Overview: CE/CZ 4042

Neural networks and deep learning

Professor Jagath Rajapakse

Biological Neural Networks

Artificial neural networks are inspired by the biological neural networks in the brain

The three pounds of jelly-like material found within our brain is the most complex machine on earth and perhaps in the universe.

It consists of a densely interconnected set of nerve cells, or basic information-processing units, called **neurons**.

Human brain incorporates nearly 10 billion neurons, each connected to about 10,000 other neurons with 60-100 trillion connections, *synapses*, between them.

By using multiple neurons simultaneously, the brain performs its functions much faster than the fastest computers in existence today.

A **neural network** is defined as a model of reasoning based on the human brain.

Concise description of brain anatomy: https://mayfieldclinic.com/pe-anatbrain.htm

Biological Neural Networks

Typical operating speeds of biological neurons is in milliseconds (10^{-3} s), while silicon chip operate in nanoseconds (10^{-9} s). But Brain makes up (for slower rate of operation of a neuron) by having significant number of neurons with massive interconnections between them.

Human brain is extremely energy efficient, using approximately 10^{-16} joules per operation per second, whereas the computers use around 10^{-6} joules per operation per second.

Brains have been evolving for tens of millions of years, computers have been evolving for tens of decades.

Biological Neural Networks

Our brain is highly complex, non-linear parallel information-processing system.

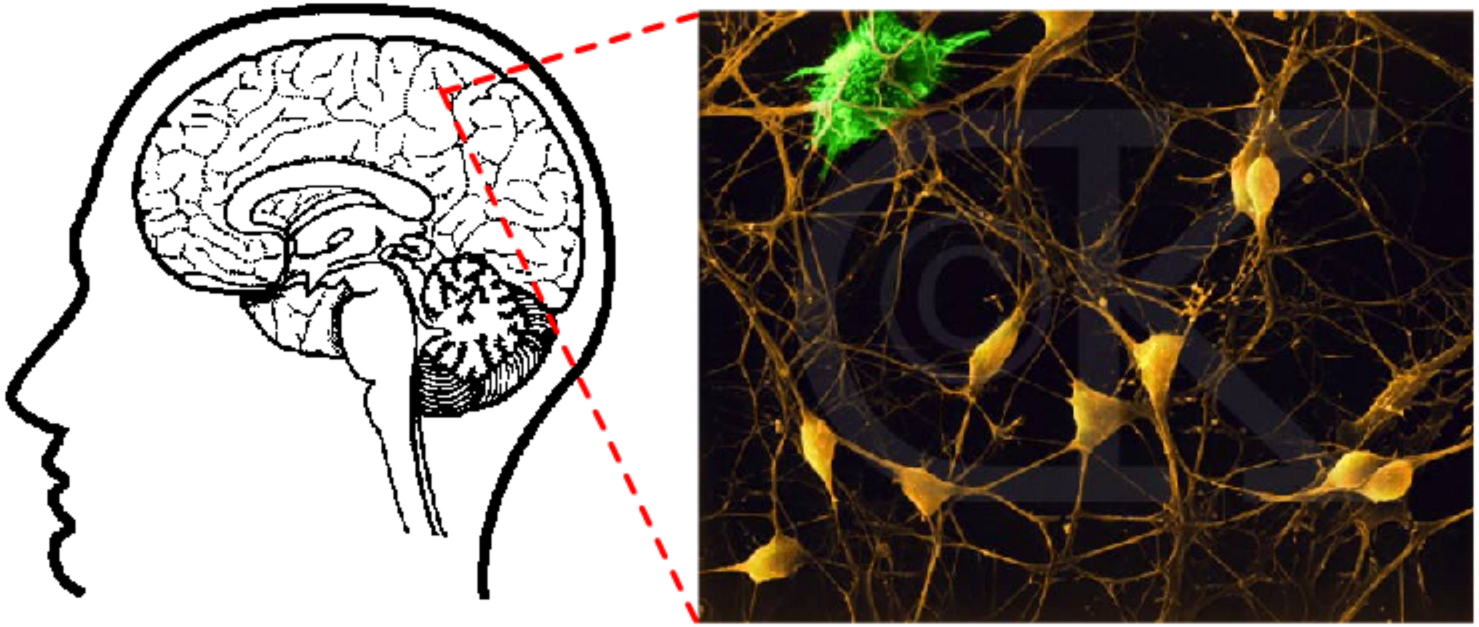
Information is stored and processed in a **neural network** simultaneously throughout the whole network, rather than at specific locations.

