EE7207 Neural Networks and Deep Learning

Mao Kezhi

School of Electrical and Electronic Engineering
Nanyang Technological University

Email: ekzmao@ntu.edu.sg
Office: S2-B2c-84

Teaching Team

- Mao Kezhi
  - Course coordinator
  - Weeks 1-7

- Feng Zijian
  - Weeks 8-10

- Zhang Handuo
  - Weeks 11-13

Continuous Assessments and Exam

- Continuous Assessments (CA)
  - 40%
  - 2 assignments (to be released on Week 3, Week 9)
  - 2 quizzes (on Week 7, Week 13),
  - 10% each

- Exam
  - 60%
  - 4 questions (MK 2, FZJ 1, ZHD 1)
  - 25 marks each

References

1. Simon Haykin, Neural Networks and Learning Machines, 3rd Edition, Prentice Hall, 2009.

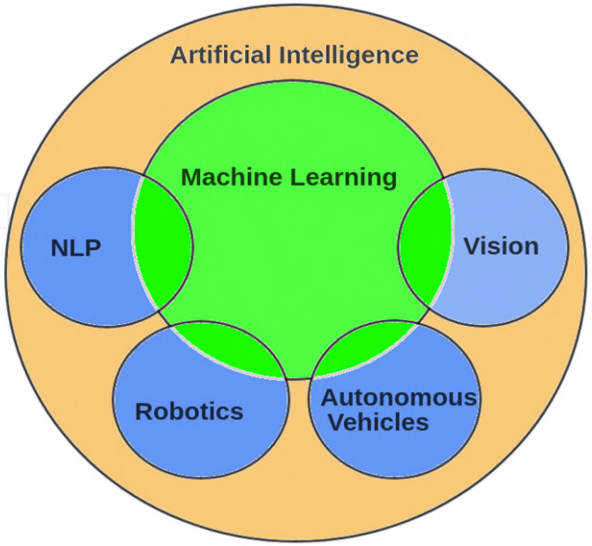
2. Deep Learning, by Ian Goodfellow, Yoshua Bengio and Aaron Courville, 2016, MIT Press.

3. Materials on the Internet, in conference proceedings and journals.

AI, Machine Learning, Neural Networks and Deep Learning: at a Glace

- **AI and Machine Learning**

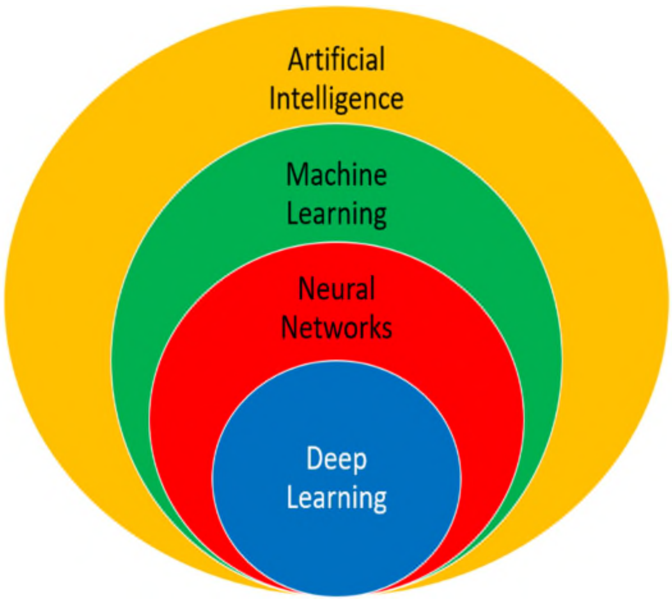
  - AI is a broad term for techniques that enable machines to mimic human behaviours
  - Machine learning is a sub-set of AI techniques that enable machines to improve with experience
    - Data-driven
    - Statistical algorithms



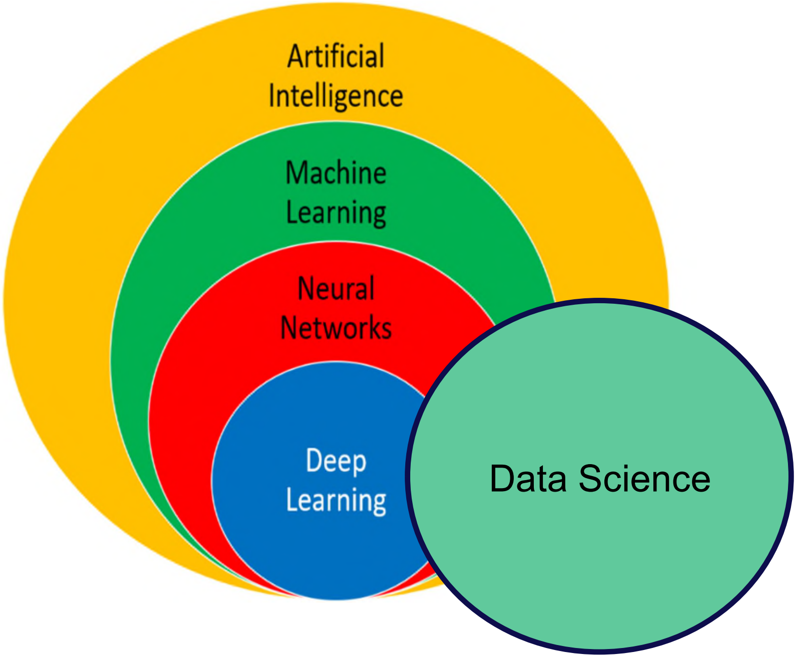
- **Machine Learning, Neural Networks and Deep Learning**

❑ Neural networks is a sub-field of machine learning

❑ Deep learning is a sub-field of neural networks

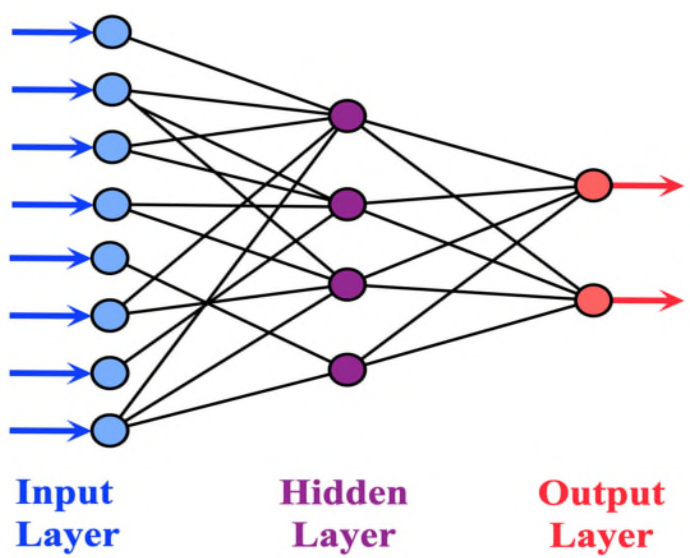


- **Data Science and Machine Learning**
  - Data science is a broad term covering everything about data
  - Machine learning focuses on learning algorithms: how to learn from data

