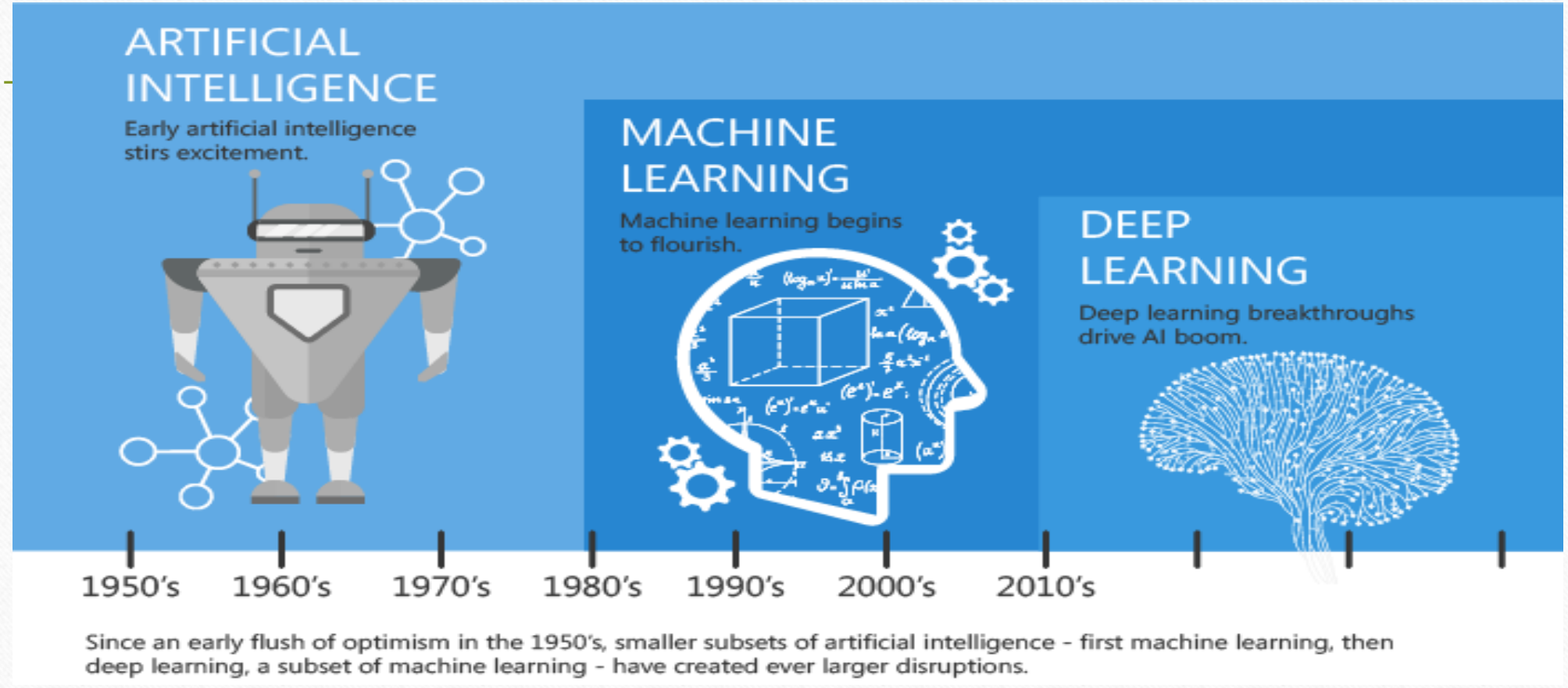
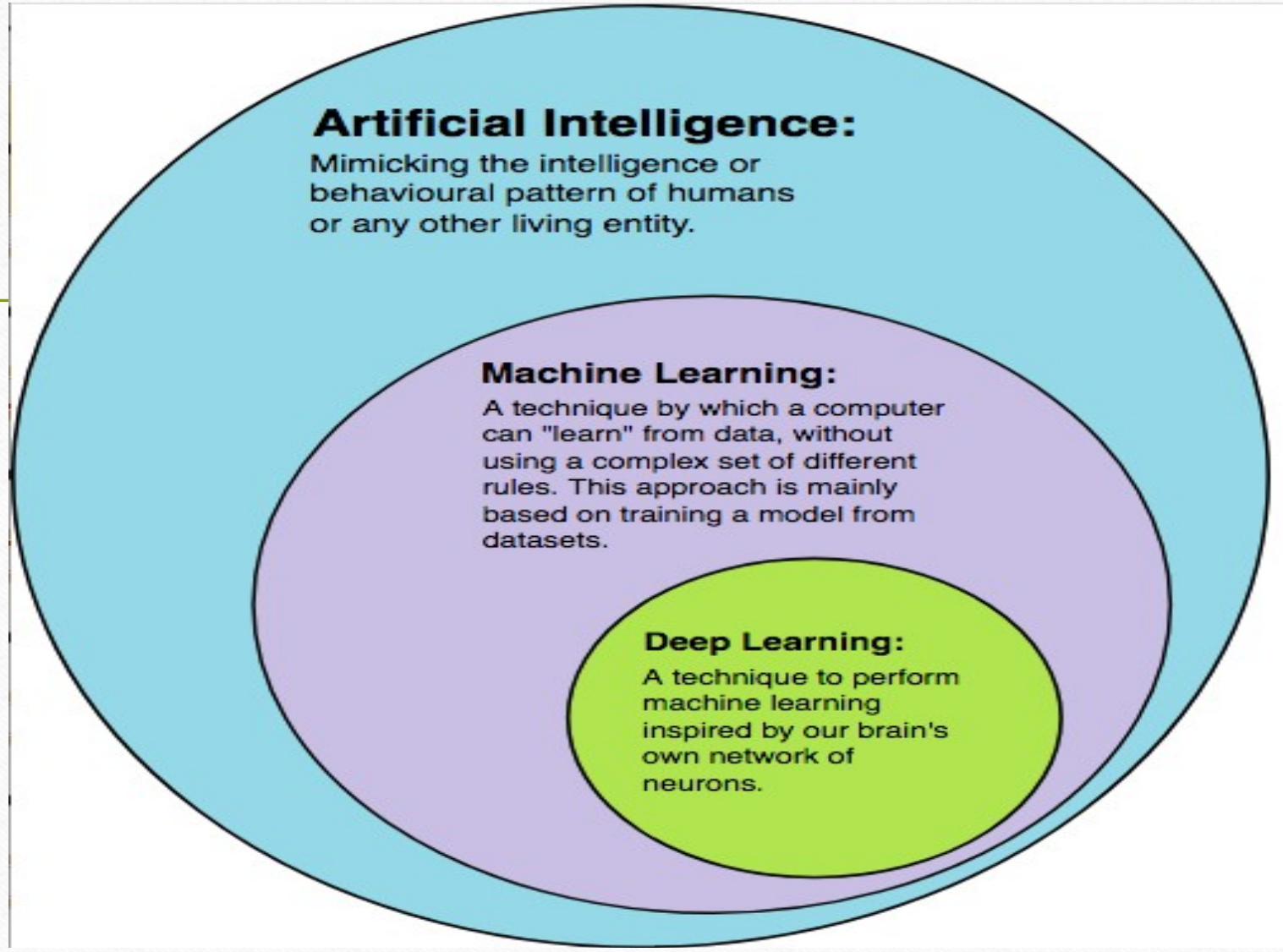
ECII/ECSI 3206:
Artificial Intelligence [and expert systems]
Topic 6: Machine Learning and ANN

