CS 4803 / 7643: Deep Learning

Website: https://www.cc.gatech.edu/classes/AY2020/cs7643_fall/

Piazza: https://piazza.com/gatech/fall2019/cs48037643

Canvas: https://gatech.instructure.com/courses/60374 (4803)
https://gatech.instructure.com/courses/60364 (7643)

Gradescope: https://www.gradescope.com/courses/56799 (4803)
https://www.gradescope.com/courses/53817 (7643)

Dhruv Batra
School of Interactive Computing
Georgia Tech
What are we here to discuss?

Some of the most exciting developments in

Machine Learning,
Vision, NLP, Speech, Robotics
& AI in general

in the last decade!
Proxy for public interest
AlphaGo seals 4-1 victory over Go grandmaster Lee Sedol

DeepMind’s artificial intelligence astonishes fans to defeat human opponent and offers evidence computer software has mastered a major challenge.

Google DeepMind’s AlphaGo program triumphed in its final game against South Korean Go grandmaster Lee Sedol to win the series 4-1, providing further evidence of the landmark achievement for an artificial intelligence program.
Outline

• What is Deep Learning, the field, about?
  – Highlight of some recent projects from my lab

• What is this class about?
  – What to expect?
  – Logistics

• FAQ
Outline

• What is Deep Learning, the field, about?
  – Highlight of some recent projects from my lab

• What is this class about?
  – What to expect?
  – Logistics

• FAQ
Demo time

vqa.cloudcv.org.
demo.visualdialog.org
Concepts
What is (general) intelligence?

• Boring textbook answer

  The ability to acquire and apply knowledge and skills
  – Dictionary

• My favorite

  The ability to navigate in problem space
  – Siddhartha Mukherjee, Columbia
What is artificial intelligence?

- Boring textbook answer

  *Intelligence demonstrated by machines*
  - Wikipedia

- My favorite

  *The science and engineering of making computers behave in ways that, until recently, we thought required human intelligence.*
  - Andrew Moore, CMU
What is machine learning?

- My favorite

Study of algorithms that improve their performance (P) at some task (T) with experience (E)
  - Tom Mitchell, CMU
Image Classification

ImageNet Large Scale Visual Recognition Challenge (ILSVRC)

1000 object classes  1.4M/50k/100k images

http://image-net.org/challenges/LSVRC/{2010,...,2015}
Image Classification

ILSVRC top-5 error on ImageNet

(C) Dhruv Batra
Tasks are getting bolder

A group of young people playing a game of Frisbee
Vinyals et al., 2015

Antol et al., 2015

Das et al., 2017
Janelle Shane @JanelleCShane · Jun 24

One fun thing I discovered about Visual Chatbot. It learned from answers that humans gave, and apparently nobody ever asked "how many giraffes are there?" when the answer was zero.

demo.visualdialog.org
Embodied Question Answering
[CVPR ’18]

Abhishek Das  
(Georgia Tech)

Samyak Datta  
(Georgia Tech)

Georgia Gkioxari  
(FAIR)

Stefan Lee  
(Georgia Tech)

Devi Parikh  
(Georgia Tech / FAIR)

Dhruv Batra  
(Georgia Tech / FAIR)
What is to the left of the shower?

Cabinet
So what is Deep (Machine) Learning?

• Representation Learning

• Neural Networks

• Deep Unsupervised/Reinforcement/Structured/ <insert-qualifier-here> Learning

• Simply: Deep Learning
So what is Deep (Machine) Learning?

• A few different ideas:
  
  • (Hierarchical) Compositionality
    – Cascade of non-linear transformations
    – Multiple layers of representations

  • End-to-End Learning
    – Learning (goal-driven) representations
    – Learning to feature extraction

  • Distributed Representations
    – No single neuron “encodes” everything
    – Groups of neurons work together
Traditional Machine Learning

VISION

hand-crafted features
SIFT/HOG

your favorite classifier

"car"

SPEECH

hand-crafted features
MFCC

your favorite classifier

\'d ē p\n
NLP

This burrito place is yummy and fun!

hand-crafted features
Bag-of-words

your favorite classifier

“+”

Slide Credit: Marc'Aurelio Ranzato, Yann LeCun
Hierarchical Compositionality

**VISION**

pixels $\rightarrow$ edge $\rightarrow$ texton $\rightarrow$ motif $\rightarrow$ part $\rightarrow$ object

