Neural Networks Beginnings

Computer Vision (CSCI 5520G)

Faisal Z. Qureshi

http://vclab.science.ontariotechu.ca

A brief history of Neural Networks

- Claude Shannon, Father of Information Theory.

  I visualise a time when we will be to robots what dogs are to humans, and I'm rooting for the machines.

- Jeff Hawkins, Founder of Palm Computing.

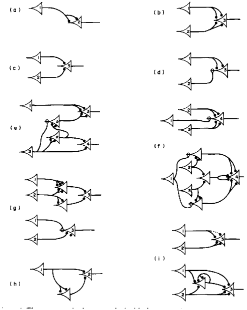
  The key to artificial intelligence has always been the representation.

Lesson Plan

- Computational models of Neurons
- Pre-deep learning
- Imagenet 2012
- Takeaways
  - What
  - How
  - Why now?
  - Impact
- Ethical and social implications

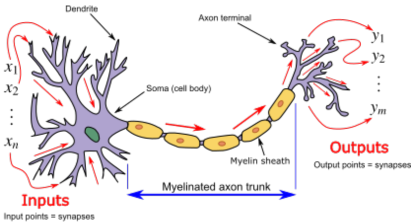
McCulloch and Pitts (1943)

- Proposed a model of nervous systems as a network of threshold units.
- Connections between simple units performing elementry operations give rise to intelligence.

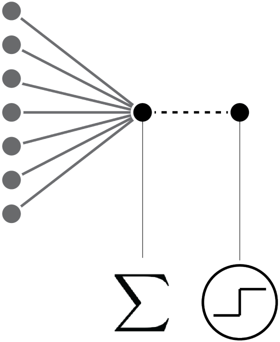


Threshold units (Neuron)

- Neuron (picture from Wikipedia)



Threshold units (Artificial neuron)



Learning via reinforcing connections between Neurons (1949 to 1982)

Hebbian Learning

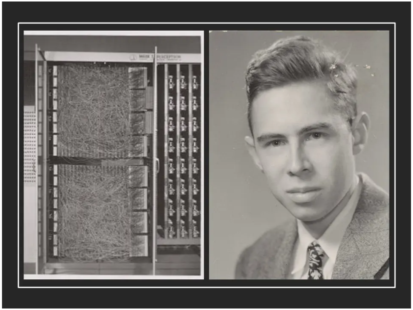
- Hebbian Learning (Donald Hebb, 1949) principle proposes to learn patterns by reinforcing connections between Neurons that tend to fire together.
  - Biologically plausible, but it is not used in practice

Hebbian Learning

- Hebbian Learning (Donald Hebb, 1949) principle proposes to learn patterns by reinforcing connections between Neurons that tend to fire together.
  - Biologically plausible, but it is not used in practice
- First artificial neural network consisting of 40 neurons (Marvin Minsky, 1951)
  - Uses Hebbian Learning

Perceptron

- Frank Rosenblatt (1958) perceptron to classify 20x20 images
  - Percpetron is neural network comprising a single neuron



Cat visual cortex

- David Hubel and Torsten Wiesel studied cat visual cortex and showed that visual information goes through a series of processing steps:
- edge detection;
- edge combination;
- motion perception; etc. (Hubeland Wiesel, 1959)

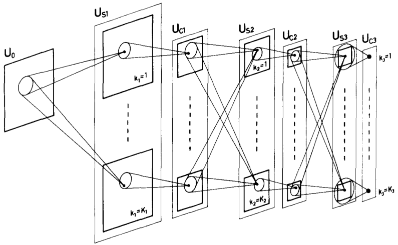
Backpropatation

- Backpropagation for artificial neural networks (Paul Werbos, 1982)
- An application of chain-rule from differential calculus

Towards (deep) neural networks

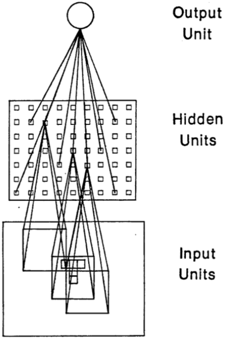
Neocognitron

- Fukushima (1980) implemented Neocognitnron that was capable of handwritten character recognition.
  - This model was based upon the findings of Hubel and Wiesel.
  - This model can be seen as a precursor of modern convolutional networks.



