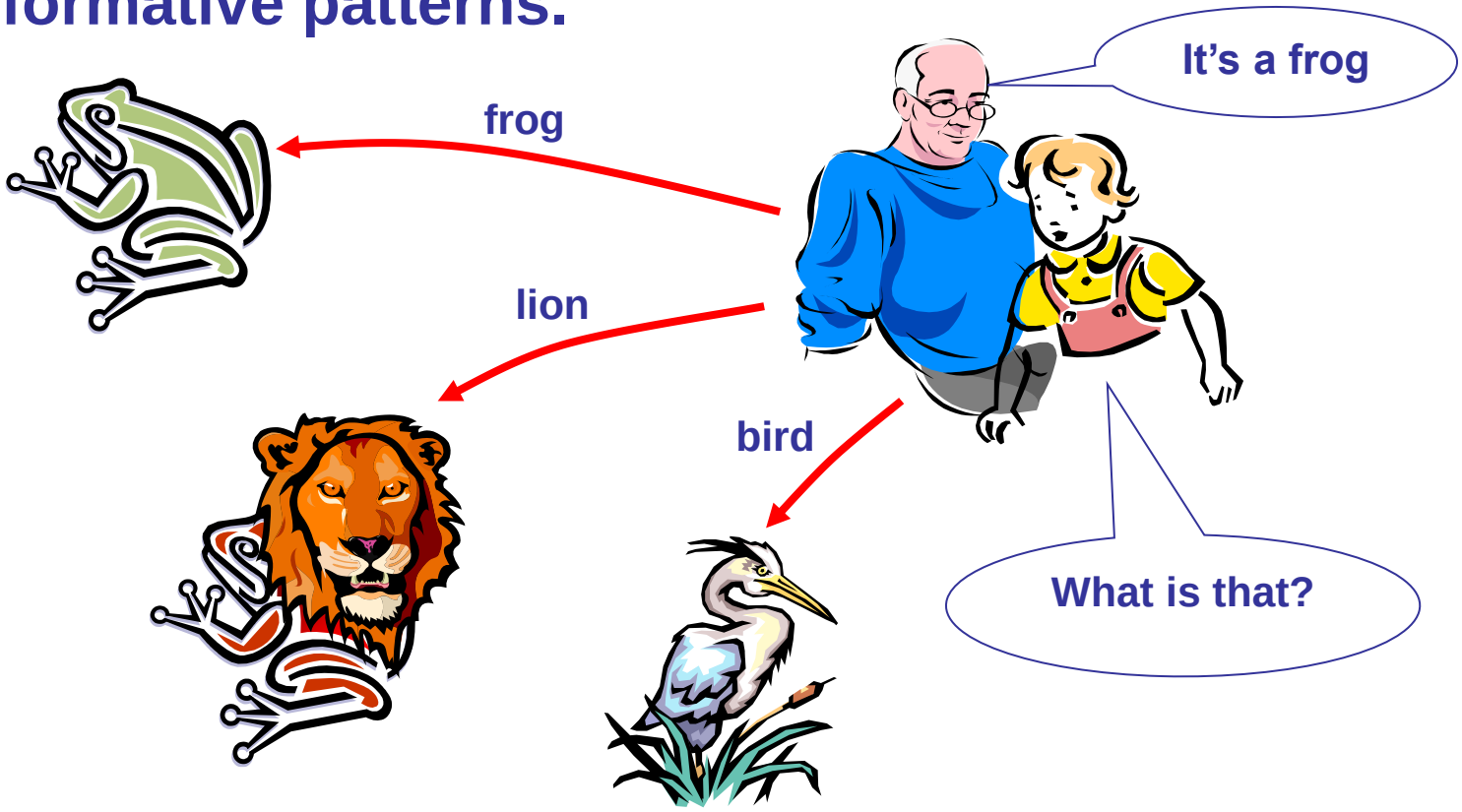
INTRODUCTION TO ARTIFICIAL NEURAL NETWORKS (ANN)

Outline

- Definition, why and how are neural networks being used in solving problems
- Human biological neuron
- Artificial Neuron
- Applications of ANN
- Comparison of ANN vs conventional AI methods

The idea of ANNs..?

- **NNs learn relationship between cause and effect or organize large volumes of data into orderly and informative patterns.**



Neural networks to the rescue…

- **Neural network:** *information processing paradigm inspired by biological nervous systems, such as our brain*
- Structure: large number of highly interconnected processing elements (*neurons*) working together
- Like people, they learn *from experience* (by example)

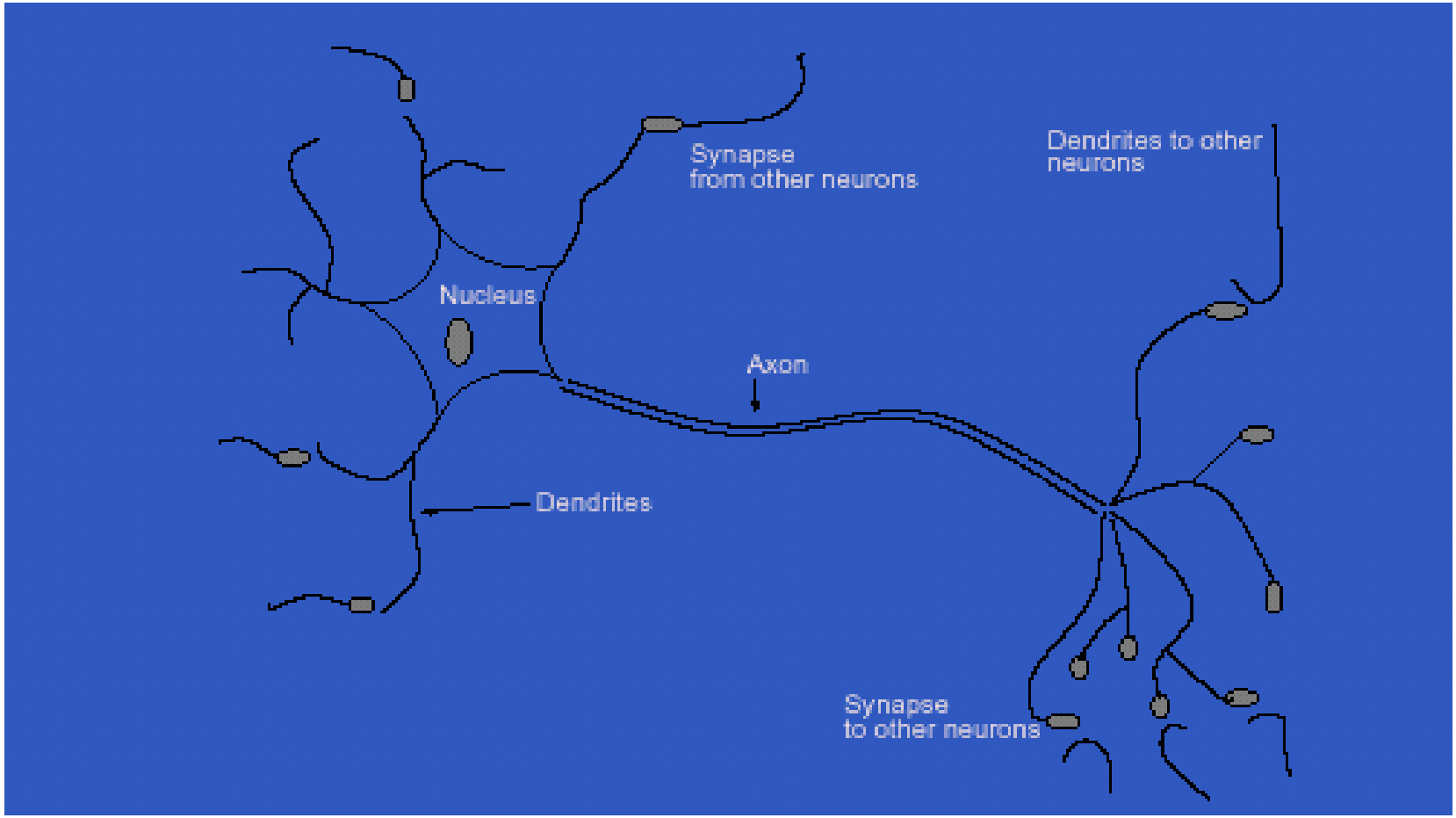
Definition of ANN

"Data processing system consisting of a large number of simple, highly interconnected processing elements (artificial neurons) in an architecture inspired by the structure of the cerebral cortex of the brain"

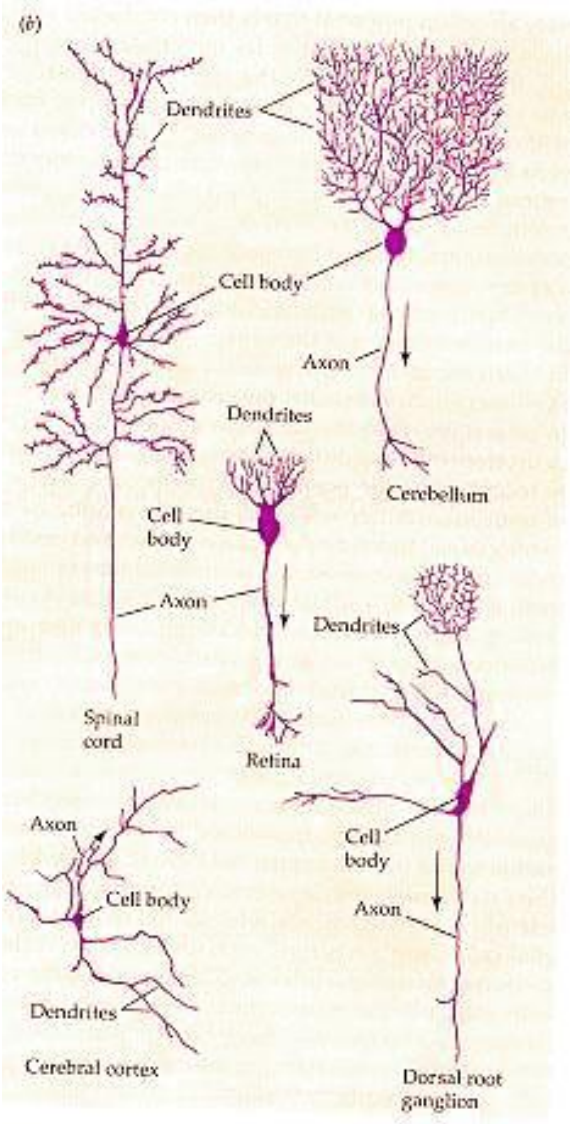
(Tsoukalas & Uhrig, 1997).