Learning is a fundamental and essential characteristic of biological neural networks.

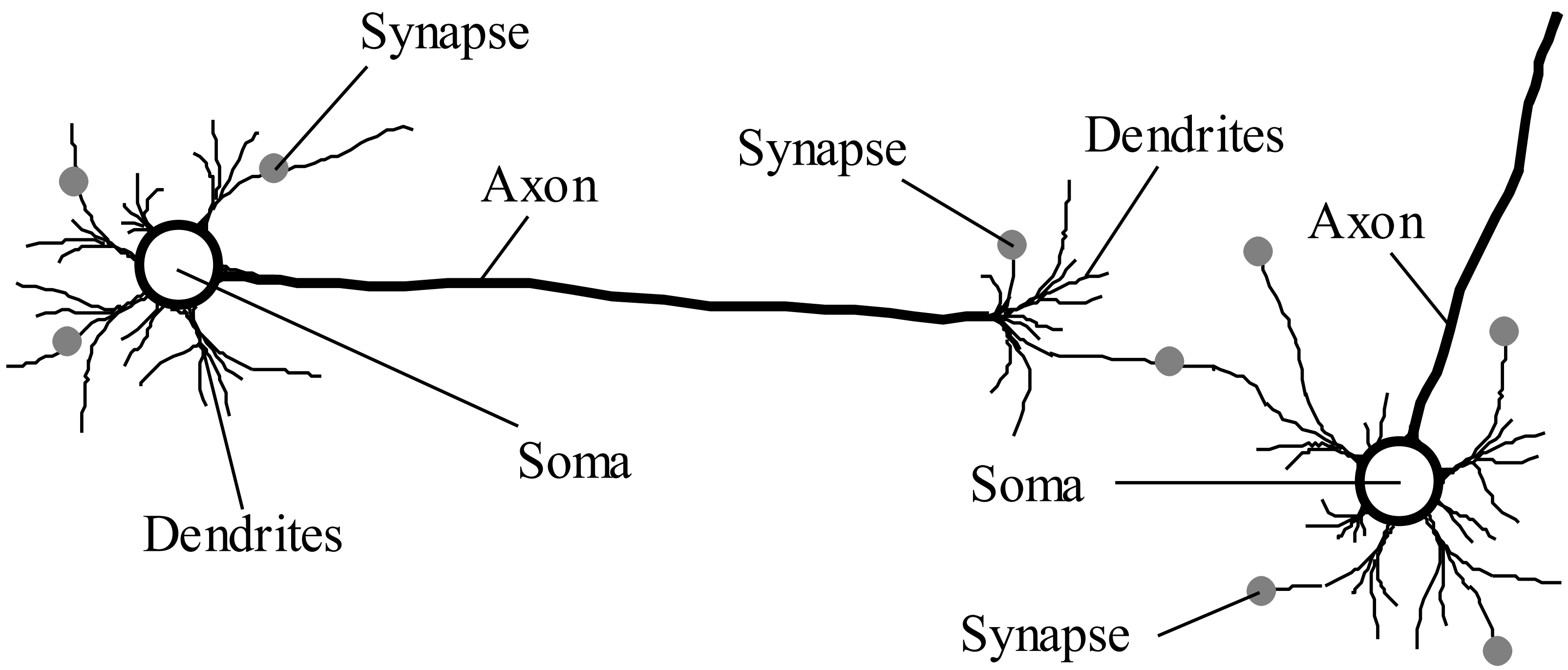
The ease with which they can learn, led to attempts to emulate a **biological neural network** in a computer.

Biological Neural Networks

Each of the yellow blobs in the picture above are neuronal cell bodies (soma), and the lines are the input and output channels (dendrites and axons) which connect them.