By: Edgar Otieno

Background/Overview




Artificial Intelligence:
Mimicking the intelligence or behavioural pattern of humans or any other living entity.
Machine Learning:
A technique by which a computer can "learn" from data, without using a complex set of different rules. This approach is mainly based on training a model from datasets.
Deep Learning:
A technique to perform machine learning inspired by our brain's own network of neurons.

Introduction to Machine Learning

- The term “machine learning” was first introduced by Arthur Samuel. He defined it as the “*field of study that gives computers the ability to learn without being explicitly programmed*”.
- It uses **statistical methods** to give computer the ability to "learn" from data, without being explicitly programmed.
- The learning process improves the machine ***model*** over time by using training data.
- The evolved model is used to make future predictions.

Cont…

- If a computer program can improve how it performs certain tasks based on past experiences, then it has learned. This differs from performing the task always the same way because it has been programmed to do so .

Cont…

- ***Statistical Model***-A model in a computer is a mathematical function that represents a relationship or mapping between a set of inputs and a set of outputs. e.g

$$f(x)=x^2$$

$$\text{Violent crime incidents per day} = \text{Average Temperature} \times 2$$

- This is an oversimplified example to explain that machine learning refers to a set of techniques for estimating functions (for example, predicting crime incidents) that is based on data sets (pairs of the day's average temperature and the associated number of crime incidents). These models can be used for predictions of future data.
- New data "X" can predict the output "Y"

Machine Learning Algorithms

- **A machine learning algorithm** is a technique through which the system extracts useful patterns from historical data. These patterns can be applied to new data.
- The objective is to have the system learn a **specific input/output transformation.**
- The data quality is critical to the accuracy of the machine learning results.

Steps of ML Algorithms

- 1. Define objective of the problem[what needs to be done]
- 2. Gather data
- 3. Prepare / Clean the data[remove missing data, redundant data]
- 4.Explore the data[E.D.A]-[check for patterns in the data]
- 5.Build a Model [Split data into training set and testing set and choose a ML to use]
- 6. Evaluate model [test and check accuracy of model e.g. confusion matrix ,cross validation e.t.c]
- 7. Make predictions[use model to make predictions]
- 8. Deploy

Machine Learning Approaches

1. Supervised learning: A machine learning technique that Train by using labeled data, and learn and predict new labels for unseen input data.

- *Classification* is the task of predicting a discrete class label, such as “black, white, or gray” and “tumor or not tumor”.
- *Regression* is the task of predicting a continuous quantity, such as “weight”, “probability”, and “cost”.

Cont…

2. Unsupervised learning: A machine learning technique that Detect patterns and relationships between data without using labeled data.

- *Clustering algorithms:* Discover how to split the data set into a number of groups such that the data points in the same groups are more similar to each other compared to data points in other groups.

Cont…

3. Semi Supervised learning-A machine learning technique that falls between supervised and unsupervised learning. It includes some labeled data with a large amount of unlabeled data.

- In semi-supervised learning, you try to get the best out of your unlabeled data
- Many real practical problems fall into this category of machine learning where you have little labeled data and the rest of the data is unlabeled. E.g labelling different face types.

Cont…

4. Reinforcement learning-It is a goal-oriented learning that is based on interaction with the environment. As the system performs certain actions, it finds out more about the world .It uses trial and error (a rewarding/punishment approach).

- The algorithm discovers an association between the goal and the sequence of events that leads to a successful outcome.
- Example reinforcement learning applications:
  - Robotics: A robot that must find its way.
  - Self-driving cars.