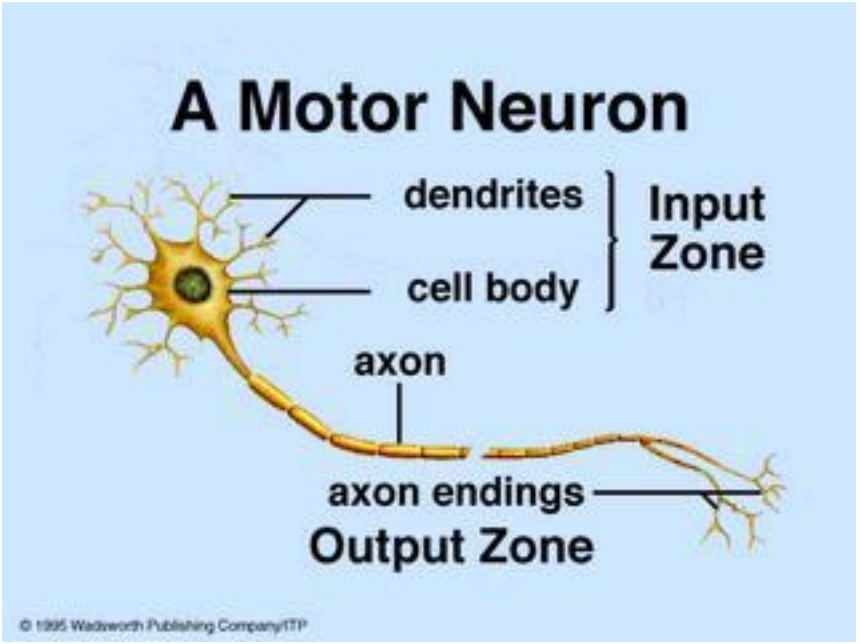
Inspiration from Neurobiology

Human Biological Neuron



Biological Neural Networks

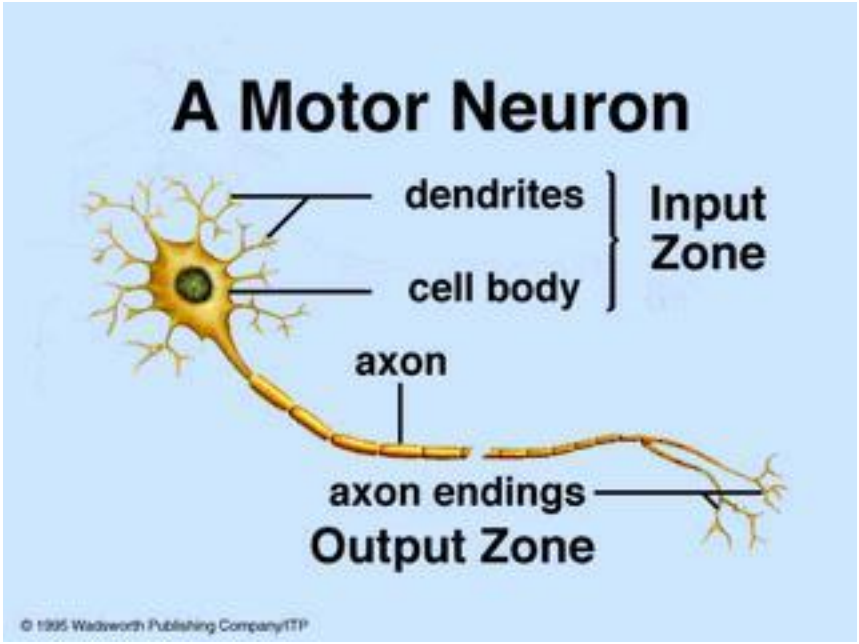




Biological neuron

Biological Neural Networks

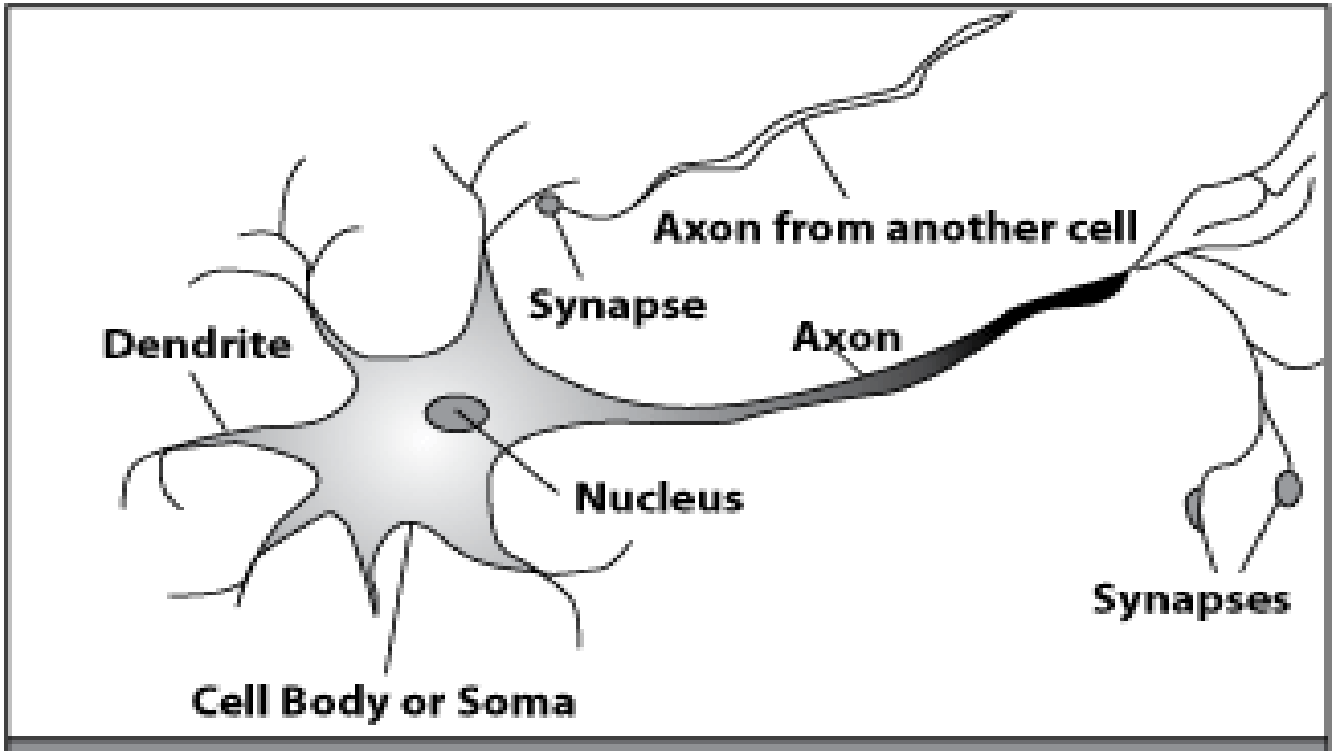
- A biological neuron has three types of main components; dendrites, soma (or cell body) and axon.
- Dendrites receives signals from other neurons.



- The soma, sums the incoming signals. When sufficient input is received, the cell fires; that is it transmit a signal over its axon to other cells.

How do our brains work?

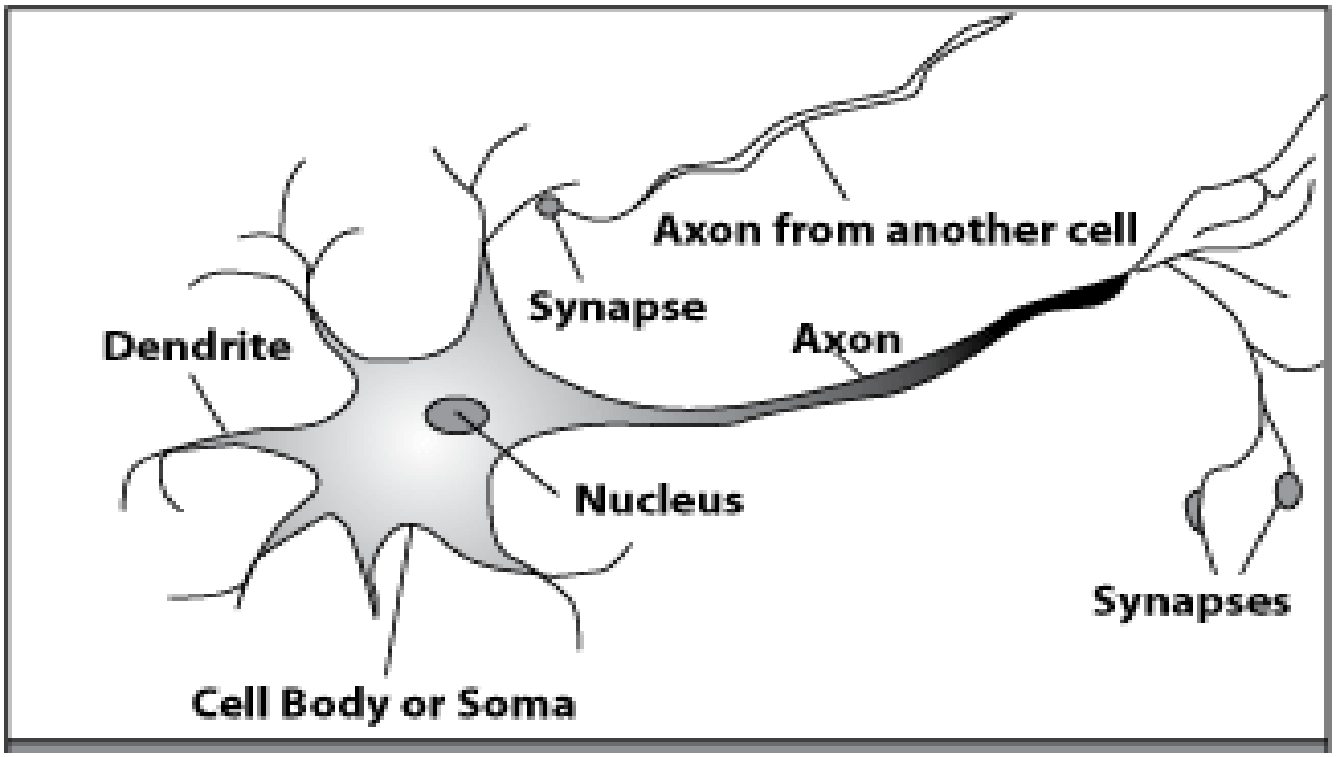
- A processing element



Dendrites: Input
Cell body: Processor
Synaptic: Link
Axon: Output

How do our brains work?

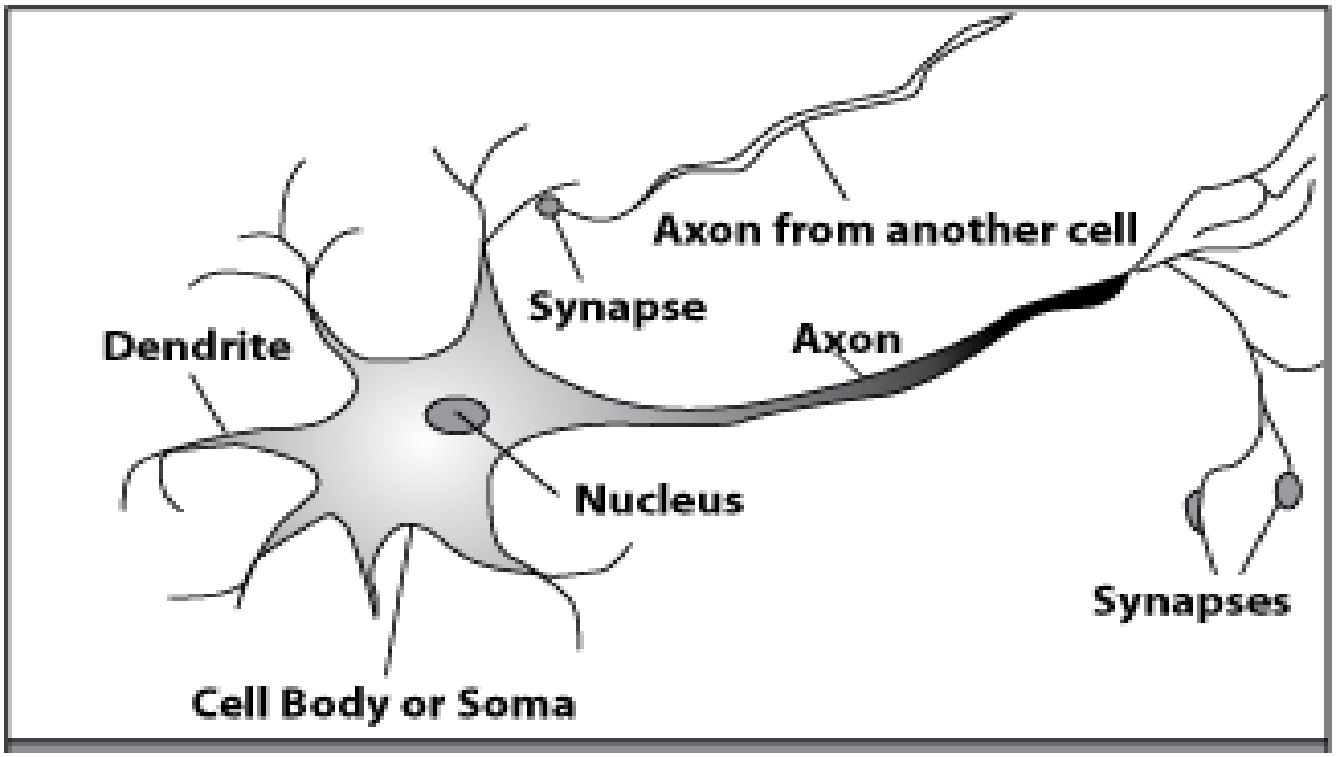
- A processing element



A neuron is connected to other neurons through about *10,000 synapses*

How do our brains work?

