Neural Networks: Biological Motivation

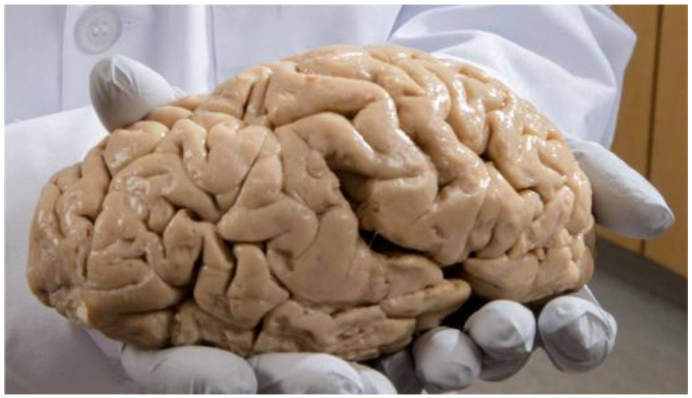
Sargur Srihari
srihari@buffalo.edu

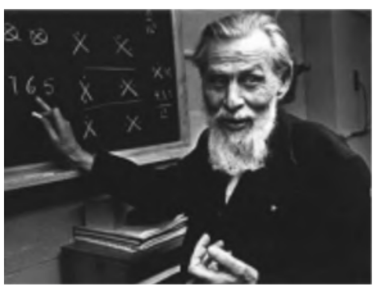
Two types of Neural Network Research

1. Study and model biological learning
   - Network of neurons in the brain provide people with ability to assimilate information.
   - Will simulations of such networks reveal the underlying mechanism of learning?
2. Obtain highly effective learning machines
   - Biological realism imposes unnecessary constraints
   - Primarily multilayer perceptron

Underpinnings of Deep Learning

Neurobiology





Warren McCulloch
1898-1969

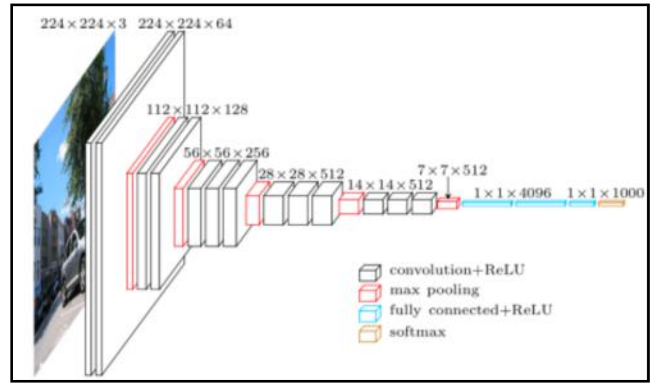
Walter Pitts
1923-1969

Statistical Physics

Ludwig Boltzmann
(1844-1906)

Josiah Willard Gibbs
(1839-1903)

