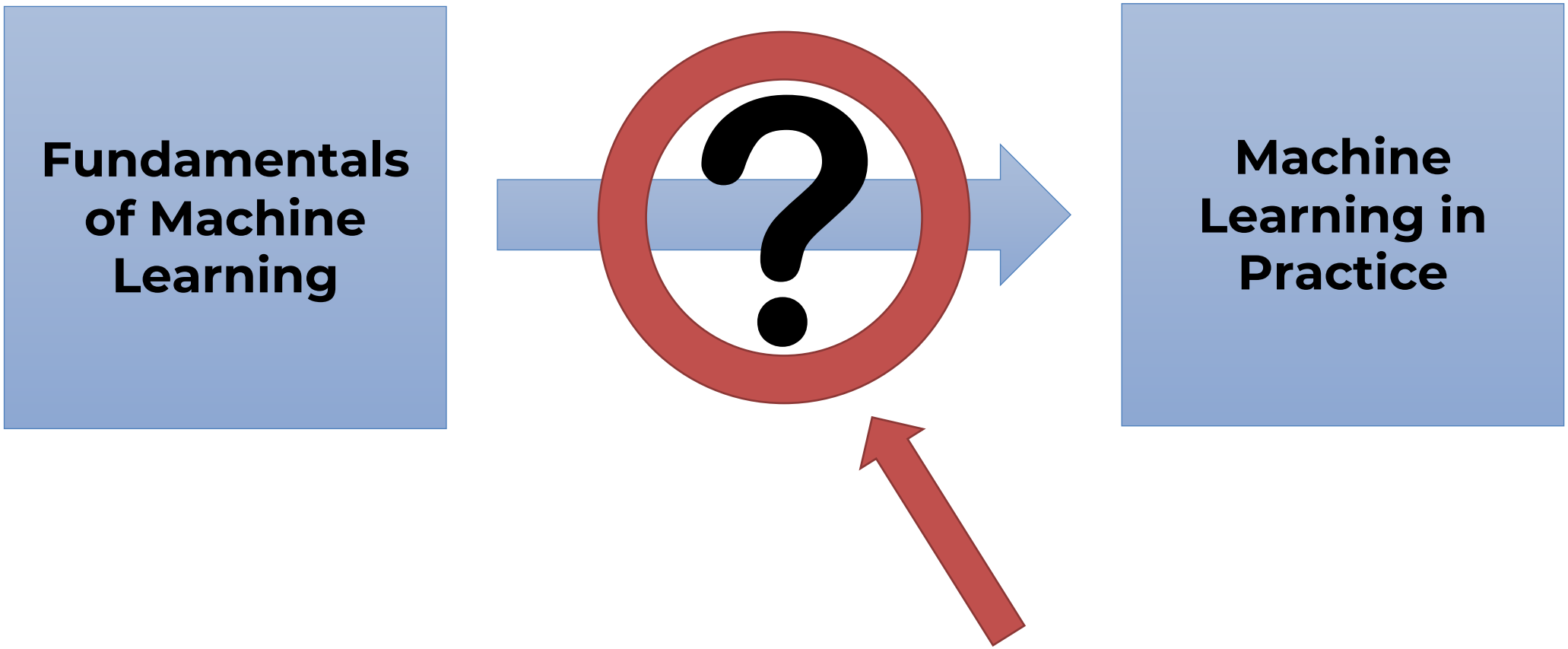
CS6787: Advanced Machine Learning Systems

CS6787 Lecture 1 — Spring 2024


Fundamentals of Machine Learning
Machine Learning in Practice
this course

What's missing in the basic stuff?

Efficiency!
Scalability!

Motivation:

Machine learning applications involve large amounts of data

More data → Better services

Better systems → More data

How do practitioners make their systems better?

How do we improve our systems?

Course outline

- Build frameworks/software that make it easy to express & train a machine learning/deep learning model. **Part 1**
- Use methods for accelerating convergence of learning algorithms — learn in fewer iterations. **Part 2**
- Automatically configure learning systems by using hyperparameter optimization **Part 3**
- Use large pre-trained models to improve performance of downstream tasks — “foundation models” **Part 4**
- Use methods for improving hardware efficiency — run each iteration faster, with less energy, etc. **Part 5**

Course Format

One half

Traditional lectures

Broad description of techniques

One half

Important papers

Presentations by **you**

In-class discussions

Reviews of each paper

Prerequisites

- Basic ML knowledge (CS 4780)
- Math/statistics knowledge
  - At the level of the entrance exam for CS 4780
- Also useful, but not a prerequisite:
  - Knowledge of computer systems, computer hardware, NLP, and computer vision

Grading

- Paper presentations
- Discussion participation
- Paper reviews
- Programming assignments
- Final project

Paper presentations

- Papers listed on the website
  - 20-minute presentation slot for each paper
  - Presenting in groups of two-to-three

- Send any papers you want to suggest we cover **by noon tomorrow (Tuesday)**

- **Signups by Thursday!**
  - Survey will be sent out tomorrow (Tuesday)

- **Learning goal:**
  - **Practice digesting, unpacking, and talking about other people's work**

Paper Reading and Discussion

- Each presentation is followed by a period of questions and breakout discussion

- Please **read at least one of the papers** before class
  - And at least skim the other paper, so you know what to expect

- Note: grade is not for attendance, but rather on participation and bringing insightful ideas to the table

- **Learning goal: practice how to deeply read and critique a paper in context**

Paper Reviews

- For each class period, **submit a mock review** of one of the two papers
  - (Only if you are not presenting.)
- Review the paper as if you were doing peer review on a newly submitted work
- Reviews due a few days after our in-class discussion
- **Learning goal: build technical reading and writing skills, and get some sense of how peer review works.**

Programming Assignments

- Two short assignments in the first part of the semester only

- **Learning goal: become familiar with ML frameworks/tools**
  - **…and the principles that underlie them**
  - **This will build skills for the final project**
  - **Especially useful for folks from non-CS background**

Final Project

- **Open-ended**: work on what you think is interesting!
  - Learning goal: **do a bit of non-trivial research on your own**
- Groups of **up to three**
- Your proposed project must include:
  - The **implementation** of a machine learning system for some task
  - Exploring one or more of the **techniques discussed in the course**
  - To **empirically evaluate performance** and compare with a baseline, using both a ML-side and systems-side metric

Late Policy

- This is a graduate level course
- Two free late days for each of the paper reviews and programming assignments
- No late days on the final project
  - To make things easy on the graders
- No late days on the presentations (for obvious reasons)

Questions?

Today's Topic

Stochastic Gradient Descent: The Workhorse of Machine Learning

CS6787 Lecture 1 — Spring 2024

But first…an icebreaker activity!

For each person:

- What is your name?
- What are you studying?
- **What do you hope to learn from CS6787?**

Then discuss together:

Why do we use stochastic gradient descent?

(And its related algorithms: Adam, AdaGrad, etc.)

Today's Topic

Stochastic Gradient Descent: The Workhorse of Machine Learning

CS6787 Lecture 1 — Spring 2024

On the board

Demo

Stochastic gradient descent is super popular.

What Does SGD Power?

- Everything!

But how SGD is implemented in practice is not exactly what I've just shown you...

...and we'll see how it's different in the upcoming lectures.

To Do

- If you have any papers you particularly want us to cover or topics you think might be interesting, send me an email before noon-ish tomorrow.

- Be on the lookout for an email with the paper presentation signup survey.