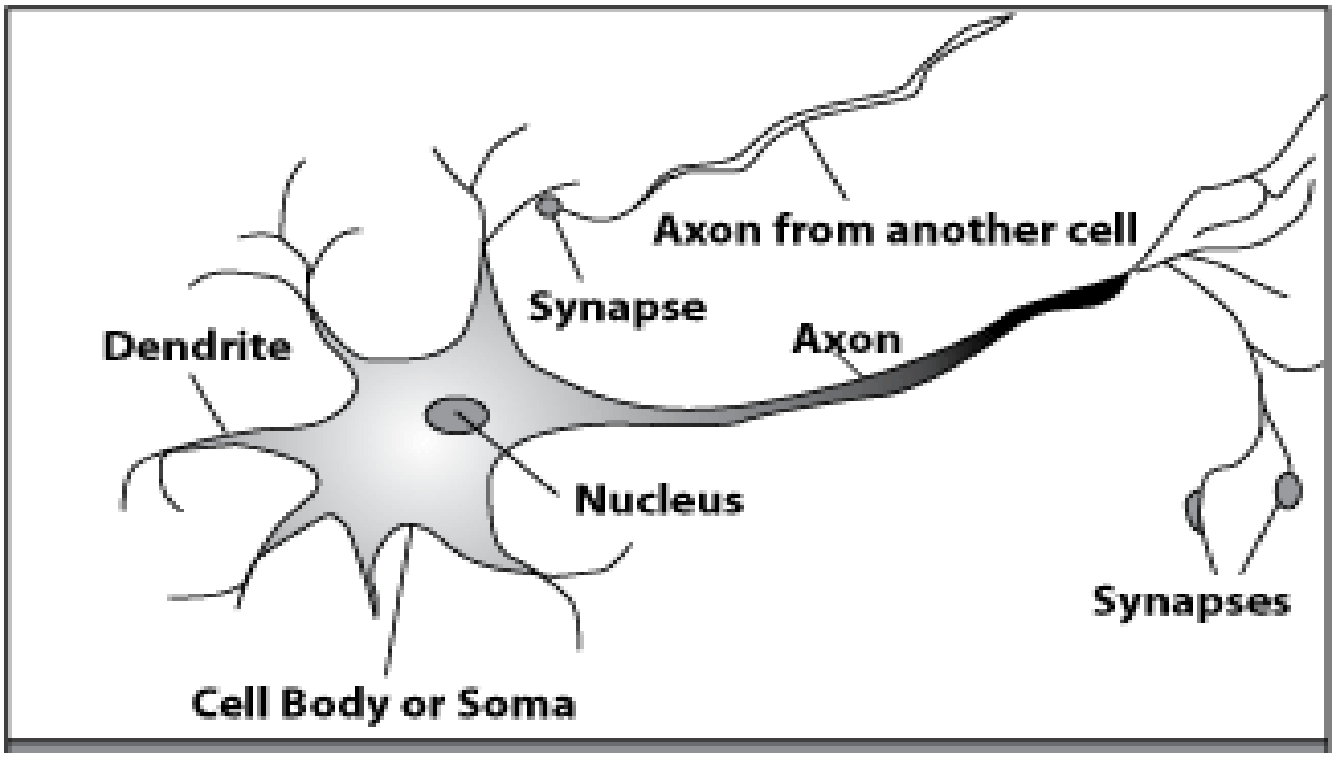
- A processing element



A neuron receives input from other neurons. Inputs are combined.

How do our brains work?

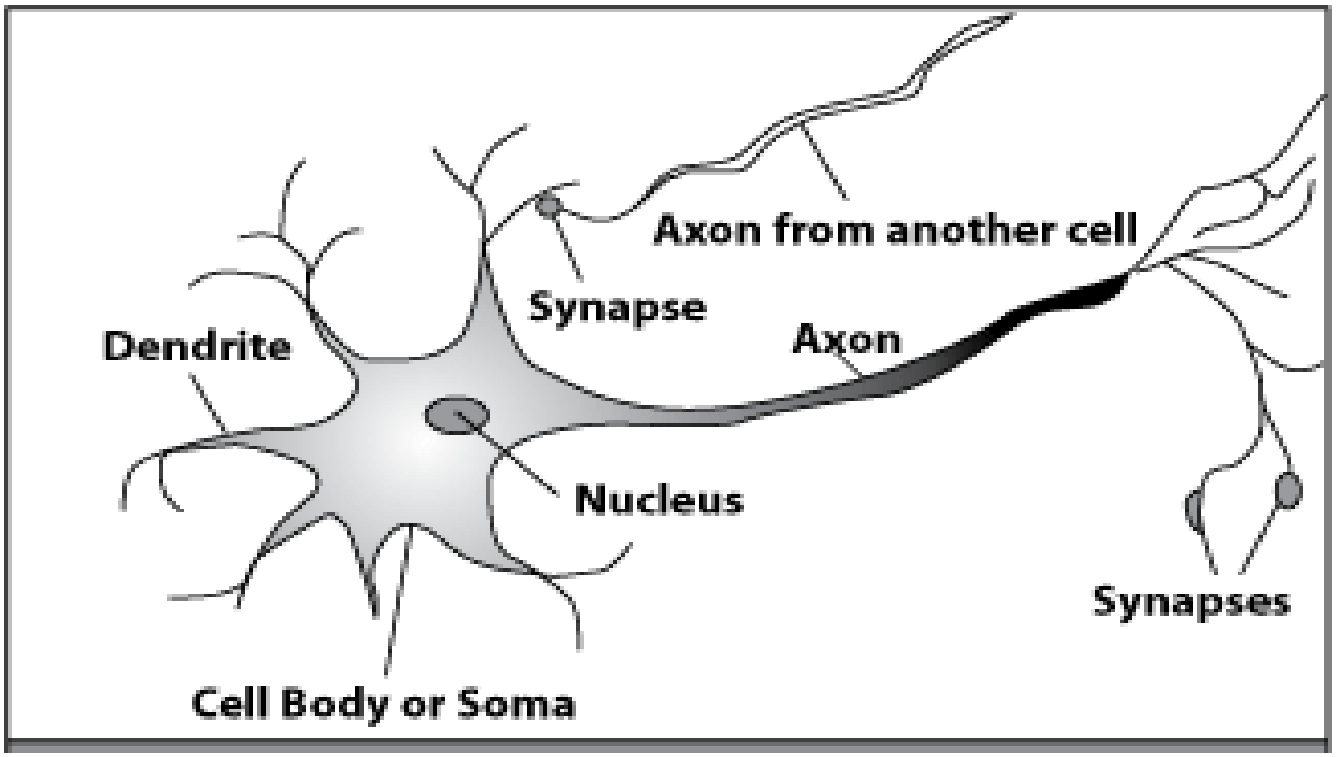
- A processing element



Once input exceeds a critical level, the neuron discharges a spike - an electrical pulse that travels from the body, down the axon, to the next neuron(s)

How do our brains work?

- A processing element



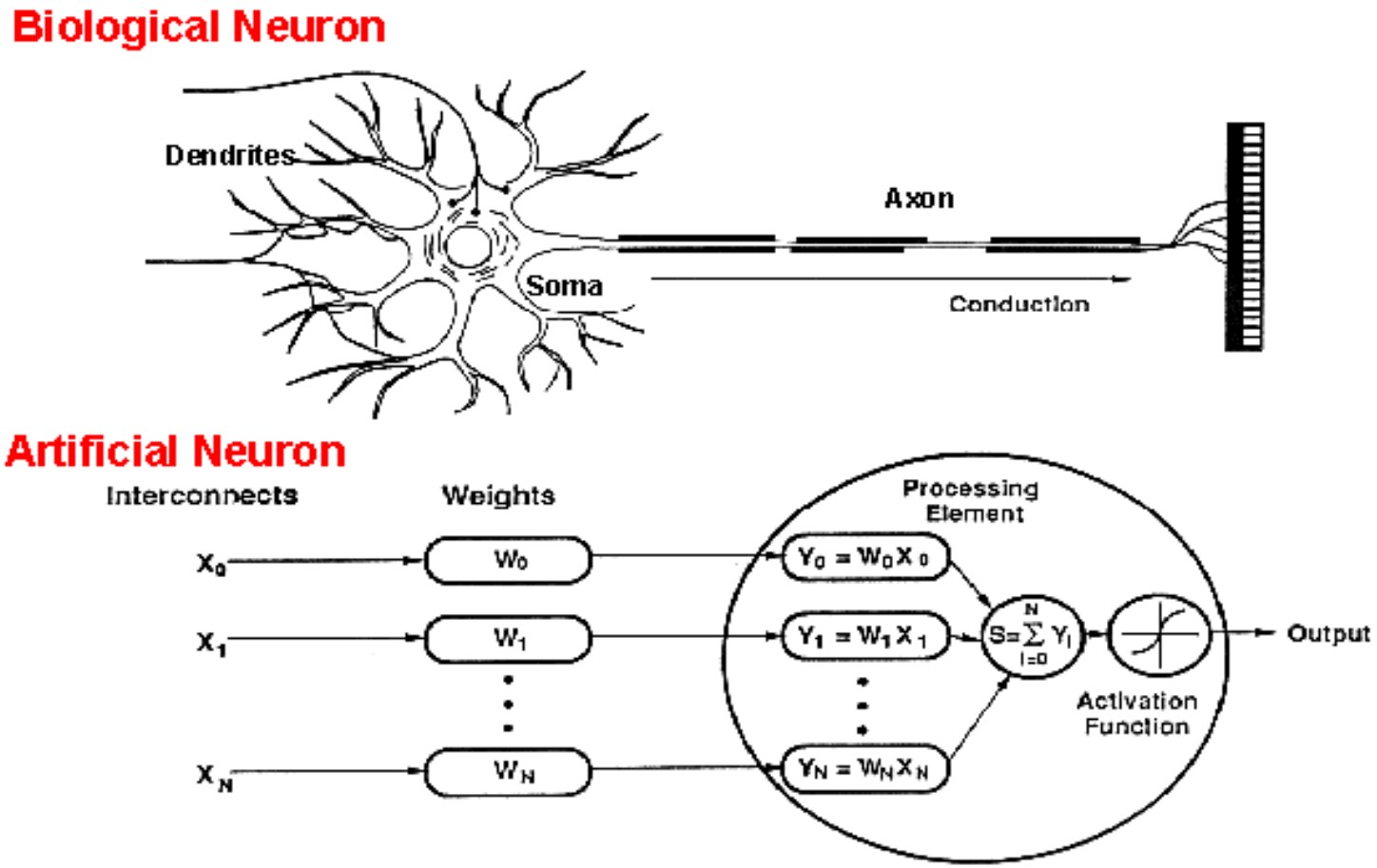
The axon-endings almost touch the dendrites or cell body of the next neuron.

Artificial Neurons

- **ANN is an information processing system that has certain performance characteristics in common with biological nets.**
- **Several key features of the processing elements of ANN are suggested by the properties of biological neurons:**

1. **The processing element receives many signals.**
2. **Signals may be modified by a weight at the receiving synapse.**
3. **The processing element sums the weighted inputs.**
4. **Under appropriate circumstances (sufficient input), the neuron transmits a single output.**
5. **The output from a particular neuron may go to many other neurons.**

How do ANNs work?

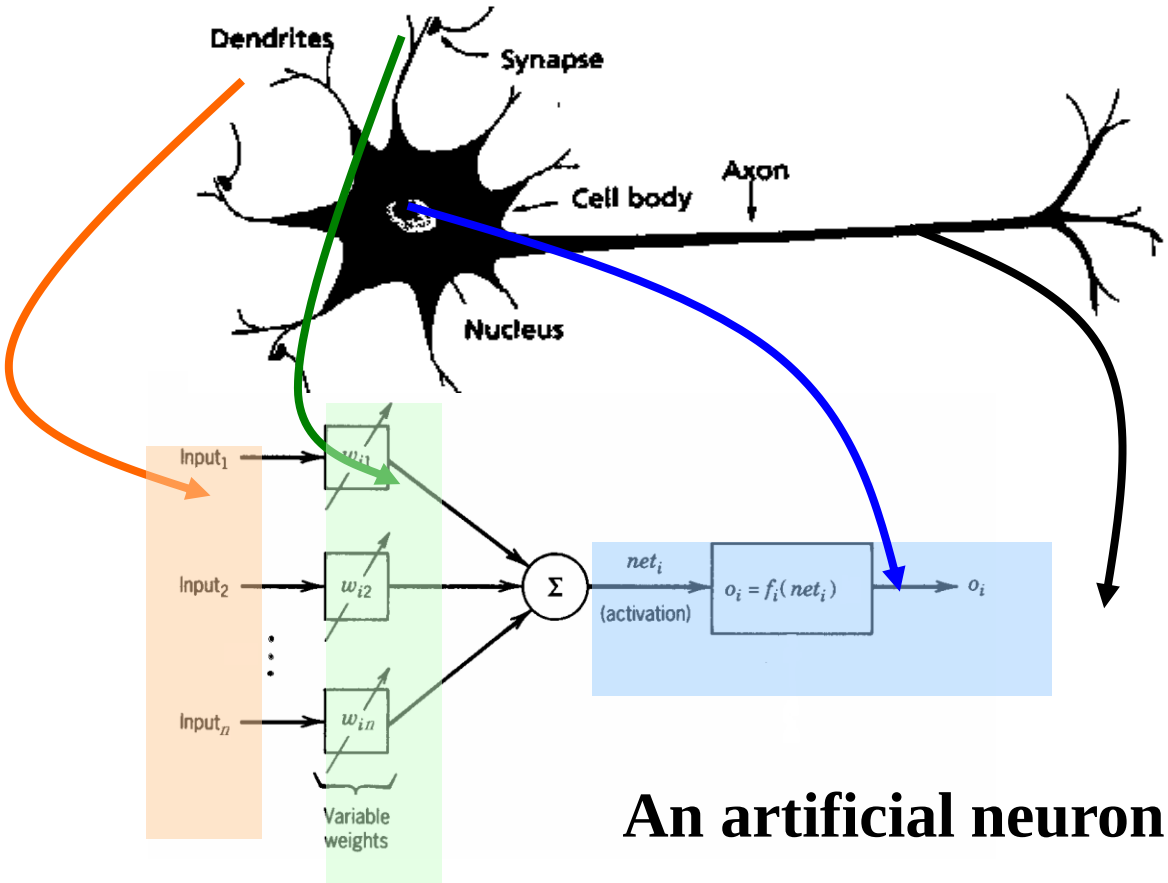


An artificial neuron is an imitation of a human neuron

Artificial Neurons

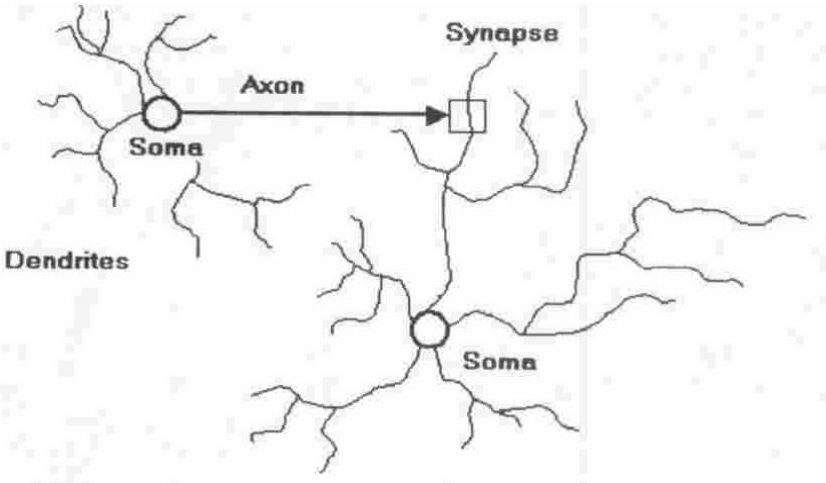
- From experience: examples / training data
- Strength of connection between the neurons is stored as a **weight-value** for the specific connection.
- Learning the solution to a problem = changing the connection weights