Distributed Computing



2019 Turing Award Winners
LeCun, Hinton, Bengio

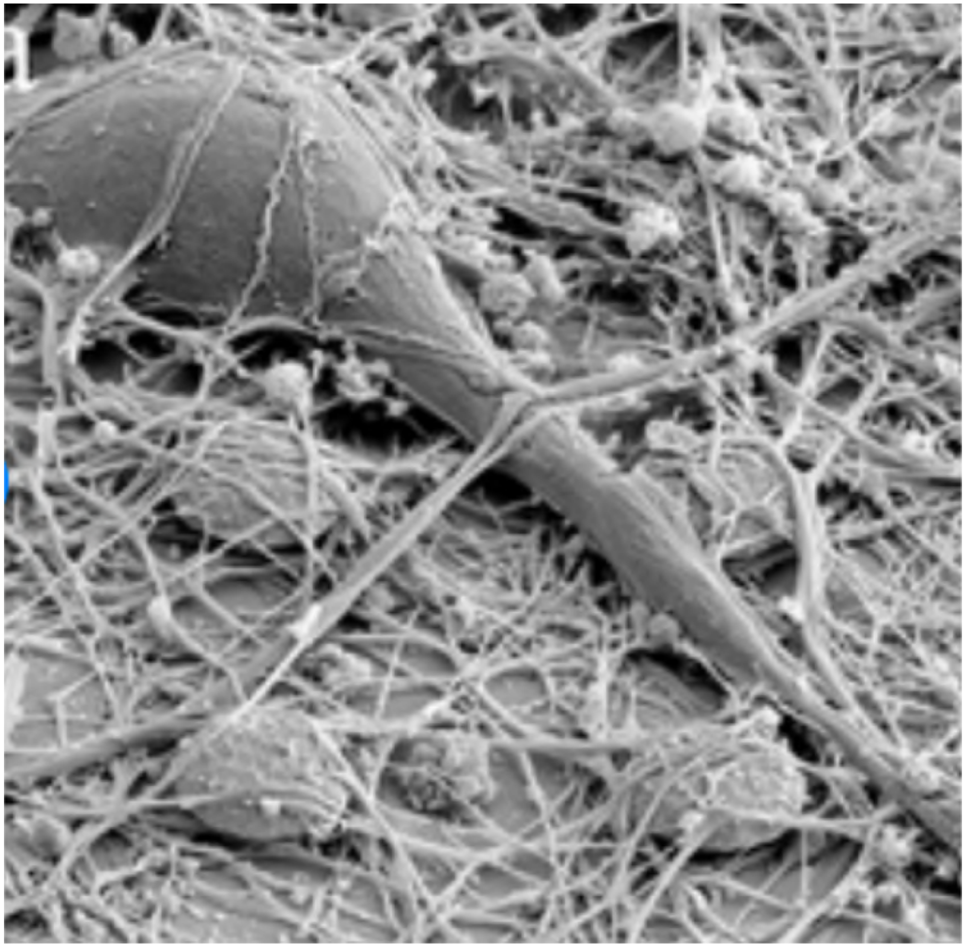
Neural Computation

Biological Motivation for Artificial Neural Networks

Biological Motivation

- Study of neural computation inspired by the observation:
  - Biological learning systems are built of very complex webs of interconnected neurons
  - Each unit takes real-valued inputs (possibly from other units)
  - Produces a single real valued output (which becomes the input to many other units)

SEM of Brain Tissue



A colored scanning electron micrograph (SEM) of a neuron (nerve cell).

Neuron is a cell

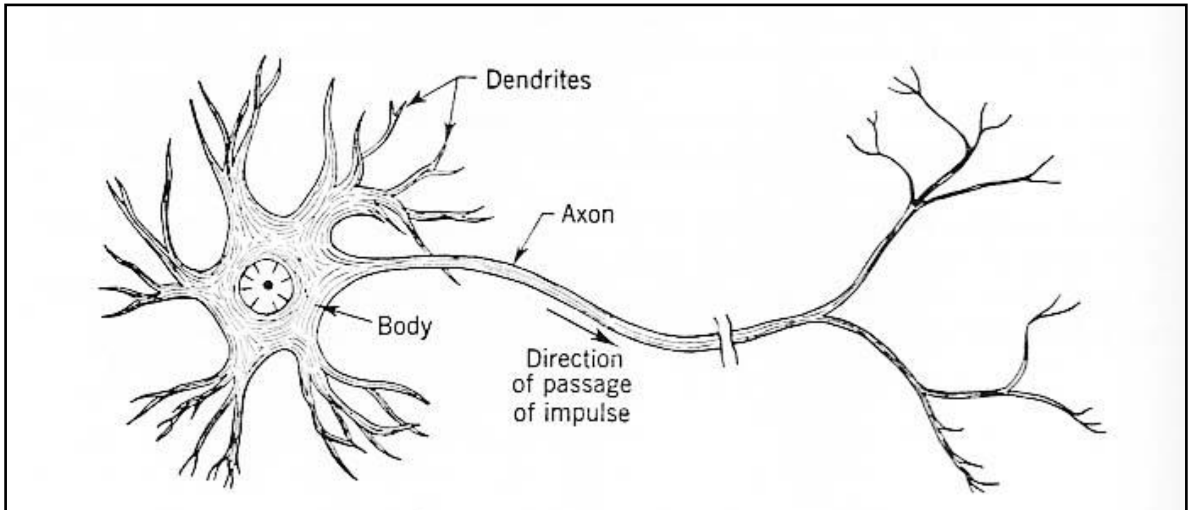
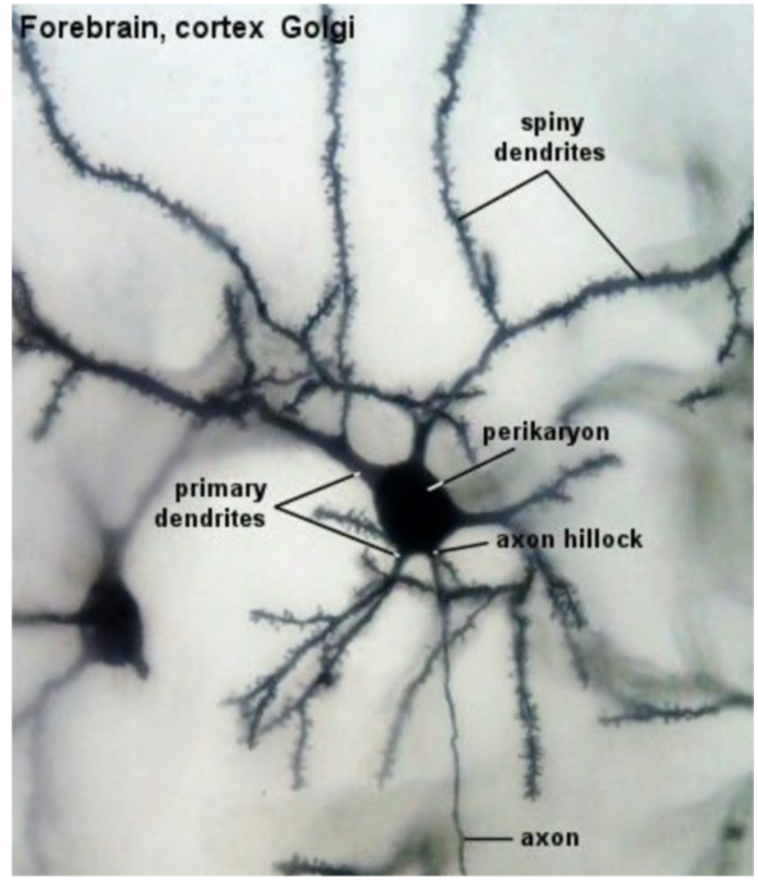


