Intro to Deep Learning

University of California, Riverside
Instructor: Yinglun Zhu

Some slides are adapted/borrowed from materials developed by Prof. Samet Oymak.

Course Overview

Course purpose: Covering the fundamental ideas behind deep learning and provides an exposition to contemporary machine learning algorithms.

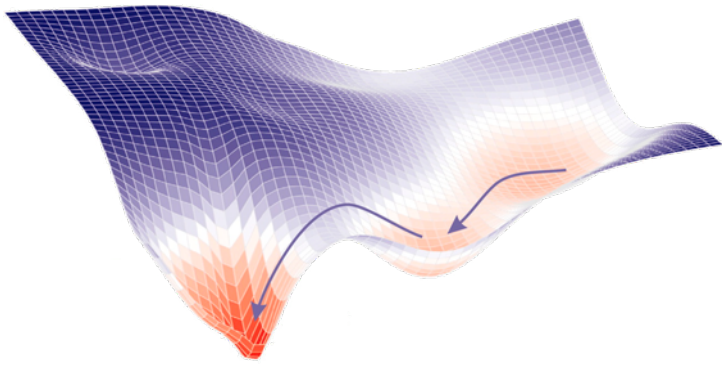
Course Topics

Machine Learning

Course Topics

Machine Learning

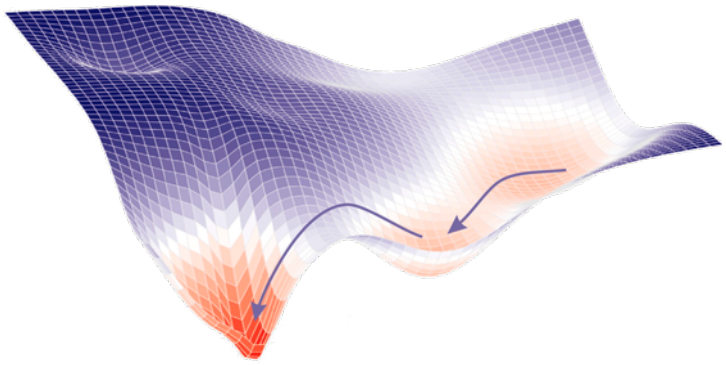
Optimization



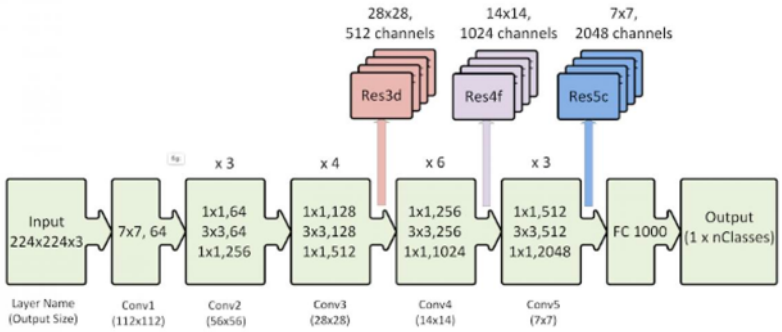
Course Topics

Machine Learning

Optimization



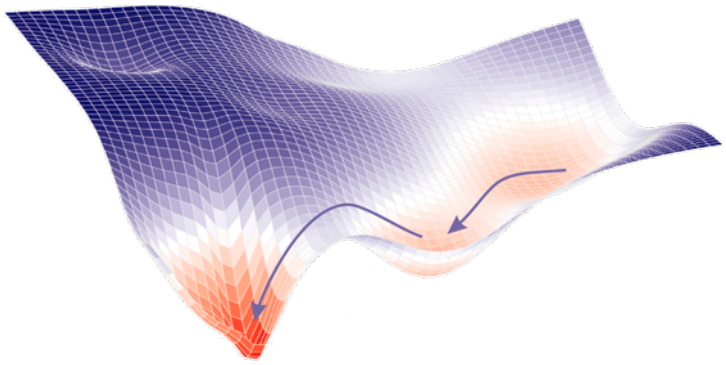
Models



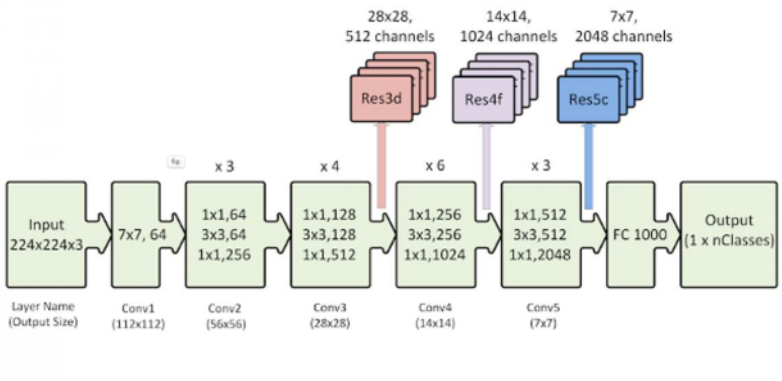
Course Topics

Machine Learning

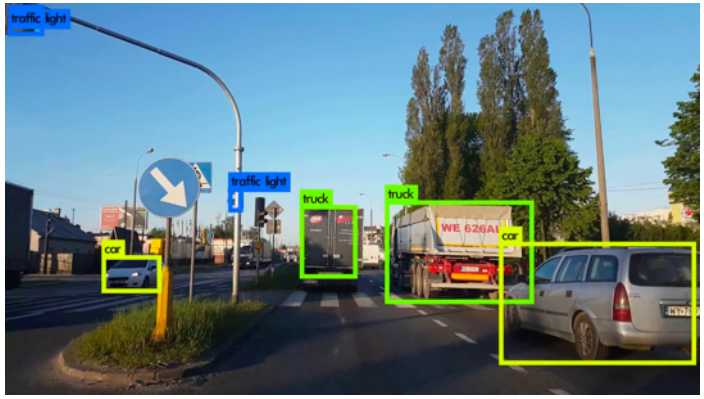
Optimization



Models



Applications



Gradings

a) **Assignments:** 30% (4-5 in total)

b) **Midterm exam:** 30% (TBD)

c) **Final project:** 40% (GPT help allowed, **but need to specify**)

- Proposal 5%
- Presentation + Demo 15%
- Report 20%

Late Policy

Assignments:

- Each late day leads to 20% grade off on top of your received grade
- You will receive a 0 grade if your submission is late for 3 or more days
- Late days are rounded up, e.g., 28 hours late accumulated 2 late days

Proposal/Report: Late submission is not allowed

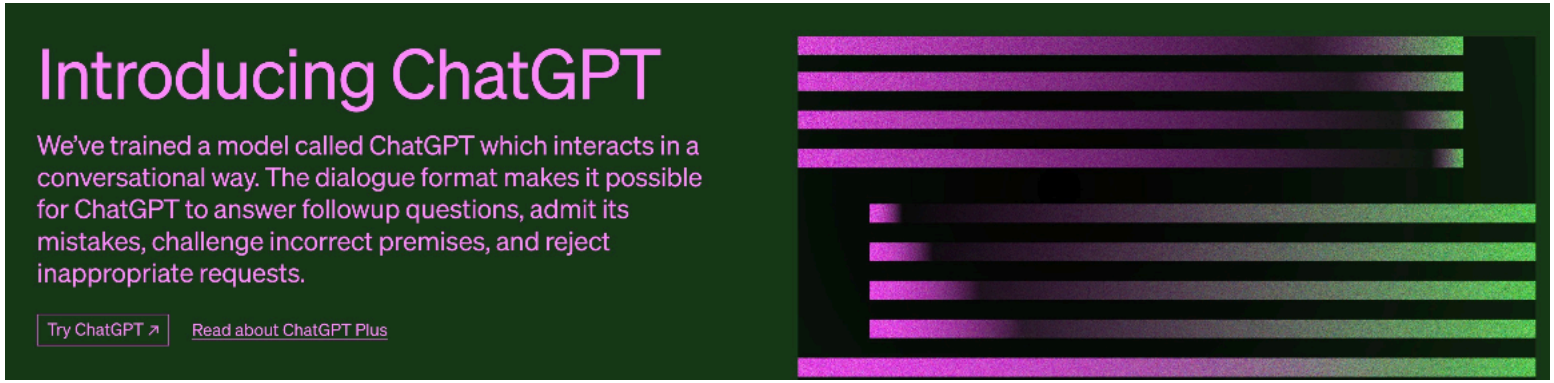
Course Project

The final project is expected to demonstrate a non-trivial application of deep learning techniques. The project will be done in groups and the group sizes should be **3-5**. The project topics are flexible however project has to include software implementation and demo.

- Students can come up with a new application idea and develop the associated software product by designing a deep learning model.
- Students can design a new algorithm improving on state-of-the-art techniques and demonstrate its benefit on machine learning tasks.

For both directions, you need to compare your approach against existing methods.

Example: Sentiment analysis w/ pretrained language models



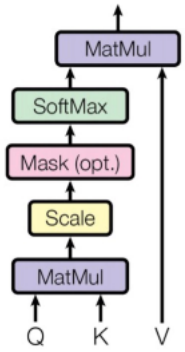
Course Project

The final project is expected to demonstrate a non-trivial application of deep learning techniques. The project will be done in groups and the group sizes should be **3-5**. The project topics are flexible however project has to include software implementation and demo.

- Students can come up with a new application idea and develop the associated software product by designing a deep learning model.
- Students can design a new algorithm improving on state-of-the-art techniques and demonstrate its benefit on machine learning tasks.

For both directions, you need to compare your approach against existing methods.

Example: Accelerate the computation of attention blocks



Attention computation:
$\text{softmax}(Q \cdot K^T) \cdot V$

$O(n^2)$ computation
n = token length

Low-rank approximation?

Sparse approximation?

Project Deliverables

- Project proposal is 1 page. Instructor/TA will provide feedback.
- Week 10 will be dedicated to presentations. Each presentation is expected to be around 10 to 15 minutes depending on the number of groups.
- Final Report will describe the design of the project, and detail the intellectual and practical merit; final report should be 6-8 pages long (excluding appendices).
- Both proposal and final report should follow NeurIPS 2024 format (see Syllabus for links).
  - Educate yourself about how to use LaTex (e.g., Overleaf)!

Prerequisites

This course requires:

- linear algebra
- multivariable calculus
- probability
- experience with Python

Prerequisites

This course requires:

- linear algebra
- multivariable calculus
- probability
- experience with Python

If you don't know matrix multiplication,
DO NOT enroll ☺

Tentative Schedule

Tentative Class Schedule:

week	Tuesday	Thursday	assigned	due
1	Introduction	Perceptron		
2	Training	Classification	PS1	
3	Optimization I	Optimization II	PS2	
4	Convnets	Architectures	PS3	PS1
5	Optimization III	Overparameterization		PS2, Proposal
6	Generalization	Regularization	PS4	PS3, Midterm
7	Transformers	NLP applications	PS5	
8	Self-supervision	Pre-training		PS4
9	Foundation Models	Recent LLM advances		PS5
10	Presentation I	Presentation II		Final project

Computational resources

- Access to a server
  1. EE server Bender
  2. CS students: Consider CS server
  3. PhD students: Consider using your lab servers
  4. Online resources: Google Colab

Go to TAs' office hour!

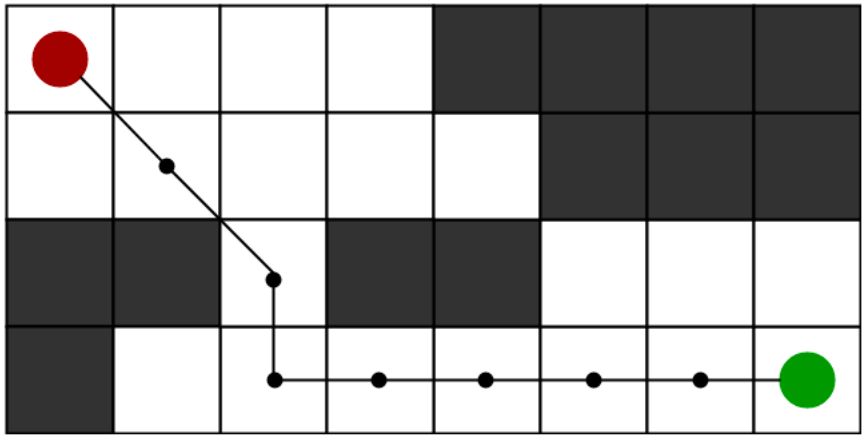
with some background...