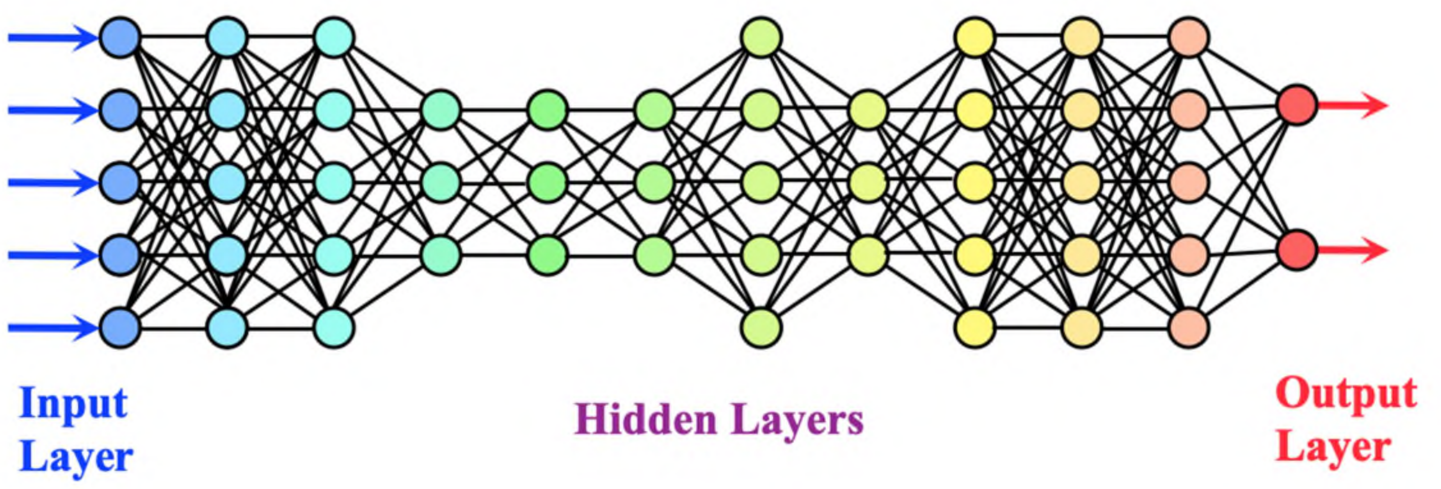


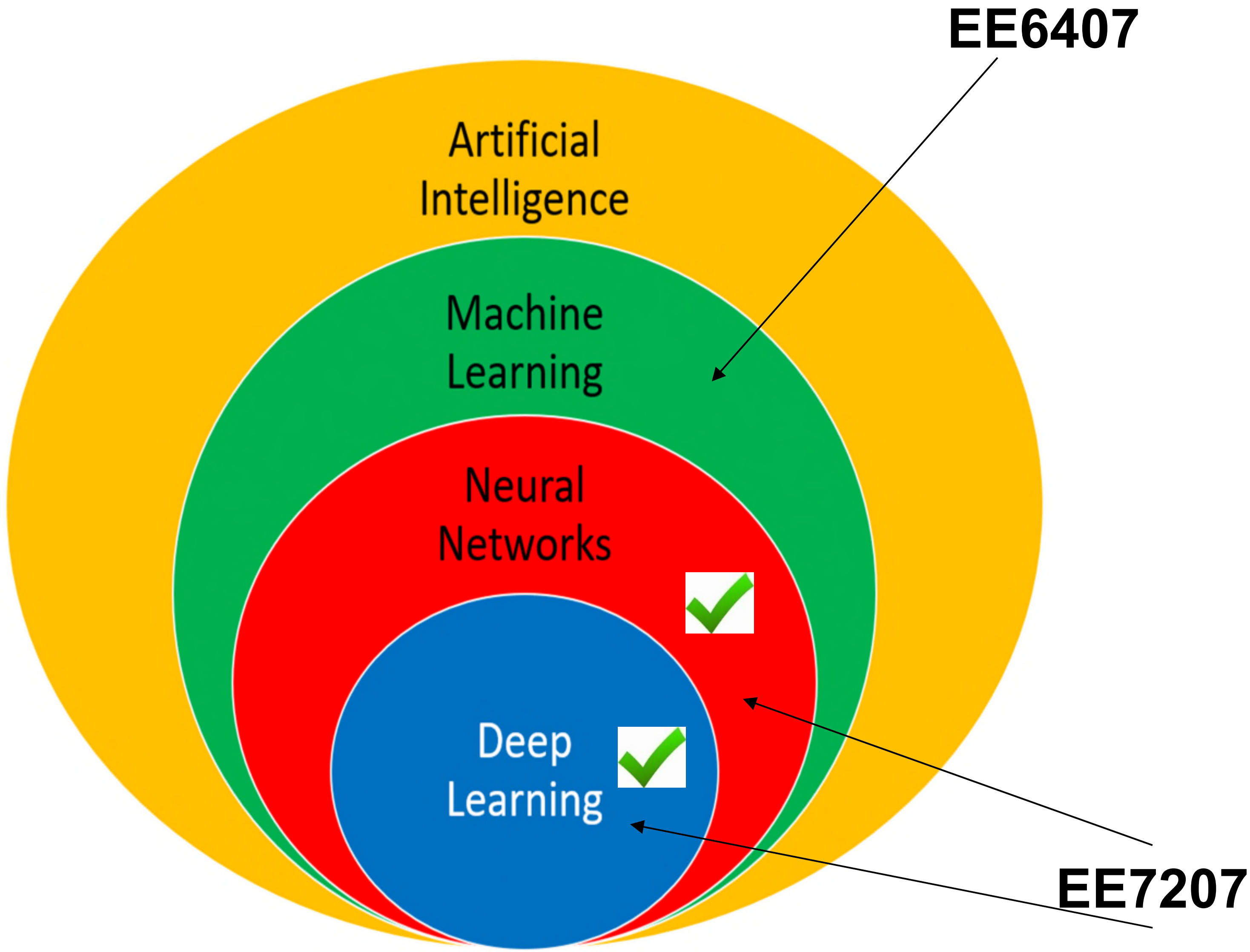
- **Shallow and Deep Neural Networks**

  - ❑ Shallow neural networks
    - ➢ **Structure**: with one or two layers, excluding input layer.
    - ➢ **Strengths**: relatively easy to train; require less computing resource, demand less training data; sufficient for some tasks.
    - ➢ **Limitations**: ability is limited when comes to capturing complex relationships or features in data like images, text, and speech.



- Deep neural networks
  - **Structure**: with many layers.
  - **Strengths**: excel at extracting features and relationships from complex data, leading to breakthroughs in fields like computer vision and natural language processing.
  - **Limitations**: training can be computationally expensive; require large amounts of training data; prone to overfitting.




EE6407
Artificial Intelligence
Machine Learning
Neural Networks
Deep Learning
EE7207

Contents

Introduction to Neural Networks

- An Overview of Neural Networks
- Neuron Models and Network Architectures
- Neural Network Learning

1. An Overview of Neural Networks

- **Why Artificial Neural Networks?**

  - Traditional AI systems are much less effective than human brains in many fields, such as visual information processing.

  - For example, a one-year-old baby is much better and faster at recognizing objects, faces and so on than many AI systems running on the fastest computer.

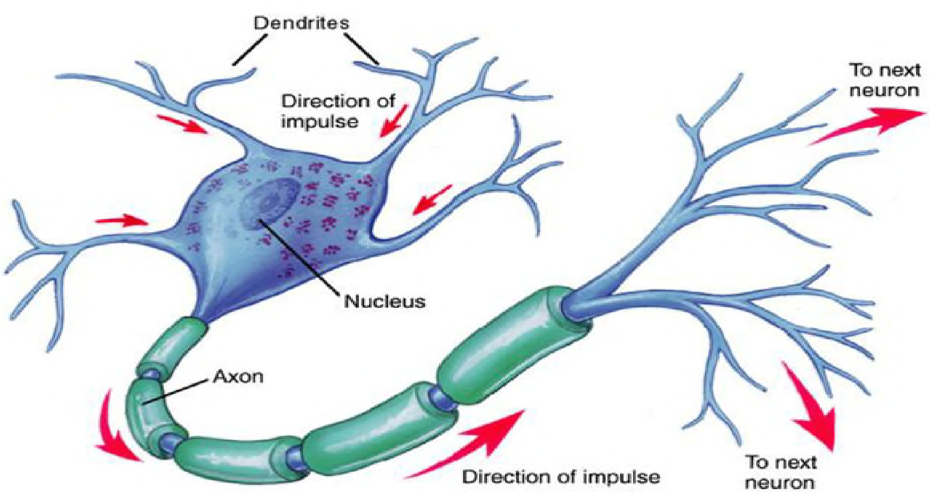
- **Features of Human Brain**

  - Human brain is robust and fault tolerant
  - Human brain is flexible
  - Human brain can deal with complex information that is fuzzy, probabilistic, noisy, or inconsistent
  - Human brain is highly parallel and highly nonlinear.

- **Biological Origin of Artificial Neural Networks**

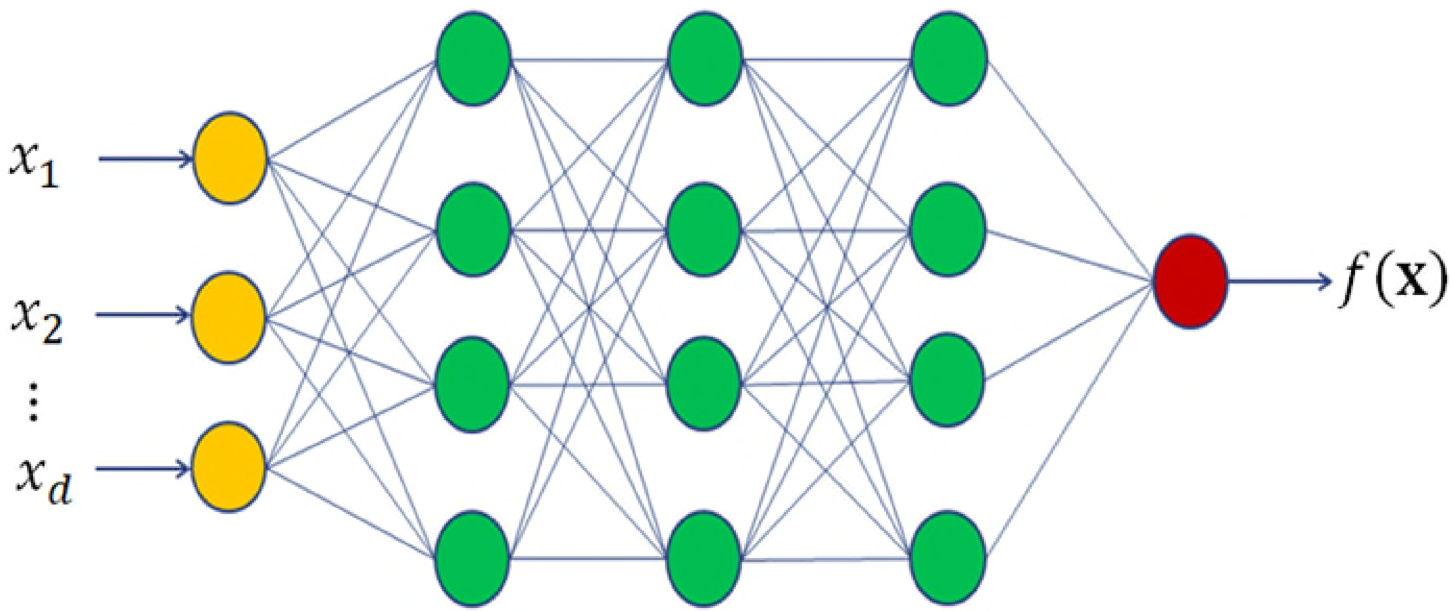
  - Two classes of cells:
    - Neurons (nerve cells)
    - Glia (glia cells)

  - A biological neuron consists of three major portions:
    - Cell body
    - Axon
    - Dendrites