A physical neuron

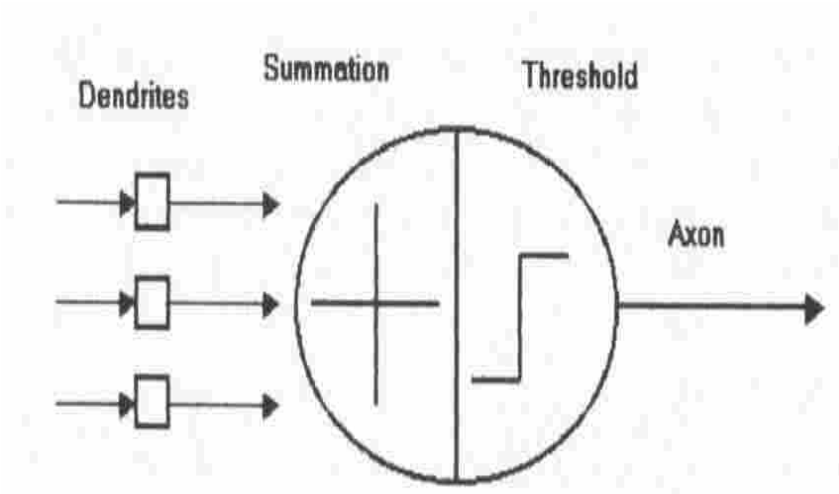


An artificial neuron

Artificial Neuron [Zoom-in view]

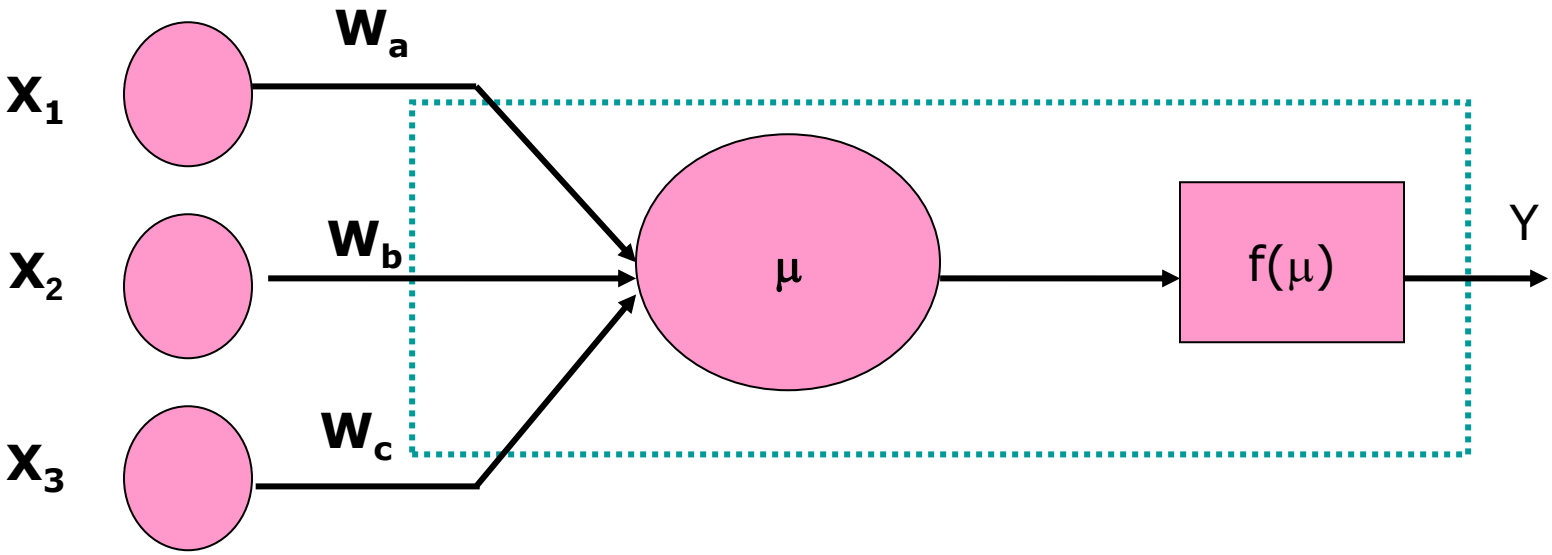


Four basic components of a human biological neuron



The components of a basic artificial neuron

Model Of A Neuron [Zoom-in view]



Input units — Connection weights — Summing function — computation

(dendrite) (synapse) (soma) (axon)

- A neural net consists of a large number of simple processing elements called neurons, units, cells or nodes.
- Each neuron is connected to other neurons by means of directed communication links, each with associated weight.
- The weight represent information being used by the net to solve a problem.

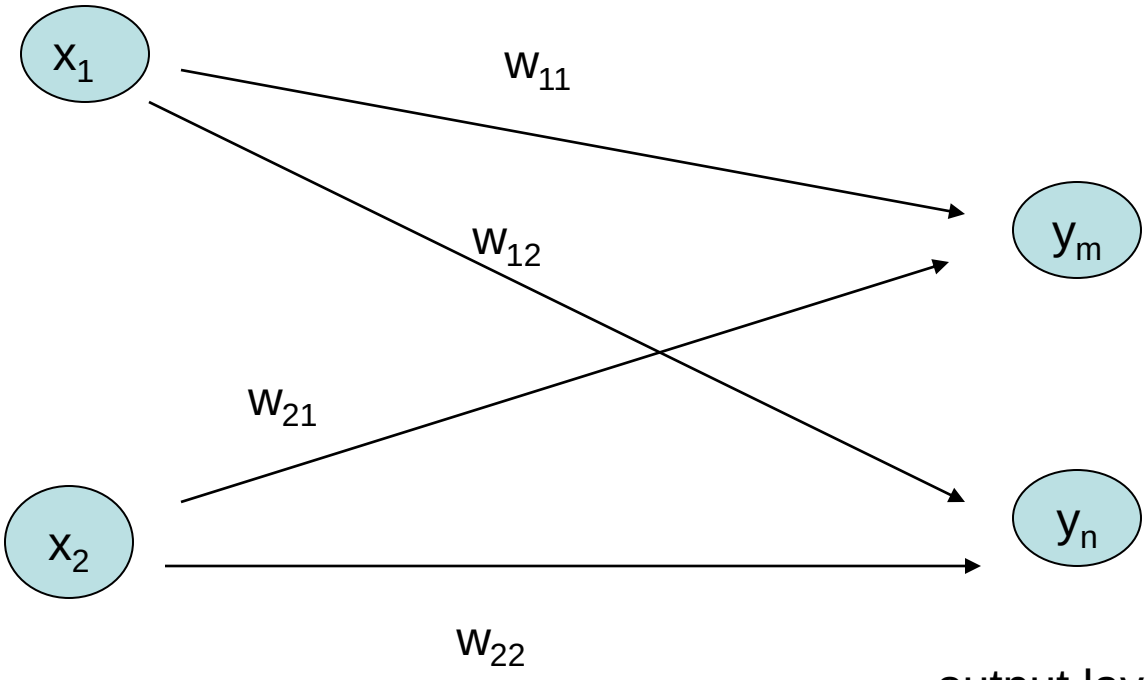
- Each neuron has an internal state, called its activation or activity level, which is a function of the inputs it has received. Typically, a neuron sends its activation as a signal to several other neurons.

- It is important to note that a neuron can send only one signal at a time, although that signal is broadcast to several other neurons.

Characterization

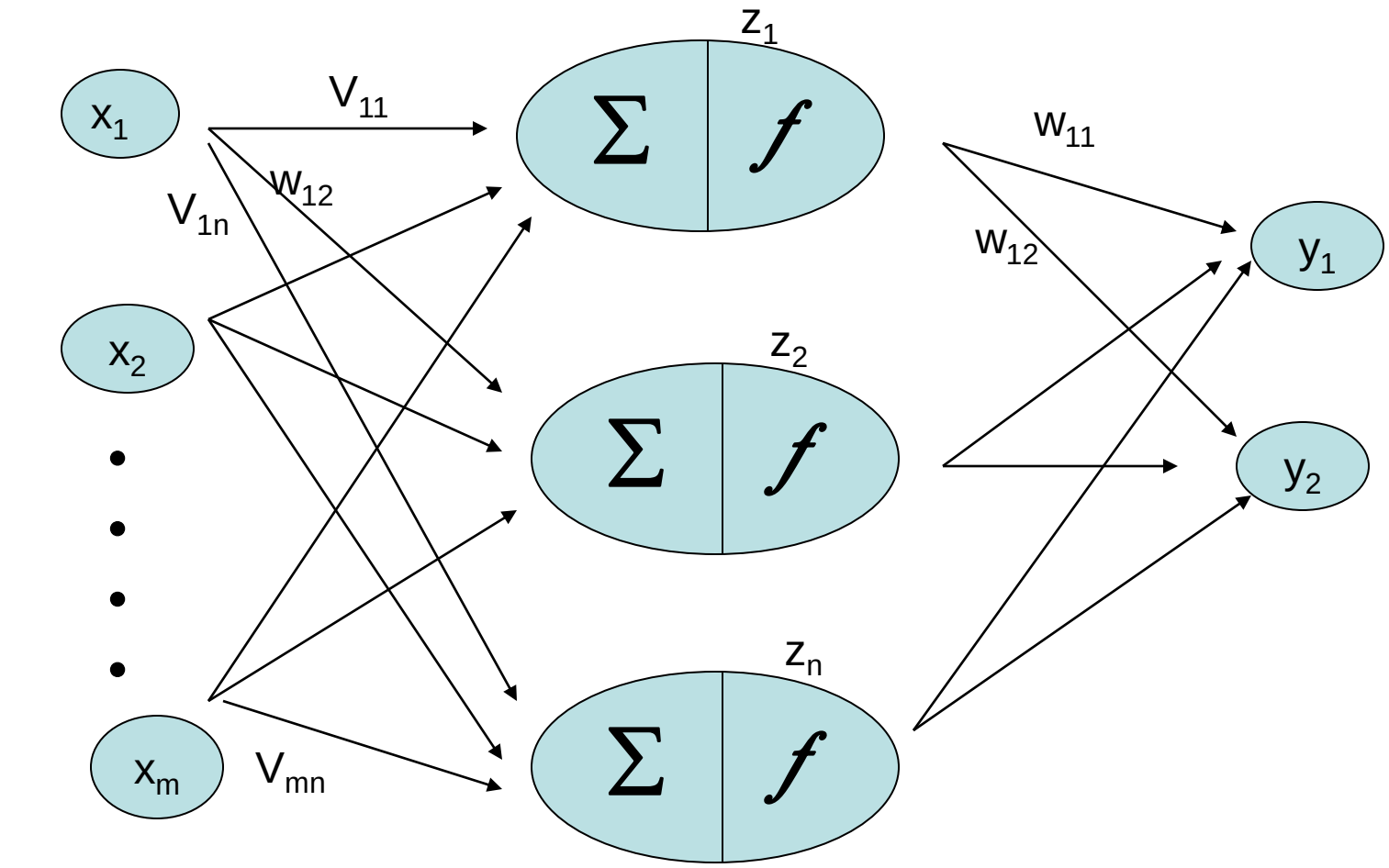
- Architecture
  - a pattern of connections between neurons
    - Single Layer Feedforward
    - Multilayer Feedforward
    - Recurrent
- Strategy / Learning Algorithm
  - a method of determining the connection weights
    - Supervised
    - Unsupervised
    - Reinforcement
- Activation Function
  - Function to compute output signal from input signal