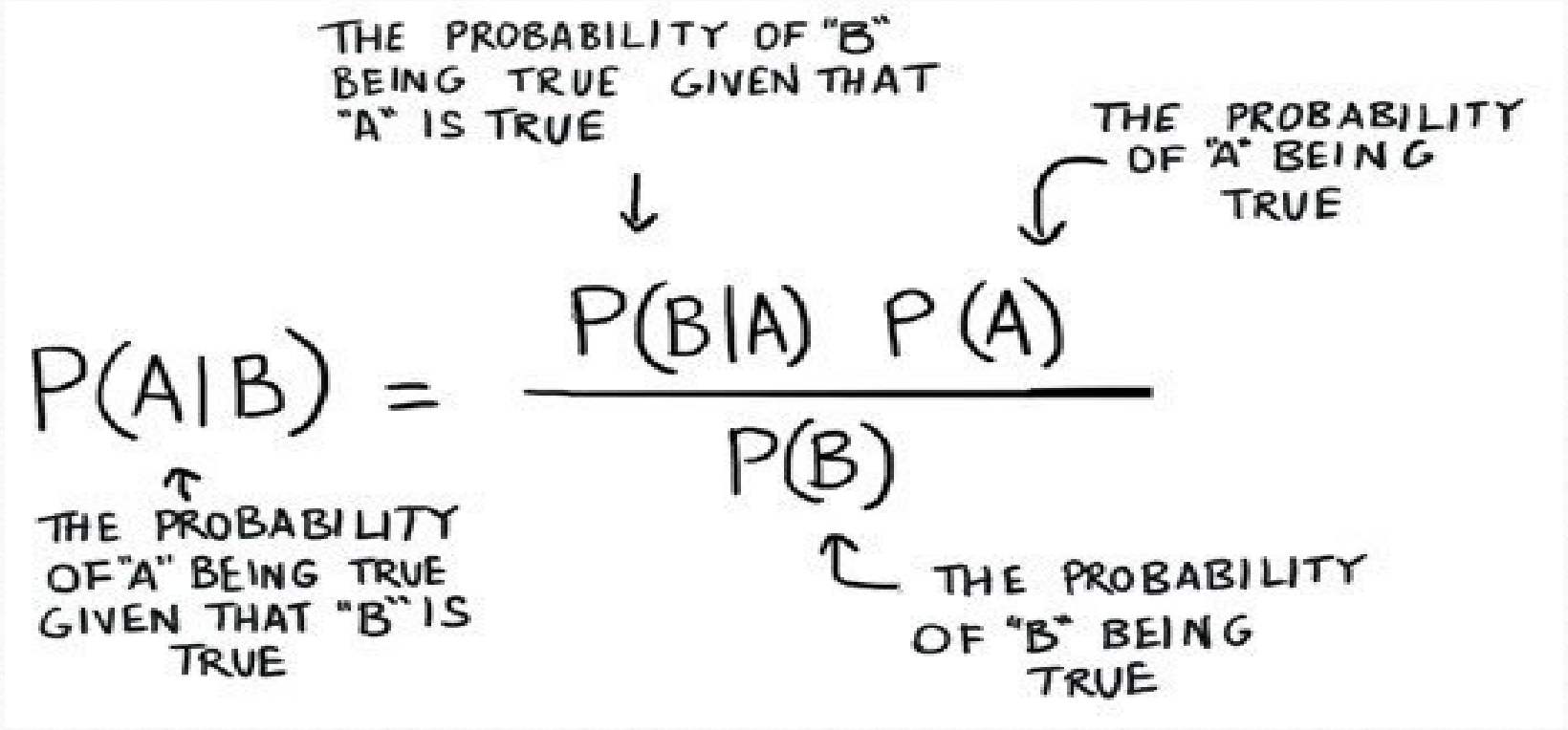
Cont…

- Understanding your problem and the different types of ML algorithms helps in selecting the best algorithm.
- Some Common machine learning algorithms include:
  - Naïve Bayes classification (supervised classification – probabilistic)
  - Linear regression (supervised regression)
  - Logistic regression (supervised classification)
  - Support vector machine (SVM) (supervised linear or non-linear classification)
  - Decision tree (supervised non-linear classification)
  - K-means clustering (unsupervised learning)

Naïve Bayes classification

- Naïve Bayes classifiers is a powerful and simple supervised machine learning algorithm. It assumes that the value of a particular feature is independent of the value of any other feature, given the class variable.
- For example, a fruit may be considered to be an apple if it is red, round, and about 10 cm in diameter.
  - Features: Color, roundness, and diameter.
  - A Naïve Bayes classifier considers each of these features to contribute independently to the probability that this fruit is an apple, regardless of any possible correlations between the color, roundness, and diameter features.