- **What are artificial neural networks?**

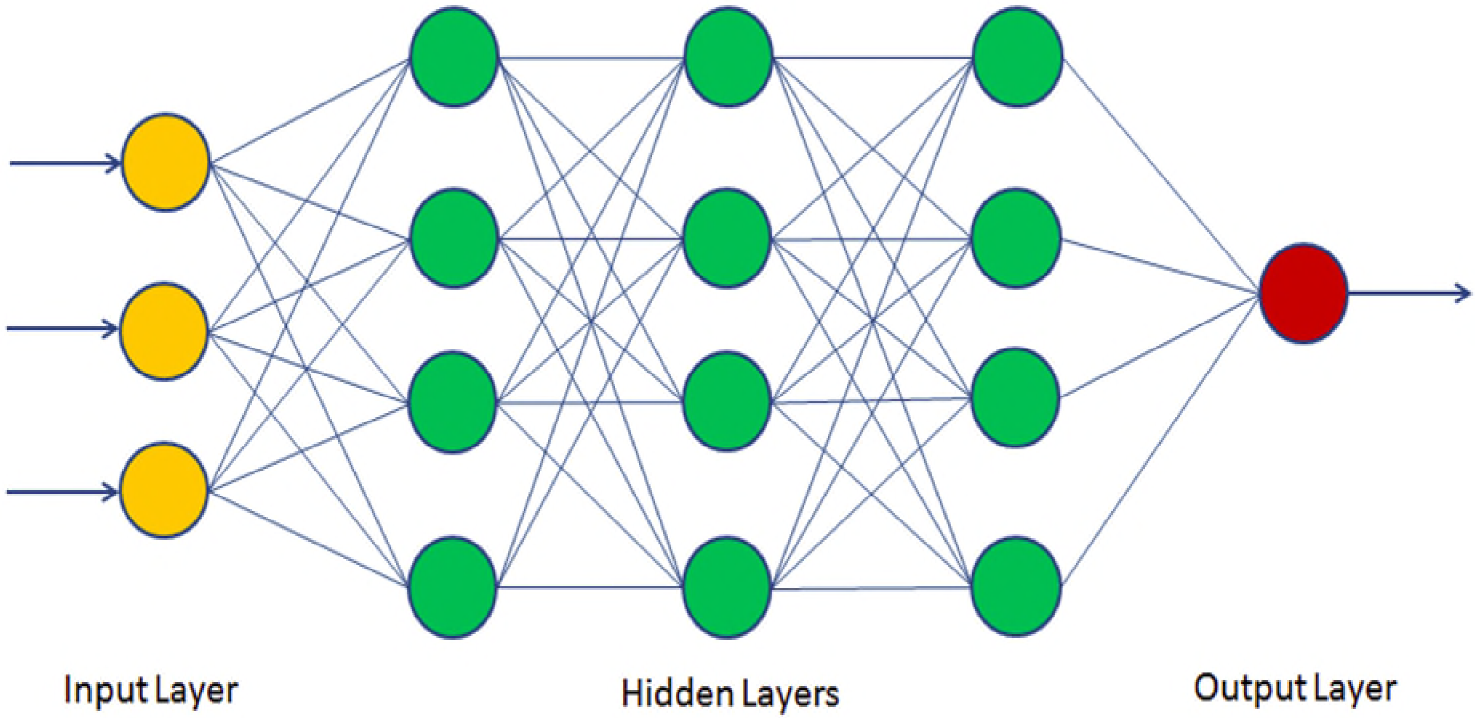
  Artificial neural networks are generalizations of mathematical models of human neural biology.



Neural networks are developed based on the following assumptions:

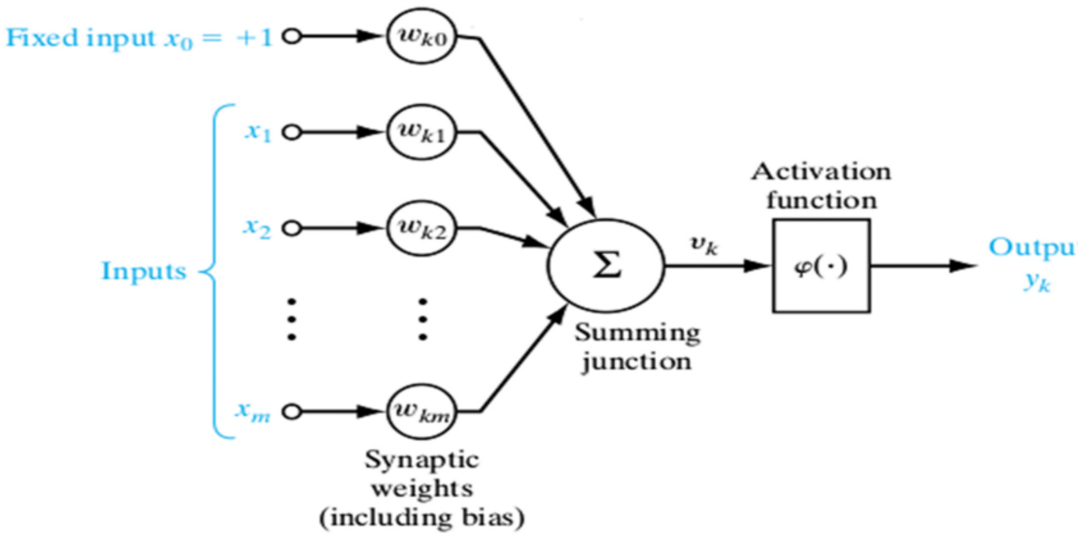
- Information processing occurs at many simple elements called neurons
- Signals are passed between neurons over connection links
- Each connection link has an associated weight, which multiplies the signal transmitted
- Each neuron applies an activation function to the input to determine the output signal

2. Neuron Models and Network Architectures



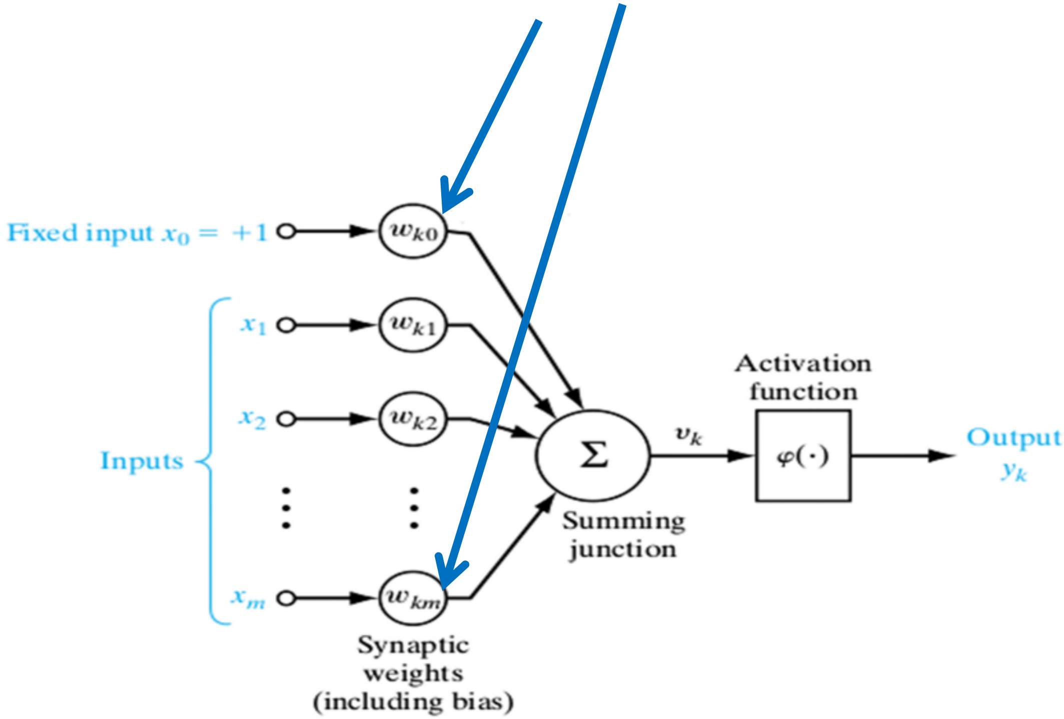
- **Neuron Models**

The model of a neuron is shown below:

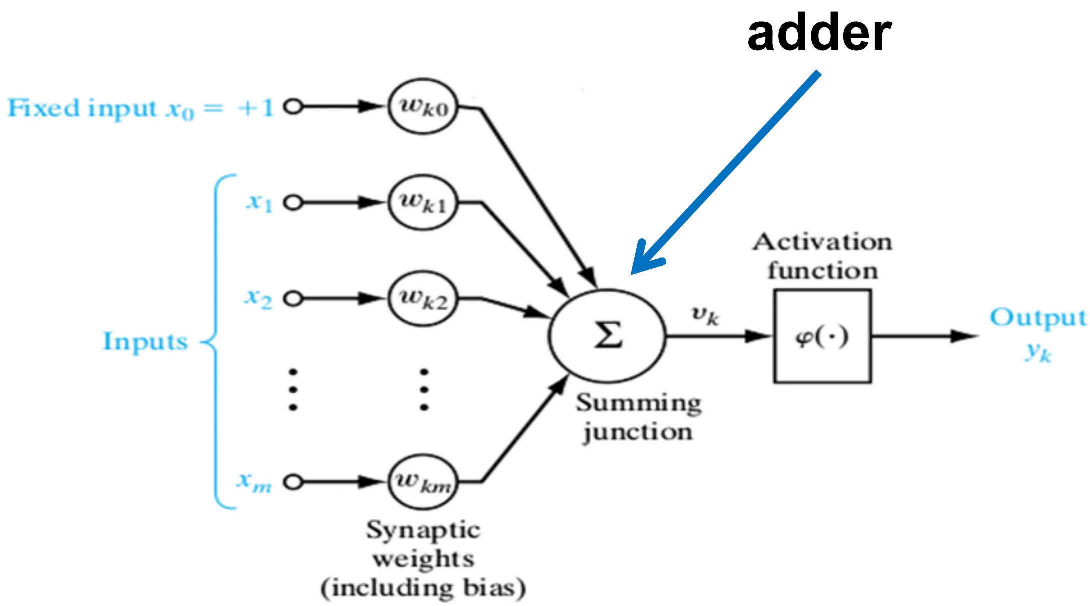


We may identify three basic elements of the neuron model:

- A set of synapses or connecting links, each of which is characterized by a <u>weight or strength</u> of its own

