CS 559 Deep Learning

Hallmarks of deep learning

Gokberk Cinbis
Recap: Supervised Learning

• Input: $x$ (images, text, emails…)
• Output: $y$ (spam or non-spam…)
• (Unknown) Target Function
  – $f: X \rightarrow Y$ (the “true” mapping / reality)
• Data: $(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)$
• Model / Hypothesis Class
  – $g: X \rightarrow Y$
  – $y = g(x) = \text{sign}(w^T x)$
• Learning = Search in hypothesis space
  – Find best $g$ in model class.
Recap: Synonyms

- Representation Learning
- Deep (Machine) Learning
- Deep Neural Networks
- Deep Unsupervised 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

Slide by Dhruv Batra
Traditional Machine Learning

VISION

hand-crafted features
SIFT/HOG

your favorite classifier

“car”

fixed

learned

SPEECH

hand-crafted features
MFCC

your favorite classifier

\d ē p\n
fixed

learned

NLP

This burrito place
is yummy and fun!

hand-crafted features
Bag-of-words

your favorite classifier

“+”

fixed

learned

Slide Credit: Marc'Aurelio Ranzato, Yann LeCun
It’s an old paradigm

• The first learning machine: the Perceptron
  • Built at Cornell in 1960
• The Perceptron was a linear classifier on top of a simple feature extractor
• The vast majority of practical applications of ML today use glorified linear classifiers or glorified template matching.
Hierarchical Compositionality

**VISION**

- pixels
- edge
- texton
- motif
- part
- object

**SPEECH**

- sample
- spectral
- band
- formant
- motif
- phone
- word

**NLP**

- character
- word
- NP/VP/..
- clause
- sentence
- story

Slide by Dhruv Batra
Building A Complicated Function

Given a library of simple functions

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

Compose into a complicate function

Slide by Dhruv Batra
Given a library of simple functions:

- $\sin(x)$
- $\cos(x)$
- $\log(x)$
- $x^3$
- $\exp(x)$

Compose into a complicated function using linear combinations:

$$f(x) = \sum_i \alpha_i g_i(x)$$

Idea 1: Linear Combinations

- Boosting
- Kernels
- ...

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

Given a library of simple functions

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

Compose into a complicate function

Idea 2: Compositions

- Deep Learning
- Grammar models
- Scattering transforms…

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

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

Idea 2: Compositions

- Deep Learning
- Grammar models
- Scattering transforms…

Slide by Dhruv Batra
Deep Learning = Hierarchical Compositionality

Slide by Dhruv Batra
Deep Learning = Hierarchical Compositionality

Feature visualization of convolutional net trained on ImageNet from [Zeiler & Fergus 2013]

Slide Credit: Marc'Aurelio Ranzato, Yann LeCun
Sparse DBNs

[Lee et al. ICML ’09]

Figure courtesy: Quoc Le
Hierarchical compositionality

- Mathematically makes sense
  
  \[
  \sin(x) \rightarrow x^3 \rightarrow \exp(x) \rightarrow \cos(x) \rightarrow \log(x)
  \]

- Works well in vision and other domains

- How about biological arguments?

  ![Diagram with steps: Low-Level Feature, Mid-Level Feature, High-Level Feature, Trainable Classifier, "car"]
Hubel & Wiesel's experiment

Simple Cells: Response to light orientation
Complex Cells: Response to light orientation & movement
Hypercomplex Cells: Response to movement with an end point

Recording electrode
Visual area of brain
Electrical signal from brain

Cell response

Hubel & Wiesel, 1959
The Mammalian Visual Cortex is Hierarchical

- The ventral (recognition) pathway in the visual cortex
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
fixed

your favorite classifier
learned

"car"

SPEECH

hand-crafted features
MFCC
fixed

your favorite classifier
learned

\'d ē p\n
NLP

hand-crafted features
Bag-of-words
fixed

your favorite classifier
learned

"+

Slide by Dhruv Batra
Feature Engineering

SIFT

Spin Images

HoG

Textons

and many many more....
Traditional approaches are not fully hard coded

Cinbis et al. PAMI 2016
Traditional Machine Learning (more accurately)

**VISION**

- SIFT/HOG
  - Fixed
  - Unsupervised
  - Supervised
  - “Learned”
  - “car”

**SPEECH**

- MFCC
  - Fixed
  - Unsupervised
  - Supervised
  - “d e p”

**NLP**

- Parse Tree Syntactic
  - Fixed
  - Unsupervised
  - Supervised
  - “+”

This burrito place is yummy and fun!