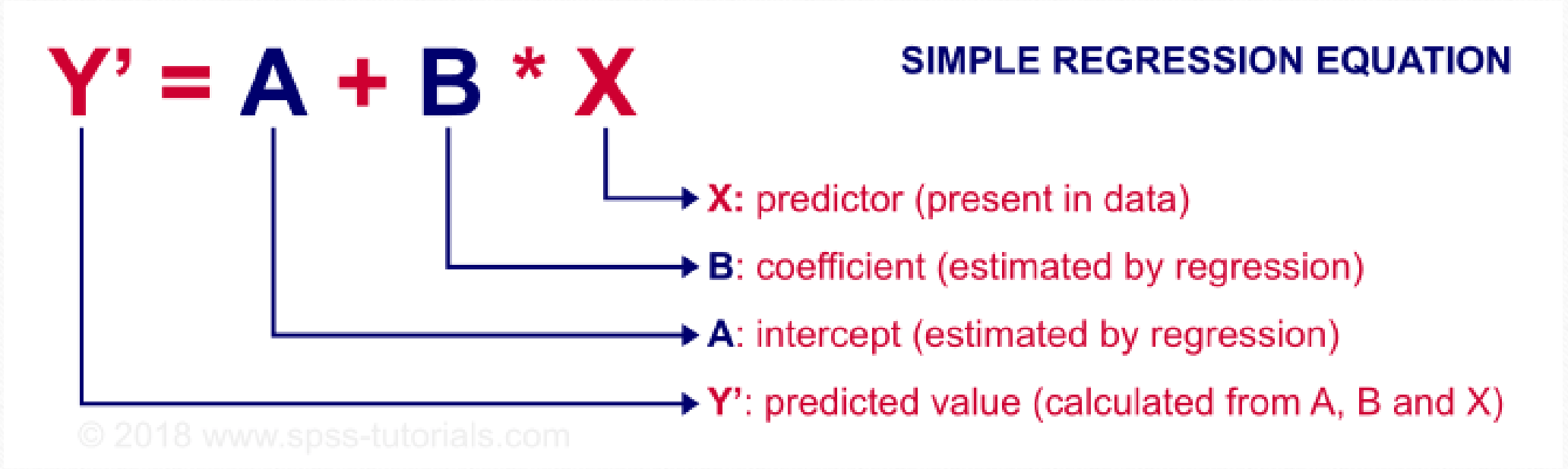
Equation for Naïve Bayes algorithm



Linear regression

- Regression algorithms help analysts to model relationships between input variables X and the output label Y for the training data points. This algorithm targets supervised regression problems, that is, the target variable is a continuous value.
- In **simple linear regression**, we establish a relationship between the target variable and input variables by fitting a line that is known as the *regression line.* E.g
  - Analyze the marketing effectiveness, pricing, and promotions on the sales of a product.
  - Forecast sales by analyzing the monthly company’s sales for the past few years.
  - Analyze the amount of hours spent in the gym and the Weight loss achieved

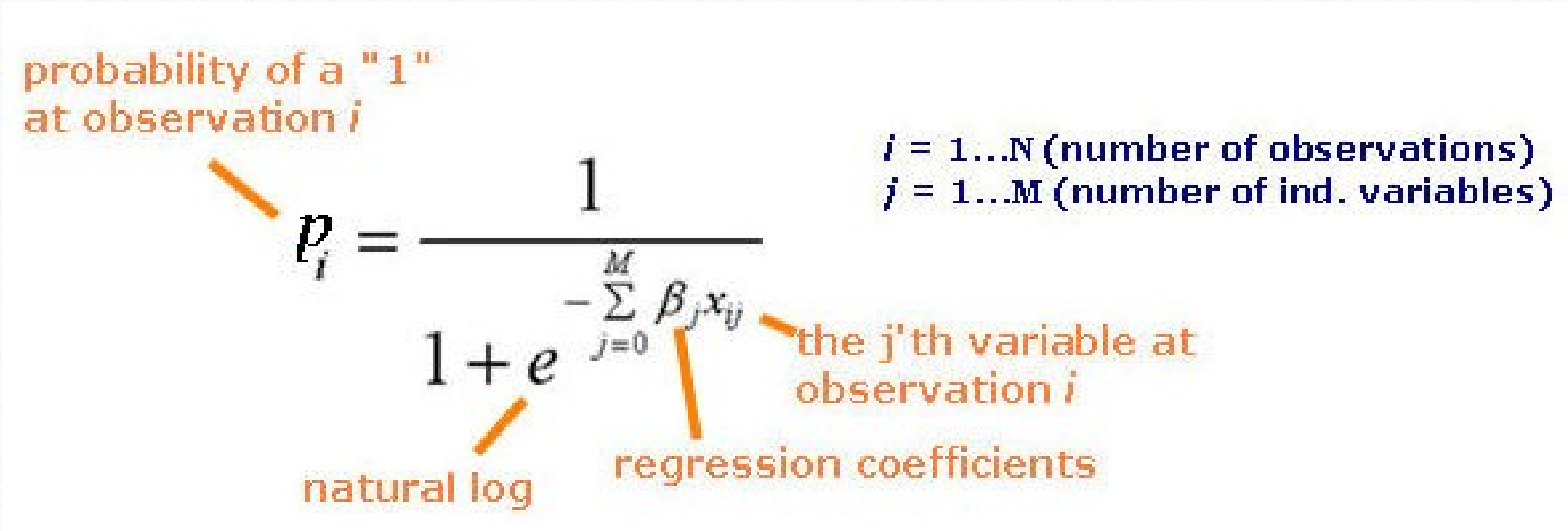
Linear regression equation



Logistic regression

- **Logistic regression** is a supervised classification algorithm. It is different from linear regression where the dependent or output variable is a *category or class*. The target is a discrete category or a class (not a continuous variable as in linear regression), for example, Class1 = cancer, Class2 = No Cancer.
- Logistic regression is named for the function that is used at the core of the algorithm, which is the logistic function *(sigmoid function).* It is an S-shaped curve for data segregation across multiple classes that can take any real value 0 - 1.