Hidden units and backpropagation

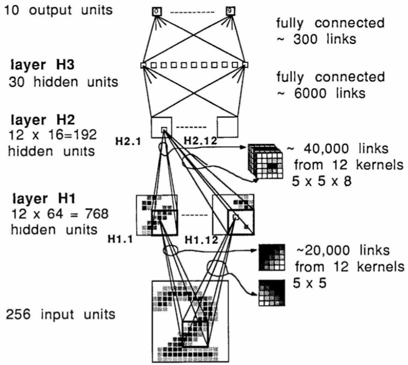
- Rumelhart et al. (1988) used backpropagation to train a network similar to Neocognitron.
  - Units in hidden layers learn meaningful representations



LeNet

- In 1989, LeCun et al. proposed LeNet, a convolution neural network very similar to networks that we see today
  - Capable for recognizing hand-written digits
  - Trained using backpropagation

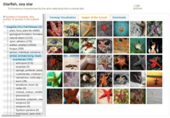
LeNet family

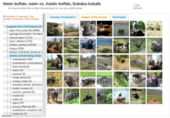


Deep learning (the beginning)

ImageNet Large Scale Visual Recognition Challenge

- Large amount of training data is critical to the success of deep learning methods
- ImageNet challenge was devised to capture the performance of various image recognition methods
  - 1 million images belonging to 1000 different classes
  - It's size was key to the development early deep learning models



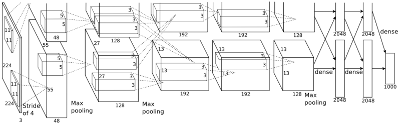


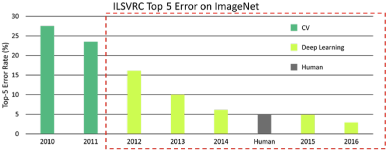
Datasets

- Datasets used for deep learning model develop are divided into three sets:
  - Training set is used train the deep learning model;
  - Validation set is used to tune the hyperparameters, implement early stopping, etc.; and
  - Test set is used to evaluate model performance.

AlexNet (2012)

- Krizhevsky et al. trained a convolution network, similar to LeNet5, but containing far more layers, neurons, and connections, on the ImageNet Challenge using Graphical Processing Units (GPUs). This model was able to beat the state-of-the-art image classification methods by a large margin.

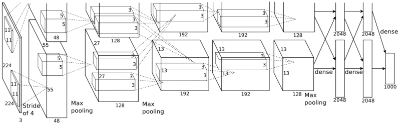


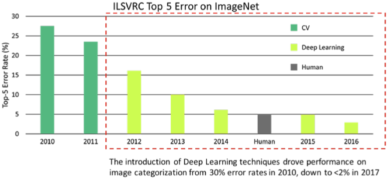


The introduction of Deep Learning techniques drove performance on image categorization from 30% error rates in 2010, down to <2% in 2017

AlexNet (2012)

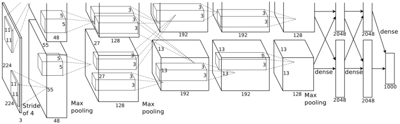
- Krizhevsky et al. trained a convolution network, similar to LeNet5, but containing far more layers, neurons, and connections, on the ImageNet Challenge using Graphical Processing Units (GPUs). This model was able to beat the state-of-the-art image classification methods by a large margin.
- GPUs are criticial to the success of deep learning methods.

