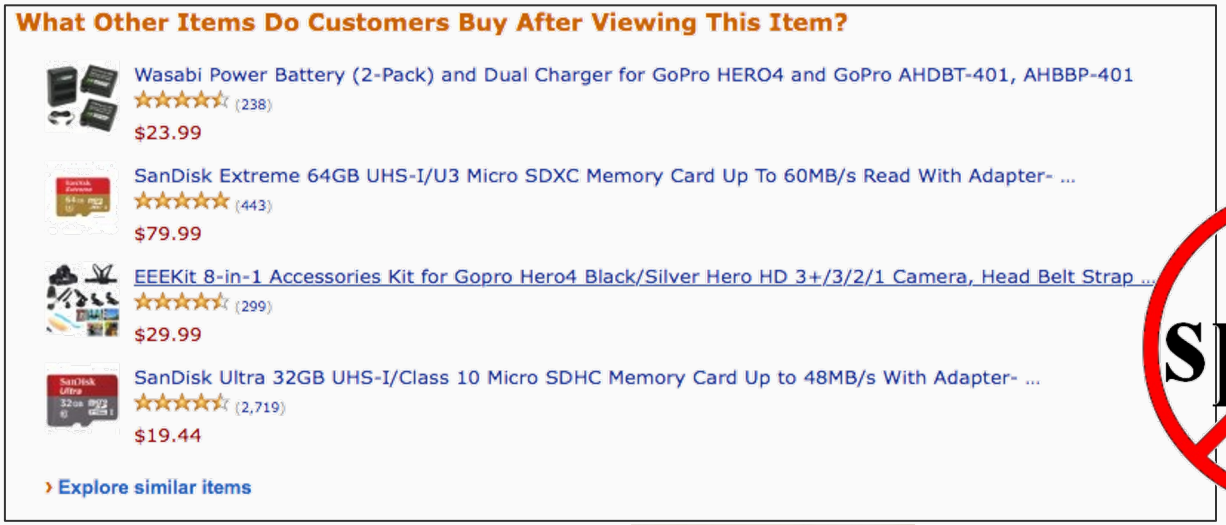
What is the label for *Indiana Jones*?

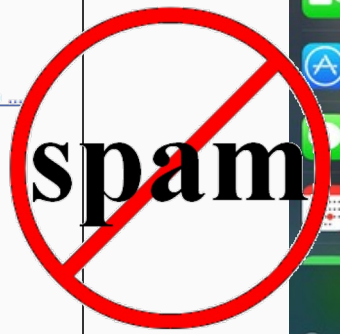
Can you guess the label for my name? Yours?

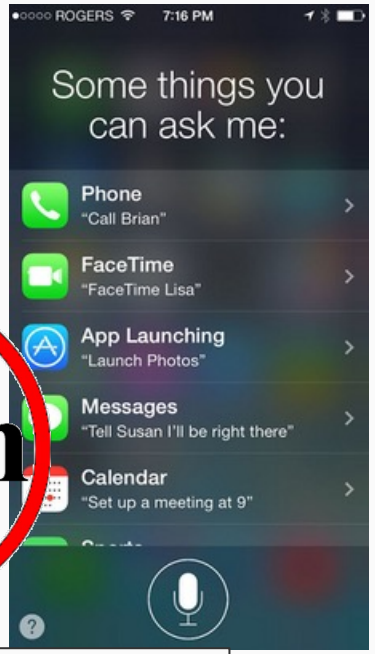
What is machine learning?