Figure 2.2 Schematic drawing of a neuron.



Perikaryon: cell body of a neuron containing nucleus

Neuron nucleus contained in the soma or body of the cell

More than 10 billion neurons in human brain

- Mosquito has 100,000 neurons to perform vision, eat, etc
  - Implies human is 100,000 mosquitos!

Switching Time

- Human Brain
  - Densely interconnected network of 10^{11} (100 billion) neurons
  - Each connected to 10^4 (10,000) others
  - Fastest neuron switching time is 10^{-3} seconds
  - Activity excited or inhibited through connections to other neurons
  - Slow compared to computer switching speed: 10^{-10} secs

Human Information Processing Speed

- Humans can make certain decisions (visually recognize your mother) in 10^{-1} secs
- Implies that in 10^{-1} sec interval cannot possibly have more than a few hundred steps, given switch speed
- Therefore
  - information processing abilities of biological systems follow from highly parallel processing operations distributed over many neurons

Neurophysiology

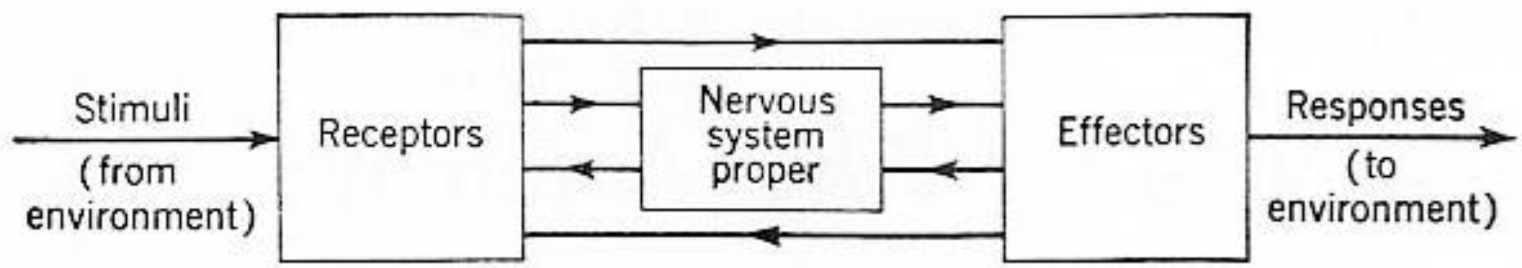


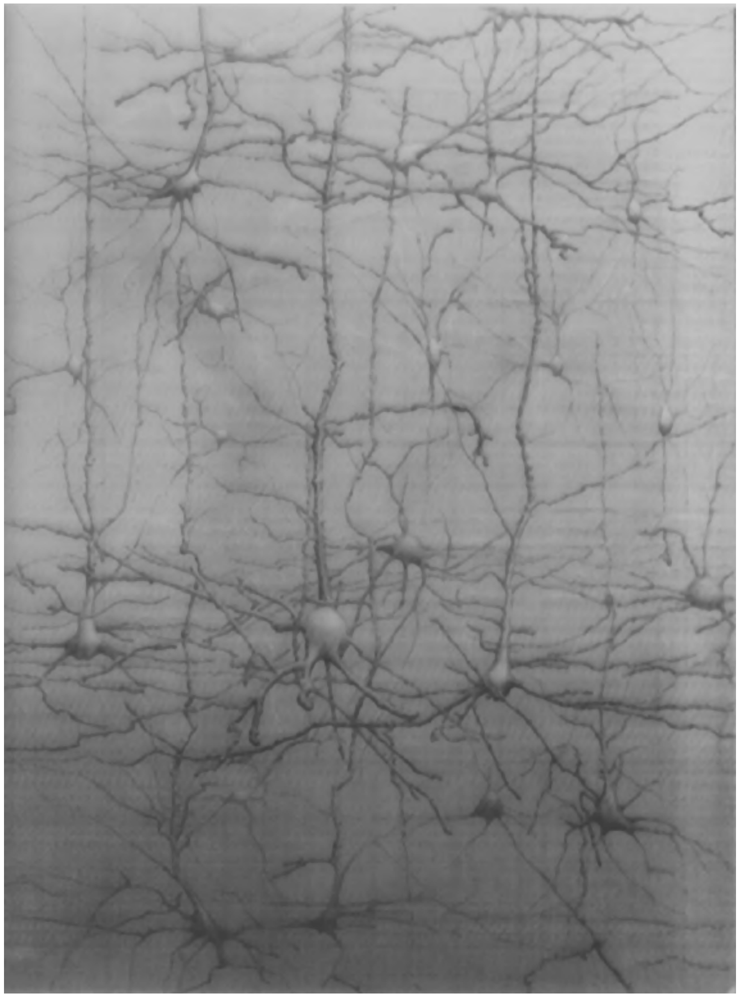
Figure 2.1 The nervous system considered as a three-stage system.