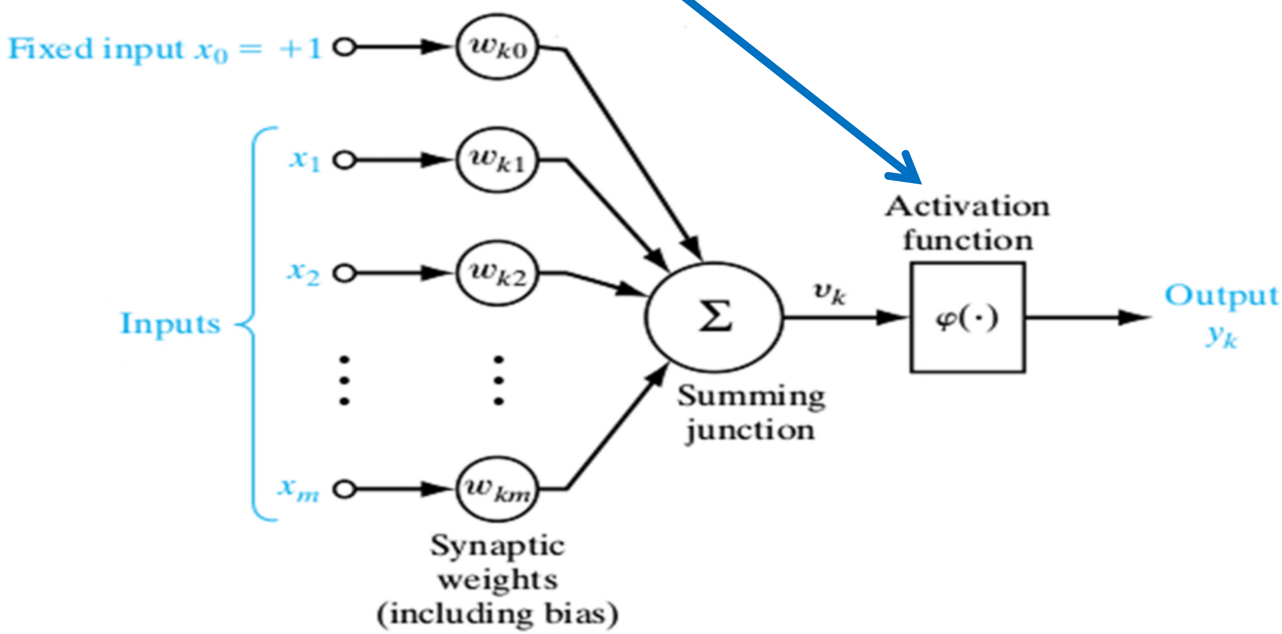


- An <u>adder</u> for summing the input signal



$$v_k = \sum_{j=0}^{m} w_{kj} x_j$$ activation signal

- An activation function for transforming the activation signal



$y_k = \varphi(v_k)$ output signal

- **Activation Functions**

(1) Binary Function:

$$f(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases}$$

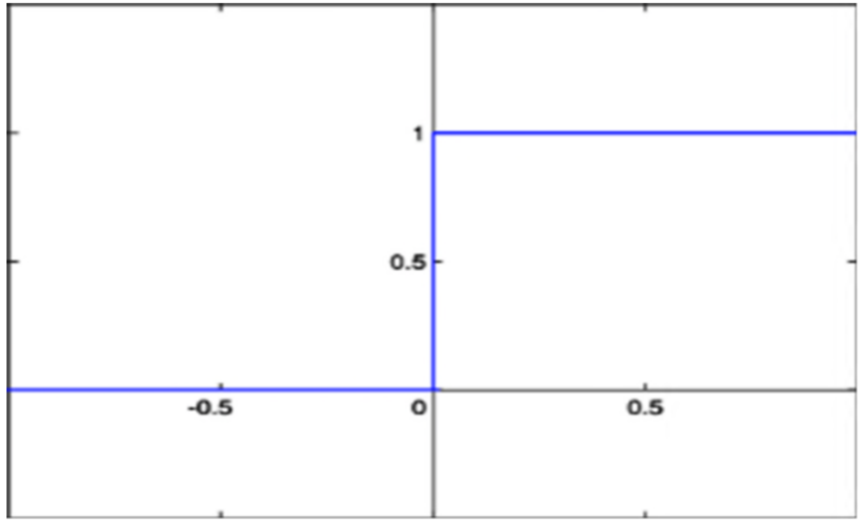
uni-polar binary

$$f(x) = \begin{cases} 1 & x \geq 0 \\ -1 & x < 0 \end{cases}$$

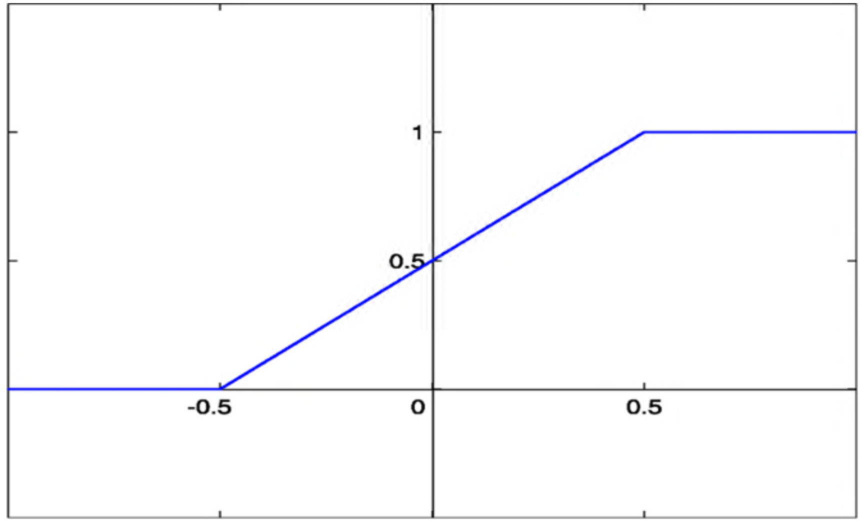
bi-polar binary

(2) Piecewise-Linear Function

$$f(x) = \begin{cases} 1 & x \geq 0.5 \\ x + 0.5 & -0.5 < x < 0.5 \\ 0 & x \leq -0.5 \end{cases}$$