Artificial Intelligence

- From Wikipedia: AI is often used to describe machines (or computers) that mimic "cognitive" functions that humans associate with the human mind, such as "learning" and "problem solving".

Artificial Intelligence

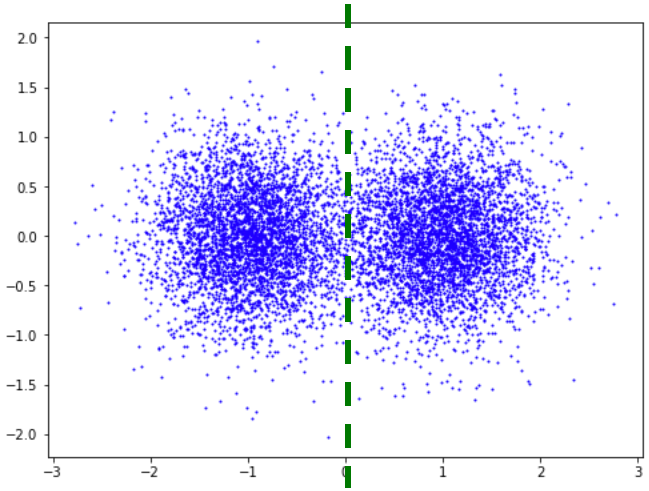
- From Wikipedia: AI is often used to describe machines (or computers) that mimic "cognitive" functions that humans associate with the human mind, such as "learning" and "problem solving".

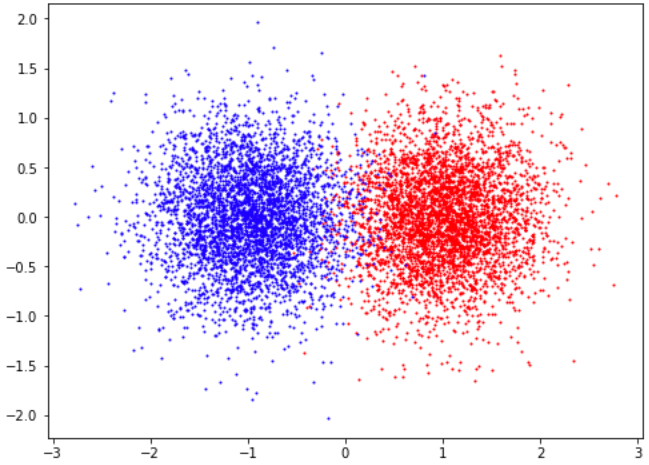


A* search algorithm

Machine Learning

- ML is the scientific study of algorithms and statistical models that computer systems use to perform a specific task without using explicit instructions, relying on patterns and inference instead.

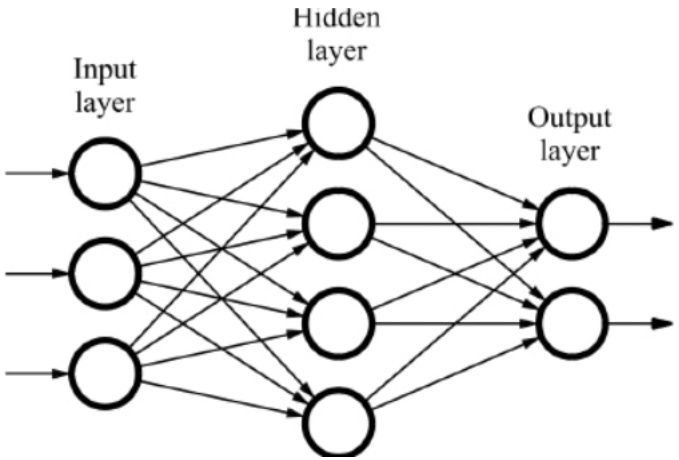




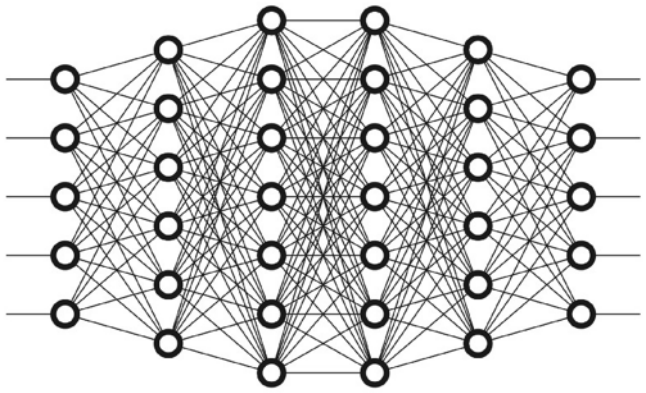
Deep Learning

- DL is part of a broader family of machine learning methods based on artificial neural networks with representation learning.