Single Layer Feedforward NN



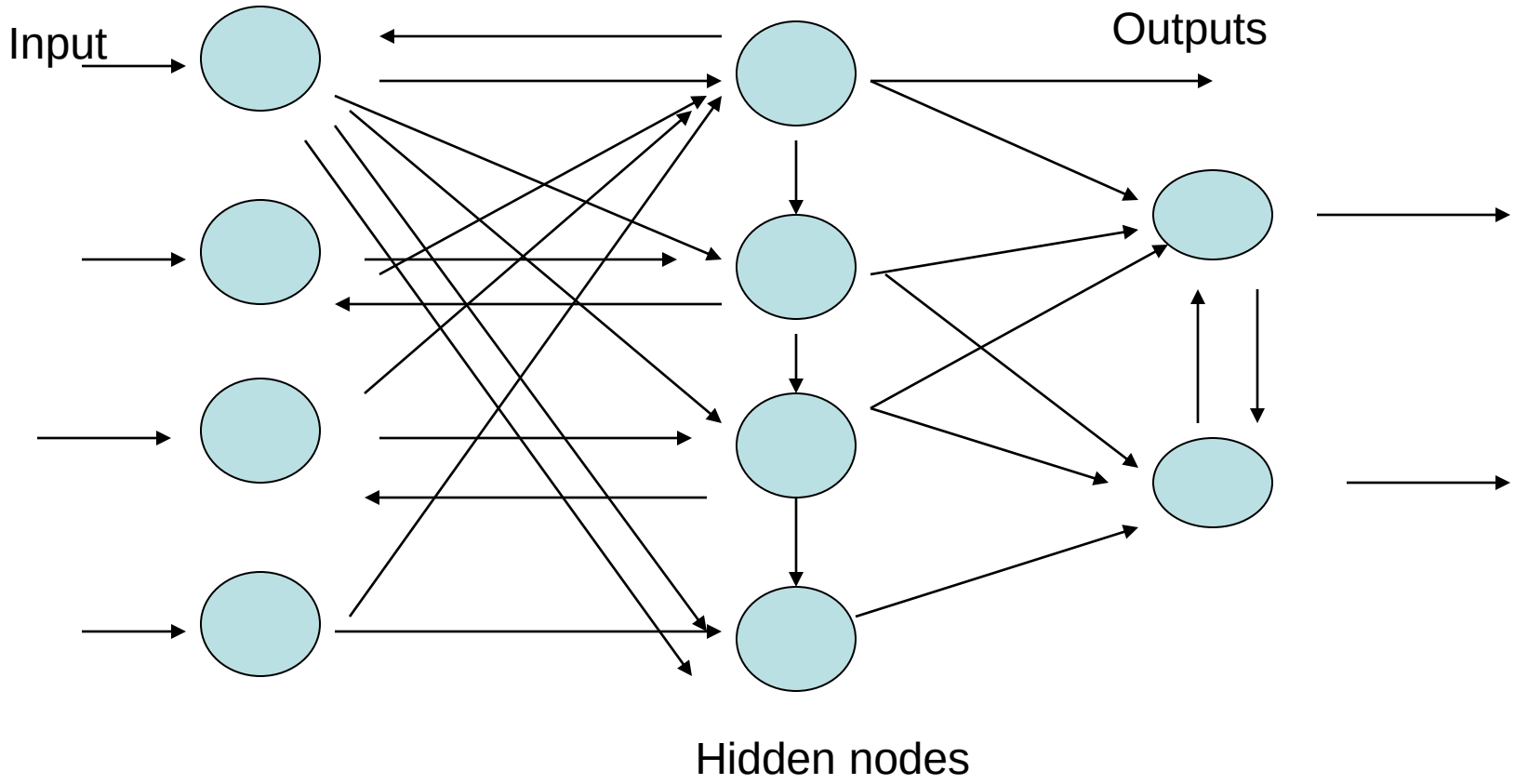
Contoh: **ADALINE, AM, Hopfield, LVQ, Perceptron, SOFM**

Multilayer Neural Network



Contoh: **CCN, GRNN, MADALINE, MLFF with BP, Neocognitron, RBF, RCE**

Recurrent NN



Contoh: **ART, BAM, BSB, Boltzman Machine, Cauchy Machine, Hopfield, RNN**

Activation Function

- An activation function in a neural network defines how the weighted sum of the input is transformed into an output from a node or nodes in a layer of the network. It also defines the non-linearity of the NN.

- Sometimes the activation function is called a "*transfer function*." If the output range of the activation function is limited, then it may be called a "*squashing function*." Many activation functions are nonlinear and may be referred to as the "*nonlinearity*" in the layer or the network design.

Activation Functions

- Identity

  $f(x) = x$

- Binary step

  $f(x) = 1$ if $x >= \theta$

  $f(x) = 0$ otherwise

- Binary sigmoid

  $f(x) = 1 / (1 + e^{-\sigma x})$

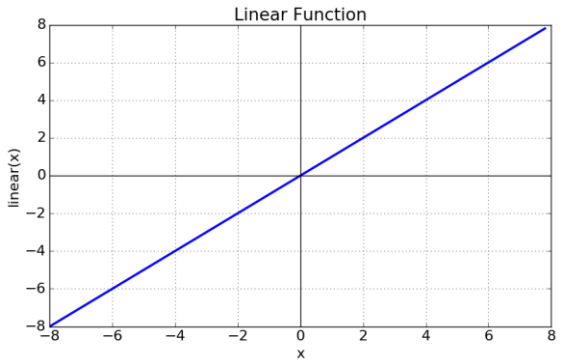
- Bipolar sigmoid

  $f(x) = -1 + 2 / (1 + e^{-\sigma x})$

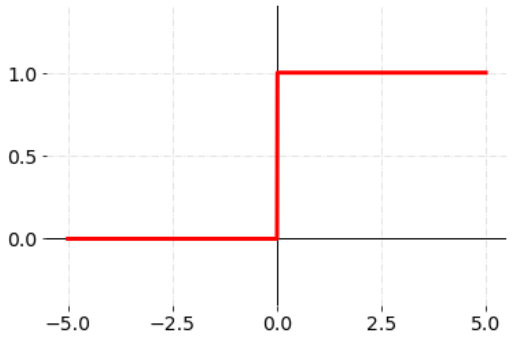
- Hyperbolic tangent

  $f(x) = (e^{x} - e^{-x}) / (e^{x} + e^{-x})$

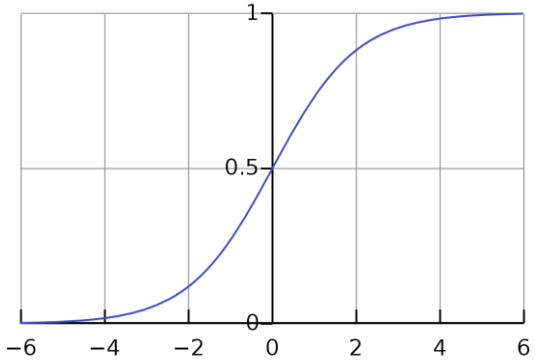
Activation Functions



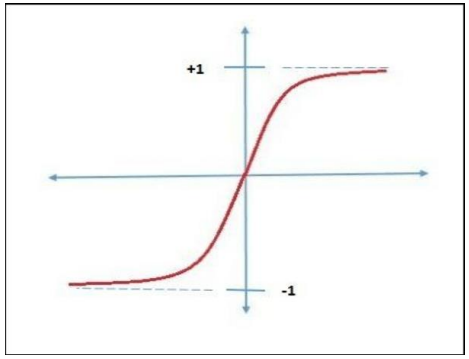
Identity: f(x) = x



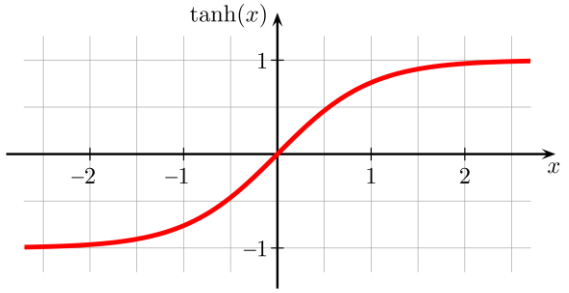
Binary Step: f(x) = 1 if x >= θ
f(x) = 0 otherwise



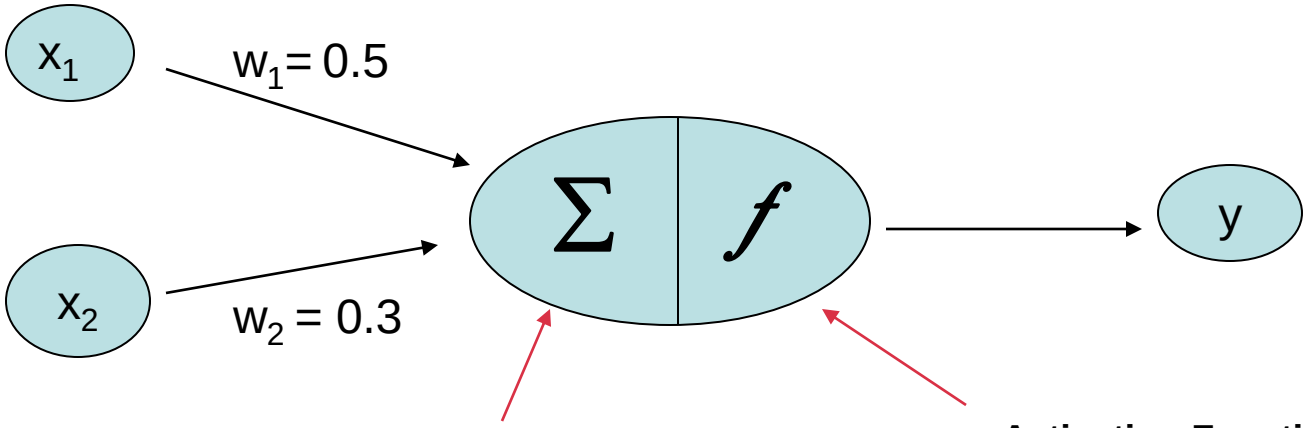
Binary Sigmoid:
$f(x) = 1 / (1 + e^{-\sigma x})$



Bipolar Sigmoid:
$f(x) = -1 + 2 / (1 + e^{-\sigma x})$



Hyperbolic tangent:
$f(x) = (e^{x} - e^{-x}) / (e^{x} + e^{-x})$

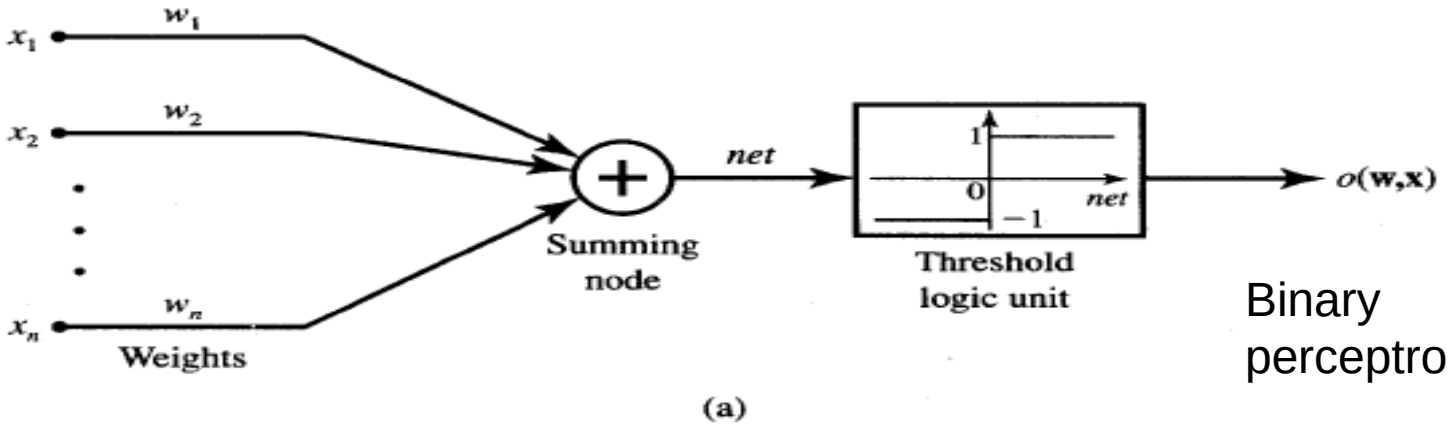


$y_{in} = x_1w_1 + x_2w_2$

Activation Function:
Binary Step Function
$\theta = 0.5$

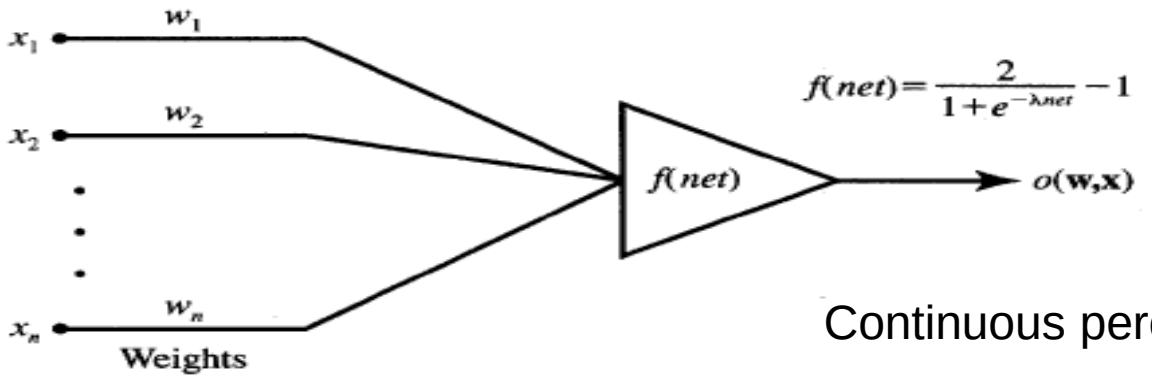
if y-in >= θ
f(y-in) = 1
else,
f(y-in) = 0