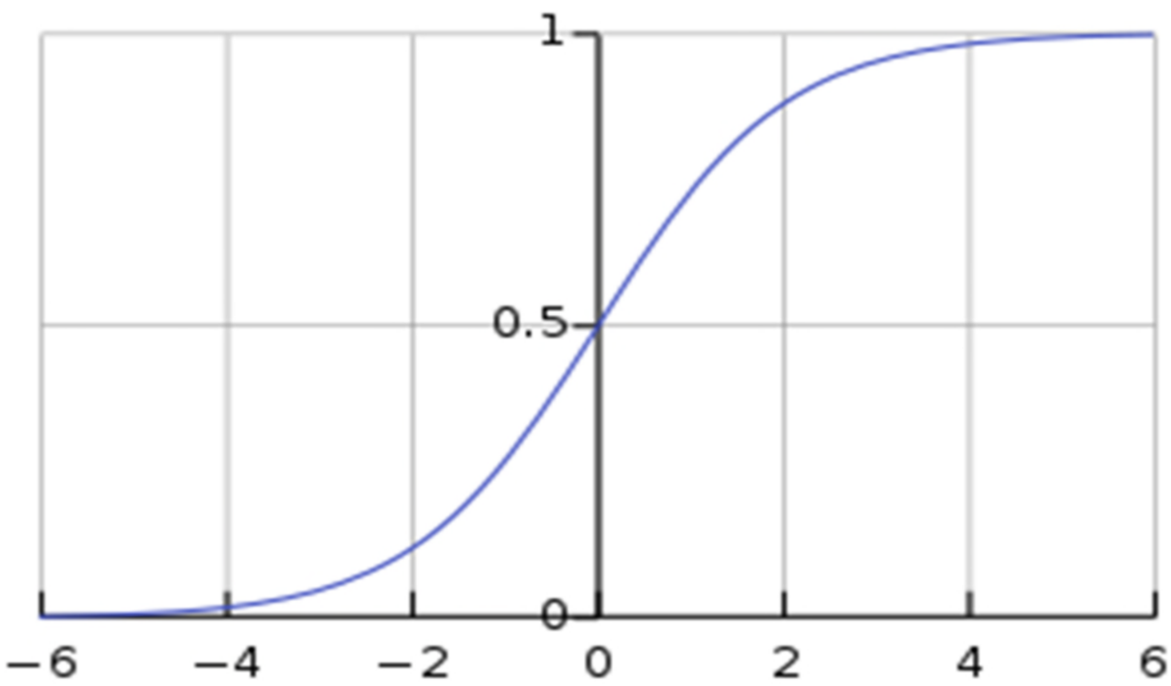
uni-polar binary



piecewise linear

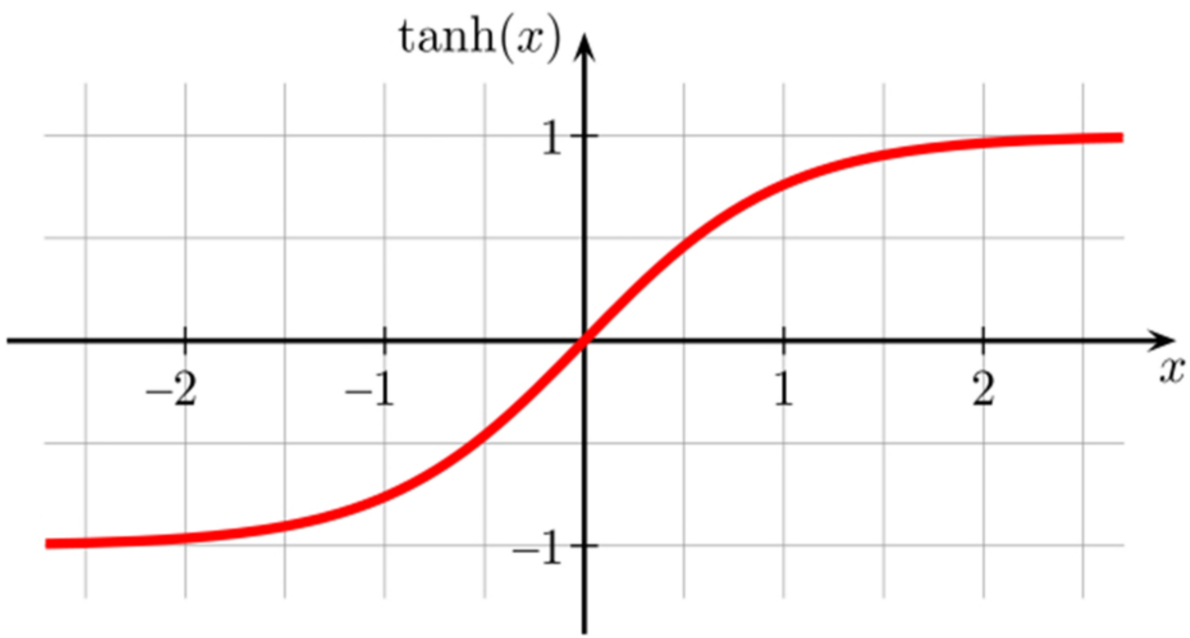
(3) Sigmoid Function

$$f(x) = \frac{1}{1 + \exp(-x)}$$



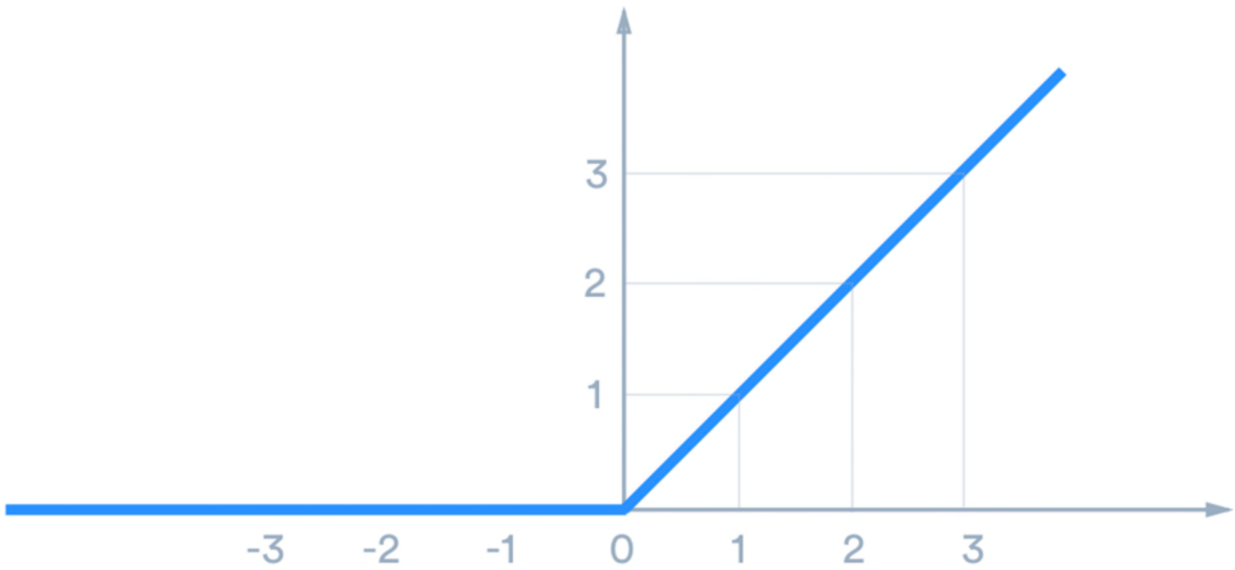
(4) Hyperbolic Tangent Function

$$\tanh(x) = \frac{\exp(x) - \exp(-x)}{\exp(x) + \exp(-x)}$$



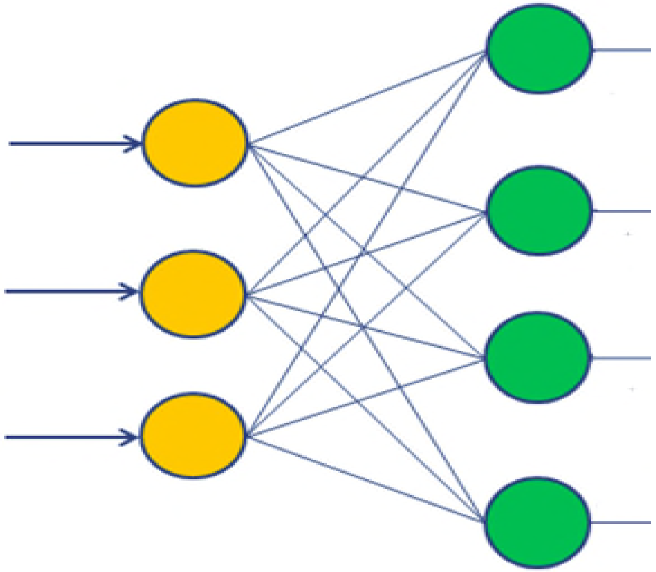
(5) Rectified Linear Unit

$$\mathbf{ReLU}(x) = \max(0, x)$$

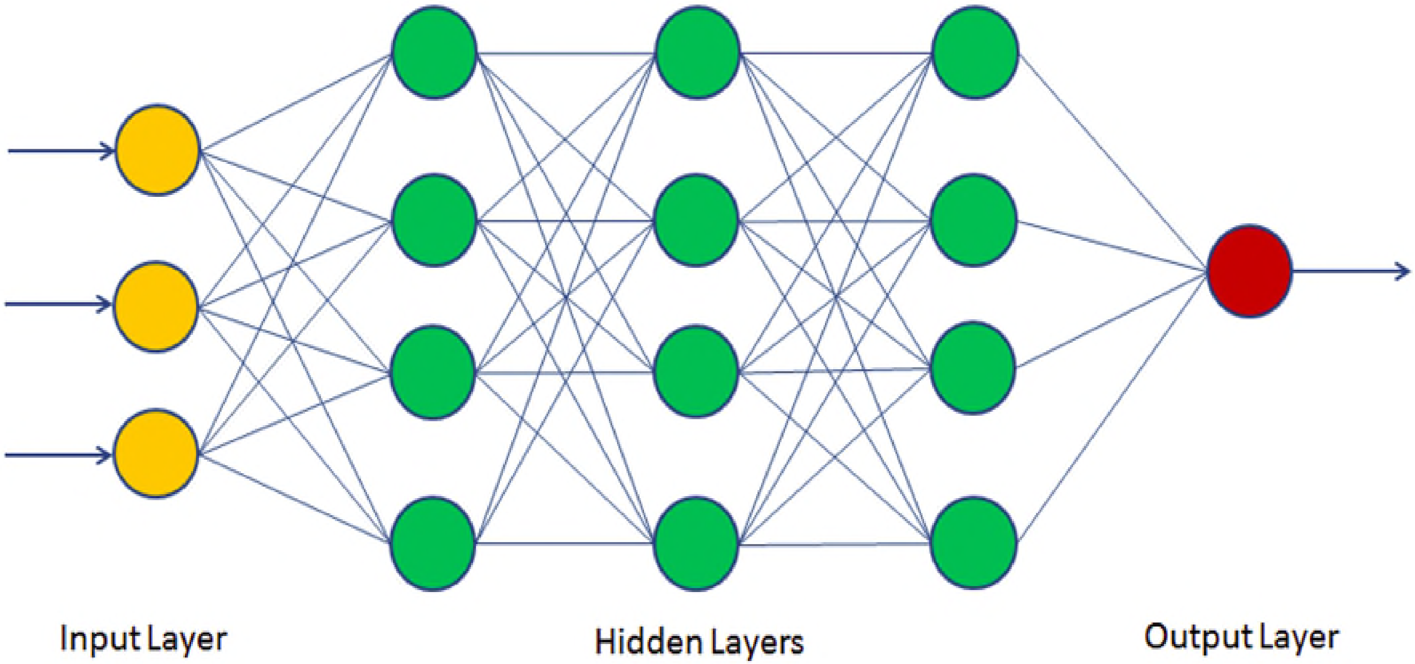


- **Neural Network Architectures**

(1) Feed-Forward Neural Networks



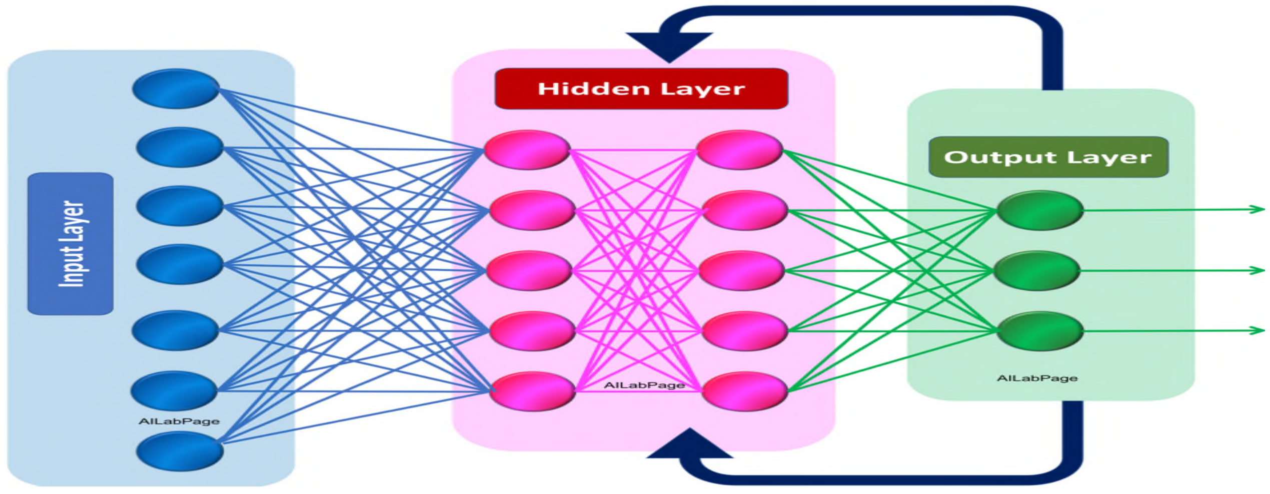
Single-layer feed-forward network



Multilayer feed-forward neural network

(2) Recurrent Neural Networks

feedback



A recurrent neural network has at least one feedback loop!