Biological Neurons



Biological neurons

A **biological neuron** consists of the following components:

- ❖ Soma: Cell body which processes incoming activation signals and converts input into output activations. The nucleus of soma contains the genetic material in the form of DNA.
- ❖ Axon: transmission lines that send activation signals to other neurons
- ❖ Dendrites: receptive zones that receive activation signals from other neurons
- ❖ Synapses: allow weighted signal transmission between the dendrites and axons. Process of transmission is by diffusion of chemicals.

Although neuronal cell body performs majority of cells function, most of the cells total volume is taken up by axons (about 90%).

Signal flow in neurons

Each neuron receives **electrochemical inputs (neurotransmitters) from other neurons at the dendrites.**

Soma sums the incoming signals. If the sum of these electrical inputs is sufficiently powerful to activate the neuron, it **transmits an electrochemical signal** along the **axon**, and passes this signal to the other neurons whose **dendrites** are attached at any of the **axon** terminals.

These attached neurons may then fire. Signals flow along the axon as a form of electric pulses.

Note that synapses could be either excitatory or inhibitory.

Signal flow in neurons

Hence a neuron fires only if the total signal received at the cell body exceeds a certain level from resting state (-70mV). The neuron either fires or it doesn't, there aren't different grades of firing.

So, our entire brain is composed of these interconnected electro-chemical transmitting neurons.

From a very large number of extremely simple processing units (each performing a weighted sum of its inputs, and then firing a binary signal if the total input exceeds a certain level) the brain manages to perform extremely complex tasks.

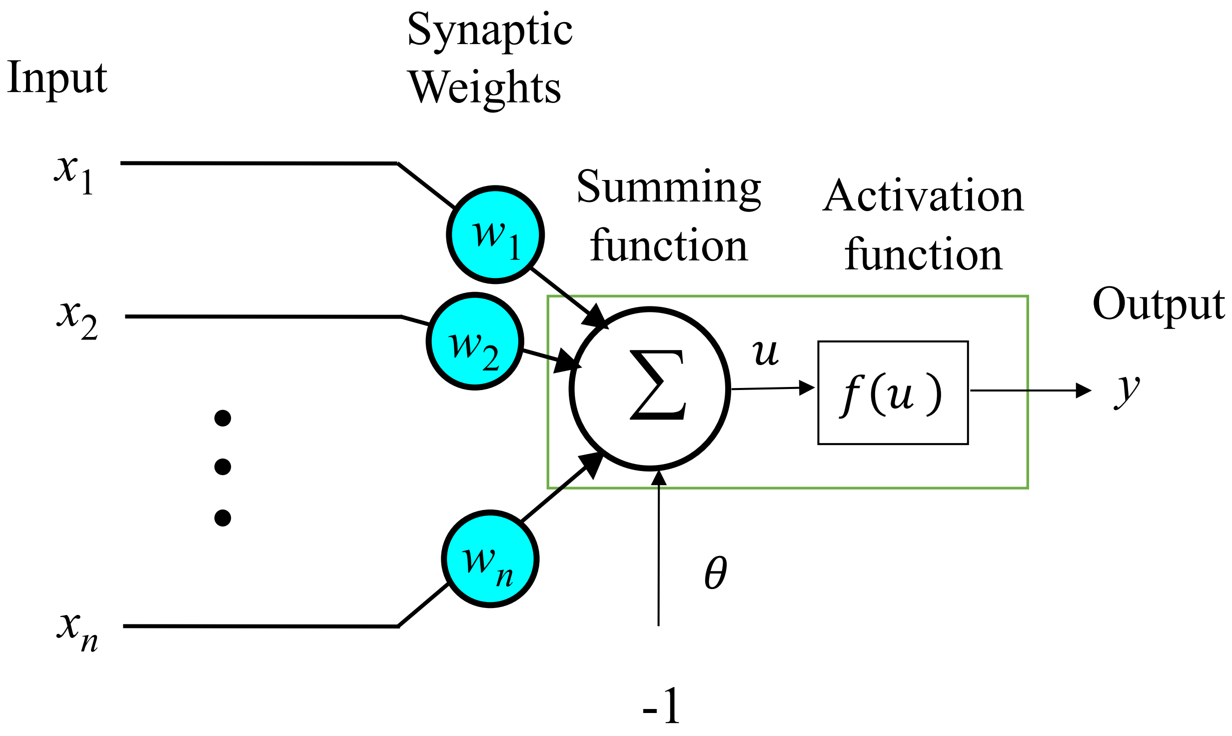
Artificial neural networks

Artificial neural networks attempt to mimic biological neural networks in the brain.