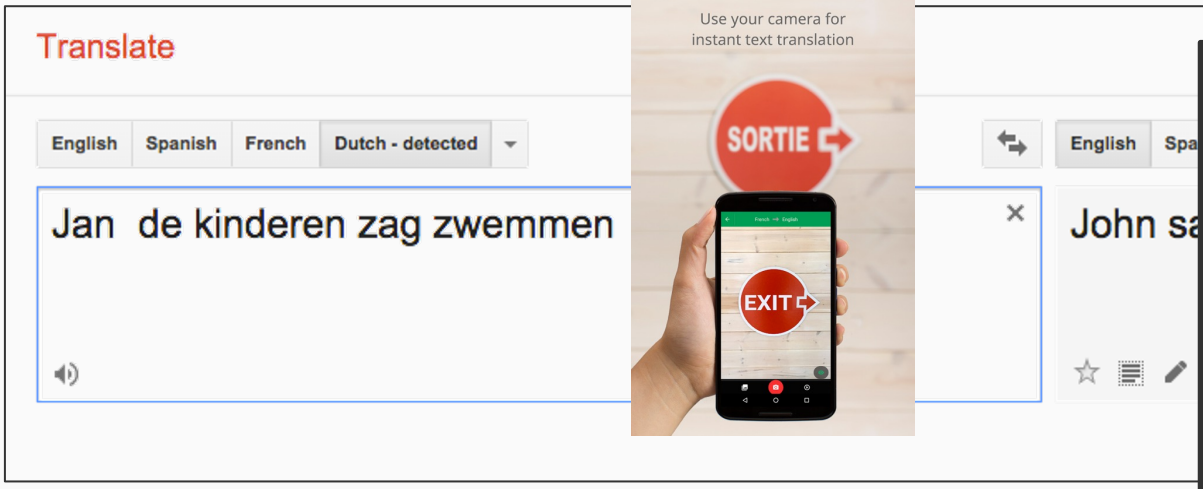
Machine learning is everywhere!

And you are probably already using it









Machine learning is everywhere!

And you are probably already impacted by it

- Is an email spam?
- Find all the people in this photo
- If I like these three movies, what should I watch next?
- Based on your purchase history, you might be interested in…
- Will a stock price go up or down tomorrow? By how much?
- Handwriting recognition
- What are the best ads to place on this website?
- I would like to read that Dutch website in English
- Does this genetic marker correspond to Alzheimer's disease?
- What is the 3 dimensional structure of this protein?

But what is learning?

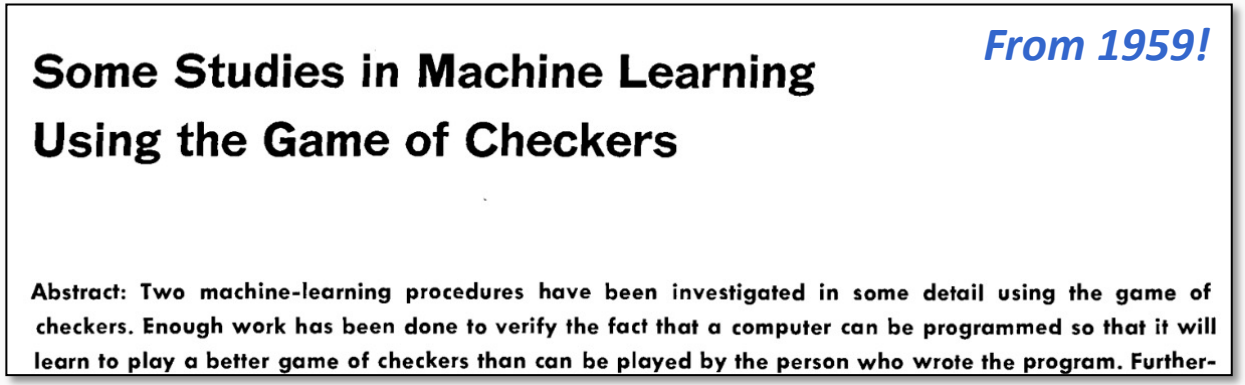
Let’s try to define (machine) learning

What is machine learning?

"Programming computers to learn from experience should eventually eliminate the need for much [...] programming effort."

"As a result of these experiments one can say with some certainty [...] such learning schemes may eventually be economically feasible as applied to real-life problems."

Arthur Samuel



Some Studies in Machine Learning Using the Game of Checkers

From 1959!

Abstract: Two machine-learning procedures have been investigated in some detail using the game of checkers. Enough work has been done to verify the fact that a computer can be programmed so that it will learn to play a better game of checkers than can be played by the person who wrote the program. Further-

Talks about the differences between *rote learning* and *generalization*

Samuel, Arthur L. 1959. "Some Studies in Machine Learning Using the Game of Checkers." *IBM Journal of Research and Development* 3 (3): 210–29.

Learning as generalization

"Learning denotes changes in the system that are adaptive in the sense that they enable the system to do the task or tasks drawn from the same population more effectively the next time."

Herbert Simon (1983)

"Old programs do not learn, they simply fade away."

Economist, psychologist, political scientist, computer scientist, sociologist, Nobel Prize (1978), Turing Award (1975)…

Simon, Herbert A. 1983. "Why Should Machines Learn?" In Machine Learning: An Artificial Intelligence Approach, 25–37.

