CPSC 452 01 (SP22): Deep Learning Theory and Applications

Deep neural networks have gained immense popularity within the last decade due to their success in many important machine learning tasks such as image recognition, speech recognition, and natural language processing. This course will provide a principled and hands-on approach to deep learning with neural networks. By the end of the course, students will have mastered the principles and practices underlying neural networks including optimization and training methods, design and architecture, generalization theory, loss landscapes, and will have applied deep learning methods to real-world problems including image recognition, natural language processing, and biomedical applications. The course will be based on homework, a final exam, and a final project (either group or individual, depending on the total number enrolled). Students' grades will be based on their homework scores and the quality of the written component of their projects. The course assumes basic prior knowledge in linear algebra and probability. This year the class will have elements of a flip format (pre-recorded lectures, in class activities).

Instructor: Prof. Smita Krishnaswamy (smita.krishnaswamy@yale.edu)

TA/ULA:

Ross Johnson: (Ross.Johnson@yale.edu)

Sumner Magruder: (Sumner.Magruder@yale.edu)

Kincaid MacDonald: (Kincaid.MacDonald@yale.edu)

Chae Young Lee: (chaeyoung.lee@yale.edu)

Charles Xu (charles.xu@yale.edu)

Additional Programming Resources:

Pytorch Bootcamp: Sunday January 30th 9am-11am

Recording: **https://yale.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=3bd5e0d8-6de4-4df1-b3ff-ae2d0145a4c8**

Colab Notebook: https://colab.research.google.com/drive/1zZIcTjCMKZ_pUDK_edEfc6NEpyWnTETE?usp=sharing

NBDev for Literate Programming: https://www.craft.do/s/DZM5DsSuMyJfMz

Instructor Office Hours:

- Wednesdays 1–2 On the class zoom link https://zoom.us/s/99183561572
- By appointment (email smita.krishnaswamy@yale.edu)

TA/ULA Office hours by ZOOM

- **Sumner:** Monday 10–12pm
- **Kincaid:** Friday 3–5pm
- **Ross:** Tuesdays 7:30—9:30pm
- **Chae Young:** Thursdays 7–9pm
- **Charles Xu:** Sundays 10–12pm

These may be subject to change.

Class Discord Server: https://discord.gg/TNy5Cc6JM6
Class Piazza: https://piazza.com/class/kyoikimyzbz6xj (you can ask questions anonymously here!)
Data Resources for Final Project: google doc with links

Textbooks (available online freely):

1. Neural Networks and Deep Learning by Michael Nielsen available at http://neuralnetworksanddeeplearning.com/
2. Deep Learning by Goodfellow, Bengio, and Courville available at http://www.deeplearningbook.org/
3. Graph Neural Networks, Hamilton: https://www.cs.mcgill.ca/~wlh/grl_book/files/GRL_Book.pdf

Class sessions: Tuesdays & Thursdays 11:35–12:50

Instructor Office Hours: After class Tuesdays, by appointment

Reading

Lecture 1: Deep Learning Overview Reading: Nielsen Chapter 1, Goodfellow Chapter 1.

Lecture 2: Machine Learning Reading: Chapter 5 Goodfellow

Lecture 3: Stochastic Gradient Descent: Nielsen Chapter 1, Goodfellow et al., Section 5.9 and chapter 6

https://towardsdatascience.com/step-by-step-tutorial-on-linear-regression-with-stochastic-gradient-descent-1d35b088a843

Lecture 4: Backpropagation: Nielsen Chapters 1-2 Goodfellow 6.5

Lecture 5: Losses and Activations Nielsen Chapter 3

Lecture 6: Regularization Nielsen Chapter 3, Goodfellow Chapter 7, Neyshambur et al. 2015. Arpit et a. 2017

- A closer look at memorization in deep networks [Arpit et al 2017] https://arxiv.org/pdf/1706.05394.pdf
- In search of the real inductive bias: on the role of implicit regularization in deep learning [Neyshabur et al, 2015] https://arxiv.org/pdf/1412.6614. pdf
- On the implicit Bias of Dropout https://arxiv.org/abs/1806.09777

Lecture 7: Neural Net Architectures and Project Proposal

Lecture 8 : Variations on SGD - momentum : Goodfellow et al., Sections 4.3, 8.1-8.3, 8.5-8.6, Nielsen book, chapter 3

- Choromanska et al, “The Loss Surfaces of Multilayer Neural Networks,” 2014.
- https://towardsdatascience.com/adam-latest-trends-in-deep-learning-optimization-6be9a291375c

Lecture 9: Non-linear dimensionality reduction and autoencoders:

- Goodfellow et al., chapter 14
- Hinton and Salakhutdinov Science 2006
- Coifman and Lafon Applied and Computational Harmonic Analysis 2006
- Belkin & Niyogi, Neural Computation 2003