There are two types of artificial neural networks. One that emulates the action potentials are referred to as ‘**spiking**’ neural networks and the other that emulates the aggregate of action potentials are ‘**rate-based**’ neural networks.

Neural networks discussed in this class are rate-based. However, spiking neural networks are more amenable for hardware implementations.

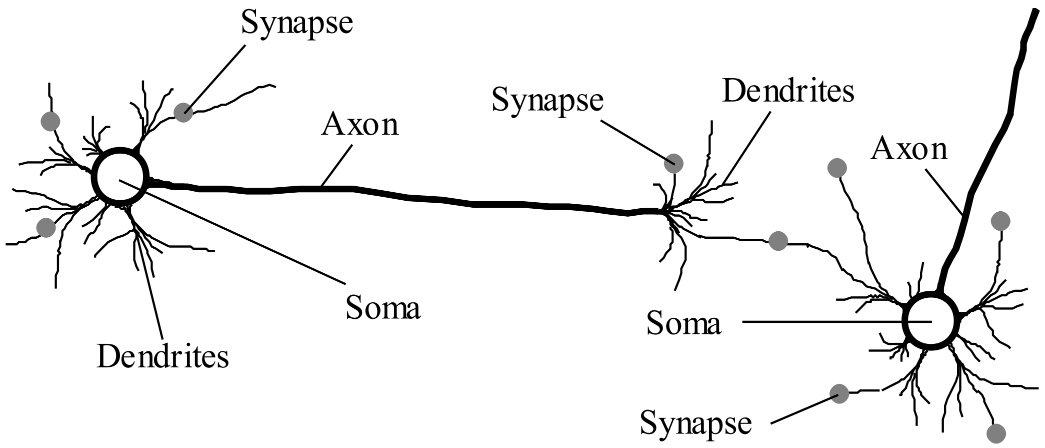
Artificial Neurons



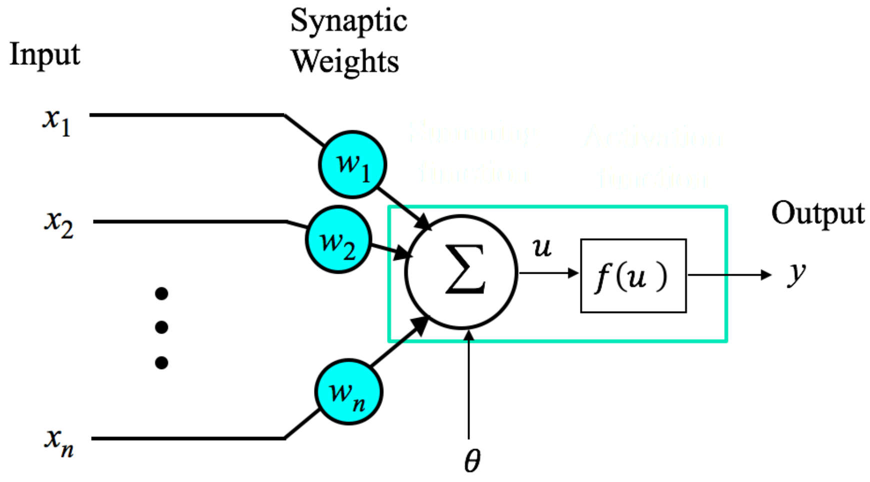
Input $\boldsymbol{x} = (x_1 \quad x_2 \quad \cdots \quad x_n)^T$
u - synaptic input
$f(u)$ - activation
y - output