Common models of neurons



(a)

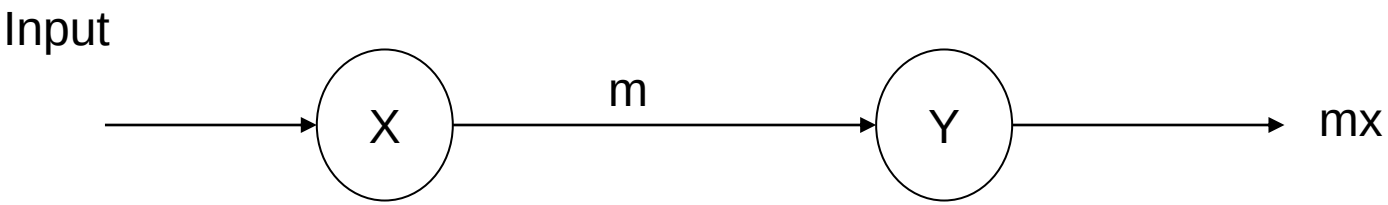
Binary perceptrons



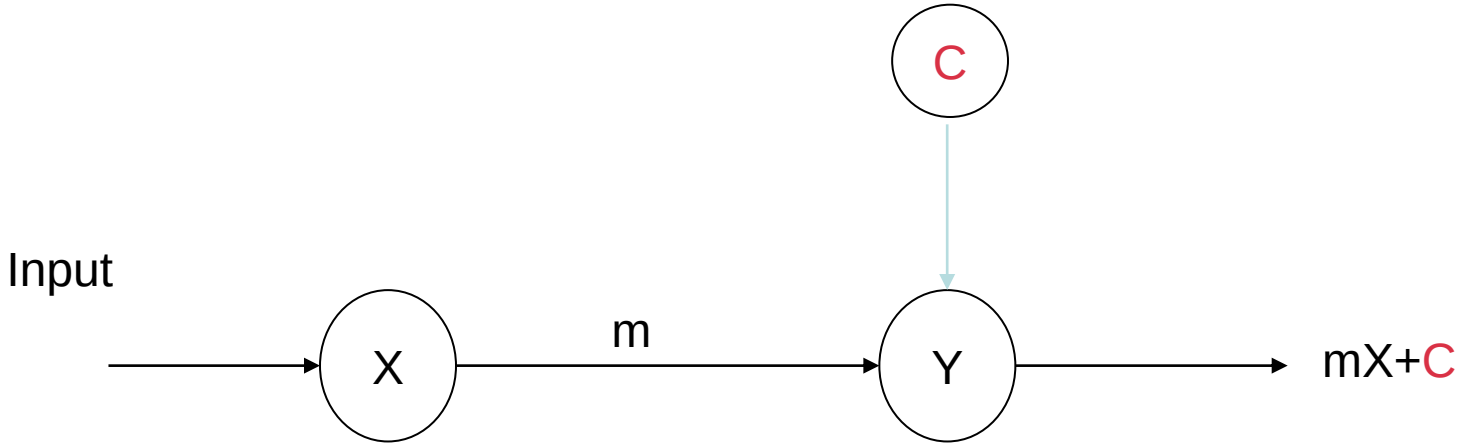
(b)

Continuous perceptrons

Neural net of pure linear eqn.

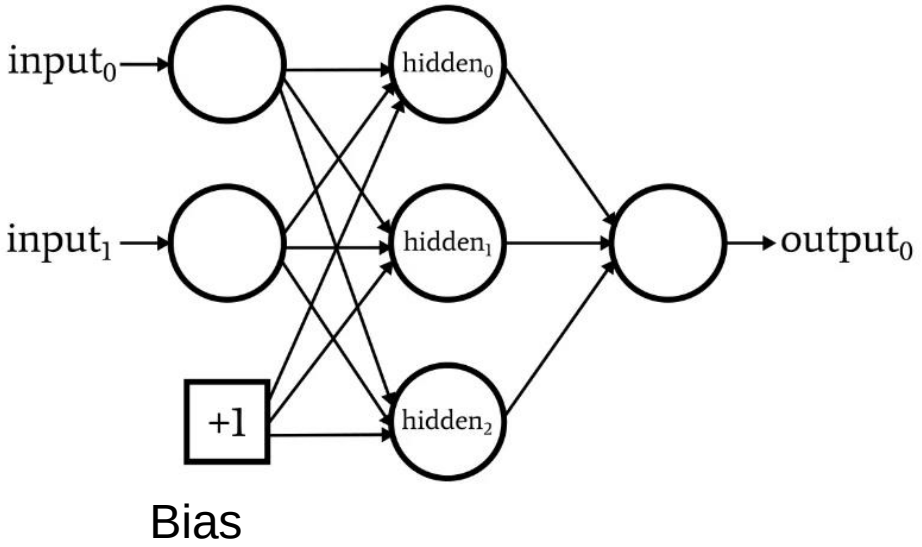


But, we know: Y=mX+C



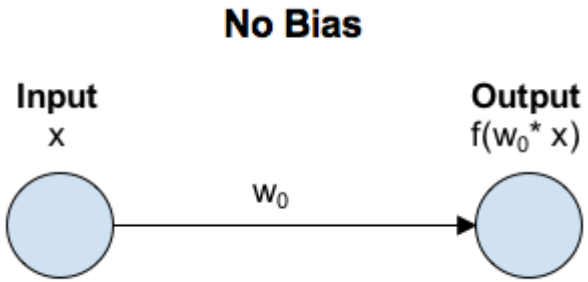
Bias in Neural Network

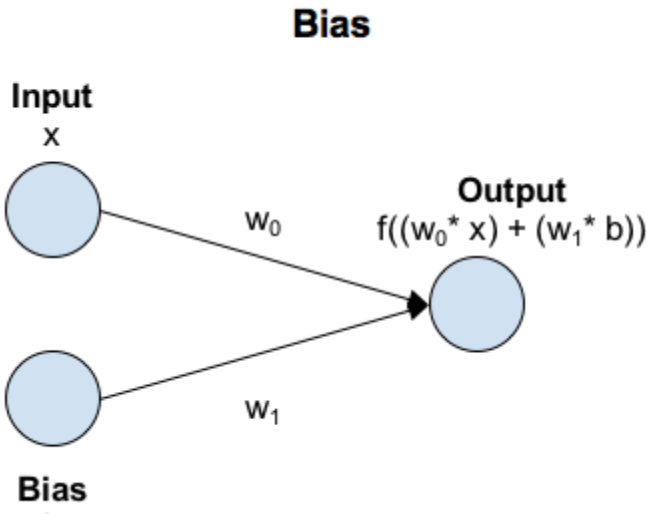
- The activation function in Neural Networks takes an input 'x' multiplied by a weight 'w'.
- Bias allows you to shift the activation function by adding a constant (i.e. the given bias) to the input.
- Bias in Neural Networks can be thought of as analogous to the role of a constant in a linear function, whereby the line is effectively transposed by the constant value.



Bias

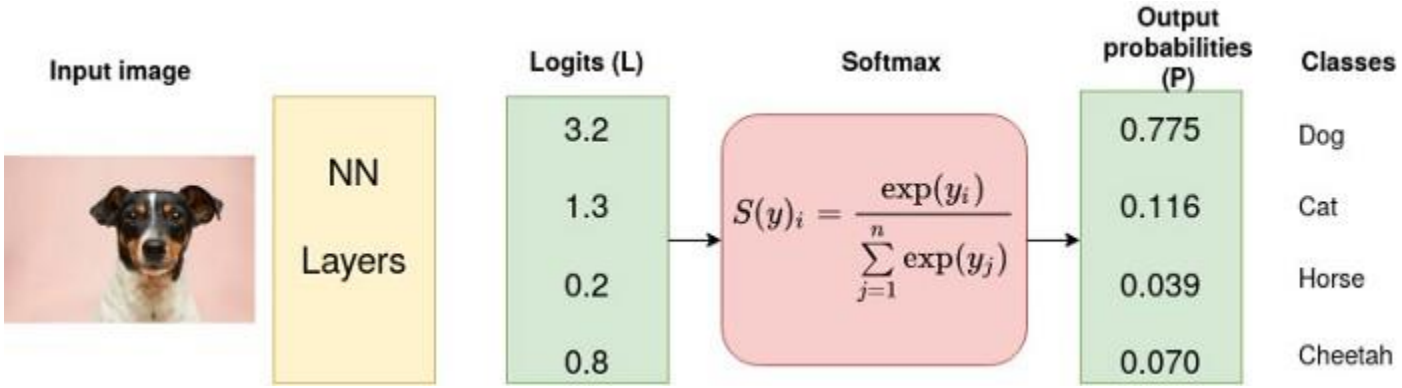
Bias In Neural Network





Logits and Softmax Layer

- Consider a CNN model which aims at classifying an image as either a dog, cat, horse or cheetah (4 possible outcomes/classes).
- The last (fully-connected) layer of the CNN outputs a vector of logits, L, that is passed through a Softmax layer that transforms the logits into probabilities, P.
- These probabilities are the model-predictions for each of the 4 classes.



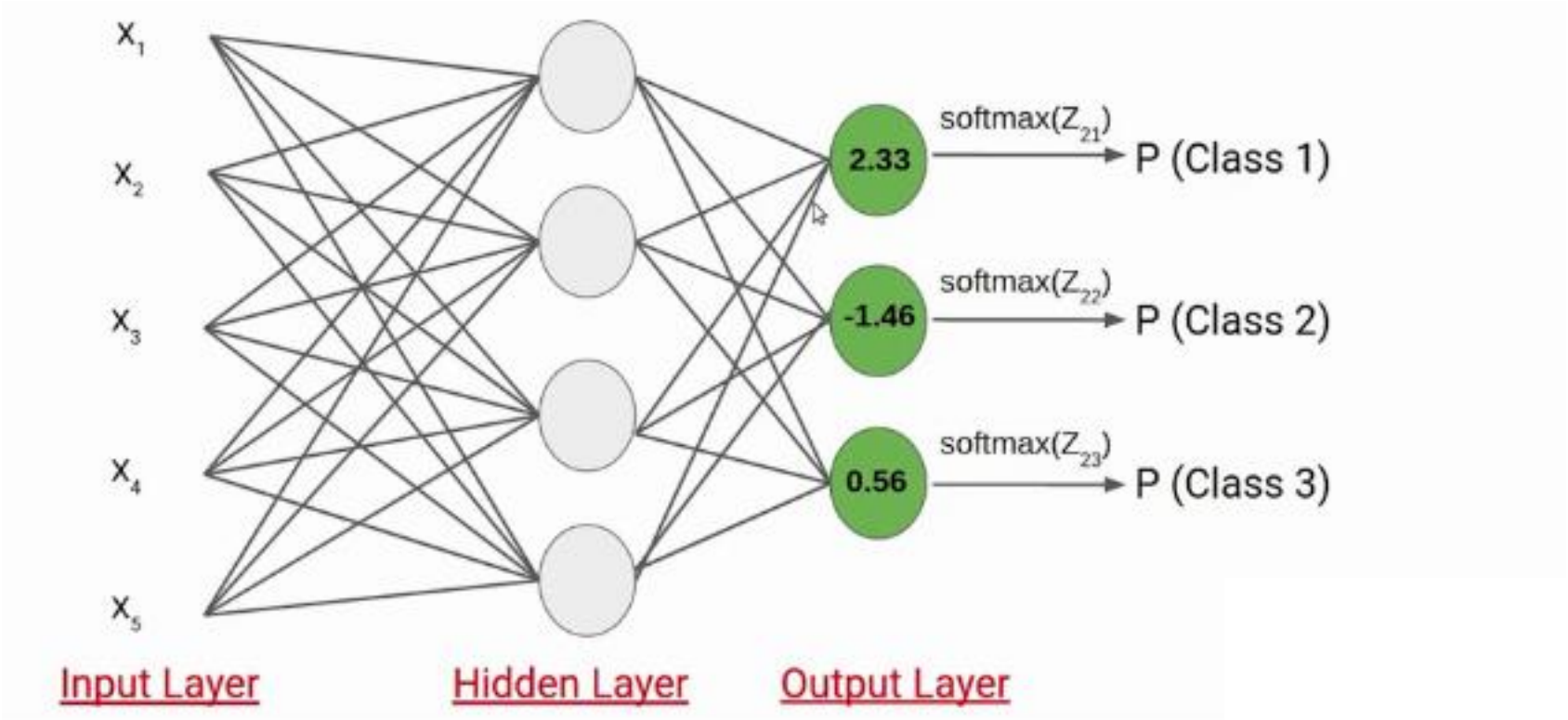
Softmax Activation

Softmax is an activation function that scales numbers/logits into probabilities.
The output of a Softmax is a vector (say *V)* with probabilities of each possible outcome. The probabilities in vector *V* sums to one for all possible outcomes or classes.

