Overview

COMS 4771
What is machine learning?

- **Machine learning**: study of computational mechanisms that “learn” from data to make predictions and decisions.
Application 1: image classification

- Birdwatcher takes photos of birds, organizes by species.

Indigo bunting
Application 1: image classification

- Birdwatcher takes photos of birds, organizes by species.
- **Goal**: automatically recognize bird species in new photos.

![Indigo bunting](image-url)
Application 2: recommender system

- Netflix users watch movies and provide ratings.

(Image credit: Koren, Bell, and Volinsky, 2009.)
Application 2: recommender system

- Netflix users watch movies and provide ratings.
- **Goal**: predict user’s rating of unwatched movie.

(Image credit: Koren, Bell, and Volinsky, 2009.)
Application 2: recommender system

- Netflix users watch movies and provide ratings.
- **Goal**: predict user’s rating of unwatched movie.
- **(Real goal)**: keep users paying customers.

(Image credit: Koren, Bell, and Volinsky, 2009.)
Application 2: recommender system

- Netflix users watch movies and provide ratings.
- **Goal**: predict user’s rating of unwatched movie.
- **(Real goal)**: keep users paying customers.
- **(Real effect)**: reinforce stereotypes found in the data?

(Image credit: Koren, Bell, and Volinsky, 2009.)
Application 3: machine translation

- Linguists provide translations of all English language books into French, sentence-by-sentence.
Application 3: machine translation

- Linguists provide translations of all English language books into French, sentence-by-sentence.
- **Goal**: translate any English sentence into French.

![Translation Example](image_url)
Basic setting: supervised learning

**Training data:** labeled examples

\[(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\]
Basic setting: supervised learning

Training data: labeled examples

\[(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\]

where

- each input \(x_i\) is a machine-readable description of an instance (e.g., image, sentence), and
Basic setting: supervised learning

**Training data:** labeled examples

\[(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\]

where

- each input \(x_i\) is a machine-readable description of an instance (e.g., image, sentence), and
- each corresponding label \(y_i\) is an annotation relevant to the task—typically not easy to automatically obtain.
Basic setting: supervised learning

**Training data:** labeled examples

\[(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\]

where

- each **input** \(x_i\) is a machine-readable description of an instance (e.g., image, sentence), and

- each corresponding **label** \(y_i\) is an annotation relevant to the task—typically not easy to automatically obtain.

**Goal:** learn a function \(\hat{f}\) from labeled examples, that accurately “predicts” the labels of **new (previously unseen)** inputs.
Learn a function?

The learned function $\hat{f}$ might look like the following:

```plaintext
1: if age $\geq$ 40 then
2:   if genre = western then
3:     return 4.3
4:   else if release date $>$ 1998 then
5:     return 2.5
6:   else
7:     ...
8:   end if
9: else if ... then
10: ...
11: end if
```
Learn a function?

The learned function $\hat{f}$ might look like the following:

```
1: if age $\geq$ 40 then
2:   if genre = western then
3:     return 4.3
4:   else if release date $> 1998$ then
5:     return 2.5
6:   else
7:     ...
8: end if
9: else if ... then
10: ...
11: end if
```

Want machine to figure out these if-else clauses from the data, rather than hand-code them ourselves.
Other kinds of functions

"Perceptron" (linear classifier)

1: if \( 0.335 \cdot x_1 + 2.5 \cdot x_2 + \cdots + 6.35 \cdot x_{10} > 4.3 \) then

2: return spam

3: else

4: return not spam

5: end if

"Neural network" (non-linear classifier)

and many others!
Other kinds of functions

“Perceptron” (linear classifier)

1: if \(0.335 \cdot x_1 + 2.5 \cdot x_2 + \cdots + 6.35 \cdot x_{10^6} > 4.3\) then
2: return spam
3: else
4: return not spam
5: end if
Other kinds of functions

“Perceptron” (linear classifier)

1: if $0.335 \cdot x_1 + 2.5 \cdot x_2 + \cdots + 6.35 \cdot x_{10^6} > 4.3$ then
2: return spam
3: else
4: return not spam
5: end if

“Neural network” (non-linear classifier)
Other kinds of functions

“Perceptron” (linear classifier)

\[\text{if } 0.335 \cdot x_1 + 2.5 \cdot x_2 + \cdots + 6.35 \cdot x_{10^6} > 4.3 \text{ then} \]

2: \text{ return spam}

3: \text{ else}

4: \text{ return not spam}

5: \text{ end if}

“Neural network” (non-linear classifier)

... and many others!
Why is machine learning challenging?

**Task**: given image, predict if it depicts indigo bunting or not.
Why is machine learning challenging?

**Task**: given image, predict if it depicts indigo bunting or not.

Why is this hard?
Why is machine learning challenging?

**Task:** given image, predict if it depicts indigo bunting or not.

**Why is this hard?**

1. Data sparsity.
Why is machine learning challenging?

Task: given image, predict if it depicts indigo bunting or not.

Why is this hard?

1. Data sparsity.
2. Complex relationship between input and label.
Why is machine learning challenging?

**Task**: given image, predict if it depicts indigo bunting or not.

**Why is this hard?**

1. Data sparsity.
2. Complex relationship between input and label.
3. Too many possible functions!
This course (COMS 4771)

Topics

- Principles of *supervised machine learning*
  - Algorithmic and statistical techniques
  - Some well-weathered algorithms and models

Course website and syllabus

http://www.cs.columbia.edu/~djhsu/coms4771-f18/

- Prerequisites, course requirements (homework, exams), rules of conduct.
- **You must read, understand, and abide by all course policies.**

Course staff

- **Instructor**: Prof. Daniel Hsu
- **Teaching assistants, office hours, online forum**: see course website
Announcements

1. First homework assignment ("Homework 0") due Friday. Required; submit on Gradescope even if you are not enrolled yet!

2. Our classroom is also being used for some talks that happen immediately after our lecture, and the speakers need some setup time. ▶ Consequence: we’ll end lecture 10 minutes early on some dates. First one: Sept 12, 11:40am.

3. Computer Science Distinguished Lecture on Sept 10, 11:30am, in Davis Auditorium. Pieter Abbeel (UC Berkeley), on “Deep Learning to Learn”
1. First homework assignment ("Homework 0") due Friday.

**Required;** submit on Gradescope even if you are not enrolled yet!
1. First homework assignment ("Homework 0") due Friday.  
   **Required**: submit on Gradescope even if you are not enrolled yet!

2. Our classroom is also being used for some talks that happen immediately after our lecture, and the speakers need some setup time.
   
   ▶ **Consequence**: we’ll end lecture 10 minutes early on some dates.
   
   First one: Sept 12, 11:40am.

   David Knowles (New York Genome Center / Columbia), on “Probabilistic models of transcriptomic (dys)regulation in human disease”
1. First homework assignment ("Homework 0") due Friday.  
   **Required**: submit on Gradescope even if you are not enrolled yet!

2. Our classroom is also being used for some talks that happen immediately after our lecture, and the speakers need some setup time.  
   ▶ **Consequence**: we’ll end lecture 10 minutes early on some dates.  
   
   First one: Sept 12, 11:40am.  
   
   David Knowles (New York Genome Center / Columbia), on “Probabilistic models of transcriptomic (dys)regulation in human disease”

3. Computer Science Distinguished Lecture on Sept 10, 11:30am, in Davis Auditorium.  
   
   Pieter Abbeel (UC Berkeley), on “Deep Learning to Learn”
1. Examples of machine learning problems and why they are challenging.
2. Course information.