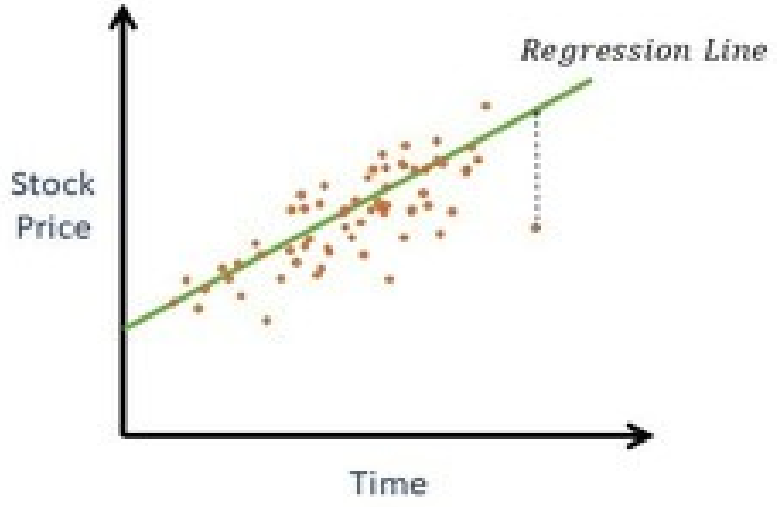
Logistic Regression Equation



Comparison between Linear and Logistic regression

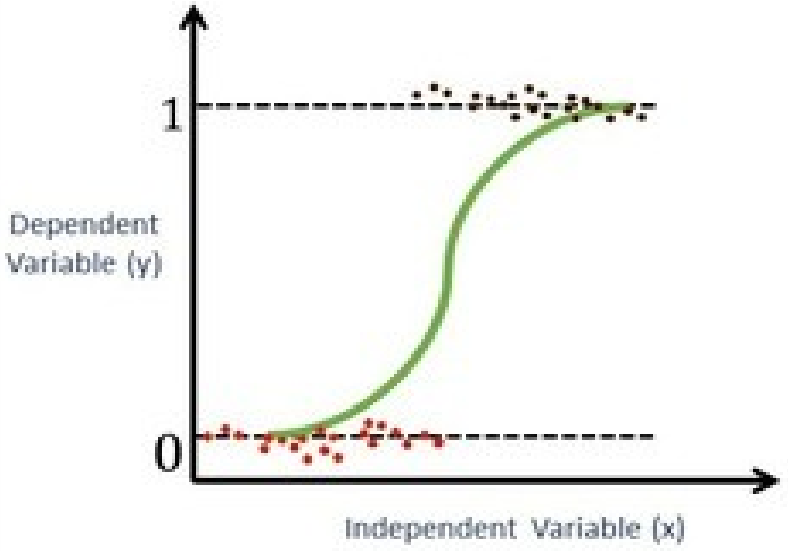
Linear Regression

- Aim is to predict continuous valued output.
- Output value can be any possible integer number.



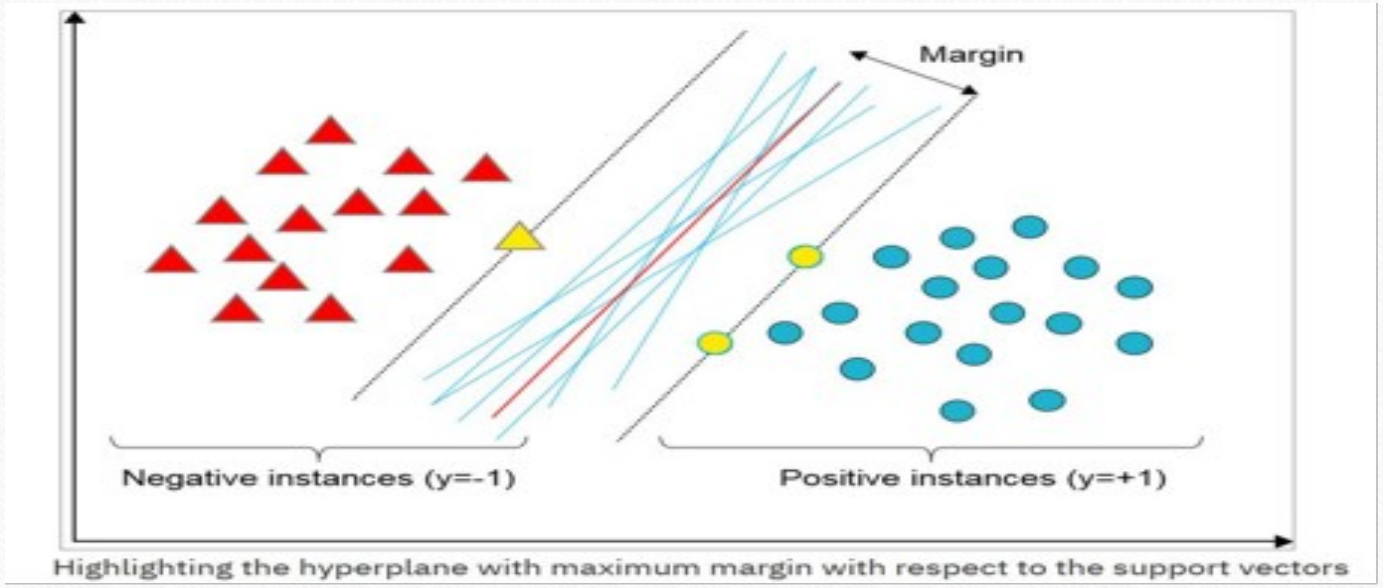
Logistic Regression

- Aim is to predict the label for input data.
- Output is categorical (Binary) i.e. 0/1, True/False, etc.