Mathematically, Softmax is defined as,

$$softmax(z_i) = \frac{exp(z_i)}{\sum_j exp(z_j)}$$

Softmax Layer



(Logits)

Softmax Layer

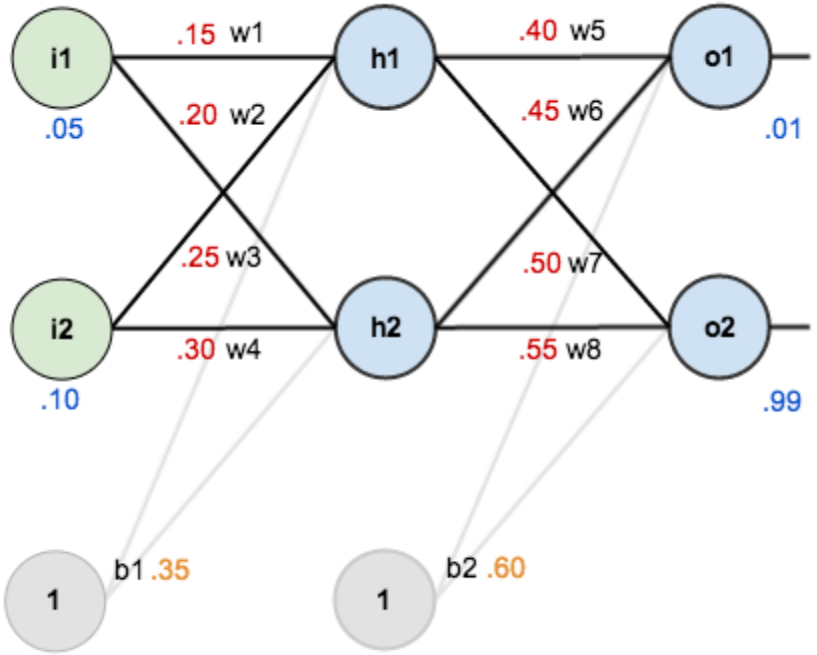
Example :

2.33 → $P(\text{Class 1}) = \frac{\exp(2.33)}{\exp(2.33) + \exp(-1.46) + \exp(0.56)} = 0.83827314$

-1.46 → $P(\text{Class 2}) = \frac{\exp(-1.46)}{\exp(2.33) + \exp(-1.46) + \exp(0.56)} = 0.01894129$

0.56 → $P(\text{Class 3}) = \frac{\exp(0.56)}{\exp(2.33) + \exp(-1.46) + \exp(0.56)} = 0.14278557$

Test



Find the values of h1, h2, o1, o2
use tanh activation function.

Strategy / Learning Algorithm

Supervised Learning

- Learning is performed by presenting pattern with target
- During learning, produced output is compared with the desired output
  - The difference between both output is used to modify learning weights according to the learning algorithm
- Recognizing hand-written digits, pattern recognition and etc.
- Neural Network models: perceptron, feed-forward, radial basis function, support vector machine.

Unsupervised Learning

- Targets are not provided
- Appropriate for clustering task
  - Find similar groups of documents in the web, content addressable memory, clustering.
- Neural Network models: Kohonen, self organizing maps, Hopfield networks.

Reinforcement Learning

- Target is provided, but the desired output is absent.
- The net is only provided with guidance (appreciate/criticize) to determine the produced output is correct or vise versa.
- Weights are modified in the units that have errors

Self Study (ANN Misc.)

Where can neural network systems help…

- when we can't formulate an algorithmic solution.
- when we ***can*** get lots of examples of the behavior we require.

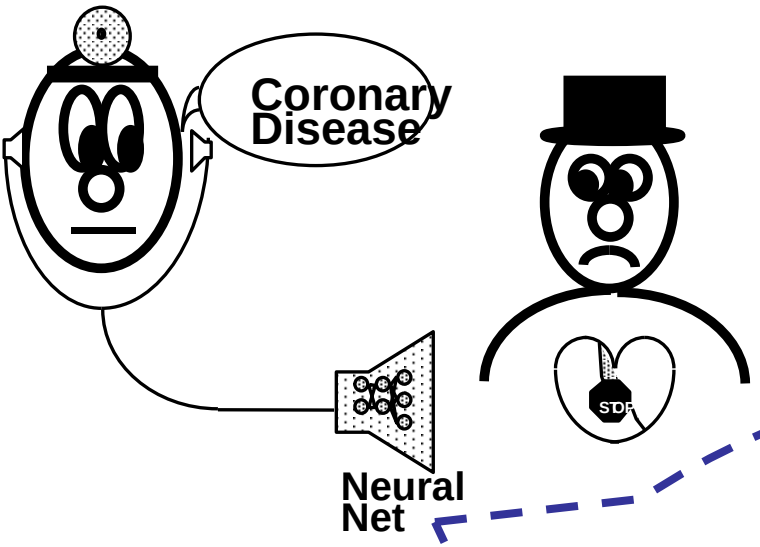
  'learning from experience'
- when we need to pick out the structure from existing data.

Who is interested?...

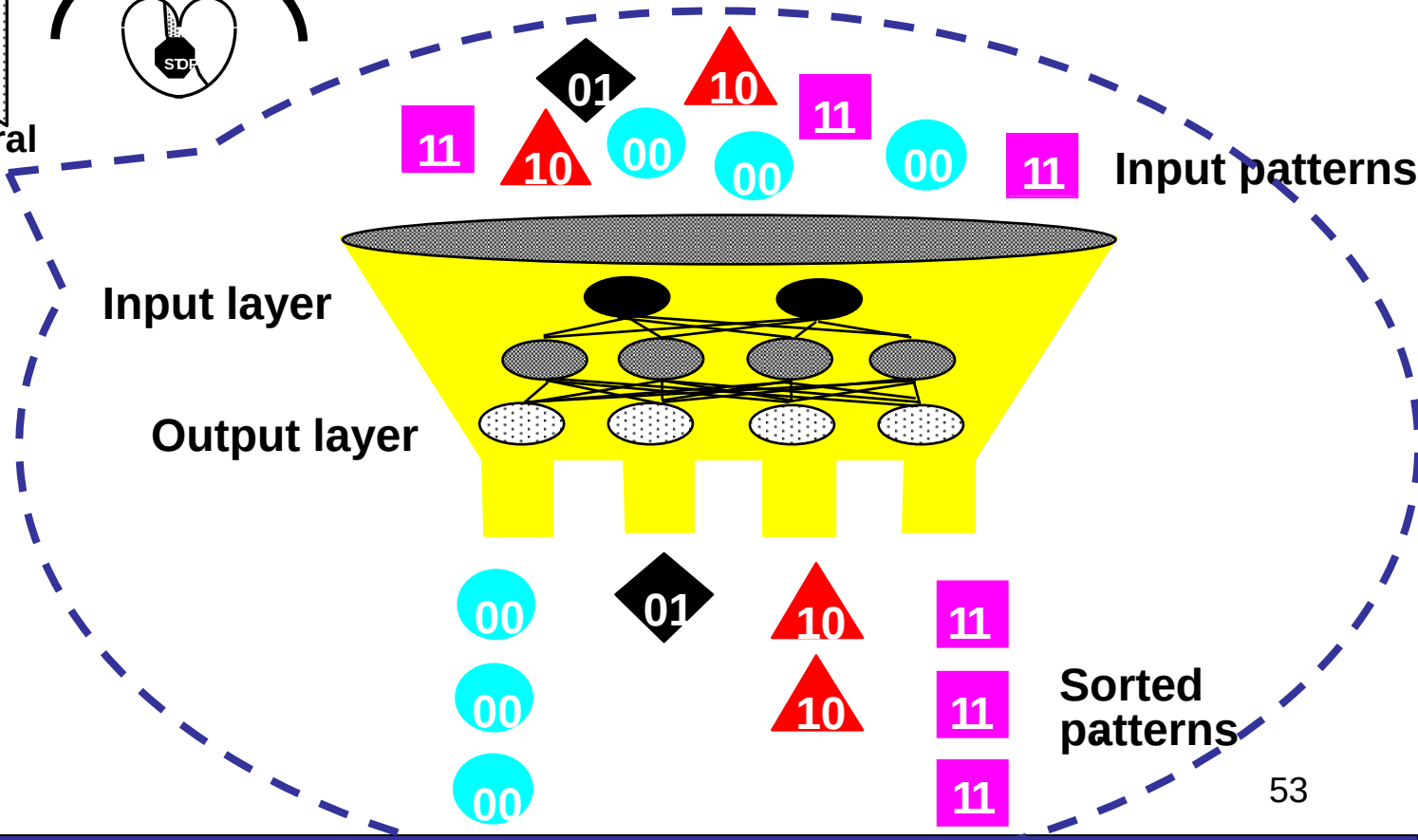
- Electrical Engineers – signal processing, control theory
- Computer Engineers – robotics
- Computer Scientists – artificial intelligence, pattern recognition
- Mathematicians – modelling tool when explicit relationships are unknown

Problem Domains

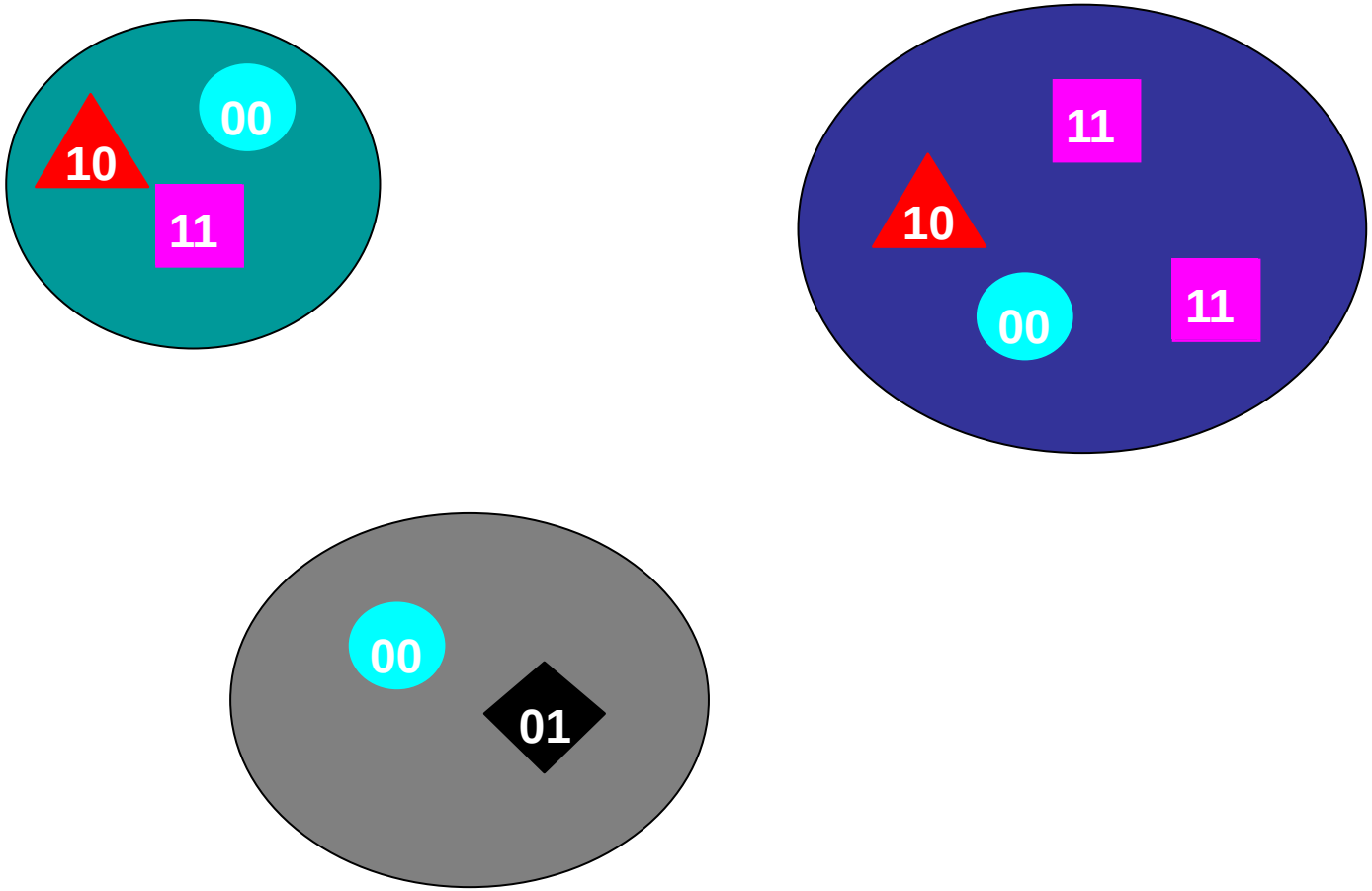
- Storing and recalling patterns
- Classifying patterns
- Mapping inputs onto outputs
- Grouping similar patterns
- Finding solutions to constrained optimization problems