**SPEECH**

sample $\rightarrow$ spectral band $\rightarrow$ formant $\rightarrow$ motif $\rightarrow$ phone $\rightarrow$ word

**NLP**

character $\rightarrow$ word $\rightarrow$ NP/VP/.. $\rightarrow$ clause $\rightarrow$ sentence $\rightarrow$ story

Slide Credit: Marc'Aurelio Ranzato, Yann LeCun
Building A Complicated Function

Given a library of simple functions

\[
\begin{align*}
\sin(x) \\
\log(x) \\
\cos(x) \\
x^3 \\
\exp(x)
\end{align*}
\]

Compose into a complicated function

Slide Credit: Marc'Aurelio Ranzato, Yann LeCun

(C) Dhruv Batra
Building A Complicated Function

Given a library of simple functions

\[
\begin{align*}
\sin(x) \\
\log(x) \\
\cos(x) \\
x^3 \\
\exp(x)
\end{align*}
\]

Compose into a complicate function

Idea 1: Linear Combinations

- Boosting
- Kernels
- ...

\[
f(x) = \sum_{i} \alpha_i g_i(x)
\]
Building A Complicated Function

Given a library of simple functions

\[ f(x) = g_1(g_2(\ldots(g_n(x)\ldots))) \]

Idea 2: Compositions

- Deep Learning
- Grammar models
- Scattering transforms…

(C) Dhruv Batra

Slide Credit: Marc'Aurelio Ranzato, Yann LeCun
Building A Complicated Function

Given a library of simple functions

\[ f(x) = \log(\cos(\exp(\sin^3(x)))) \]

Idea 2: Compositions

- Deep Learning
- Grammar models
- Scattering transforms...

(C) Dhruv Batra

Slide Credit: Marc'Aurelio Ranzato, Yann LeCun
Deep Learning = Hierarchical Compositionality

Slide Credit: Marc'Aurelio Ranzato, Yann LeCun
Deep Learning = Hierarchical Compositionality

Feature visualization of convolutional net trained on ImageNet from [Zeiler & Fergus 2013]
So what is Deep (Machine) Learning?

- A few different ideas:
  - (Hierarchical) Compositionality
    - Cascade of non-linear transformations
    - Multiple layers of representations
  - End-to-End Learning
    - Learning (goal-driven) representations
    - Learning to feature extraction
  - Distributed Representations
    - No single neuron “encodes” everything
    - Groups of neurons work together

(C) Dhruv Batra
Traditional Machine Learning

VISION

hand-crafted features
SIFT/HOG

your favorite classifier

“car”

SPEECH

hand-crafted features
MFCC

your favorite classifier

\'d ē p\n
NLP

This burrito place is yummy and fun!

hand-crafted features
Bag-of-words

your favorite classifier

“+”
Feature Engineering

SIFT

Spin Images

HoG

Textons

and many many more….
Traditional Machine Learning (more accurately)

VISION

SIFT/HOG
fixed

K-Means/pooling
unsupervised

classifier
supervised

“Learned”

“car”

SPEECH

MFCC
fixed

Mixture of Gaussians
unsupervised

classifier
supervised

\'d ē p\n
NLP

This burrito place is yummy and fun!

Parse Tree Syntactic
fixed

n-grams
unsupervised

classifier
supervised

“+”

(C) Dhruv Batra

Slide Credit: Marc'Aurelio Ranzato, Yann LeCun
Deep Learning = End-to-End Learning

VISION

SIFT/HOG → K-Means/pooling → classifier

*SPEECH*

MFCC → Mixture of Gaussians → classifier

NLP

Parse Tree Syntactic → n-grams → classifier

This burrito place is yummy and fun!

“Learned”

“car”

\'d ē p\`

“+”

(C) Dhruv Batra

Slide Credit: Marc'Aurelio Ranzato, Yann LeCun
“Shallow” vs Deep Learning

• “Shallow” models

- hand-crafted
  Feature Extractor
  fixed

- “Simple” Trainable
  Classifier
  learned

• Deep models

- Trainable
  Feature-Transform / Classifier

- Trainable
  Feature-Transform / Classifier

- Trainable
  Feature-Transform / Classifier

Learned Internal Representations

Slide Credit: Marc'Aurelio Ranzato, Yann LeCun
So what is Deep (Machine) Learning?

