Algorithmic and Statistical Aspects of Deep Learning

Lecture 0: Logistics and course overview

Instructor: Hongyang R. Zhang

Lecture plan

- Course structure
  - Course logistics
  - What do we hope to teach?
  - Connection to existing courses in Northeastern
  - Course work and grading policy
    - High-level plan for problem sets
    - High-level plan for projects
- An overview of course syllabus
  - Theme one: neural networks
  - Theme two: weak supervision
- Basics of neural net: MNIST, two-layer neural net

Lecture plan

- Course structure
  - Course logistics
  - What do we hope to teach?
  - Connection to existing courses in Northeastern
  - Course work and grading policy
    - High-level plan for problem sets
    - High-level plan for projects
- An overview of course syllabus
  - Theme one: neural networks
  - Theme two: weak supervision

Course syllabus

- Instructor: Hongyang Zhang
- TAs: Minghao Liu, John Park
- Time: Mon/Wed 2:50pm – 4:30pm. NUFlex/Zoom

- Other information: see the class webpage or canvas site
  - http://www.hongyangzhang.com/CS7180_Fall2020.html
  - Syllabus, office hours
  - Slides/lecture materials uploaded before every lecture

What do we hope to teach?

1. An understanding of the effective methods of modern neural networks
2. A big picture understanding of how neural networks well
    1. Optimization theory
    2. Generalization theory
    3. Expressivity
3. An understanding of and ability to deal with practical challenges from using neural networks
    1. Training techniques
    2. Getting labeled data more efficiently

Connection to existing courses

- This is the first time this course is taught within Khoury
- Some of the materials overlap with previous offerings of CS 7180, and other courses in deep learning
- Using pytorch rather than tensorflow
- No midterm or final exam

Course work and grading policy

- 3 2-week assignments: 15% x 2 + 10% x 1, total 40%
- Research presentation 15%
  - How to read papers and learn a research area quickly
- Final project 40%
  - ML project management skills, how to quickly get project started and obtain results
- Attendance 5%
- Late day policy: 3 late days in total allowed; after that, grade depreciates by 20% per day
- Collaboration policy: see course webpage

High-level plan for problem sets

- HW1 is hopefully an easy on ramp – an IPython notebook
  - Implement a two-layer neural net
  - Recognize hand-written digits on MNIST
  - Get familiar with gradient calculation
- HW2 touches on theme one
  - IPython notebook: Role of over-parametrization in escaping bad local minima
  - Generalization bounds of multi-layer neural net
- HW3 touches on theme two
  - IPython notebook: Implement a commonly used multi-task learning method
  - Generalization bounds of MTL

High-level plan for research presentation and project

- Research presentation
  - 20 min presentation about a chosen paper
- Final project
  - In-class project pitch
  - In-class project plan/update
  - Final presentation
  - Project report

Lecture plan

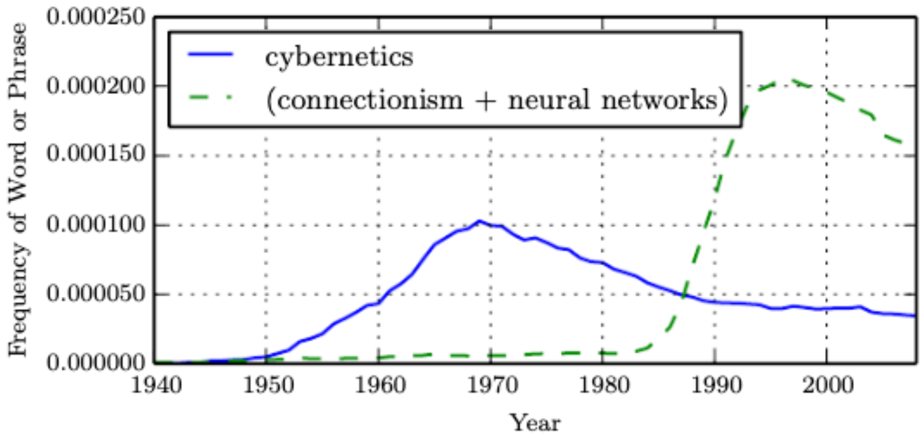
- Course structure
  - Course logistics
  - What do we hope to teach?
  - Connection to existing courses in Northeastern
  - Course work and grading policy
    - High-level plan for problem sets
    - High-level plan for projects
- An overview of course syllabus
  - Theme one: neural networks
  - Theme two: weak supervision

An overview of theme one

- Basics of neural nets
  - Architecture, learning algorithm, loss function
- Optimization theory
  - Concepts of non-convex optimization
  - Convergence analysis of ultra-wide neural net
- Generalization theory
  - Over-parametrization, inductive bias, linear models
  - Matrix sensing, inductive bias of initialization
- Expressivity
  - Guest lecture