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

• “Shallow” models

- hand-crafted Feature Extractor (fixed)
- “Simple” Trainable Classifier (learned)

• Deep models (especially supervised deep learning)

- Trainable Feature-Transform / Classifier
- Trainable Feature-Transform / Classifier
- Trainable Feature-Transform / Classifier

Learned Internal Representations

Slide by Dhruv Batra
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

Slide by Dhruv Batra
Distributed Representations Toy Example

- Can we interpret each dimension?
Power of distributed representations!

Local

\[
\text{● ● ● ○ ●} = \text{VR} + \text{HR} + \text{HE} = ?
\]

Distributed

\[
\text{● ● ● ○ ●} = \text{V} + \text{H} + \text{E} \approx \bigcirc
\]

Slide by Dhruv Batra

Slide Credit: Moontae Lee
Power of distributed representations!

- United States:Dollar :: Mexico:?
Power of distributed representations!

- Example: all face images of a person
  - 1000x1000 pixels = 1,000,000 dimensions
  - But the face has 3 cartesian coordinates and 3 Euler angles
  - And humans have less than about 50 muscles in the face
  - Hence the manifold of face images for a person has <56 dimensions

- The perfect representations of a face image:
  - Its coordinates on the face manifold
  - Its coordinates away from the manifold

---

Slide by Dhruv Batra
Power of distributed representations!

The Ideal Disentangling Feature Extractor

Pixel n

Ideal Feature Extractor

Pixel 2

View

Pixel 1

Expression

Slide Credit: Marc'Aurelio Ranzato, Yann LeCun

Slide by Dhruv Batra
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

Slide by Dhruv Batra
Benefits of Deep/Representation Learning

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

• New domains without “experts”
  – RGBD
  – 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

- “Maybe the molecule didn’t go to graduate school”
  - Will Welch defending the success of his approximate molecular screening algorithm, given that he’s a computer scientist, not a chemist

Problems with Deep Learning

• **Problem#1: Non-Convex! Non-Convex! Non-Convex!**
  – Depth \( \geq 3 \): 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!”
Problems with Deep Learning

• Problem#2: 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: Hard to track down what's failing

Slide by Dhruv Batra
Problems with Deep Learning

- **Problem#2: 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
    – Caffe, Theano, Torch, TensorFlow, PyTorch...

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

Slide by Dhruv Batra
Yes it works, but how?

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

Slide by Dhruv Batra
NEW NAVY DEVICE LEARNS BY DOING

Psychologist Shows Embryo of Computer Designed to Read and Grow Wiser

WASHINGTON, July 7 (UPI) — The Navy revealed the embryo of an electronic computer today that it expects will be able to walk, talk, see, write, reproduce itself and be conscious of its existence.

The embryo—the Weather Bureau’s $2,000,000 “704” computer—learned to differentiate between right and left after fifty attempts in the Navy’s demonstration for newsmen.

The service said it would use this principle to build the first of its Perceptron thinking machines that will be able to read and write. It is expected to be finished in about a year at a cost of $100,000.

Dr. Frank Rosenblatt, designer of the Perceptron, conducted the demonstration. He said the machine would be the first device to think as the human brain. As do human beings, Perceptron will make mistakes at first, but will grow wiser as it gains experience, he said.

Dr. Rosenblatt, a research psychologist at the Cornell Aeronautical Laboratory, Buffalo, said Perceptrons might be fired to the planets as mechanical space explorers.

Without Human Controls

The Navy said the perceptron would be the first non-living mechanism “capable of receiving, recognizing and identifying its surroundings without any human training or control.”