Support Vector Machine (SVM)

- SVM is a supervised learning model that can be a linear or non-linear classifier. SVM is also called a “large Margin Classifier” because the algorithm seeks the hyperplane with the largest margin, that is, the



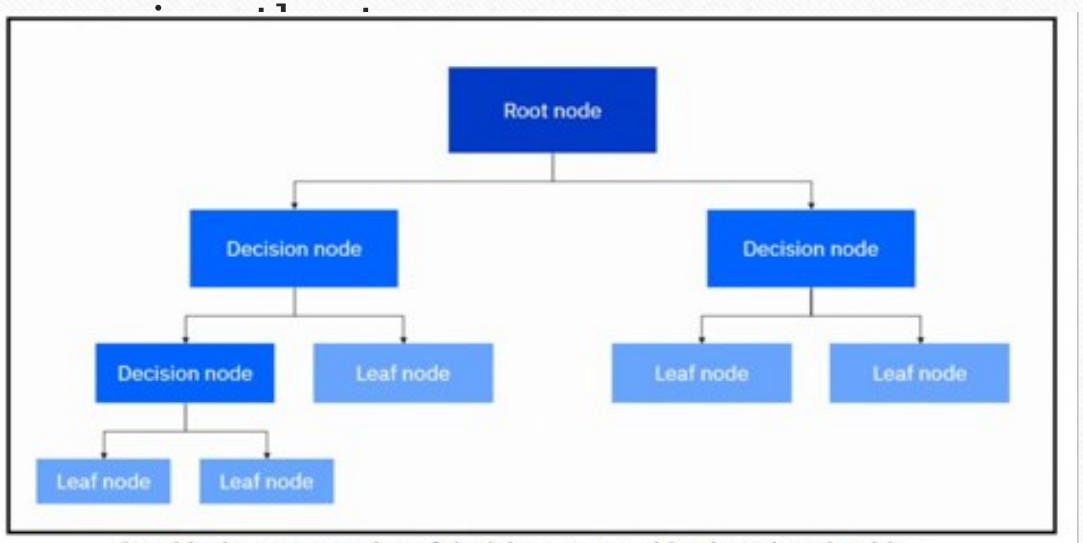
Highlighting the hyperplane with maximum margin with respect to the support vectors

Although many lines (in light blue) do separate all instances correctly, there is only one optimal hyperplane (red line) that maximizes the distance to the closest points (in yellow).

Decision Tree

- **A decision tree** is a popular supervised learning algorithm that can be used for classification and regression problems. Decision trees are a popular prediction method. Decision trees can explain why a specific prediction was made by tr

A decision tree includes three main entities: ***root node***, ***decision nodes***, and ***leaves.*** The figure shows the graphical representation of these entities.
A decision tree builds the classification or regression model in the form of a tree structure.



Graphical representation of decision tree machine learning algorithm

Decision tree cont..

- The algorithm works by recursively splitting the data based on the value of a feature. After each split, the portion of the data becomes more homogeneous.
- Now, the algorithm needs to decide:
  - Which feature to choose as the root node.
  - What conditions to use for splitting.
  - When to stop splitting

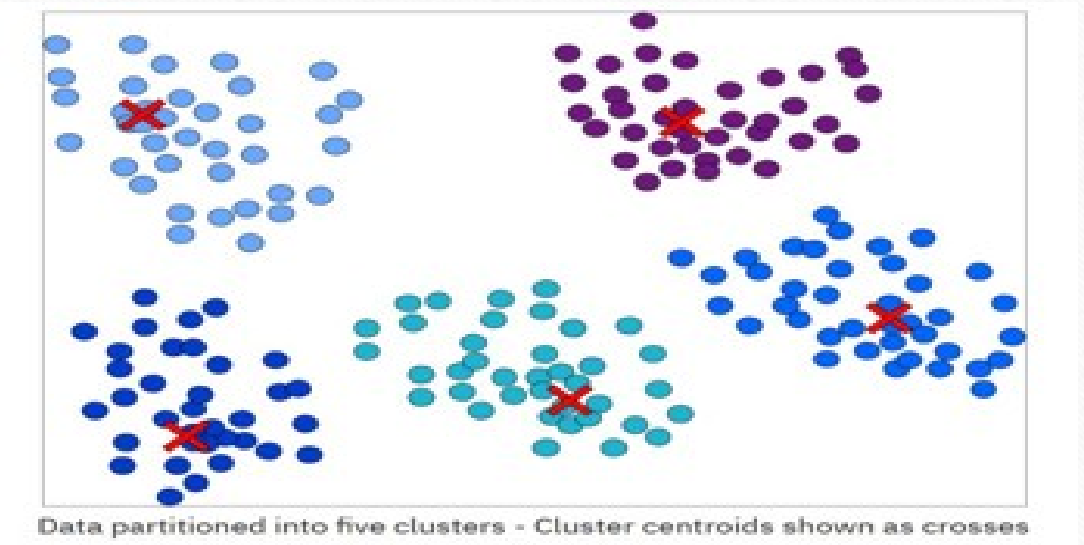
Decision tree cont..

- The Iterative Dichotomiser3 (ID3) is a decision tree algorithm that works by using *entropy* and *information gain* to construct a decision tree.
- *Entropy* is the measure of the amount of uncertainty and randomness in a set of data for the classification task. Entropy is maximized when all points have equal probabilities. Entropy zero means that there is no randomness for this attribute.
- *Information gain* is a metric that is used for ranking the attributes or features to split at given node in the tree. It defines how much information a feature provides about a class. The feature with the highest information gain is used for the first split.

K-means Clustering

- K-means clustering is an unsupervised machine learning technique. It groups a set of objects in such a way that objects in the same group (called a cluster) are more similar to each other than those in other groups (other clusters).