A small neural net

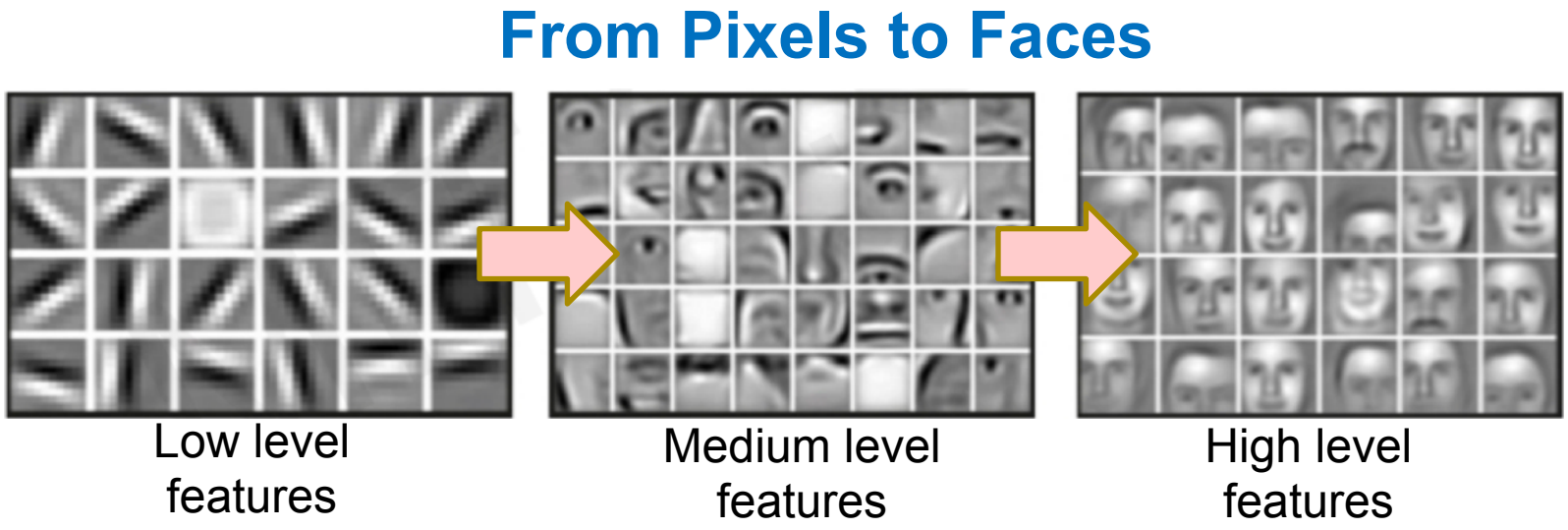


A large neural net



Deep Learning

- DL is part of a broader family of machine learning methods based on artificial neural networks with representation learning.



Representation Matters

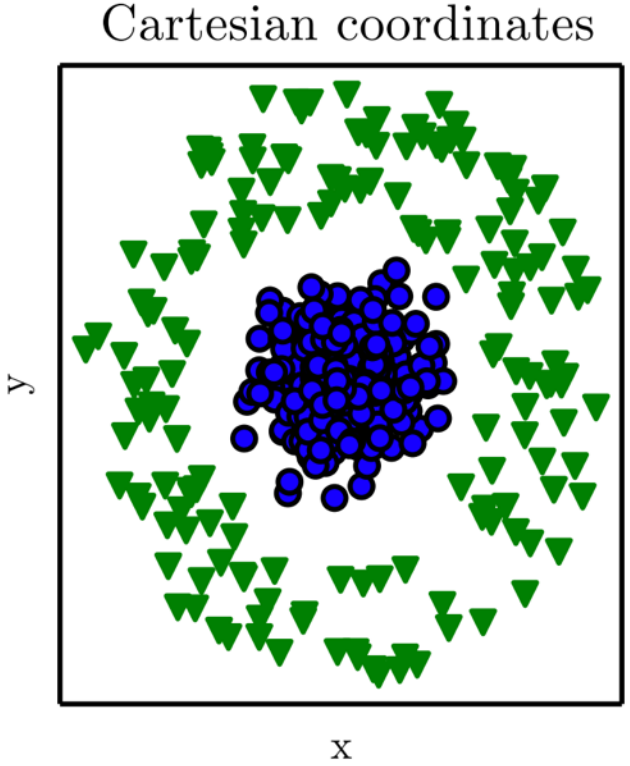
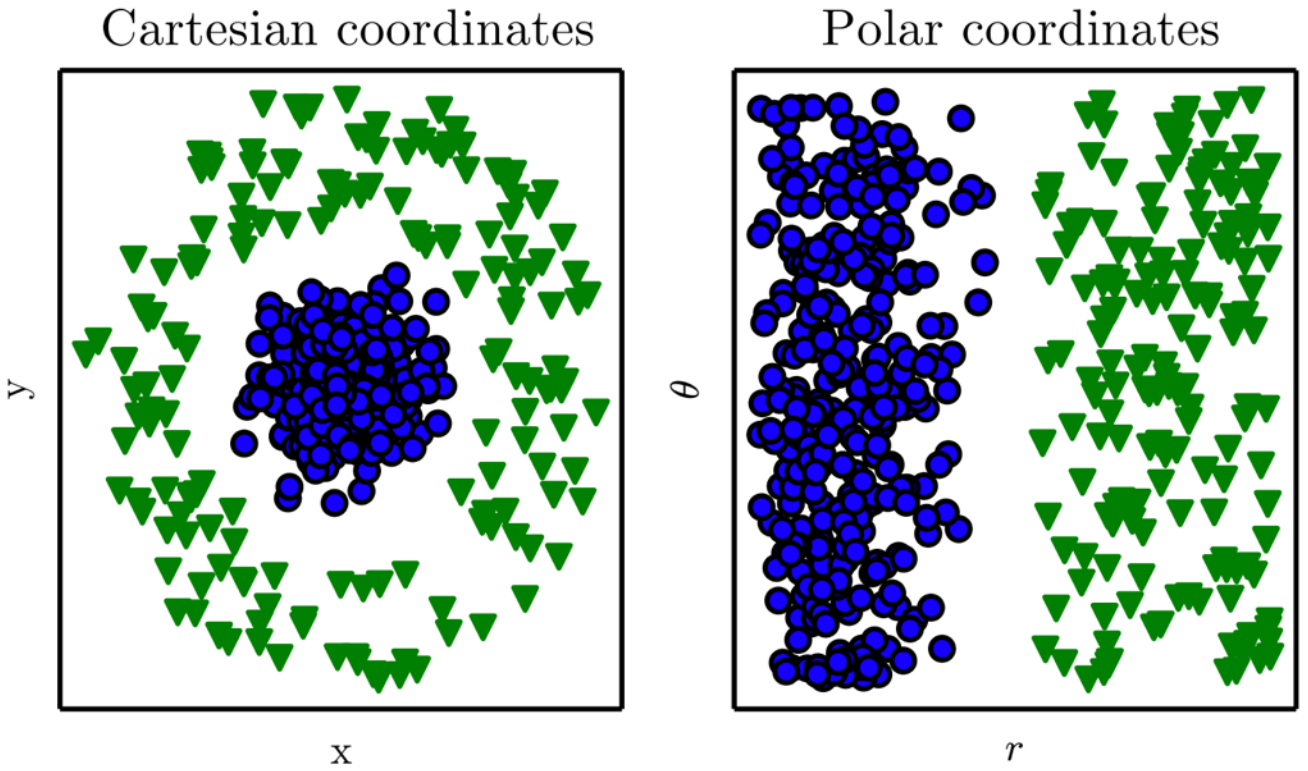


Figure 1.1

(Goodfellow 2016)

Representation Matters



$$r^2 = x^2 + y^2$$
$$\tan(\theta) = \frac{y}{x}$$

Figure 1.1

(Goodfellow 2016)

Polar coordinates can be hand-coded by a programmer.
What about more complex representations?