A historical perspective: First wave

- 1940-1960s: first wave started with cybernetics, with the development of theories of biological learning
  - Central idea: artificial neural net, brain, neuronscience
  - Linear models: $f(x, w) = x_1w_1 + x_2w_2 + \cdots + x_nw_n$
    - Input: n values $x_1, x_2, \ldots, x_n$
    - Output: y
  - Perception (Rosenblatt, 1958, 1962)
  - However, linear models are limited (Minsky and Papert 1969)
- Frequency of phrases (Google book)



A historical perspective: Second wave

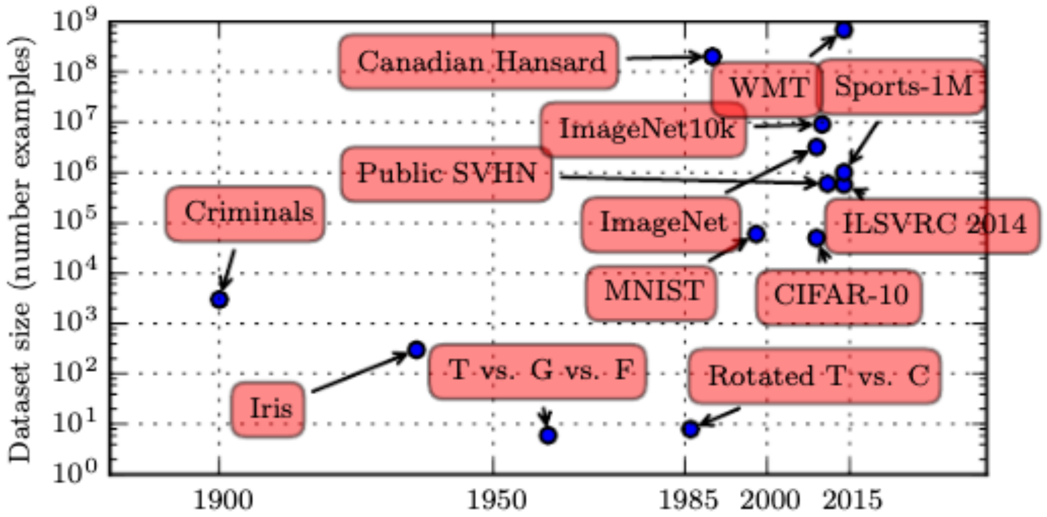
- 1980s – 1990s: second wave emerged via a movement called connectionism or parallel distributed processing
  - Central idea: a large number of simple computational units can achieve intelligent behavior when networked together
  - Distributed representation (Hinton et al 1986)
  - Backpropagation (Rumelhard et al 1986, LeCun 1987)
  - Modeling sequences with neural nets (Hochreiter et al 1991, Bengio et al 1994, Hochreiter and Schumidhuber 1997)

A historical perspective: Third wave

- 2006 to ??: led by a breakthrough that deep belief network can be efficiently trained using a strategy called greedy layer-wise pretraining (Hinton et al 2006)
- 2012: AlexNet achieves SoTA at that time on ImageNet
- Why does neural net research suddenly takeoff?

Dataset sizes

- Dataset sizes have greatly increased over time



Model sizes

- Model sizes have also greatly increased over time
- Fueled by progress in both hardware and software

Year	CNN	Developed by	Place	Top-5 error rate	No. of parameters
1998	LeNet(8)	Yann LeCun et al			60 thousand
2012	AlexNet(7)	Alex Krizhevsky, Geoffrey Hinton, Ilya Sutskever	1st	15.3%	60 million
2013	ZFNet()	Matthew Zeiler and Rob Fergus	1st	14.8%	
2014	GoogLeNet(19)	Google	1st	6.67%	4 million
2014	VGG Net(16)	Simonyan, Zisserman	2nd	7.3%	138 million
2015	ResNet(152)	Kaiming He	1st	3.6%	

Theory of neural net

- While there has been significant progress of using neural net, their theoretical understanding lags behind!
- Recent trends in the theory of deep learning
  - Optimization
  - Generalization
  - Expressivity (or representation)