Receptors:
Rods and Cones of eyes,
Pain, touch, hot and cold receptors of skin,
Stretch receptors of muscles
Effectors:
Muscles and glands, speech generators

Neurobiology

- Basic morphology of neurons including axons, dendrites, cell bodies or somata and synapses
- Chemical transmitters at synapses and how connection of nerve impulses is affected by the actions of various ions in and around the cells

Network of Neurons



Photomicrograph of Section of cerebral cortex

Dendrites

- Dendrites: form a fine filamentary bush each fiber thinner than an axon

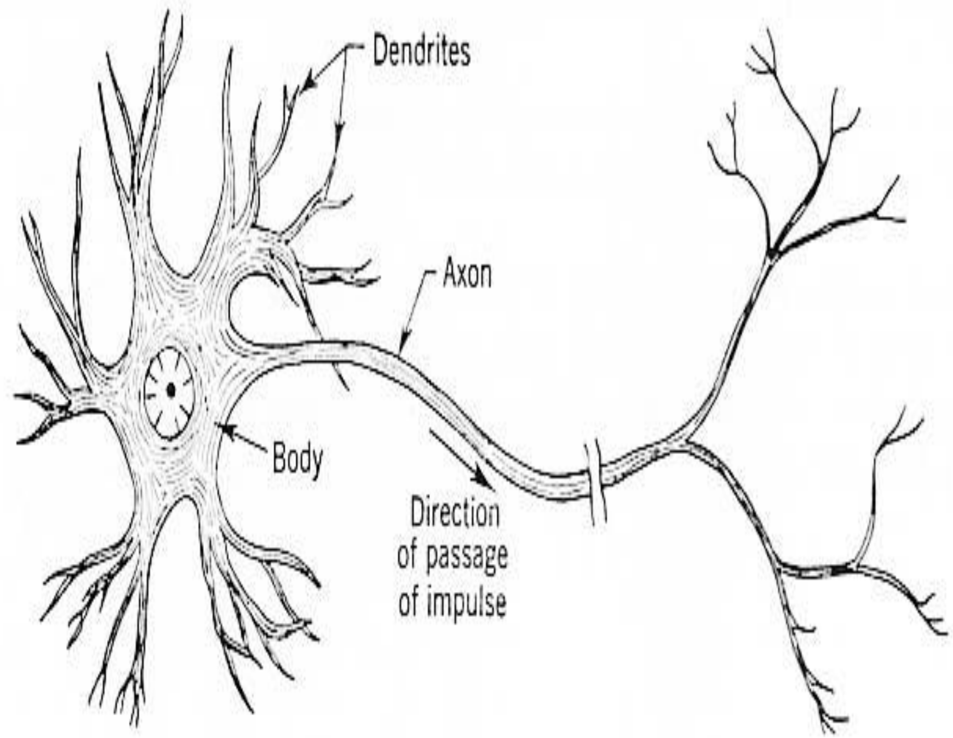


Figure 2.2 Schematic drawing of a neuron.

Axon

- Long thin cylinder carrying impulses from soma to other cells
- Splits into endbulbs
  - almost touching dendrites
- Place of near contact is a synapse

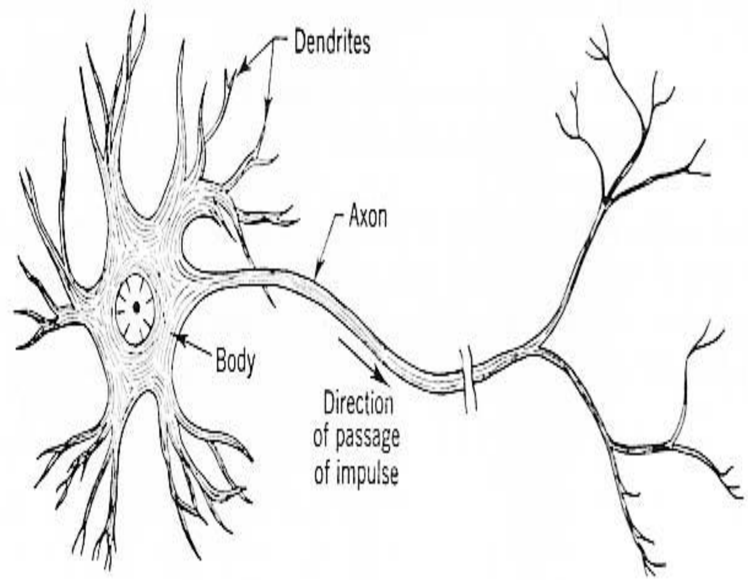


Figure 2.2 Schematic drawing of a neuron.