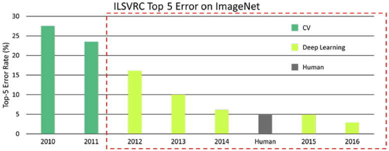




AlexNet (2012)

- Krizhevsky et al. trained a convolution network, similar to LeNet5, but containing far more layers, neurons, and connections, on the ImageNet Challenge using Graphical Processing Units (GPUs). This model was able to beat the state-of-the-art image classification methods by a large margin.
- GPUs are criticial to the success of deep learning methods.





The introduction of Deep Learning techniques drove performance on image categorization from 30% error rates in 2010, down to <2% in 2017

- Models may outperform humans!?

Deep learning takes over (2012 onwards)

- Large datasets and vast GPU compute infrastructures led to larger and more complex deep learning models for solving problems in a variety of domains ranging
  - from computer vision to speach recognition,
  - from medical imaging to text understanding,
  - from computer graphics to industrial design,
  - from autonomous driving to drug discovery, etc.

Takeaways

What

- Deep learning is a natural extension of artificial neural networks of the 90s.
  - Extracts useful patterns from data
  - Learns powerful representations
  - Reduces the “semantic gap”

How

- Chain rule (or backpropagation)
  - Computes how error (or more generally, the quantity to optimize) changes when model parameters change

How

- Chain rule (or backpropagation)
  - Computes how error (or more generally, the quantity to optimize) changes when model parameters change
- Stochastic gradient descent
  - Iteratively update network parameters to "minimize the error" **(How)**

How

- Chain rule (or backpropagation)
  - Computes how error (or more generally, the quantity to optimize) changes when model parameters change
- Stochastic gradient descent
  - Iteratively update network parameters to "minimize the error" **(How)**
- Convolutions
  - Bakes in the intuition that signal is structured and often has some stationary properties
  - Allows processing of large signals

How

- Chain rule (or backpropagation)
  - Computes how error (or more generally, the quantity to optimize) changes when model parameters change
- Stochastic gradient descent
  - Iteratively update network parameters to "minimize the error" **(How)**
- Convolutions
  - Bakes in the intuition that signal is structured and often has some stationary properties
  - Allows processing of large signals
- Hidden layers

Why now?

- GPUs that support vectorized processing (tensor operations)
- Large datasets

Engineering advances

- Computationally speaking, a deep learning model can be formalized as a graph of tensor operations:
  - Nodes perform tensor operations; and
  - Results propagate along edges between nodes.

Engineering advances

- Computationally speaking, a deep learning model can be formalized as a graph of tensor operations:
  - Nodes perform tensor operations; and
  - Results propagate along edges between nodes.
- Provides new ways of thinking about deep learning models.
  - Recursive nature: each node is capable of sophisticated, non-trivial computation, perhaps leveraging another neural network

Engineering advances

- Computationally speaking, a deep learning model can be formalized as a graph of tensor operations:
  - Nodes perform tensor operations; and
  - Results propagate along edges between nodes.
- Provides new ways of thinking about deep learning models.
  - Recursive nature: each node is capable of sophisticated, non-trivial computation, perhaps leveraging another neural network
- Autodiff
  - Techniques to evaluate the "derivative of a computer program"

Engineering advances

- Computationally speaking, a deep learning model can be formalized as a graph of tensor operations:
  - Nodes perform tensor operations; and
  - Results propagate along edges between nodes.
- Provides new ways of thinking about deep learning models.
  - Recursive nature: each node is capable of sophisticated, non-trivial computation, perhaps leveraging another neural network
- Autodiff
  - Techniques to evaluate the "derivative of a computer program"
- Deep learning frameworks
  - PyTorch
  - TensorFlow
  - etc.

Impact

- Image classification
- Face recognition
- Speech recognition
- Text-to-speech generation
- Handwriting transcription
- Medical image analysis and diagnosis
- Ads
- Cars: lane-keeping, automatic cruise control

Social and ethical implications

- Myth
  - Killer robots will enslave us
- Reality
  - Deep learning (and more generally, artificial intelligence) will have a profound effect on our society
    - Legal, social, philosophical, political, and personal

Copyright and License