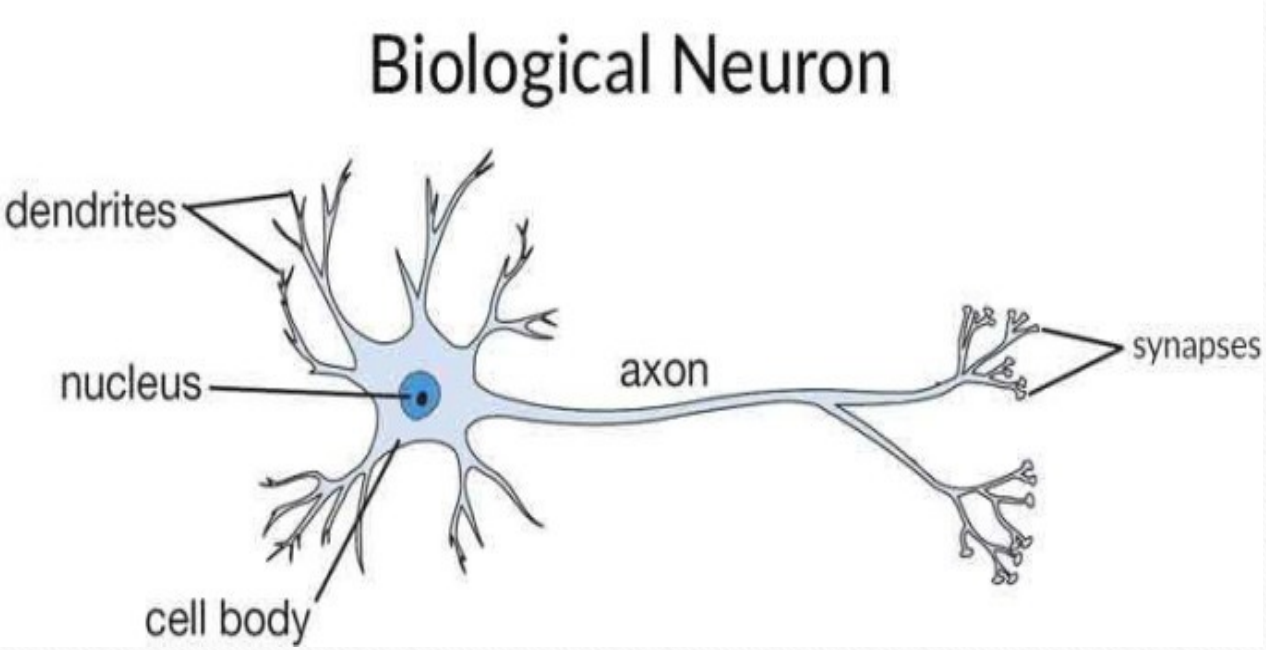
The main goal of the algorithm is to group the data observations into k clusters, where each observation belongs to the cluster with the nearest mean