Inter-Neuronal Transmission

- Impulses reaching a synapse
  - set up graded electrical signals in the dendrites of neuron on which synapse impinges
- Inter-neuronal transmission
  - is sometimes electrical
  - usually by diffusion of chemical transmitters

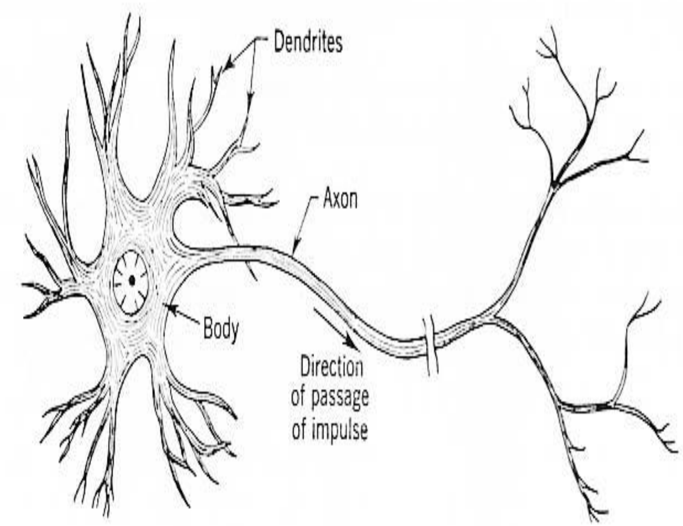


Figure 2.2 Schematic drawing of a neuron.

Synapses

- When a spike travels along an axon and arrives at a synapse it causes vesicles of transmitter chemical to be released
  - There are several kinds of transmitter
- The transmitter molecules diffuse across the synaptic cleft and bind to receptor molecules in the membrane of the post-synaptic neuron thus changing their shape.
  - This opens up holes that allow specific ions in or out.
- The effectiveness of the synapse can be changed
  - vary the number of vesicles of transmitter
  - vary the number of receptor molecules.
- Synapses are slow, but they have advantages over RAM
  - Very small
  - They adapt using locally available signals (but how?)

Chemical synapse operation

- Transmitting neuron, or presynaptic cell
  - liberates transmitter substance that diffuses across synaptic junction
- Electrical signal converted to chemical signal
- Changes postsynaptic cell membrane potential
- Chemical signal converted back to electrical signal