- **How Are Neural Networks Used?**

❑ **Phase 1: neural network development**

➢ An appropriate architecture is selected
➢ Train the neural network
➢ Test the neural network trained.

❑ **Phase 2: neural network deployment**

The neural network developed in phase 1 is applied to newly acquired data

- **How to Train Neural Networks?**

The major task of a neural network is to learn a model of the world based on two types of information of the world using suitable learning algorithms.

(1) Prior information: Any knowledge or assumptions we have about a problem, data, or model before observing the data.

(2) Observations. The observations are often referred to as *examples/samples/data*.

- **Data: labelled and unlabelled**

(a) In labeled samples, each sample is paired with a target value, which is called class label in pattern classification.

Example 1: Object Recognition

Dog

Class label

Cat

Example 2: Sentiment Analysis

Class label

Sentiment	Tweets
Negative	@united is the worst. Nonrefundable First class tickets? Oh because when you select Global/FC their system auto selects economy w/upgrade.
	@united I will not be flying you again
Neutral	@VirginAmerica my drivers license is expired by a little over a month. Can I fly Friday morning using my expired license?
	@VirginAmerica any plans to start flying direct from DAL to LAS?
Positive	@VirginAmerica done! Thank you for the quick response, apparently faster than sitting on hold ;)
	@united I appreciate your efforts getting me home!

(b) In unlabeled data, just the input signal is available, the target value is unknown.

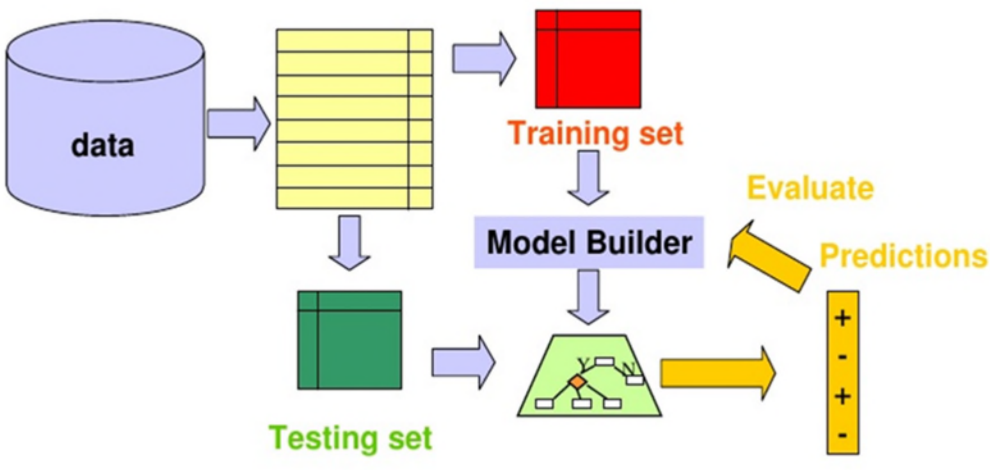
For example, a basket of fruits without naming

- **Training and Testing Data**

  Data is often divided into two subsets:

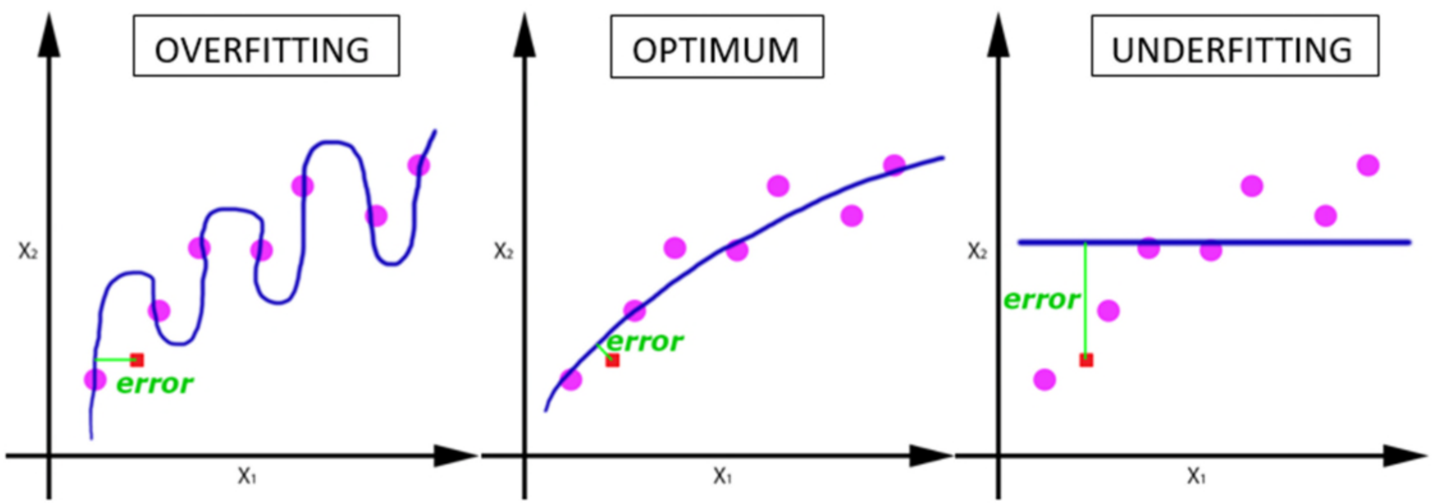
  ❑ Training set: to train (build) the model

  ❑ Testing set: to evaluate the performance of the model



Quite often, part of the data is used as *validation set* for determining hyper-parameters of the models or the learning algorithms.

- **Overfitting and Under-fitting**
  - ❑ If a network performs well on the training data but very badly on the testing set, the network might be *over-trained (overfitting)*.
  - ❑ If the network performs bad on the training data, the network might be *under-trained (under-fitting)*. An under-trained network also performs badly on the testing set.



3. Neural Network Learning/Training

- **What is learning/training?**

- In the context of neural networks, *learning* is defined as a process by which the free parameters of a neural network are determined
- Learning is through a continuing process of stimulation by the environment

The definition implies that:

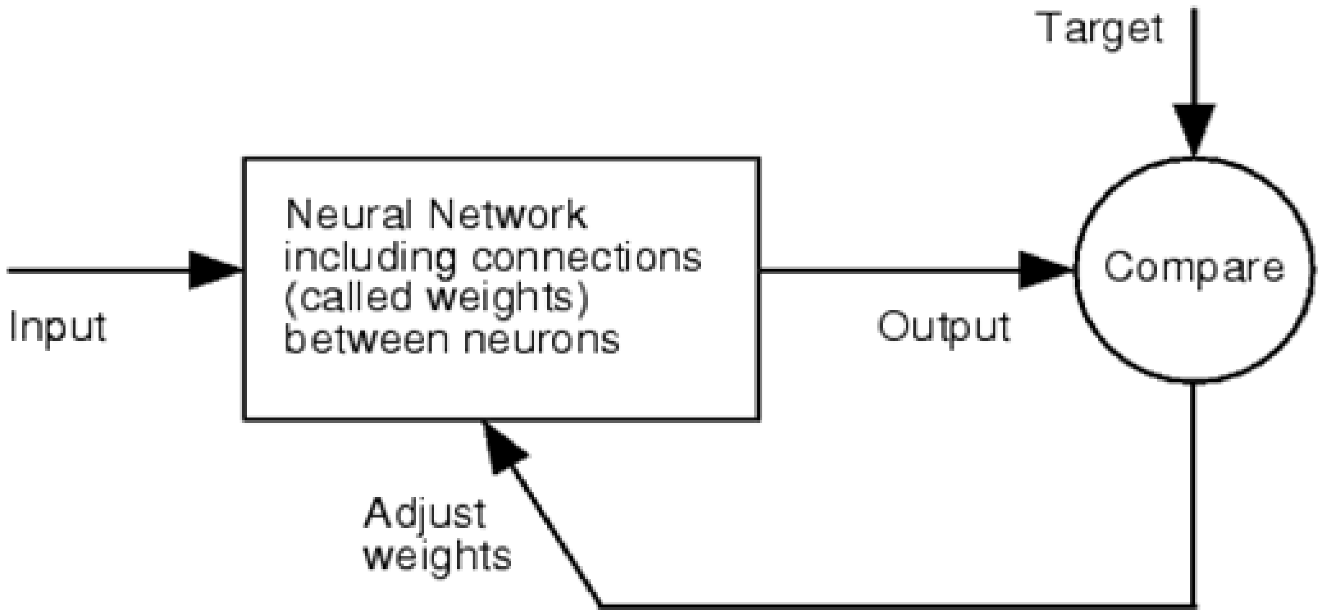
- The network is stimulated by the environment
- The network changes as a result of stimulation
- The network responds to the environment in a new way after the change.

- **Three learning rules**
  - Error-correction learning
  - Hebbian learning
  - Competitive learning

- **Two learning paradigms**
  - Supervised learning
  - Unsupervised learning

(1) Error-correction Learning

In error correction learning, the weight adjustment is based on the error, which is defined as the difference between the target value and the network output.



One example error-correction learning rule is of the following form:

$$\Delta w_{kj}(n) = \eta e_k(n) x_j(n)$$

$$w_{kj}(n+1) = w_{kj}(n) + \Delta w_{kj}(n)$$

where η is a positive constant called *learning rate*. $e_k(n)$ is the error signal for neuron k at step n.