Representation Learning

Can you write a code to distinguish a CAT and a DOG?

Machine Learning and AI

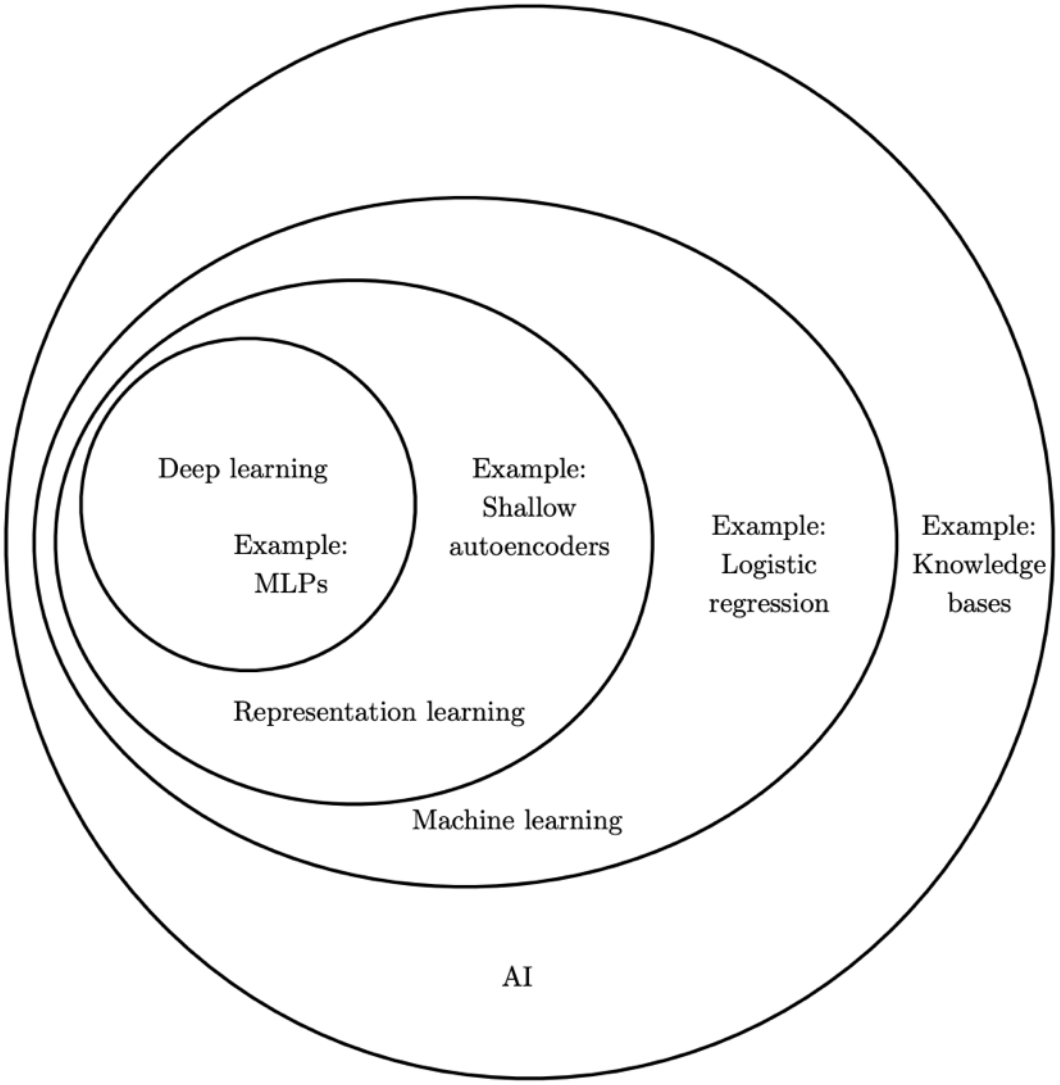
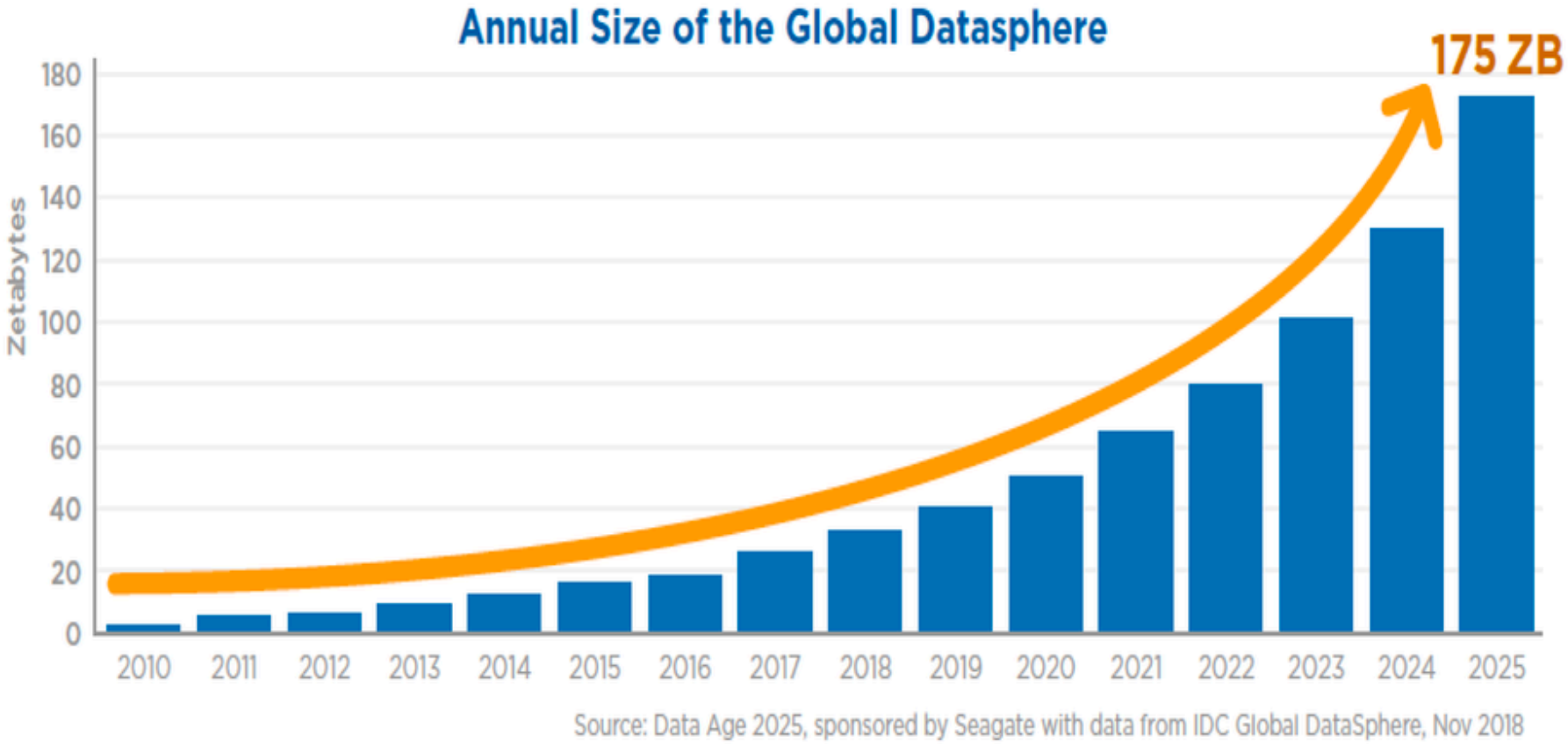
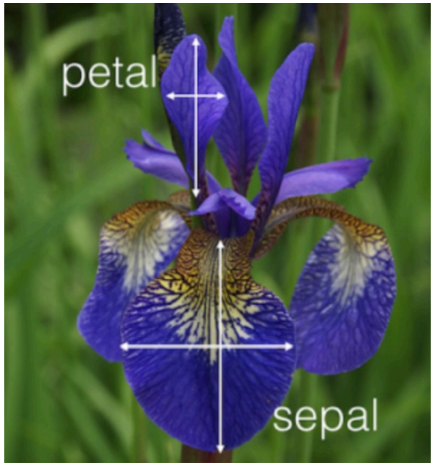