Analogy between biological and artificial neuron

Biological neuron



Artificial neuron

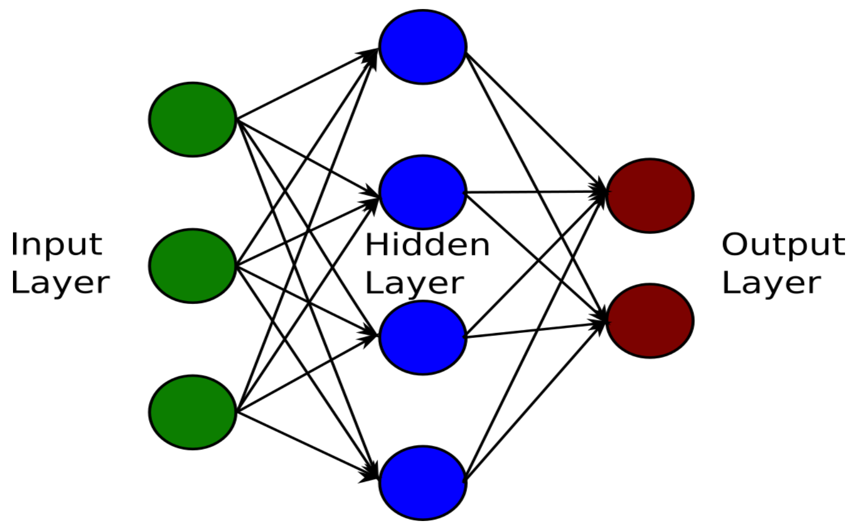


Analogy between biological and artificial neuron

Biological Neuron	*Artificial Neuron*
Soma	Sum + Activation function
Dendrite	Input
Axon	Output
Synapse	Weight

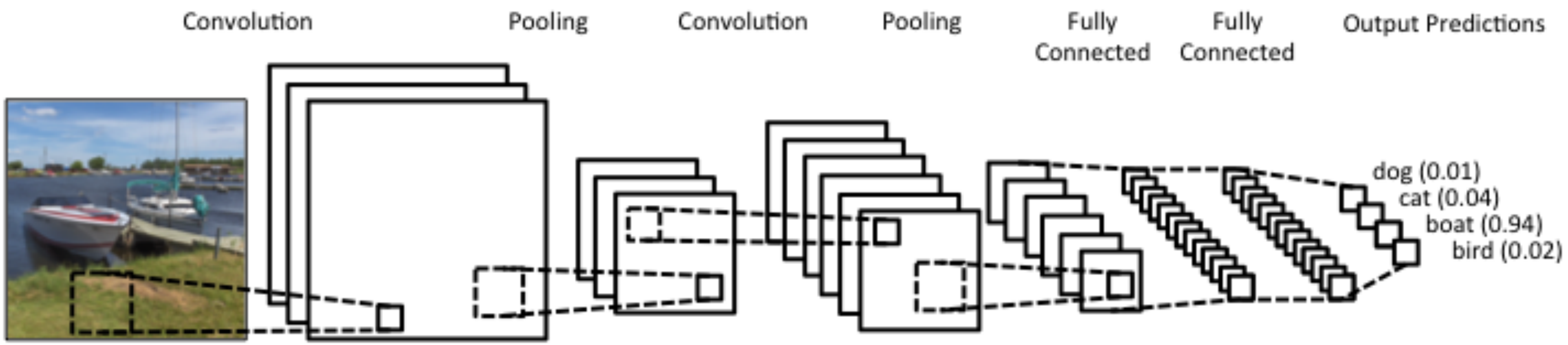
- *McCulloch-Pitts neuron* (~ 1940) is an artificial neuron with binary inputs and outputs
- *Perceptron* (~ 1950) is another name for an artificial neuron with analog inputs

Artificial Neural Networks



Three-layer feedforward neural network

Deep Neural Networks



Deep convolution neural network

Applications of Neural Networks

Neural network is a computational paradigm for machine learning and data analytics.