• A few different ideas:

• (Hierarchical) Compositionality
  – Cascade of non-linear transformations
  – Multiple layers of representations

• End-to-End Learning
  – Learning (goal-driven) representations
  – Learning to feature extraction

• Distributed Representations
  – No single neuron “encodes” everything
  – Groups of neurons work together
Distributed Representations Toy Example

- Local vs Distributed

\[(a)\]

- No pattern
- Line pattern
- Square pattern
- Circle pattern
- Oval pattern
Distributed Representations Toy Example

- Can we interpret each dimension?

(a) no pattern
- (a) no pattern
- (a) no pattern
- (a) no pattern
- (a) no pattern

(b) no pattern
- (b) no pattern
- (b) no pattern
- (b) no pattern
- (b) no pattern

vertical, horizontal, rectangle, ellipse
Power of distributed representations!

Local: \[ \bullet \bullet \ O \ \bullet = VR + HR + HE = ? \]

Distributed: \[ \bullet \bullet \ O \ \bullet = V + H + E \approx \bigcirc \]
Power of distributed representations!

- United States:Dollar :: Mexico:?
ThisPlusThat.me

the matrix - thoughtful + dumb

mbiguated into +1 the_matrix -1 thoughtful +1 dumb in 0.0 seconds from ip-10-32-114-31

FILM, W FILM, NETFLIX TITLE,

Blade II

Blade II is a 2002 American vampire superhero action film base Marvel Comics character Blade. It is the sequel of the first film a part of the Blade film series. It was written by David S. Goyer, w previous film. Guillermo del Toro was signed in to d...

Horror Film

Image Credit: http://insightdatascience.com/blog/thisplusthat_a_search_engine_thatLets_you_add_words_as_vectors.html
So what is Deep (Machine) Learning?

• A few different ideas:

  • (Hierarchical) Compositionality
    – Cascade of non-linear transformations
    – Multiple layers of representations

  • End-to-End Learning
    – Learning (goal-driven) representations
    – Learning to feature extraction

  • Distributed Representations
    – No single neuron “encodes” everything
    – Groups of neurons work together
Benefits of Deep/Representation Learning

• (Usually) Better Performance
  – “Because gradient descent is better than you”
    Yann LeCun

• New domains without “experts”
  – RGBD/Lidar
  – Multi-spectral data
  – Gene-expression data
  – Unclear how to hand-engineer
“Expert” intuitions can be misleading

• “Every time I fire a linguist, the performance of our speech recognition system goes up”
  – Fred Jelinik, IBM ’98
Benefits of Deep/Representation Learning

- Modularity!
- Plug and play architectures!
Any DAG of differentiable modules is allowed!
Key Computation: Forward-Prop

\[ X \rightarrow \theta \rightarrow Z \]
Key Computation: Back-Prop

\[
\frac{\partial L}{\partial X} \rightarrow \left\{ \frac{\partial Z}{\partial X}, \frac{\partial Z}{\partial \theta} \right\} \rightarrow \frac{\partial L}{\partial Z}
\]

\[\frac{\partial L}{\partial \theta}\]
Any DAG of differentiable modules is allowed!
Problems with Deep Learning

• **Problem#1: Non-Convex! Non-Convex! Non-Convex!**
  - Depth>=3: most losses non-convex in parameters
  - Theoretically, all bets are off
  - Leads to stochasticity
    • different initializations $\rightarrow$ different local minima

• **Standard response #1**
  - “Yes, but all interesting learning problems are non-convex”
  - For example, human learning
    • Order matters $\rightarrow$ wave hands $\rightarrow$ non-convexity

• **Standard response #2**
  - “Yes, but it often works!”

(C) Dhruv Batra
Problems with Deep Learning

• Problem#2: Lack of interpretability
  – Hard to track down what’s failing
  – Pipeline systems have “oracle” performances at each step
  – In end-to-end systems, it’s hard to know why things are not working
Problems with Deep Learning

• Problem#2: Lack of interpretability

[Fang et al. CVPR15]

[Vinyals et al. CVPR15]
Problems with Deep Learning

• Problem#2: Lack of interpretability
  – Hard to track down what’s failing
  – Pipeline systems have “oracle” performances at each step
  – In end-to-end systems, it’s hard to know why things are not working

• Standard response #1
  – Tricks of the trade: visualize features, add losses at different layers, pre-train to avoid degenerate initializations…
  – “We’re working on it”

• Standard response #2
  – “Yes, but it often works!”
Problems with Deep Learning

• Problem#3: Lack of easy reproducibility
  – Direct consequence of stochasticity & non-convexity

• Standard response #1
  – It’s getting much better
  – Standard toolkits/libraries/frameworks now available
  – PyTorch, TensorFlow, MxNet…

• Standard response #2
  – “Yes, but it often works!”
Yes it works, but how?