- Alain & Bengio, JMLR 2014

Lecture 10: Autoencoders part 2

- Goodfellow Chapter 14
- SAUCIE https://www.biorxiv.org/content/biorxiv/early/2018/08/27/237065.1.full.pdf
- AAnet https://arxiv.org/pdf/1901.09078.pdf

Lecture 11: VAEs and data generation

- Kingma & Welling ICLR 2014
- Bengio et al. NeurIPS 2013
- Lopez et al. Nature Methods 2018
- Goodfellow Chapter 20
- https://arxiv.org/pdf/2103.01327.pdf

Lecture 12: Generative models and GANs

- Dziugaite et al. UAI 2015
- Goodfellow et al., Section 20.10.4
- Lectures on GANs CS 11-785 at CMU
- https://lilianweng.github.io/lil-log/2017/08/20/from-GAN-to-WGAN.html
- https://wiseodd.github.io/techblog/2017/01/26/kl-mle/
- Goodfellow et al 2014 paper
- Arjovsky et al 2017 paper

Lecture 13: GANs 2 domain transfer

- StyleGAN: https://arxiv.org/pdf/1812.04948.pdf
- InfoGAN: https://arxiv.org/pdf/1606.03657.pdf
- DiscoGAN: https://arxiv.org/abs/1703.05192
- CycleGAN: https://arxiv.org/abs/1703.10593
    - https://machinelearningmastery.com/what-is-cyclegan/#:~:text=The%20CycleGAN%20is%20a%20technique,be%20related%20in%20any%20way.
- MAGAN: http://proceedings.mlr.press/v80/amodio18a.html
- TraVeLGAN: https://arxiv.org/abs/1902.09631
- Conditional GAN: https://arxiv.org/pdf/1411.1784.pdf

Lecture 14: CNNs

- Chapter 6 Nielsen
- Szegedy et al. Going deeper with convolutions 2014
- He et al. Deep Residual Learning for Image Recognition 2015
- Ronneberger et al. U-Net: Convolutional Networks for Biomedical Image Segmentation 2014
- http ://cs231n.github.io/convolutional-networks /
- https://ujjwalkarn.me/2016/08/11/intuitive-explanation-convnets /
- https://towardsdatascience.com/a-simple-guide-to-the-versions-of-the-inception-network-7fc52b863202

Lecture 15: RNNs & LSTMs

- Chapter 10 Goodfellow
- Mikolov et al. 2013 (Word2Vec paper) https://arxiv.org/abs/1301.3781
- Luong et al. 2015 https://arxiv.org/abs/1508.04025
- Hochreiter and Schmidhuber. 1997 https://www.bioinf.jku.at/publications/older/2604.pdf
- Attention in RNNs blog https://medium.datadriveninvestor.com/attention-in-rnns-321fbcd64f05
- Understanding LSTMs blog http://colah.github.io/posts/2015-08-Understanding-LSTMs/
- https://lena-voita.github.io/nlp_course/seq2seq_and_attention.html
- Bahadaneu et al. 2015 https://arxiv.org/pdf/1409.0473.pdf

Lecture 16: Transformers

- https :// towardsdatascience.com/attention-and-its-different-forms-7fc3674d14dc
- http ://jalammar.github.io/illustrated-transformer /
- Vaswani et al. 2019 https://arxiv.org/abs/ 1706.03762
- Radford et al. (GPT2 paper)
- Brown et al. (GPT3 paper)
- https://jalammar.github.io/illustrated-gpt2 /
- https://demo.allennlp.org/next-token-lm

- https://openai.com/blog/musenet/

Lecture 17: Message Passing GNNs

- Chapter 5 from Hamilton Book
- Hamilton et al. Inductive Representation Learning on Large Graphs
- https://towardsdatascience.com/expressive-power-of-graph-neural-networks-and-the-weisefeiler-lehman-test-b883db3c7c49
- GNNs and WL test: https://arxiv.org/pdf/1810.00826.pdf
- GCN paper: https://arxiv.org/pdf/1609.02907.pdf
- Graph Sage: https://arxiv.org/pdf/1706.02216.pdf

Lecture 18: Graph Convolutional Networks

eigenvector of graph Laplacian give freq on graph, load signal to freq domain and filter signal

- Defferrard et al. Convolutional Neural Networks on Graphs
- Kipf & Welling Semisupervised Graph Classification
- https://towardsdatascience.com/spectral-graph-convolution-explained-and-implemented-step-by-step-2e495b57f801
- Graph wavelet neural network https://arxiv.org/pdf/1904.07785.pdf
- Graph scattering Autoencoder: https://arxiv.org/abs/2006.06885
- Learnable scattering network: https://arxiv.org/abs/2010.02415
- Diffusion wavelets: https://www.sciencedirect.com/science/article/pii/S106352030600056X