Inputs $\rightarrow$ NN $\rightarrow$ Outputs

Applications: for example, predictive analytics:

- Regression: outputs are continuous variables
- Classification: outputs are discrete variables

Types of input data

- Statistical/Financial
- Text
- Images

History

Modern view of Neural Networks (NN) began in the 1940s. – Warren McCulloch, Walter Pitts, Donald Hebb.

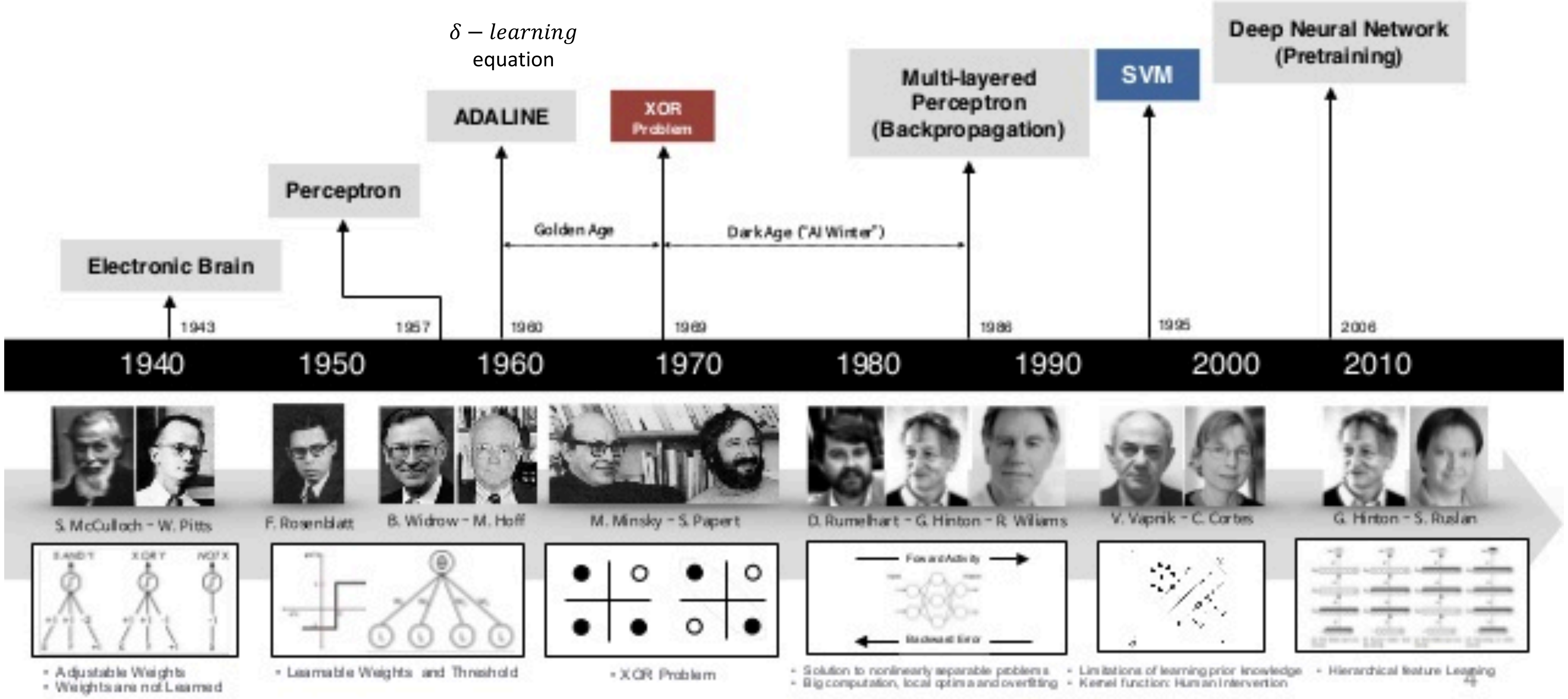
By late Sixties, most of the basic ideas and concepts necessary for neural computing had already been formulated.

Practical solution emerges only in the mid-eighties. Major reason for the delay was technological: no powerful workstations to model and experiment with ANN; algorithms for learning large neural networks were unknown.

Emergence of deep neural networks since 2012: human like performance was achieved in object recognition, using deep convolutional neural networks.