Good work -- but I think we might need a little more detail right here.
Outline

• What is Deep Learning, the field, about?
  – Highlight of some recent projects from my lab

• What is this class about?
  – What to expect?
  – Logistics

• FAQ
Outline

• What is Deep Learning, the field, about?
  – Highlight of some recent projects from my lab

• What is this class about?
  – What to expect?
  – Logistics

• FAQ
What is this class about?
What was F17 DL class about?

- Firehose of arxiv
Arxiv Fire Hose

Deep Learning papers

PhD Student

(C) Dhruv Batra
What was F17 DL class about?

• Goal:
  – After taking this class, you should be able to pick up the latest Arxiv paper, easily understand it, & implement it.

• Target Audience:
  – Junior/Senior PhD students who want to conduct research and publish in Deep Learning.

  (think ICLR/CVPR papers as outcomes)
What is the F19 DL class about?

• Introduction to Deep Learning

• Goal:
  – After finishing this class, you should be ready to get started on your first DL research project.
    • CNNs
    • RNNs
    • Deep Reinforcement Learning
    • Generative Models (VAEs, GANs)

• Target Audience:
  – Senior undergrads, MS-ML, and new PhD students
What this class is NOT

• NOT the target audience:
  – Advanced grad-students already working in ML/DL areas
  – People looking to understand latest and greatest cutting-edge research (e.g. GANs, AlphaGo, etc)
  – Undergraduate/Masters students looking to graduate with a DL class on their resume.

• NOT the goal:
  – Teaching a toolkit. “Intro to TensorFlow/PyTorch”
  – Intro to Machine Learning
Caveat

• This is an ADVANCED Machine Learning class
  – This should NOT be your first introduction to ML
  – You will need a formal class; not just self-reading/coursera

  – If you took CS 7641/ISYE 6740/CSE 6740 @GT, you’re in the right place

  – If you took an equivalent class elsewhere, see list of topics taught in CS 7641 to be sure.
Prerequisites

- Intro Machine Learning
  - Classifiers, regressors, loss functions, MLE, MAP

- Linear Algebra
  - Matrix multiplication, eigenvalues, positive semi-definiteness...

- Calculus
  - Multi-variate gradients, hessians, jacobians...
Prerequisites

• Intro Machine Learning – Classifiers, regressors, loss functions, MLE, MAP

• Linear Algebra – Matrix multiplication, eigenvalues, positive semi-definiteness…

• Calculus – Multivariable gradients, hessians, jacobians…

(C) Dhruv Batra
Prerequisites

• Intro Machine Learning
  – Classifiers, regressors, loss functions, MLE, MAP

• Linear Algebra
  – Matrix multiplication, eigenvalues, positive semi-definiteness...

• Calculus
  – Multi-variate gradients, hessians, jacobians...

• Programming!
  – Homeworks will require Python!
  – Libraries/Frameworks: PyTorch
  – HW0+4 (pure python), HW1 (python + PyTorch), HW2+3 (PyTorch)
  – Your language of choice for project
Course Information

- Instructor: Dhruv Batra
  - dbatra@gatech
  - Location: 219 CCB
My Research Group + Collaborators

Research Scientists

Stefan Lee
Peter Anderson
Zhile Ren

Postdoc

Ramakrishna Vedantam (Ph.D. Student)
Arjun Chandrasekaran (Ph.D. Student)
Jiasen Lu (Ph.D. Student)
Jianwei Yang (Ph.D. Student)

PhD Students

Aishwarya Agrawal (2014 – Present)
Yash Goyal (2014 – Present)
Michael Cogswell (2015 – Present)
Abhishek Das (2016 – Present)
Ashwin Kalyan (2016 – Present)
Nirbhay Modhe (2017 – Present)

Samyak Datta (Ph.D. Student)
Prithvijit Chattopadhyay (M.S. Student)
Viraj Prabhu (M.S. Student)
Ramprasaath Selvaraju (Ph.D. Student)
Erik Wijmans (2017 – Present)
Harsh Agrawal (2018 – Present)
Deshraj Yadav (2017 – Present)
TAs