The “brain” is designed to remember images and information it has perceived itself. Ordinary computers remember only what is fed into them on punch cards or magnetic tape.

Later Perceptrons will be able to recognize people and call out their names and instantly translate speech in one language to speech or writing in another language, it was predicted.

Mr. Rosenblatt said in principle it would be possible to build brains that could reproduce themselves on an assembly line and which would be conscious of their existence.

In today’s demonstration, the “704” was fed two cards, one with squares marked on the left side and the other with squares on the right side.

Learns by Doing

In the first fifty trials, the machine made no distinction between them. It then started registering a “Q” for the left squares and “O” for the right squares.

Dr. Rosenblatt said he could explain why the machine learned only in highly technical terms. But he said the computer had undergone a “self-induced change in the wiring diagram.”

The first Perceptron will have about 1,000 electronic “association cells” receiving electrical impulses from an eye-like scanning device with 400 photo-cells. The human brain has 10,000,000,000 responsive cells, including 100,000,000 connections with the eyes.
COMPUTER SCIENTISTS STYMIED IN THEIR QUEST TO MATCH HUMAN VISION

By WILLIAM J. BROAD
Published: September 25, 1984

EXPERTS pursuing one of man’s most audacious dreams - to create machines that think - have stumbled while taking what seemed to be an elementary first step. They have failed to master vision.

After two decades of research, they have yet to teach machines the seemingly simple act of being able to recognize everyday objects and to distinguish one from another.

Instead, they have developed a profound new respect for the sophistication of human sight and have scoured such fields as mathematics, physics, biology and psychology for clues to help them achieve the goal of machine vision.
Researchers Announce Advance in Image-Recognition Software

By JOHN MARKOFF  NOV. 17, 2014

MOUNTAIN VIEW, Calif. — Two groups of scientists, working independently, have created artificial intelligence software capable of recognizing and describing the content of photographs and videos with far greater accuracy than ever before, sometimes even mimicking human levels of understanding.

Until now, so-called computer vision has largely been limited to recognizing individual objects. The new software, described on Monday by researchers at Google and at Stanford University, teaches itself to identify entire scenes: a group of young men playing Frisbee, for example, or a herd of elephants marching on a grassy plain.

The software then writes a caption in English describing the picture. Compared with human observations, the researchers found, the computer-written descriptions are surprisingly accurate.
Captioned by Human and by Google’s Experimental Program

**Human:** “A group of men playing Frisbee in the park.”

**Computer model:** “A group of young people playing a game of Frisbee.”
a surfboard attached to the top of a car.
this appears to be a small bedroom in the snow.
ImageNet Large Scale Visual Recognition Challenge (ILSVRC)

Classification:
- 1000 object classes
- 1.4M/50k/100k images

Detection:
- 200 object classes
- 400k/20k/40k images

http://image-net.org/challenges/LSVRC/{2010,…,2014}
Data Enabling Richer Models

- [Krizhevsky et al. NIPS12]
  - 54 million parameters; 8 layers (5 conv, 3 fully-connected)
  - Trained on 1.4M images in ImageNet
  - Better Regularization (Dropout)

Slide by Dhruv Batra
ImageNet Classification 2012

- [Krizhevsky et al. NIPS12]: 16.4% error
- Next best team: 26.2% error
Other Domains & Applications

- Vision
- NLP
- Speech
- Robotics
- Game playing

Slide by Dhruv Batra
Figures from Levine et al 2016, Taigman et al., Shelhamer et al, Johnson et al, van den Oord et al, Silver et al.
Why are things working today?

- More compute power
  - GPUs are ~50x faster

- More data
  - $10^8$ samples (compared to $10^3$ in 1990s)

- Better algorithms/models/regularizers
  - Dropout
  - ReLu
  - Batch-Normalization
  - ...
Incoming slides and lectures

● Some fundamentals of machine learning
  ○ A traditional image classification pipeline (Use raw pixels as features and nearest neighbor classifier)
  ○ Linear classification
  ○ Loss functions
  ○ Training by numerical optimization

● Then, deep learning will be a natural extension.