Deep Neural Nets (& Security)
from ZIP codes to Autonomous Vehicles

Matthew Merzbacher

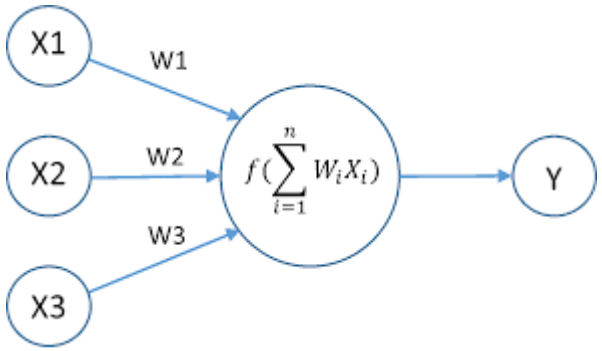
/ November 16, 2016 /

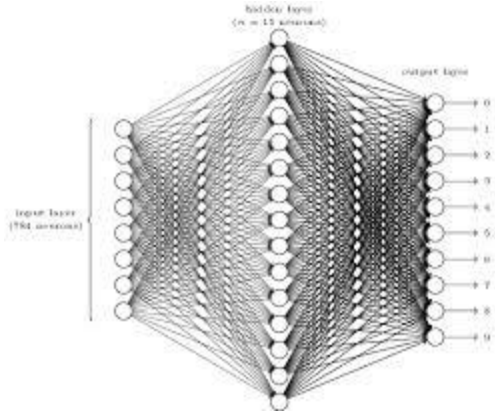
WILL DEEP LEARNING WORK FOR SECURITY?

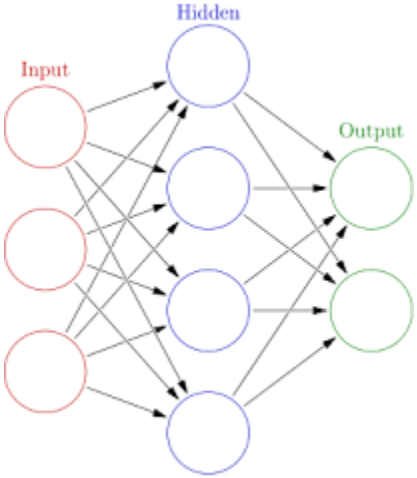
- **Promising in a myriad of fields**
  - Automated & Tunable
- **But…**
  - No transfer function → no explanations or understanding of “why”
  - Domain may not allow adaptive algorithms
  - Small & thin objects challenging
- **Better in closed-world**
- **Still…**
  - Needs to be explored and assessed
- **Outline**
  - Introduction to Deep Learning
  - Security Questions

BRIEF INTRO TO NEURAL NETWORKS

- **A gift that keeps on giving**
- **Simple Model (1965)**
- **Training by Backpropagation**
  - Requires limited model
- **Postal addresses (1997)**
  - 10% initially, now 95%