Figure 1.4

Why Deep Learning?


Annual Size of the Global Datasphere
175 ZB
Zetabytes
180
160
140
120
100
80
60
40
20
0
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
Source: Data Age 2025, sponsored by Seagate with data from IDC Global DataSphere, Nov 2018

Some Examples



Iris dataset, 1936
150 data points

ImageNet, 2009
15 million data points

Dataset	Quantity (tokens)	Weight in training mix	Epochs elapsed when training for 300B tokens
Common Crawl (filtered)	410 billion	60%	0.44
WebText2	19 billion	22%	2.9
Books1	12 billion	8%	1.9
Books2	55 billion	8%	0.43
Wikipedia	3 billion	3%	3.4

Table 2.2: Datasets used to train GPT-3. "Weight in training mix" refers to the fraction of examples during training that are drawn from a given dataset, which we intentionally do not make proportional to the size of the dataset. As a result, when we train for 300 billion tokens, some datasets are seen up to 3.4 times during training while other datasets are seen less than once.

Dataset to train GPT3, 2020
300 billion tokens

Rise of GPUs

Market Summary > NVIDIA Corp

123.51 USD

+123.47 (308,675.00%) ↑ all time

Follow

Closed: Sep 25, 7:59 PM EDT • Disclaimer

After hours 124.60 +1.09 (0.88%)