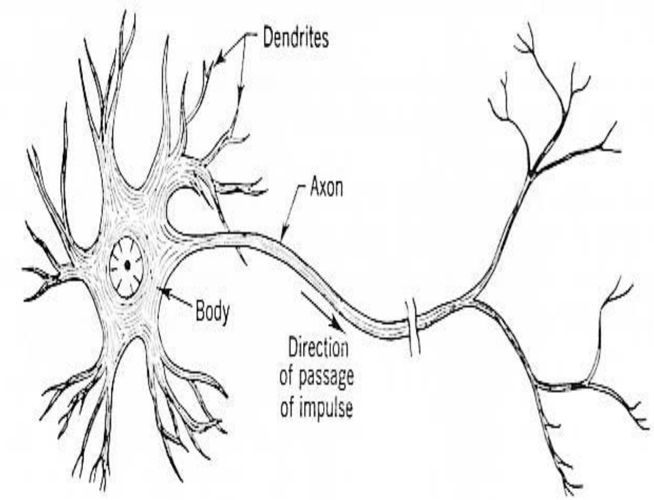
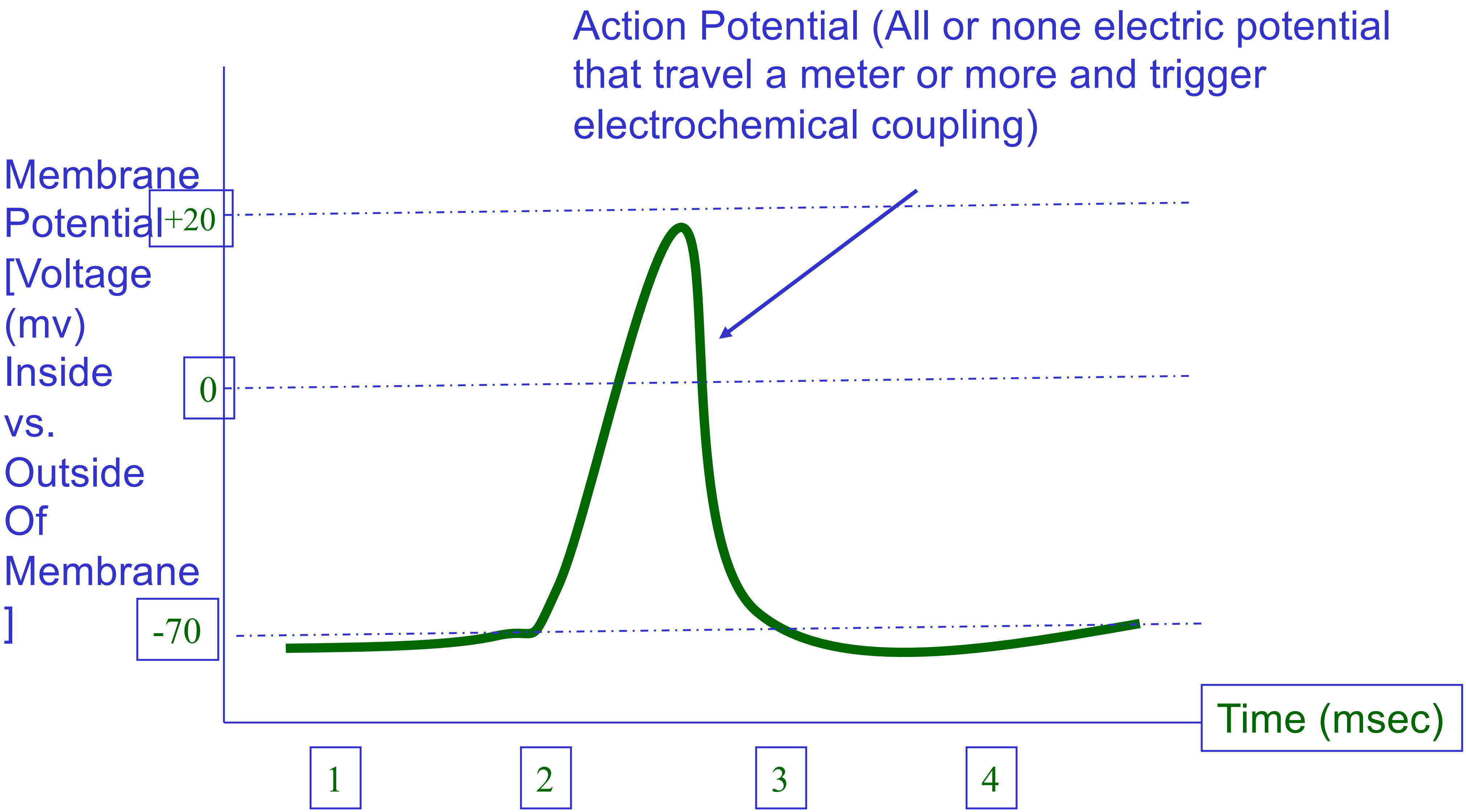


Figure 2.2 Schematic drawing of a neuron.

Nerve Impulse Waveform

As appears on oscilloscope by placing microelectrode near an axon



Neuron Firing

- A neuron will only fire an electrical impulse along its axon only if sufficient impulses reach endbulbs impinging on its dendrites in a short period of time,called period of latent summation

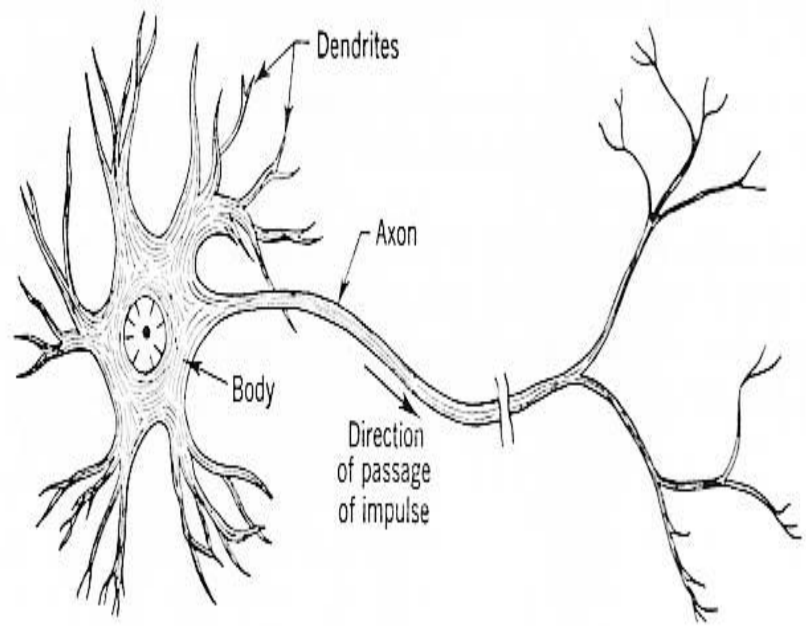


Figure 2.2 Schematic drawing of a neuron.

