Machine Learning: From Shallow to Deep

Padhraic Smyth
Professor, Department of Computer Science
Director, UCI Data Science Initiative
Research at UC Irvine in Machine Learning
Research at UC Irvine in Machine Learning

Distance in Kilometers (Log-Scale, Base 10)

Histogram of Distances between Tweets

UCI Data Science Initiative

Machine Learning Repository

Center for Machine Learning and Intelligent Systems

Bren School of Information and Computer Science
University of California, Irvine
Progress in Neural Networks
Progress in Neural Networks

Early Progress
Progress in Neural Networks

Early Progress

AI Winter

Progress in Neural Networks


Early Progress

AI Winter

Neural Networks Round 2
Progress in Neural Networks

- Early Progress
- AI Winter
- Neural Networks Round 2
- Realism

Year:
- 1960
- 1970
- 1980
- 1990
- 2000
- 2010
- 2020
Progress in Neural Networks

- Early Progress
- AI Winter
- Neural Networks Round 2
- Realism
- Deep Learning
Progress in Neural Networks

- Early Progress
- AI Winter
- Neural Networks Round 2
- Realism
- Deep Learning
- Where next?

Timeline:
- 1960
- 1970
- 1980
- 1990
- 2000
- 2010
- 2020
Examples of Machine Learning

..... many more applications of this nature: “mapping” raw inputs to semantic output
Figure from Krizhevsky, Sutskever, Hinton, 2012
Figure from Krizhevsky, Sutskever, Hinton, 2012
ImageNet Error Rate

- 2010: 28.2%
- 2011: 25.8%
- 2012: 16.4%
- 2013: 11.1%
- 2014: 6.6%
- 2015: 3.57%

Human: 5.1%

Figure from Kevin Murphy, Google, 2016
Word Error Rates in Speech Recognition

Figure from Microsoft Research
How are these Classifiers built?

• Collect (large) number of input/output pairs, e.g.,
  Inputs = pixel images
  Outputs = labels of objects provided by humans

• Optimization algorithms learn classifier parameters
  Built on basic ideas from calculus
  ....with many bells and whistles

• Significant amount of exploration and trial-and-error
  How many layers in the network?
  How many units at each layer?
  What non-linearities to use?
  .....and many more options
"Shallow Classifier"

The x’s could be pixels in a 2 x 2 image

Each x gets multiplied by a weight

Weights are like a filter or template

Key idea: the weights are learned
A Neural Network with 1 Hidden Layer

Multiple sets of weights (multiple filters)

Nonlinearity for each filter

Next layer combines filter outputs to produce a predicted label
Deep Network Classifiers: 2 or More Hidden Layers

These network models can represent very flexible highly non-linear functions
Visualizing what is Learned in the Weights

Figure from Lee et al., ICML 2009
This has resulted in many practical applications

Google Photos app (2015) can automatically predict labels from a set of ~18k. Trained on ImageNet and (noisy) internal dataset.

Figure courtesy of Kevin Murphy, Google, 2016
Results on CityScapes

Figure courtesy of Kevin Murphy, Google, 2016
The Good

A woman is throwing a **frisbee** in a park.

A **dog** is standing on a hardwood floor.

A little **girl** sitting on a bed with a teddy bear.

A group of **people** sitting on a boat in the water.

Figure courtesy of Kevin Murphy, Google, 2016
The Not so Good

A woman holding a clock in her hand.

A man wearing a hat and a hat on a skateboard.

Figure courtesy of Kevin Murphy, Google, 2016
An Image Recognition Network

From Nguyen, Yosinski, Clune, ArXiv preprint, 2014

Images used for Training
An Image Recognition Network

From Nguyen, Yosinski, Clune, ArXiv preprint, 2014

Images used for Training

New Images
An Image Recognition Network

From Nguyen, Yosinski, Clune, ArXiv preprint, 2014

Images used for Training

New Images
An Image Recognition Network

Images used for Training

New Images

From Nguyen, Yosinski, Clune, ArXiv preprint, 2014
SAY BIG DATA
ONE MORE TIME
The Dark Side of Deep Learning

• Algorithms require massive amounts of training data
  (Getting labeled data is expensive)

• Model building and optimization is a dark art
  (true experts are hard to find)
Deep Network architecture for GoogLeNet network, 27 layers

Don’t try this at home....
The Dark Side of Deep Learning

• Algorithms require massive amounts of training data
  (Getting labeled data is expensive)

• Model building and optimization is a dark art
  (true experts are hard to find)

• On some problems much simpler models often work just as well
  (but cost a lot less in consulting fees)

• Results on standard data sets are not the same as the real-world
Figure from cacm.acm.org
Where has Deep Learning been most Successful?

- Very large amounts of human-labeled training data

- Mapping/transformation problems
  - E.g., pixels -> words or labels

- Problems where the mappings don’t change
  - Train once, predict often

- Problems where it’s worth a very large investment to solve the problem once
  - Image classification
  - Speech recognition
Closing Advice

Determine if a problem is suitable for machine learning

Start in the “shallow end”

Proceed to the “deep end” with caution

Beware the hype 😊
BACKUP SLIDES
Learning Mappings End to End

Traditional Pattern Recognition: Fixed/Handcrafted Feature Extractor

Mainstream Pattern Recognition (until recently)

Deep Learning: Multiple stages/layers trained end to end
Learning Features from Pixels
Multiple Layers = Learned Templates

Figure courtesy of Kevin Murphy, Google, 2016
What’s a Deep Neural Network Classifier?

- Multiple transformation layers from input to output
  - Each layer multiplies the outputs of the previous layer by a set of weights (like a filter)
  - Passes the result through a non-linearity

- In principle can represent very complex mappings
  - ......but require a lot of data to build

Figure courtesy of Kevin Murphy, Google, 2016
Terminology

Large-scale Data Analysis
Data Mining
Data Science
Big Data
Machine Learning
Computational Statistics

......
Terminology

Large-scale Data Analysis
Data Mining
Data Science
Big Data
Machine Learning
Computational Statistics

......Using computer algorithms to analyze data sets that are too large and complex for humans to work with