Lecture 19: Neural ODEs

- Chen et al. NeurIPS 2019 Neural Ordinary Differential Equations
- Tong et al. ICML 2020 TrajectoryNet
- Errico 1997 Adjoint Model
- https://towardsdatascience.com/ the-story-of-adjoint-sensitivity-method-from-meteorology-906ab2796c73

  https://towardsdatascience.com/neural-odes-breakdown-of-another-deep-learning-breakthrough-3e78c7213795

Lecture 20: Universality of Neural Networks proof: universal

- Nielsen book, chapter 4
- https://www.mathematik.uni-wuerzburg.de/fileadmin/10040900/2019/Seminar__Artificial_Neural_Network__24_9__.pdf

Lecture 21: Representation Learning in Neural Networks

- Tishby et al. DNNs and Information Bottleneck https://arxiv.org/abs/1503.02406
- Frankle & Carbin 2019 **THE LOTTERY TICKET HYPOTHESIS: FINDING SPARSE, TRAINABLE NEURAL NETWORKS** https://arxiv.org/pdf/1803.03635.pdf
- Johnson and Lindenstrauss Lamma: random projection is good, ||v||=k/d: DasGupta & Gupta 2002 https://cseweb.ucsd.edu/~dasgupta/papers/jl.pdf
- Visualizing the PHATE of neural networks https://arxiv.org/abs/1908.02831
- Information flow in neural nets https://proceedings.mlr.press/v97/goldfeld19a/goldfeld19a.pdf

Lecture 22: Generalization and Memorization

- Belkin, et al. 2019 PNAS Reconciling Modern Machine-Learning Practice and The Classical Bias-Variance Tradeoff
- Understanding Deep Learning requires rethinking generalization, Zhang et al ICLR 2017 https://arxiv.org/pdf/1611.03530.pdf
- A closer look at memorization in deep networks [Arpit et al 2017] https://arxiv.org/pdf/1706.05394.pdf
- In search of the real inductive bias: on the role of implicit regularization in deep learning [Neyshabur et al, 2015] https://arxiv.org/pdf/1412.6614.pdf
- The role of over-parametrization in generalization of neural networks [Neyshabur et al, 2019] https://arxiv.org/pdf/1805.12076.pdf
- Train faster, generalize better: stability of stochastic gradient descent [Hardt et al, 2016] https://arxiv.org/pdf/1509.01240.pdf
- Stability and generalization [Bousquet and A. Elisseeff 2002] https://www.academia.edu/13743279/Stability_and_generalization
- Rademacher complexity: http://www.cs.cmu.edu/~ site.)ninamf/ML11/lect1117.pdf
- Chatterjee S. ICLR 2020 Coherent Gradients: An approach to understanding generalization in gradient descent-based optimization https://arxiv.org/abs/2002.10657

Lecture 23: GNN Applications (Rex Ying)

- Pinsage: https://arxiv.org/pdf/1806.01973.pdf
- Drug discovery (blog): https://towardsdatascience.com/drug-discovery-with-graph-neural-networks-part-1-1011713185eb
- GNN simulator: https://arxiv.org/pdf/2002.09405.pdf

Lecture 24: Neural Tangent Kernels wide NN not deviate from initialization very much, can use gradient approx. at beginning as fixed estimate, SGD go to global minima

- Jacot et al. **Neural Tangent Kernel: Convergence and Generalization in Neural Networks**, NeurIPS 2018
- Chizat et al. **On Lazy Training in Differentiable Programming** NeurIPS 2019
- Arora, Sanjeev, et al. **On exact computation with an infinitely wide neural net** NeurIPS 2019
- Li, Zhiyuan, et al. **Enhanced Convolutional Neural Tangent Kernels** NeurIPS 2019
- https://towardsdatascience.com/kernel-function-6f1d2be6091
- https://rajatvd.github.io/NTK/

Lecture 25: Loss landscapes and visualizations NN architecture change loss landscape, wider minima better than narrow minima

- Draxler et al. 2019 Essentially No Barriers in Neural Network Energy Landscape
- Li et al. 2018 Visualizing the Loss Landscape of Neural Nets
- Dinh et al 2019 Sharp Minima Can Generalize For Deep Nets
- Keskar, Nocedal, Mudigere, Smelyanskiy, and Tang. On Large-Batch Training for Deep Learning: Generalization Gap and Sharp Minima. https://arxiv.org/pdf/1609.04836.pdf
- Li, Xu, Taylor, Studer, and Goldstein. Visualizing the Loss Landscape of Neural Nets. https://papers.nips.cc/paper/7875-visualizing-the-loss-landscape-of-neural-nets.pdf .
- Horoi et al. Exploring the Geometry and Topology of Loss Landscapes https://arxiv.org/abs/2102.00485

Lecture 26: Review Session for Final Exam