Excitatory and Inhibitory Impulses

- Impulses may help or hinder firing of impulse
- Excitation should exceed inhibition by critical amount called threshold of the neuron
- A neuron fires only if the total weight of the synapses that receive impulses in the period of latent summation exceeds the threshold

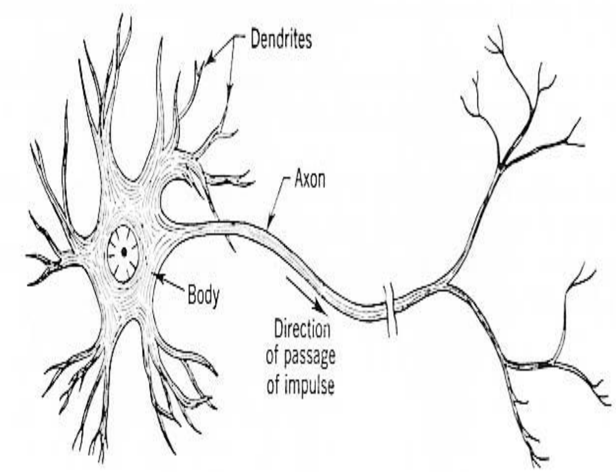
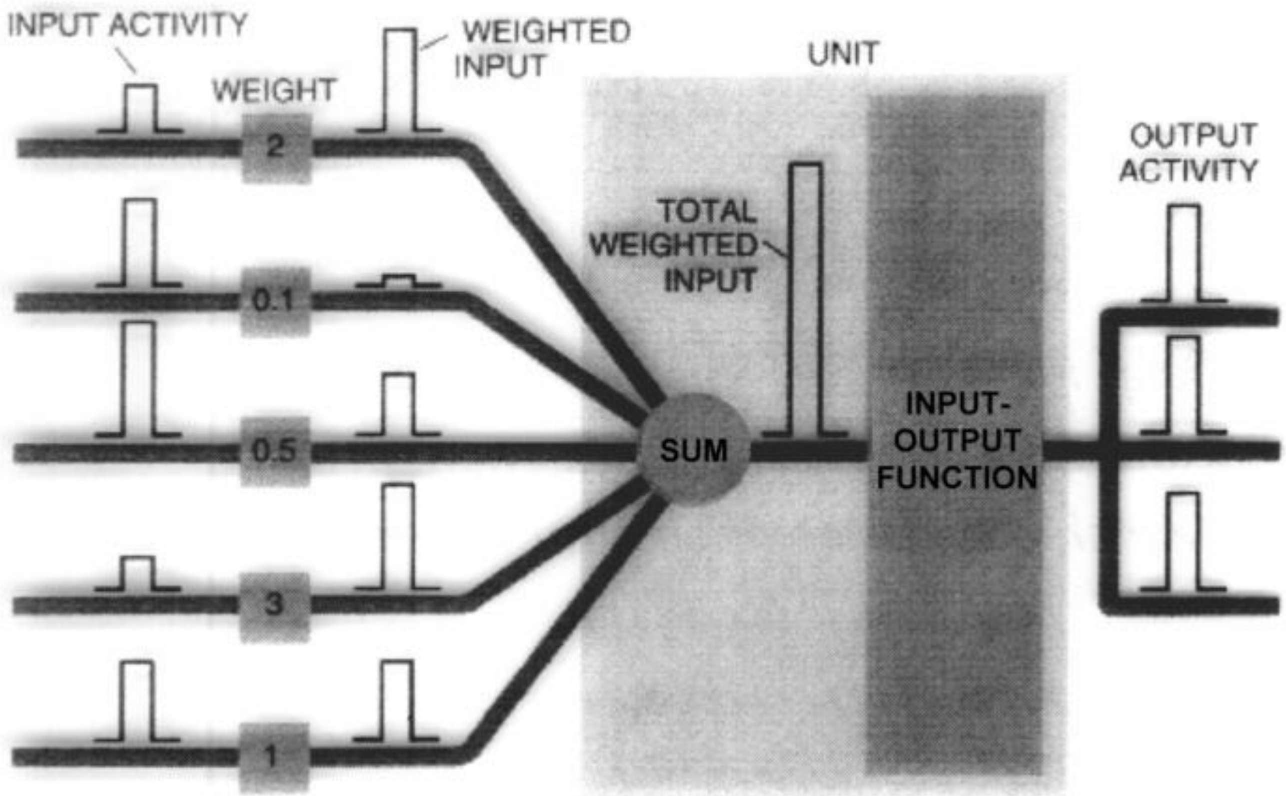


Figure 2.2 Schematic drawing of a neuron.