1D | 5D | 1M | 6M | YTD | 1Y | 5Y | Max

ML Models

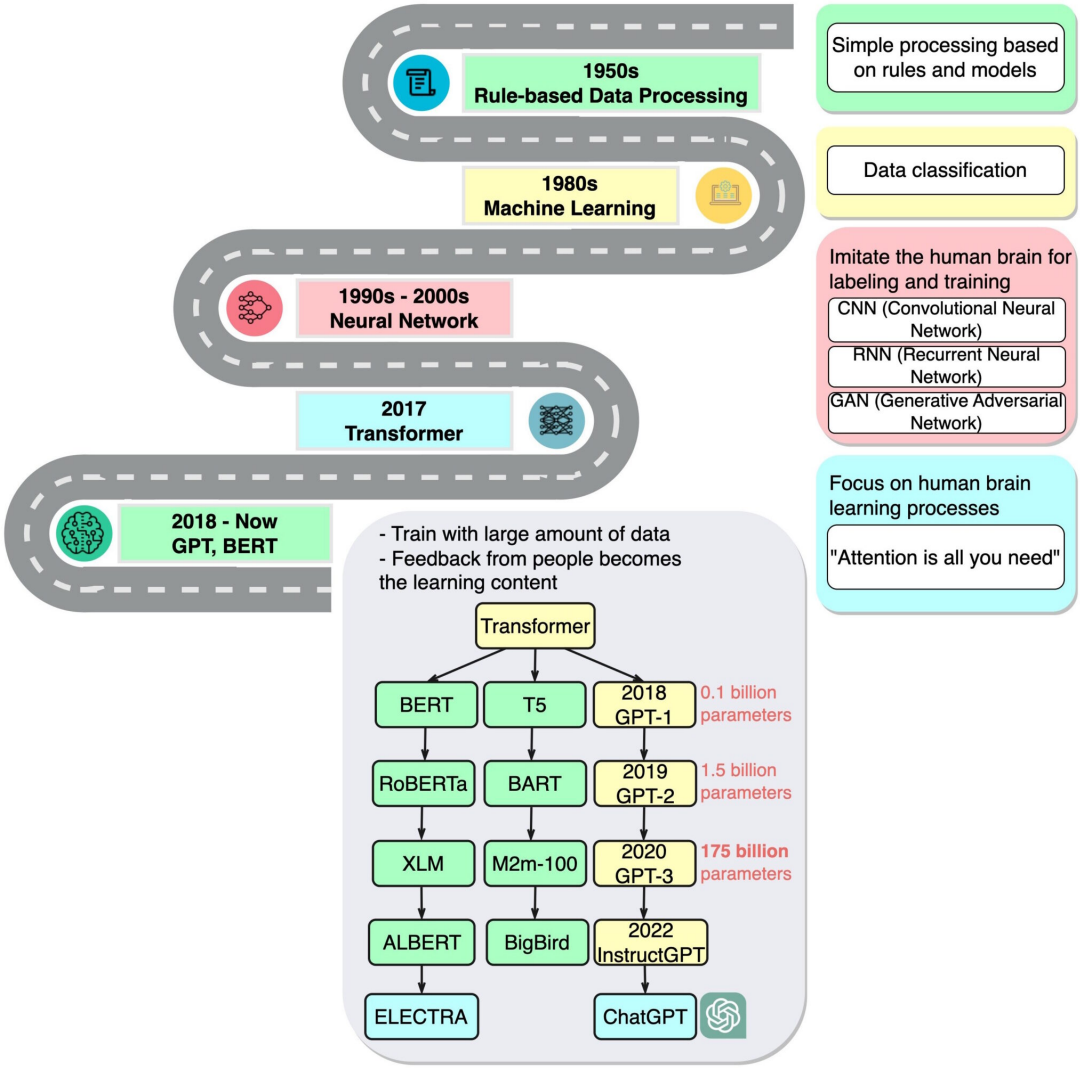
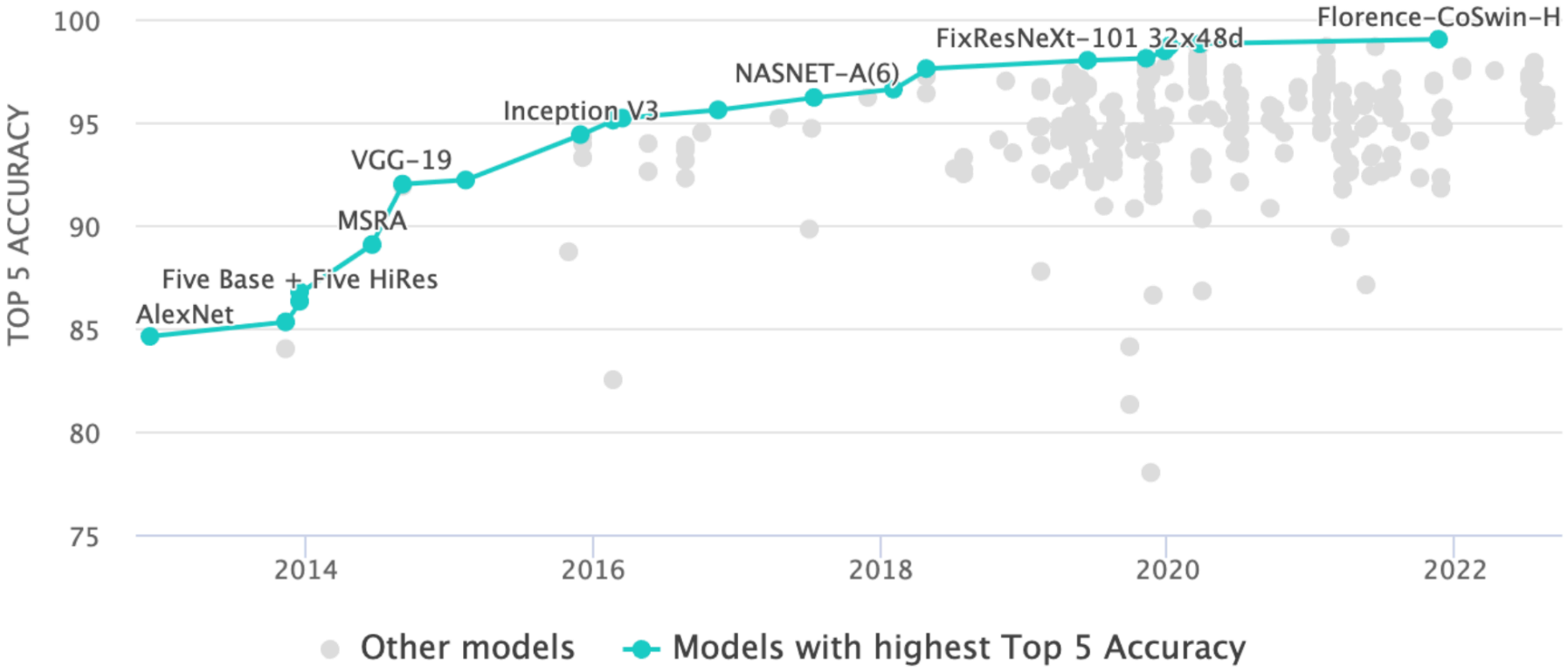

1950s
Rule-based Data Processing
Simple processing based on rules and models
1980s
Machine Learning
Data classification
1990s - 2000s
Neural Network
Imitate the human brain for labeling and training
CNN (Convolutional Neural Network)
RNN (Recurrent Neural Network)
GAN (Generative Adversarial Network)
2017
Transformer
Focus on human brain learning processes
"Attention is all you need"
2018 - Now
GPT, BERT
- Train with large amount of data
- Feedback from people becomes the learning content
Transformer
BERT
T5
2018 GPT-1
0.1 billion parameters
RoBERTa
BART
2019 GPT-2
1.5 billion parameters
XLM
M2m-100
2020 GPT-3
175 billion parameters
ALBERT
BigBird
2022 InstructGPT
ELECTRA
ChatGPT