The adjustment of the weight, Δw_{kj}, in error-correction learning is proportional to the product of the error signal e_k and the corresponding input signal x_j.

(2) Hebbian Learning

Hebbian Learning is inspired by the biological neural weight adjustment mechanism:

- If two neurons on either side of a connection are activated simultaneously, then the strength of that connection (i.e. weight) is selectively increased.
- If two neurons on either side of a connection are activated asynchronously, then the strength of that connection is selectively weakened or eliminated

The adjustment to the weight at step n is expressed in the general form:

$$\Delta w_{kj}(n) = f[x_j(n), y_k(n)]$$

Where f denotes a function.

The function f has many forms, the simplest form is as follow:

$$\Delta w_{kj}(n) = \eta x_j(n) y_k(n)$$

Where η is the *learning rate*.

The adjustment of the weight in Hebbian Learning, Δw_{kj}, is proportional to the product of the input signal x_j and output signal y_k.

(3) Competitive Learning

In competitive learning, the neurons of a neural network compete for being the winning neuron.

3 basic elements in a competitive learning rule:

- A set of neurons
- A mechanism of competition
- A mechanism allowing the winning neuron to update its weights

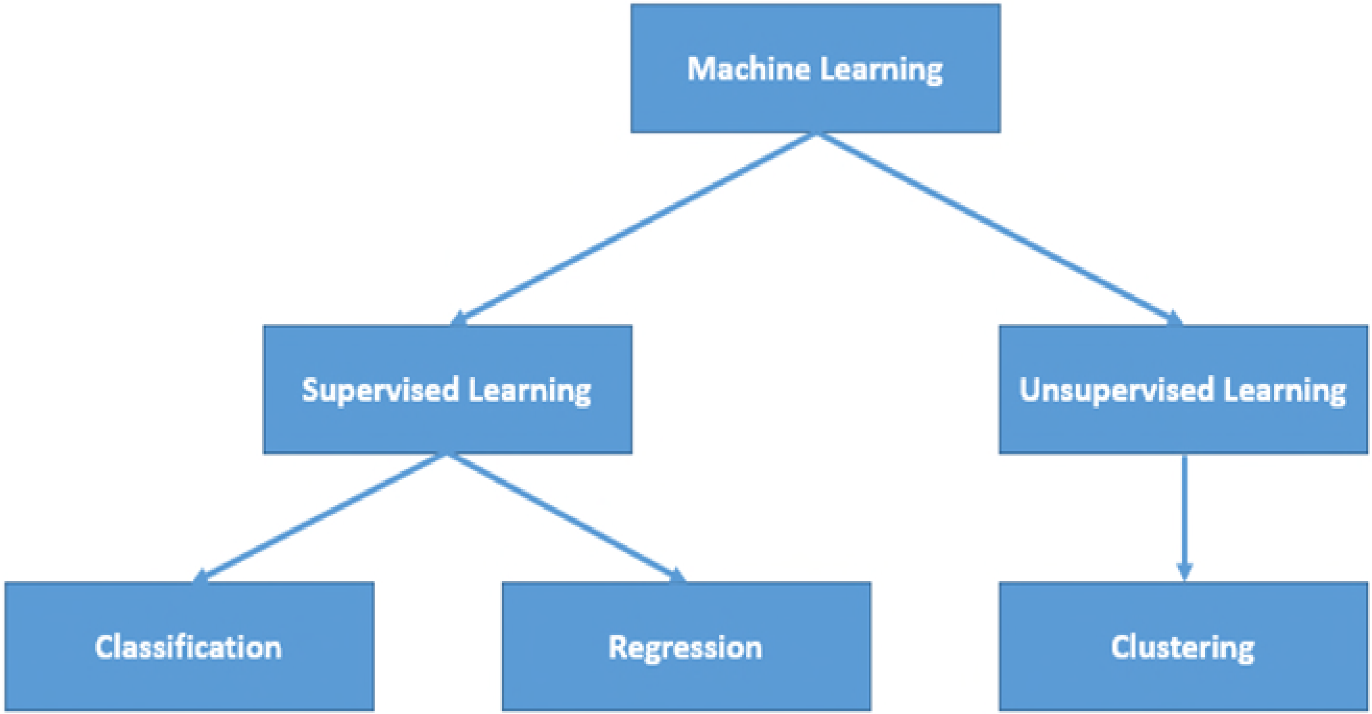
The weights of the winning neuron k are updated in the following way:

$$\Delta w_{kj}(n) = \begin{cases} \eta(x_j(n) - w_{kj}(n)) & \text{if neuron } k \text{ wins} \\ 0 & \text{otherwise} \end{cases}$$

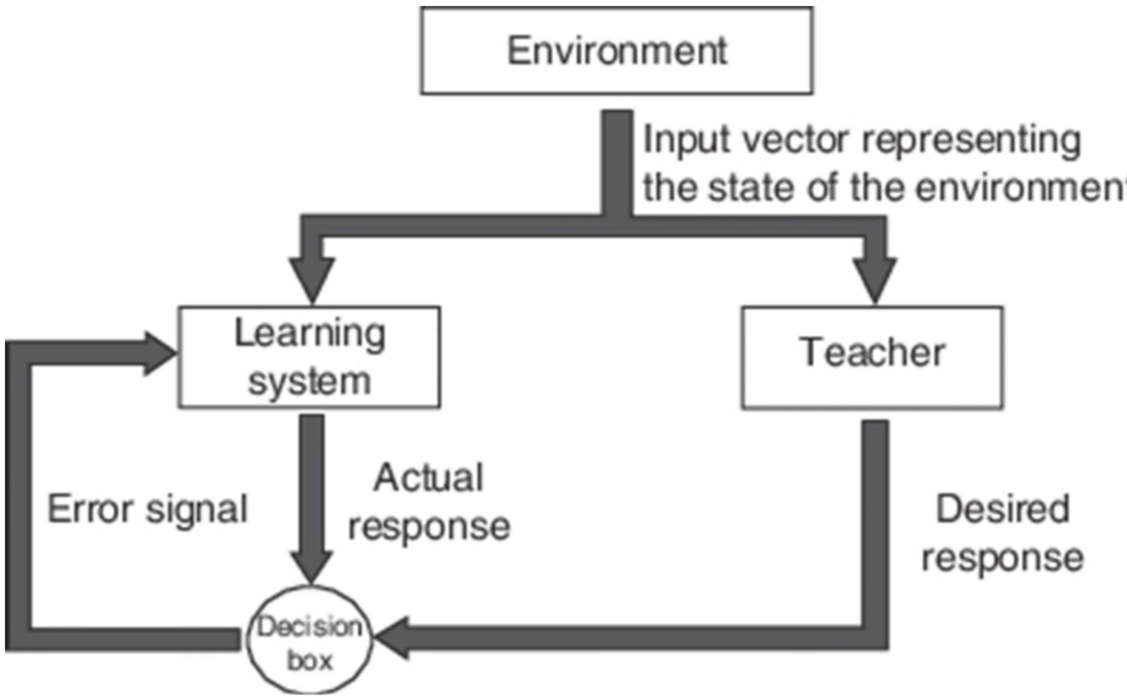
Where η is the learning rate.

The adjustment of the weight, Δw_{kj}, in competitive learning is proportional to the difference of the input signal x_j and the weight w_{kj}.