Harsh Agrawal
Harish Kamath
Michael Piseno
Neha Jain
Anishi Mehta
Viraj Prabhu
Ashwin Kalyan
Nirbhay Modhe
Sarah Wiegrefe
Organization & Deliverables

• 4 problem-sets+homeworks (80%)
  – Mix of theory (PS) and implementation (HW)
  – First one goes out next week
    • Start early, Start early, Start early, Start early, Start early, Start early,
      Start early, Start early, Start early, Start early

• Final project (20%)
  – Projects done in groups of 3-4

• (Bonus) Class Participation (5%)
  – Contribute to class discussions on Piazza
  – Ask questions, answer questions
Late Days

• “Free” Late Days
  – 7 late days for the semester
    • Use for HWs
    • Cannot use for project related deadlines

  – After free late days are used up:
    • 25% penalty for each late day
PS0

• Out today; due Aug 22
  – Available on class webpage + Canvas

• Grading
  – Not counted towards your final grade, but required
  – <=50% means that you might not be prepared for the class

• Topics
  – PS: probability, calculus, convexity, proving things
Project

• Goal
  – Chance to try Deep Learning
  – Encouraged to apply to your research (computer vision, NLP, robotics,…)
  – Must be done this semester.
  – Can combine with other classes
    • get permission from both instructors; delineate different parts
  – Extra credit for shooting for a publication

• Main categories
  – Application/Survey
    • Compare a bunch of existing algorithms on a new application domain of your interest
  – Formulation/Development
    • Formulate a new model or algorithm for a new or old problem
  – Theory
    • Theoretically analyze an existing algorithm
Computing

• Major bottleneck
  – GPUs

• Options
  – Your own / group / advisor’s resources
  – Google Cloud Credits
    • $50 credits to every registered student courtesy Google
  – Google Colab
    • jupyter-notebook + free GPU instance
4803 vs 7643

• Level differentiation

• HWs
  – Extra credit questions for 4803 students, necessary for 7643

• Project
  – Higher expectations from 7643
Outline

• What is Deep Learning, the field, about?
  – Highlight of some recent projects from my lab

• What is this class about?
  – What to expect?
  – Logistics

• FAQ
Waitlist / Audit / Sit in

• Waitlist
  – Class is full. Size will not increase further.
  – Do PS0. Come to first few classes.
  – Hope people drop.

• “I need this class to graduate”
  – Talk to your degree program advisor. They control the process of making sure you have options to graduate on time.

• Audit or Pass/Fail
  – We will give preference to people taking class for credit.

• Sitting in
  – Talk to instructor.
Research

• “Can I work with your group for funding/credits/neither?”

  – I am not taking new advising duties.

  – If you can find one of my students to supervise you,
    I am happy to sign off on the paperwork.

  – Your responsibility to approach them and ask.
    It will help if you know what they are working on.
What is the re-grading policy?

• Homework assignments
  – **Within 1 week** of receiving grades: see the TAs

• This is an advanced grad class.
  – The goal is understanding the material and making progress towards our research.
What is the collaboration policy?

• Collaboration
  – Only on HWs and project (not allowed in HW0).
  – You may discuss the questions
  – Each student writes their own answers
  – Write on your homework anyone with whom you collaborate
  – Each student must write their own code for the programming part

• Zero tolerance on plagiarism
  – Neither ethical nor in your best interest
  – Always credit your sources
  – Don’t cheat. We will find out.
How do I get in touch?

• Primary means of communication -- Piazza
  – No direct emails to Instructor unless private information
  – Instructor/TAs can provide answers to everyone on forum
  – Class participation credit for answering questions!
  – No posting answers. We will monitor.

• Staff Mailing List
  – f19-cs4803-cs7643-staff@googlegroups.com

• Links:
  – Website: https://www.cc.gatech.edu/classes/AY2020/cs7643_fall/
  – Piazza: https://piazza.com/gatech/fall2019/cs48037643
  – Canvas: https://gatech.instructure.com/courses/60374 (4803)
    https://gatech.instructure.com/courses/60364 (7643)
  – Gradescope: https://www.gradescope.com/courses/56799 (4803)
    https://www.gradescope.com/courses/53817 (7643)
Todo

• PS0
  – Due: Aug 22 11:00am
Welcome

Who is Stoker?
(For one welcomed our new computer overlords)

$18,200