Lecture plan

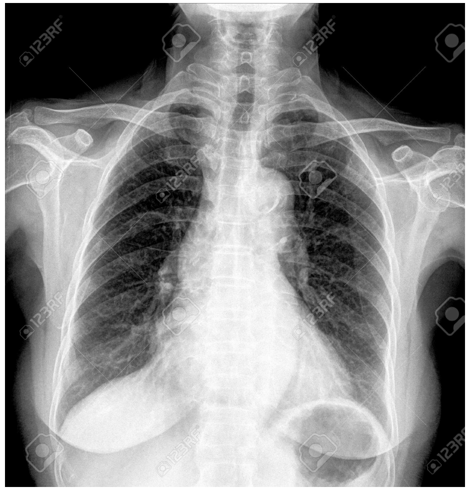
- Course structure
  - Course logistics
  - What do we hope to teach?
  - Connection to existing courses in Northeastern
  - Course work and grading policy
    - High-level plan for problem sets
    - High-level plan for projects
- An overview of course syllabus
  - Theme one: neural networks
  - Theme two: weak supervision

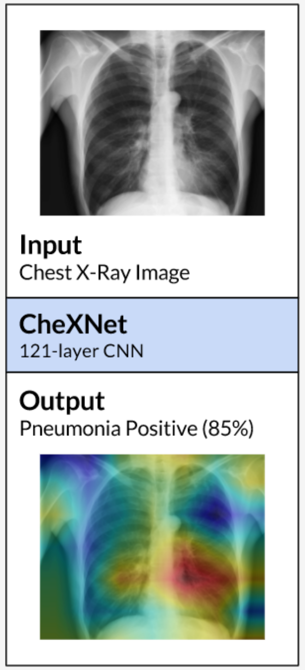
An overview of theme two

- Getting labeled data for neural network training
- Data programming
- Data augmentation
- Transfer learning
- Multi-task learning

Practical challenges in applying neural nets

- Curating large-scale labeled datasets is extremely expensive
  - Medical imaging [CheXNet, Rajpurkar et al 2017]

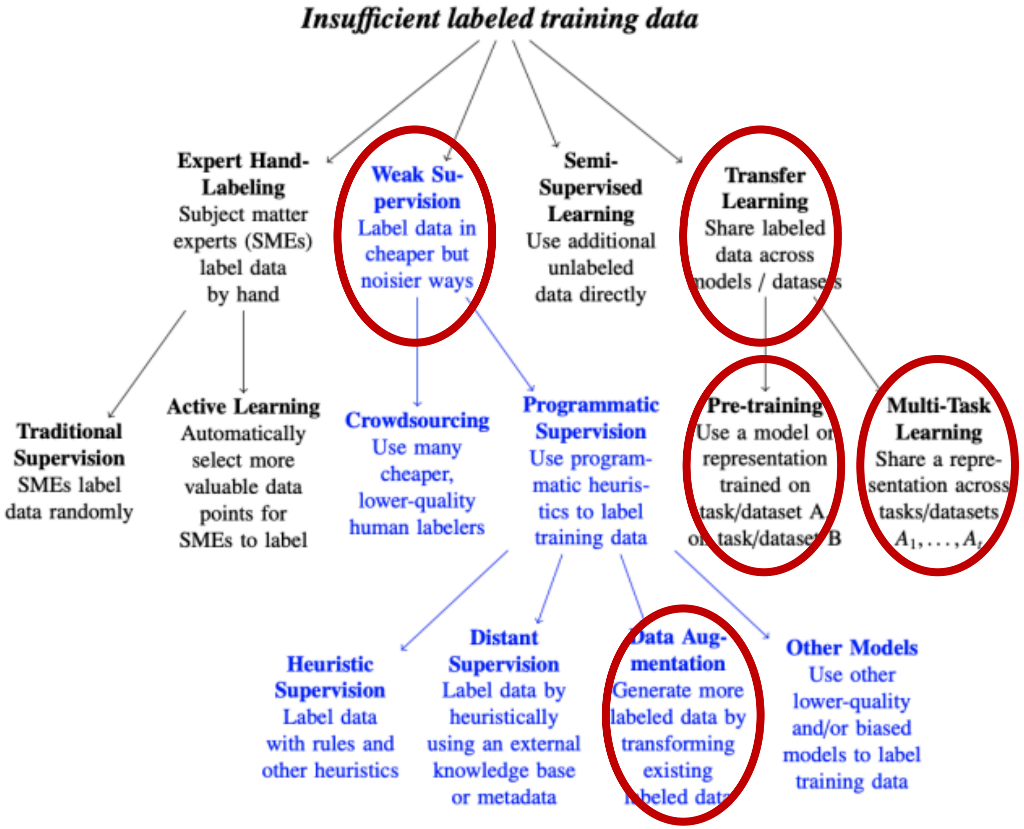




- A prevalent challenge in industrial applications: often data comes in messy formats
  - E.g. tables, pdf files, logs

A chart of weak supervision

- Methods for dealing with insufficient labeled training data



Borrowed from Chris Ré's CS229 lecture slides