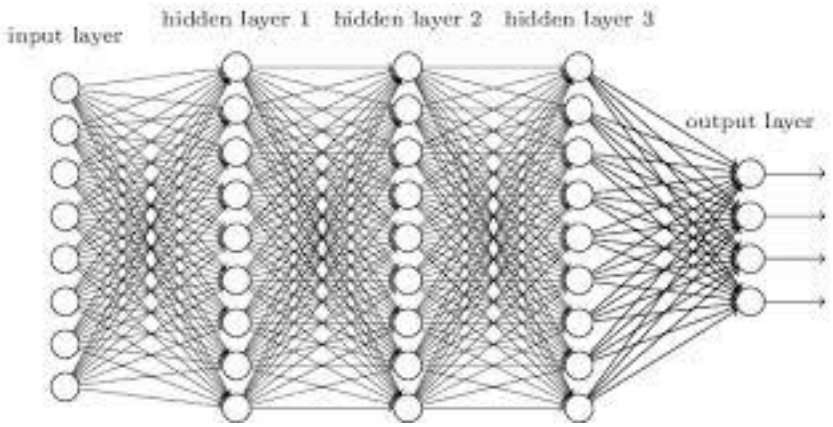






IF ONE HIDDEN LAYER IS GOOD…

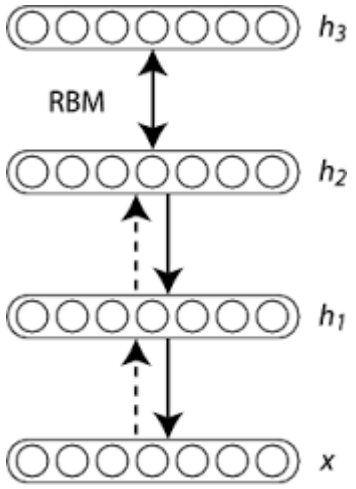
- **Multi-Layer Networks**
- **Problems**
  - Curse of Dimensionality
  - Training critical, extremely hard
    - Computationally expensive
    - Easy to overfit fully-connected network
    - Requires lots of training data
  - Vanishing Gradient problem
  - Can be solved by network architecture, but that requires domain expertise
- **Answer: Deep Learning**
  - Abstraction of layers
  - May model neuroscience



A COUPLE OF COOL IDEAS FROM 2006 – 2007

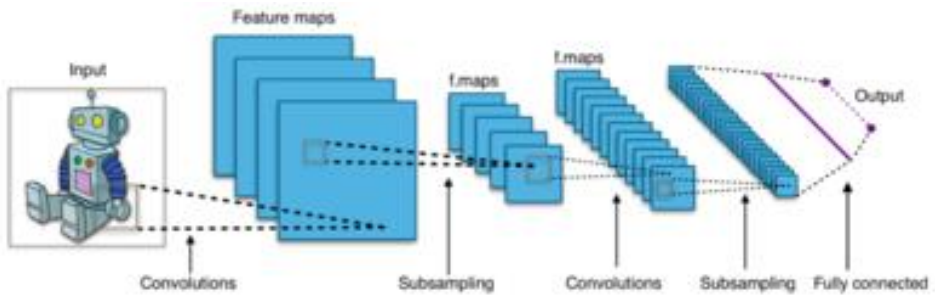
➔ Deep Belief Network

- Hinton [U. Toronto -> Google]
- Forward train one layer at a time and then touch up with backpropagation
- Dramatic reduction in training data needed
- Can be adaptive over time



➔ Convolutional Neural Nets

- LeCun [NYU -> Facebook]
- Inspired by Biology
  - Repeated convolution layer of local neurons [Depth]
  - Locality of connection
  - Pooling for abstraction
  - ReLu layer for non-linearity
- Repeat, as needed
- Final fully connected layer



APPLICATION: WHERE'S WALDO'S BACKPACK?

Backpack

Flute

Strawberry

Traffic light

Backpack

Matchstick

Sea lion

Bathing cap

Racket


WAIT

RESULTS

Image Recognition

- ImageNet Large Scale Visual Recognition Challenge
  - 1.4M images
  - Trying to locate 1000 features
- Performance close to humans
- Precision 0.44, Classification Error 6.7%
- Challenges:
  - Small & thin objects
  - Filtered images

NLP

- Other approaches (perhaps hybrid) may be better

Having consistent feedback invaluable

- Data is still King!

WILL IT WORK FOR SECURITY?

- **Promising**
  - Automated & Tunable
- **But…**
  - No transfer function → no explanations or understanding
  - Security domain may not allow adaptive algorithms
  - Small & thin objects challenging
  - Better in closed-world
- **Given recent spectacular failures of Predictive Analytics, how do we proceed prudently?**

THANK YOU!

➔ Some Resources

- DeepLearning.TV (YouTube)
- KDNuggets
- Deeplearning.net
- Image-net.org