Brief History of Neural Network

DEVIEW 2015



CE/CZ 4042 Learning Objectives

1. Interpret artificial neuron as an abstraction of biological neuron and explain how it can be used to build deep neural networks that are trained to perform various tasks such as regression and classification

2. Identify the underlying principles, architectures, and learning algorithms of various types of neural networks;

3. Select and design a suitable neural network for a given application;

4. Implement deep neural networks that can efficiently run on computing machines.

Pre-requisites: CE/CZ1011, CE/CZ1012, CE/CZ1003, CE/CZ1007
Comfortable with some Mathematics. Linear Algebra. Basic Calculus.

Course Hours

- **Lectures: Friday 11:30am – 1:30pm (ONLINE)**
- **Tutorial: Wednesday 2:30 – 3:30pm (ONLINE),**

  Tutorials start from 3rd week

- **Part-time**
  - **Lecture: Thursday 6:30 – 8:30pm (ONLINE)**
  - **Tutorials: Thursday 8:30 - 9:30pm (ONLINE),**

    Tutorials start from 2nd week

Course Topics

First Half:

1. Neural network basics
2. Regression
3. Classification
4. Layers of neurons
5. Feedforward networks
6. Model selection and overfitting

Second Half:

7. Convolution neural networks (CNN)
8. Recurrent neural networks (RNN)
9. Gated RNN
10. Autoencoders
11. Generative adversarial networks (GAN)

Python 3.8.x and Tensorflow 2.x

- **Python 3.8.x** is the programming language
- **Tensorflow 2.2 Libraries:**
  - TensorFlow: https://www.tensorflow.org/
- Codes of lecture examples and tutorials will be provided.

Assessment

- **Programming Assignments (50%)** – Individual
  - **Assignment 1:** handout Sept 21, deadline Oct 16
  - **Assignment 2:** handout Oct 19, deadline Nov 13

Codes and a report are to be submitted to NTULearn by the deadline

Assessment

- **Project (50%)** – Group (up to Three)
  - Project ideas handout: August 28
  - Deadline: Nov 6

The students are to propose the project and form project groups. The topic could also be selected from the ideas given.

The project includes potential research issue related to neural networks theory/application, literature survey, and design and implementation of a potential solution. Comparisons with existing solutions are to be presented.

A report, codes, and a video presentation are to be submitted to NTU Learn by the deadline by one of the group members. The project report should contain the names of all the project members.

Assignments and Projects

- Python and Tensorflow are recommended for assignments and projects
- PC with at least 1 GPU is recommended
- Access to SCSE GPU-TC server for those who needs computational power. Students will have accounts after add-and-drop period is over. Email: scsegpu-tc@ntu.edu.sg
- Reports are to be submitted in pdf format and codes are to be submitted in a .zip file to NTU Learn before the deadline.
- The cover page of the project report should have the names of all the group members.
- Late submissions are penalized (each day at 5% up to 3 days)
- Assessment criteria are indicated in the handout.

Text and References

Text (for additional reading)

Deep Learning

I. Goodfellow, Y. Bengio, and A. Courville, MIT Press, 2016

http://www.deeplearningbook.org/

References

Deep learning tutorials:

http://deeplearning.net/tutorial/

http://deeplearning.stanford.edu/wiki/index.php/UFLDL_Tutorial

TensorFlow:

https://www.tensorflow.org/

Keras:

https://keras.io/

Lecturers

- Prof Jagath Rajapakse (First Half)
- Assoc. Prof. Chen Change Loy (Second Half)

Instructors (First Half)

Professor Jagath RAJAPAKSE
Office: Block N4, Room 2a-06
Telephone Number: (65) 6790 5802
E-mail: asjagath@ntu.edu.sg
http://www.ntu.edu.sg/home/asjagath/
Office Hours: Thursdays 3:30 – 4:30pm

TA: **Dr. Kasun Chamara**
Email: chamarakasun@ntu.edu.sg
Office hours: 3:30 – 5:30pm on Fridays
Office: SCALE@NTU, N4-B3b-13a