Learning as generalization

"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**."

Tom Mitchell (1997)

Mitchell, Tom M. 1997. *Machine Learning*. McGraw-Hill.

Learning = generalization

Learning = generalization

Machine learning is the future

- Gives a system the ability to perform a task in a situation which has never been encountered before
  - New way to think about programming
  - Programs that can acquire new capabilities!
- Learning allows programs to interact more robustly with messy data
- Has made inroads into user facing applications

Related fields All very active research areas!

- Artificial intelligence: Computers that are as intelligent as humans
  - Machine learning closely tied to AI
- Theoretical CS and mathematics
  - Formalizing and understanding learning mathematically
  - Uses ideas from probability and statistics, linear algebra, theory of computation
- Philosophy, cognitive psychology, neuroscience, linguistics,…
- Many, many application areas
  - AI, medicine, engineering, psychology, marketing,…

Reflected in the diversity in this class and other classes like this!

Overview of this course

The main question through the semester

What is learning?

Different **formal** answers to this problem will give us:

Various families of learning algorithms

Techniques for developing new learning algorithms

We will see…

1. Different kinds of models
2. Different learning protocols
3. Learning algorithms
4. Computational learning theory
5. Representing data

We will see different “models”

Or: functions that a learner learns

- Decision trees
- Linear classifiers, linear regressors
- Non-linear classifiers, neural networks, kernels (if time permits)
- Ensembles of classifiers

Different learning protocols

- Supervised learning
  - A *teacher* supplies a collection of examples with labels
  - The *learner* has to learn to label new examples using this data
- Unsupervised learning
  - No *teacher*, *learner* has only unlabeled examples
  - Data mining
- Semi-supervised learning
  - *Learner* has access to both labeled and unlabeled examples
- Active learning
  - *Learner* and *teacher* interact with each other
  - *Learner* can ask questions
- Reinforcement learning
  - Learner learns by interacting with the environment

Different learning protocols

- Supervised learning
  - A *teacher* supplies a collection of examples with labels
  - The *learner* has to learn to label new examples using this data
- Unsupervised learning
  - No *teacher*, *learner* has only unlabeled examples
  - Data mining
- Semi-supervised learning
  - *Learner* has access to both labeled and unlabeled examples
- Active learning
  - *Learner* and *teacher* interact with each other
  - *Learner* can ask questions
- Reinforcement learning
  - Learner learns by interacting with the environment

Who has seen or used supervised learning before in some capacity?

Learning algorithms

- Online algorithms: Learner can access only one labeled at a time
  - Perceptron, Winnow
- Batch algorithms: Learner can access to the entire dataset
  - Naïve Bayes
  - Support vector machines, logistic regression, neural networks
  - Decision trees and nearest neighbors
  - Boosting
- Unsupervised/semi-supervised algorithms
  - Expectation maximization
  - K-Means

Learning algorithms

- Online algorithms: Learner can access only one labeled at a time
  - Perceptron, Winnow
- Batch algorithms: Learner can access to the entire dataset
  - Naïve Bayes
  - Support vector machines, logistic regression, neural networks
  - Decision trees and nearest neighbors
  - Boosting
- Unsupervised/semi-supervised algorithms
  - Expectation maximization
  - K-Means

Some of you may have used some of these algorithms as black boxes in the past

Representing data

What is the best way to represent data for a particular task?

- The importance of the right features
- Learning such features from data
- Dimensionality reduction (if time permits)

The theory of machine learning

What does it mean to learn?

- *Online learning*
  - Learner sees examples in a stream and stop making mistakes as we go along (or minimize regret in our decisions).
- *Probably Approximately Correct (PAC) Learning*
  - After seeing a collection of examples, the learner will (with high probability) produce a function that makes small error.
- *Bayesian learning*
  - Based on our observations, what is the probability distribution over possible functions that produced the data?

This course

Focuses on the underlying concepts and algorithmic ideas in the field of machine learning

This course is not about

- Using a specific machine learning tool/framework
- Any single learning paradigm

What will you learn?

1. A broad theoretical and practical understanding of machine learning paradigms and algorithms

2. Ability to implement learning algorithms

3. Identify where machine learning can be applied and make the most appropriate decisions (about algorithms, models, supervision, etc)

How will you learn?

or: Course information