Classification

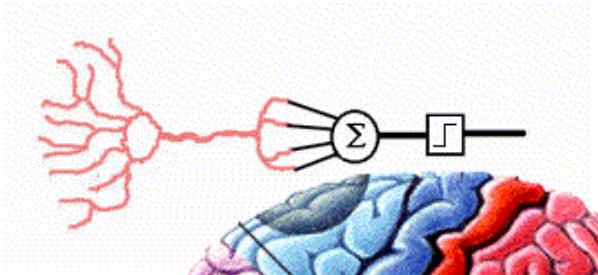


Clustering



ANN Applications

Medical Applications



Chemistry

Information
Searching & retrieval

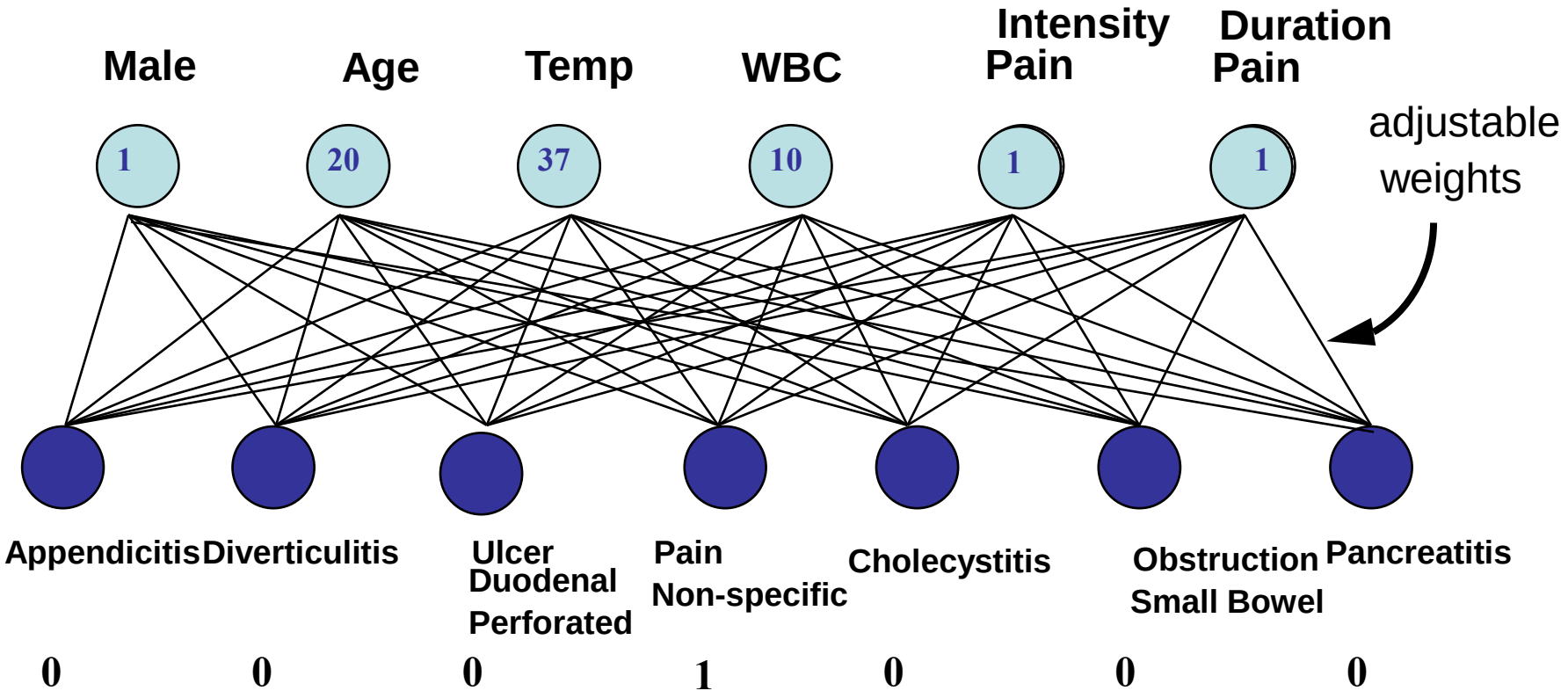
Education

Business & Management

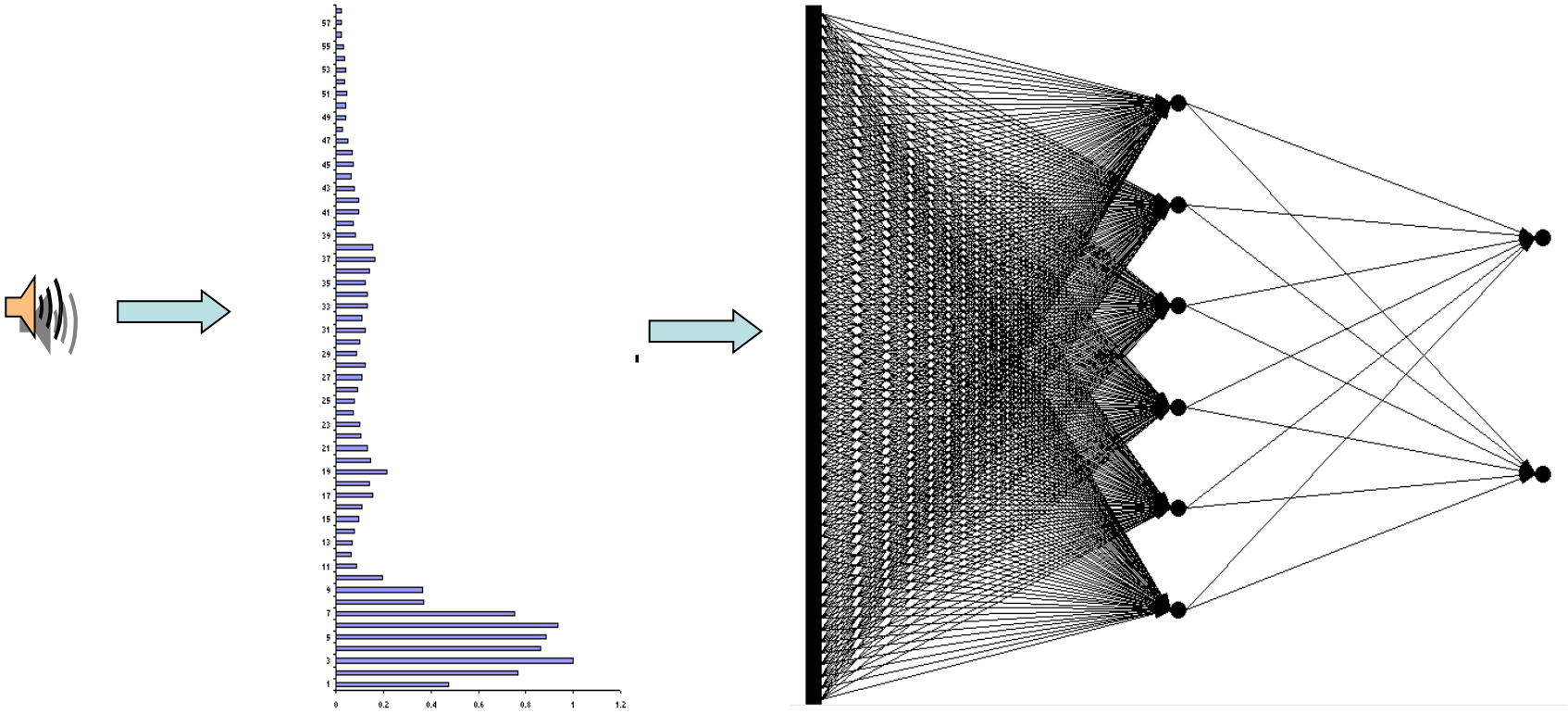
Applications of ANNs

- Signal processing
- Pattern recognition, e.g. handwritten characters or face identification.
- Diagnosis or mapping symptoms to a medical case.
- Speech recognition
- Human Emotion Detection
- Educational Loan Forecasting

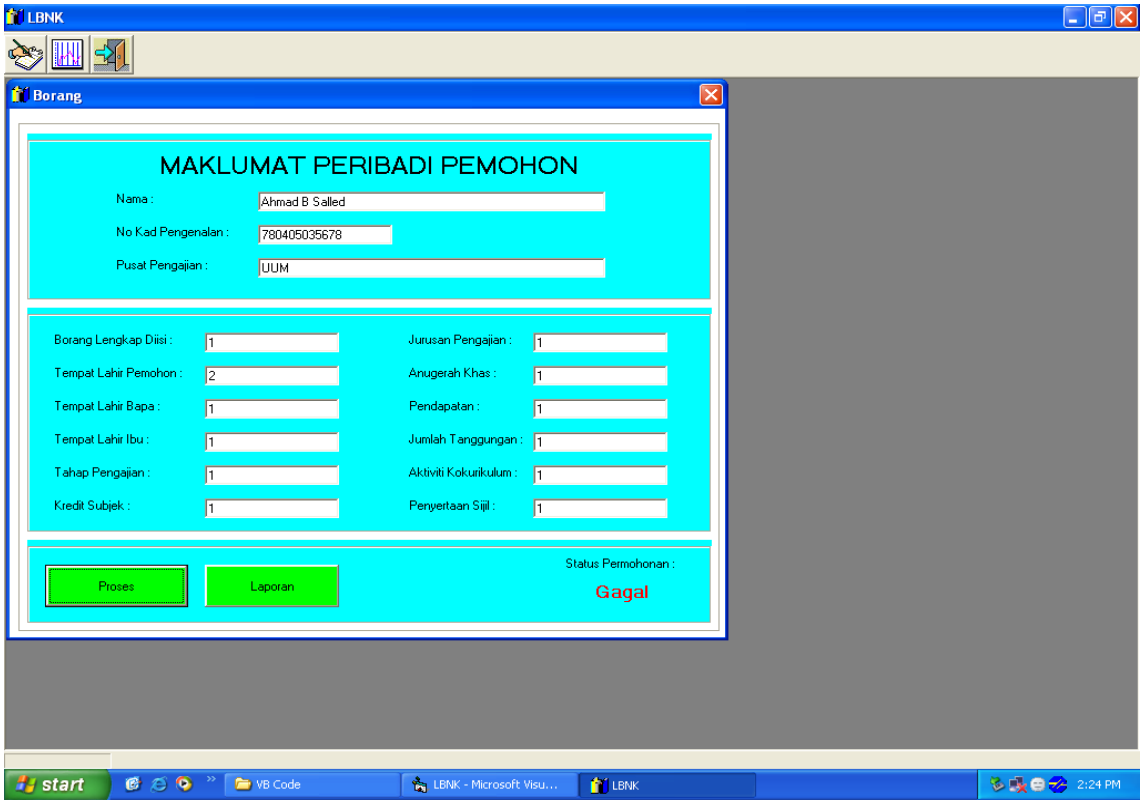
Abdominal Pain Prediction



Voice Recognition



Educational Loan Forecasting System



Advantages Of NN

NON-LINEARITY

It can model non-linear systems

INPUT-OUTPUT MAPPING

It can derive a relationship between a set of input & output responses

ADAPTIVITY

The ability to learn allows the network to adapt to changes in the surrounding environment

EVIDENTIAL RESPONSE

It can provide a confidence level to a given solution

Advantages Of NN

CONTEXTUAL INFORMATION

Knowledge is presented by the structure of the network. Every neuron in the network is potentially affected by the global activity of all other neurons in the network. Consequently, contextual information is dealt with naturally in the network.

FAULT TOLERANCE

Distributed nature of the NN gives it fault tolerant capabilities

NEUROBIOLOGY ANALOGY

Models the architecture of the brain

Comparison of ANN with conventional AI methods

CHARACTERISTICS	TRADITIONAL COMPUTING (including Expert Systems)	ARTIFICIAL NEURAL NETWORKS
Processing style	Sequential	Parallel
Functions	Logically (left brained) via Rules Concepts Calculations	Gestault (right brained) via Images Pictures Controls
Learning Method	by rules (didactically)	by example (Socratically)
Applications	Accounting, word processing, math, inventory, digital communications	Sensor processing, speech recognition, pattern recognition, text recognition