The goals of neural computation

- To understand how the brain actually works
  - Its big and very complicated and made of yukky stuff that dies when you poke it around
- To understand a new style of computation
  - Inspired by neurons and their adaptive connections
  - Very different style from sequential computation
    - should be good for things that brains are good at (e.g. vision)
    - Should be bad for things that brains are bad at (e.g. 23 x 71)
- To solve practical problems by using novel learning algorithms
  - Learning algorithms can be very useful even if they have nothing to do with how the brain works

Idealization of a Neuron

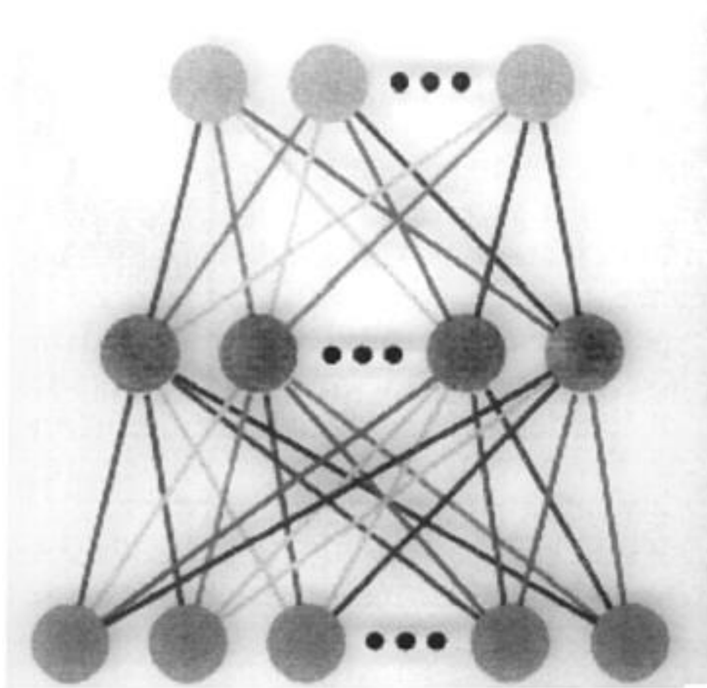


IDEALIZATION OF A NEURON processes activities, or signals. Each input activity is multiplied by a number called the weight. The "unit" adds together the weighted inputs. It then computes the output activity using an input-output function.

ANN

- ANNs are built of
  - densely interconnected set of simple units
  - each unit
    - takes several real-valued inputs
    - Produces single-valued output

Common ANN



COMMON NEURAL NETWORK consists of three layers of units that are fully connected. Activity passes from the input units (*green*) to the hidden units (*gray*) and finally to the output units (*yellow*). The reds and blues of the connections represent different weights.

ANNs

- One motivation is to capture highly parallel computations on distributed processes
- Most ANN software run on sequential machines emulating distributed processes

Use of ANNs

- General practical method
- Robust approach
- Used to learn functions that are
  - real-valued,
  - discrete-valued
  - vector-valued

Limitations of Neural Networks

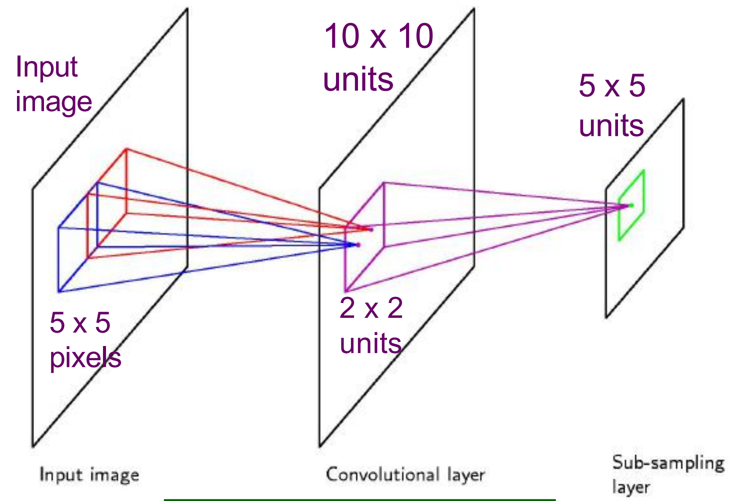
- Need substantial number of training samples
- Slow learning (convergence times)
- Inadequate parameter selection techniques that lead to poor minima

Three Mechanisms of Convolutional Neural Networks

1. Local Receptive Fields
2. Subsampling
3. Weight Sharing

Convolution and Sub-sampling

- Instead of treating input to a fully connected network
- Two layers of Neural networks are used
  1. Layer of convolutional units
     - which consider overlapping regions
  2. Layer of subsampling units
- Several feature maps and sub-sampling
  - Gradual reduction of spatial resolution compensated by increasing no. of features
- Final layer has softmax output
- Whole network trained using backpropagation



Each pixel patch is 5 x 5

This plane has 10 x 10=100 neural network units (called a feature map). Weights are same for different planes. So only 25 weights are needed. Due to weight sharing this is equivalent to convolution. Different features have different feature maps

CNN Inspired by Visual Neuroscience

- Classic notions;simple cells and complex cells
- Architecture similar to LGN-V1-V2-V4-IT hierarchy in visual cortex ventral pathway
  - LGN: lateral geniculate nucleus receives input from retina
  - 30 areas of visual cortex: V1 and V2 are principal
  - Infero-Temporal cortex performs object recognition