- **Supervised vs Unsupervised Learning**

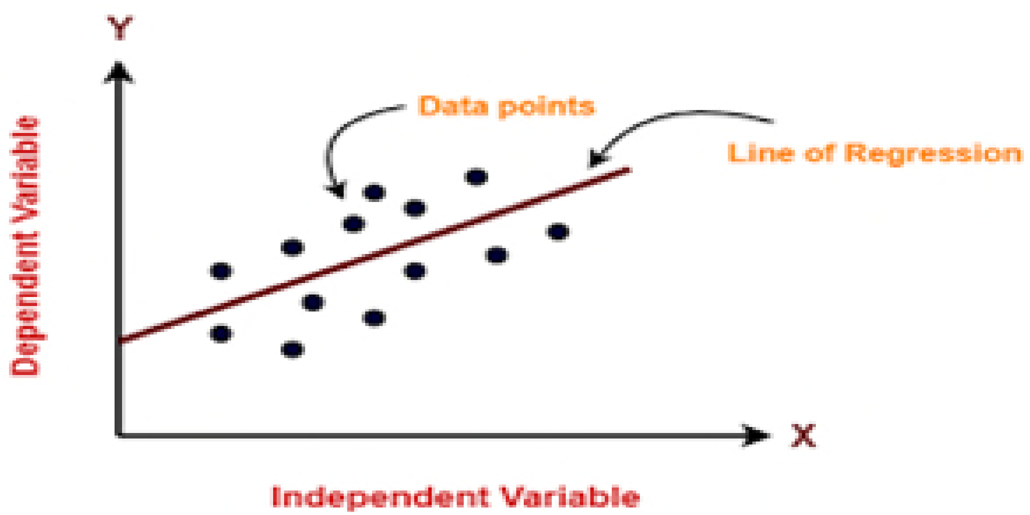


(1) Supervised Learning



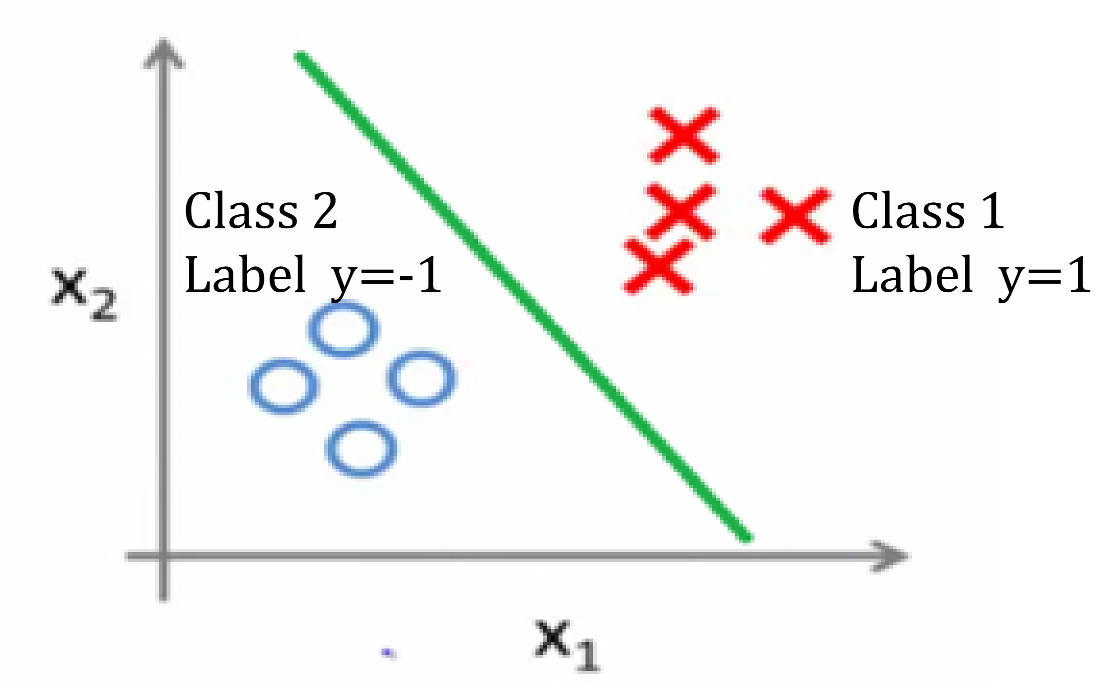
- **Regression vs Classification**
  - Regression aims to approximate a mapping function from input $\mathbf{x}$ to a continuous output y:

$$y = \mathbf{w}^T\mathbf{x} + b$$



❑ Classification aims to map input **x** to a discrete output y. It can also be interpreted as finding a decision boundary to separate samples in different classes:

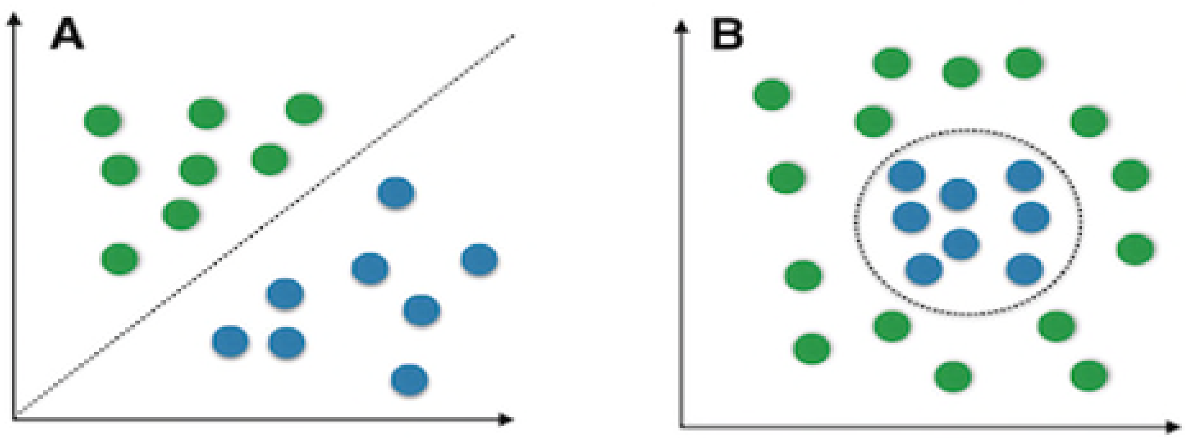
$$y = w_1 x_1 + w_2 x_2 + b = \mathbf{w}^T \mathbf{x} + b$$



- **Linear vs Nonlinear (e.g. neural network) Classifier**

❑ Linear classifier: a linear equation (a hyperplane)

❑ If a hyperplane is unable to separate, then a nonlinear classifier (hypersurface) is needed

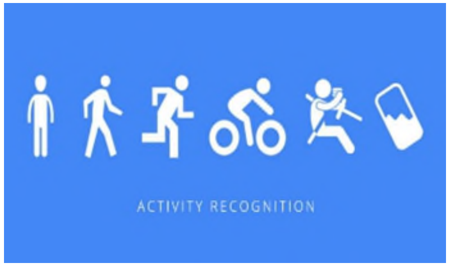


Linear classifier

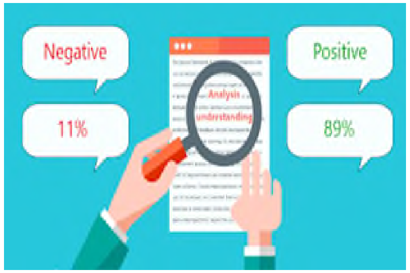
Nonlinear classifier
(e.g. neural networks)

Examples of Classification Tasks

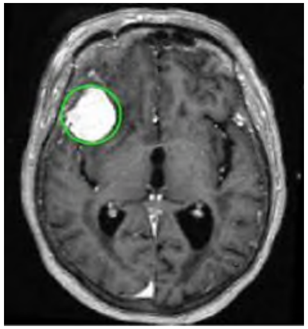
Activity Recognition



Sentiment Analysis



Brain Tumour Detection

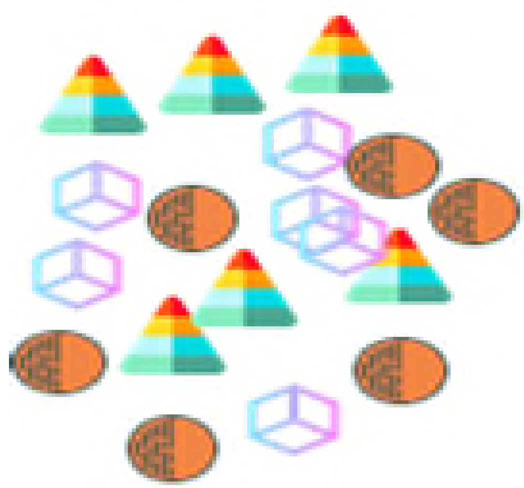


Object Recognition

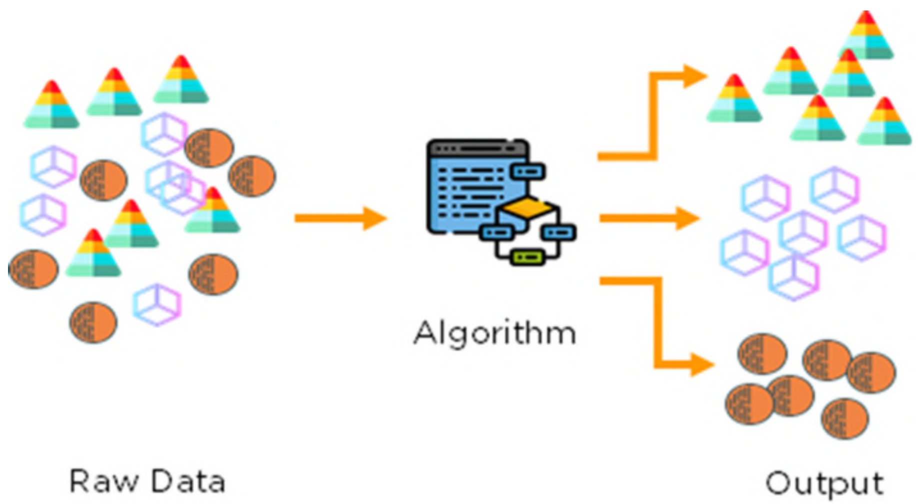
(2) Unsupervised Learning

- Unsupervised learning does not require a teacher, *i.e.* there is no target value
- The goal of unsupervised learning is to group data based on their intrinsic properties

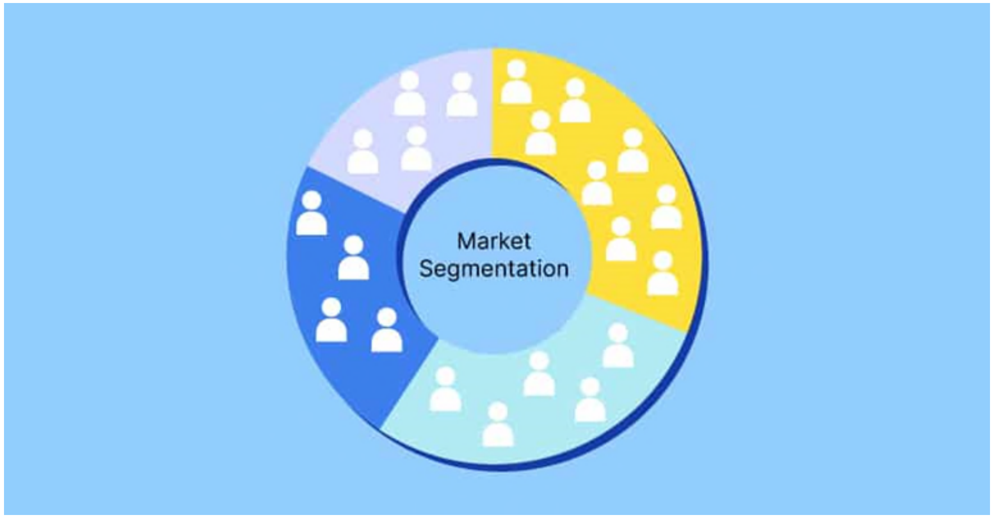
For example, a set of objects (without knowing their names):



The unsupervised learning algorithm organizes the objects into three groups:



- **Application Examples of Unsupervised Learning**
  - ❑ Market and customer segmentation based on needs, location, interests or demographics etc.



❑ Image compression

❑ Phenotype clustering in health care