Image Classification Accuracy



Advances in NLP

GPT-4

Input

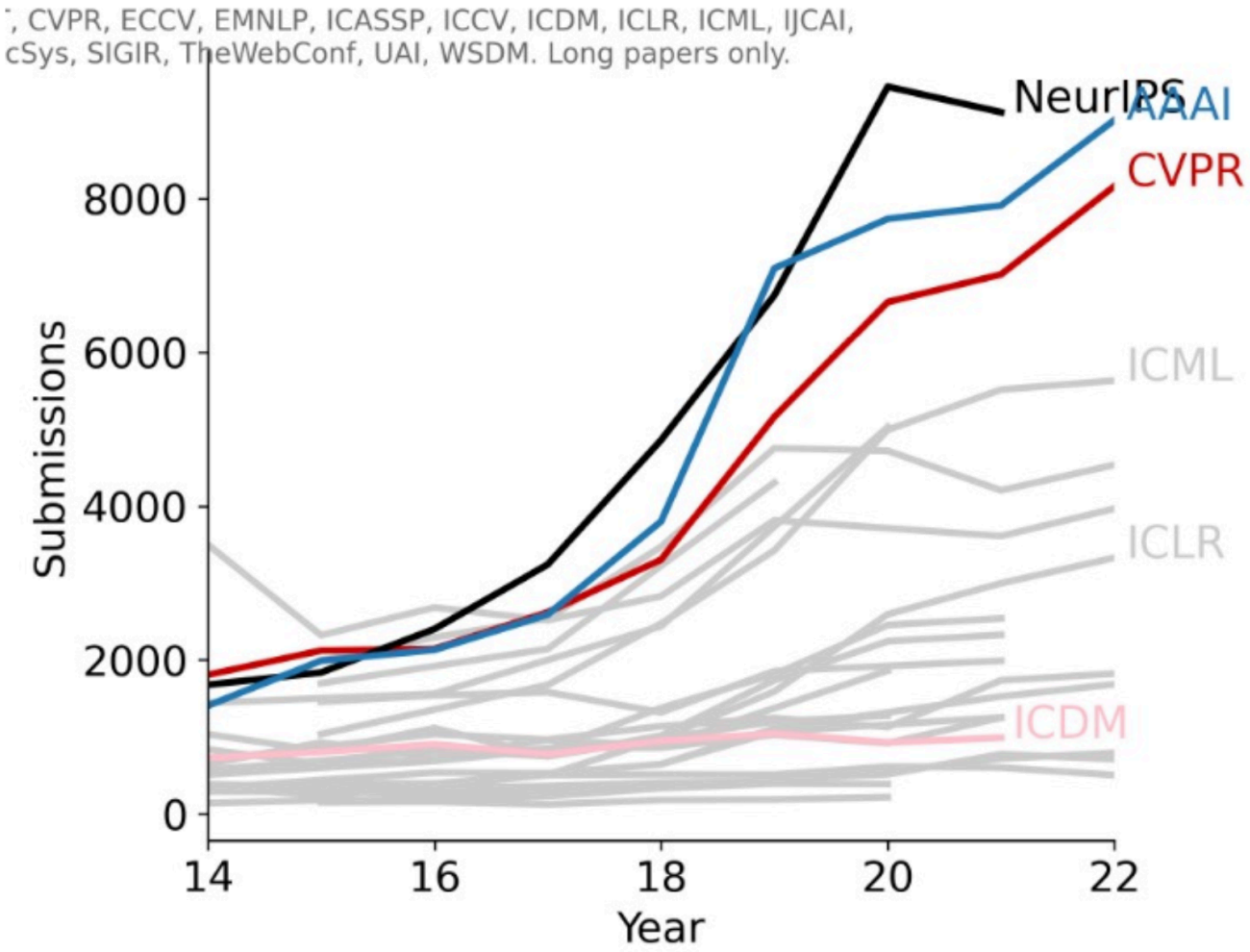
Andrew is free from 11 am to 3 pm, Joanne is free from noon to 2 pm and then 3:30 pm to 5 pm. Hannah is available at noon for half an hour, and then 4 pm to 6 pm. What are some options for start times for a 30 minute meeting for Andrew, Hannah, and Joanne?

Output

Andrew: 11 am - 3 pm
Joanne: 12 pm - 2 pm, 3:30 pm - 5 pm
Hannah: 12 pm - 12:30 pm, 4 pm - 6 pm

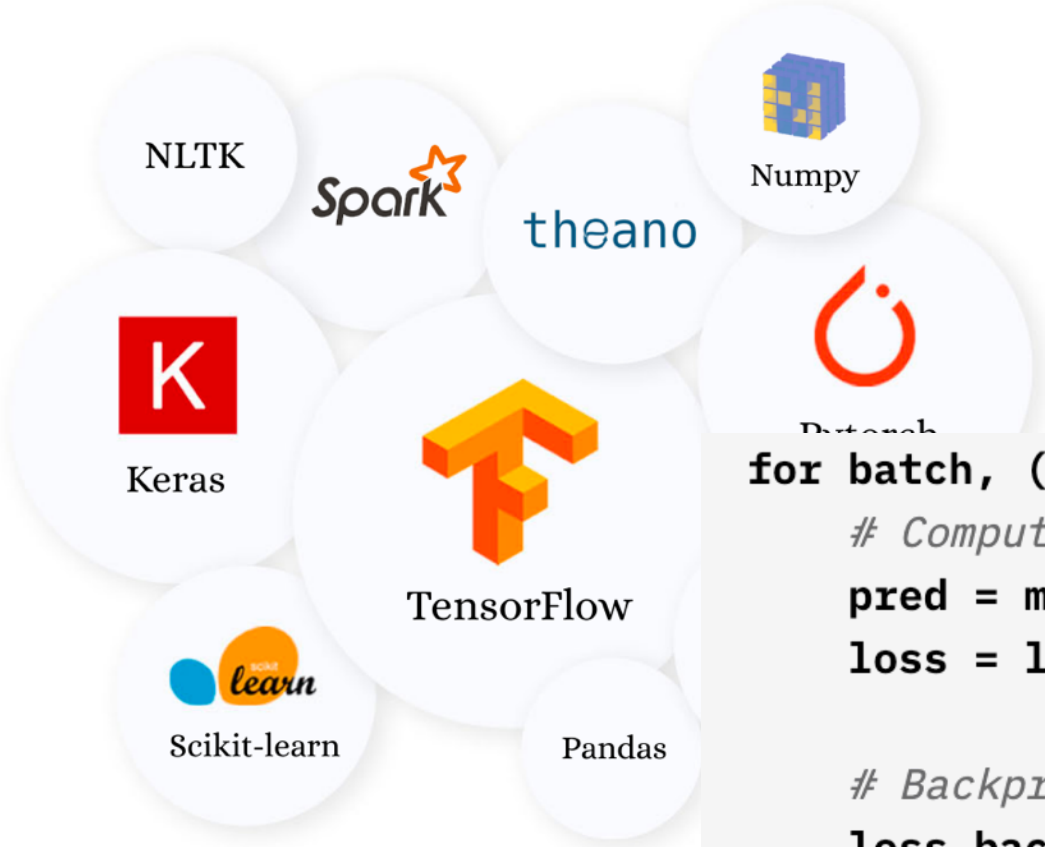
Common availability for a 30-minute meeting: 12 pm - 12:30 pm

Academic Interests



Growth of ML/AI/CV/NLP papers

Growing ML/DL software




```
for batch, (X, y) in enumerate(dataloader):
    # Compute prediction and loss
    pred = model(X)
    loss = loss_fn(pred, y)

    # Backpropagation
    loss.backward()
    optimizer.step()
    optimizer.zero_grad()
```

Summary

- Course logistics
- Importance of representation
- Motivations behind deep learning