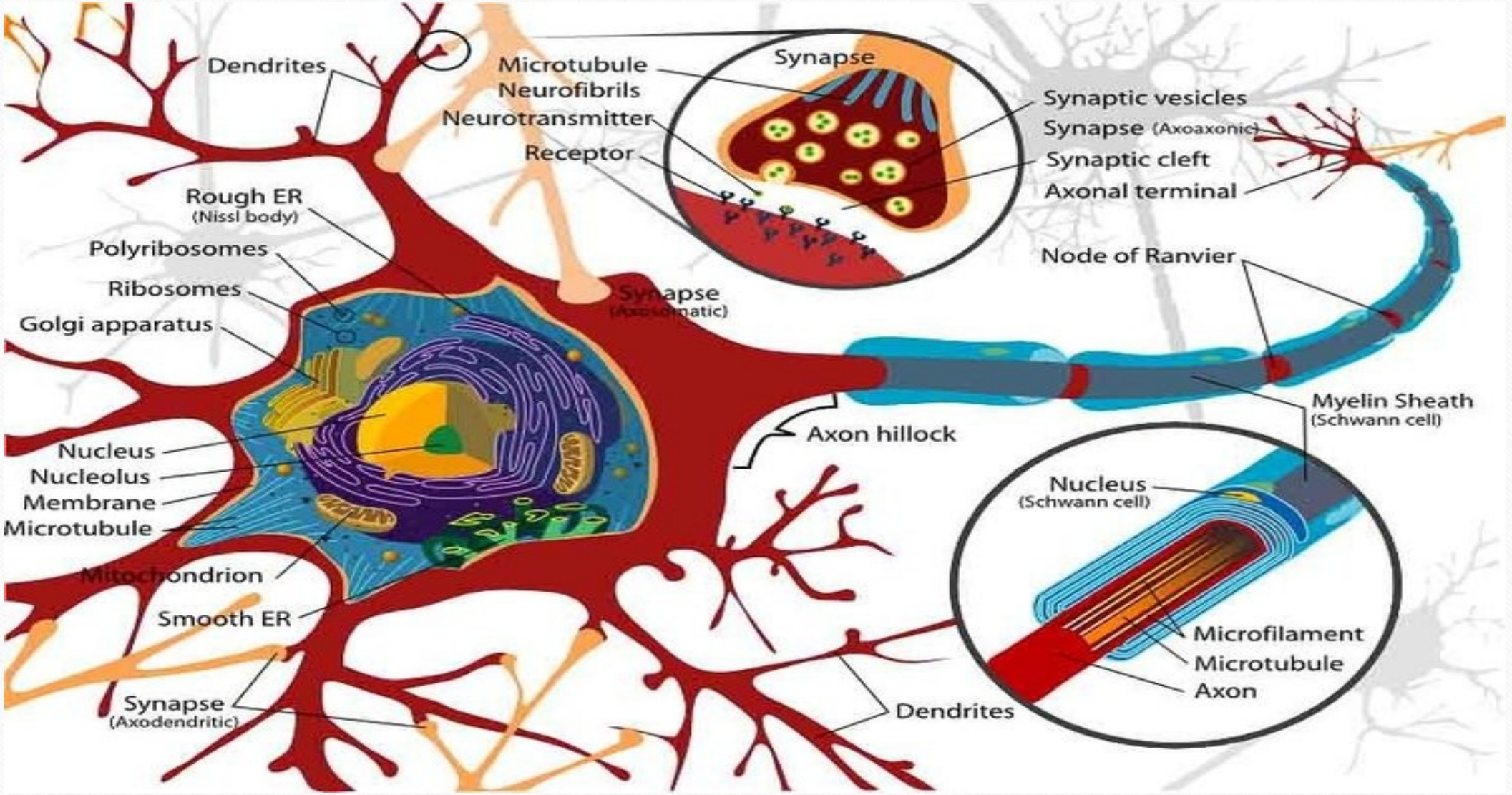
Artificial Neural Networks

- Human brain is comprised of cells called ***Neurons***.
- Neurons are the building blocks of Neural Networks
- Neurons are highly connected and communicate chemical signals through the **synapses** (a junction between two nerve cells) between the **axons**(Connect neurons to other cells) and **dendrites**(Accept input/stimuli from the external environment). These inputs create electronic impulses(spikes) which are passed quickly across the neural network. The effect of this is an action/response

Structure of Biological Neuron



The human brain is estimated to have 100 billion neurons, with each neuron connected to up to 10,000 other neurons.


Dendrites
Microtubule
Neurofibrils
Neurotransmitter
Receptor
Synapse
Synaptic vesicles
Synapse (Axoaxonic)
Synaptic cleft
Axonal terminal
Rough ER
(Nissl body)
Polyribosomes
Ribosomes
Golgi apparatus
Synapse
(Axosomatic)
Node of Ranvier
Myelin Sheath
(Schwann cell)
Axon hillock
Nucleus
Nucleolus
Membrane
Microtubule
Mitochondrion
Smooth ER
Nucleus
(Schwann cell)
Microfilament
Microtubule
Axon
Synapse
(Axodendritic)
Dendrites

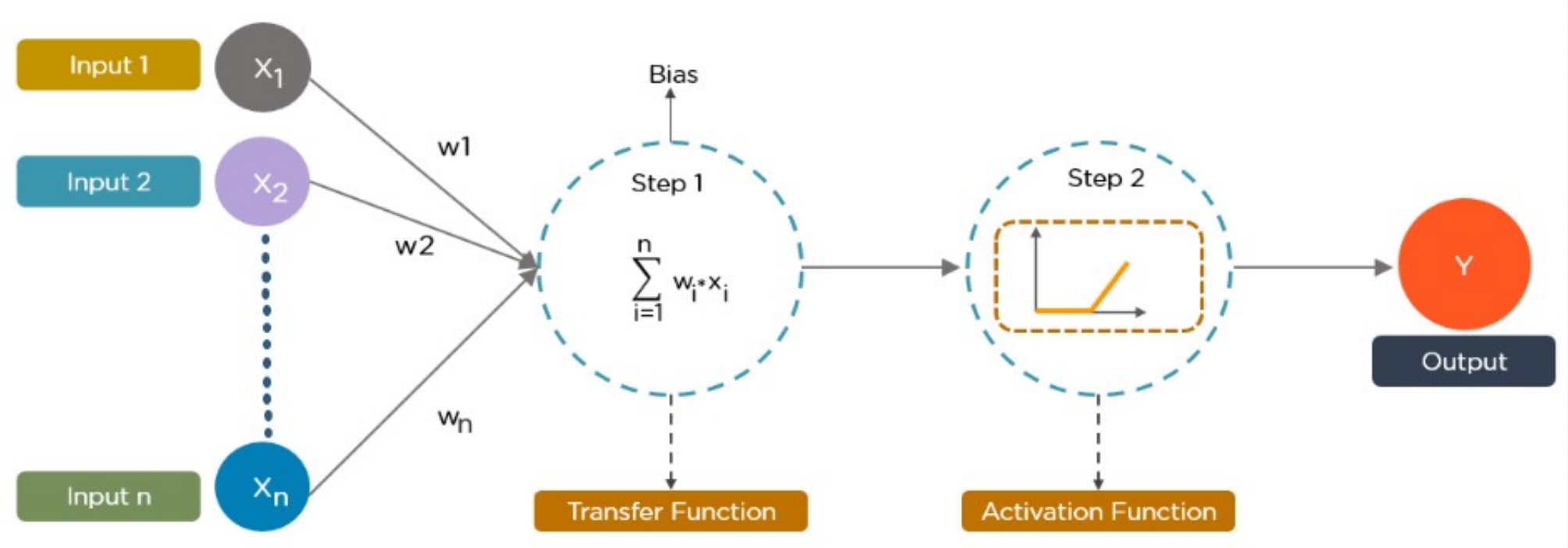
Artificial Neural Networks

- **ANN** are a hardware and software implementation which tries to simulate information processing capabilities of biological NN.
- They are collections of interconnected ***"neurons"*** (called nodes) that work together to transform input data to output data.
- Each node applies a mathematical transformation to the data it receives; it then passes its result to the other nodes in its path.
- Examples for applications:
  - Object detection, tracking, and image analysis (Convolutional Neural Network (CNN))
  - Natural language processing /machine Translation (recurrent neural network (RNN))
  - Autonomous cars and robots (more complex neural networks)

Cont....

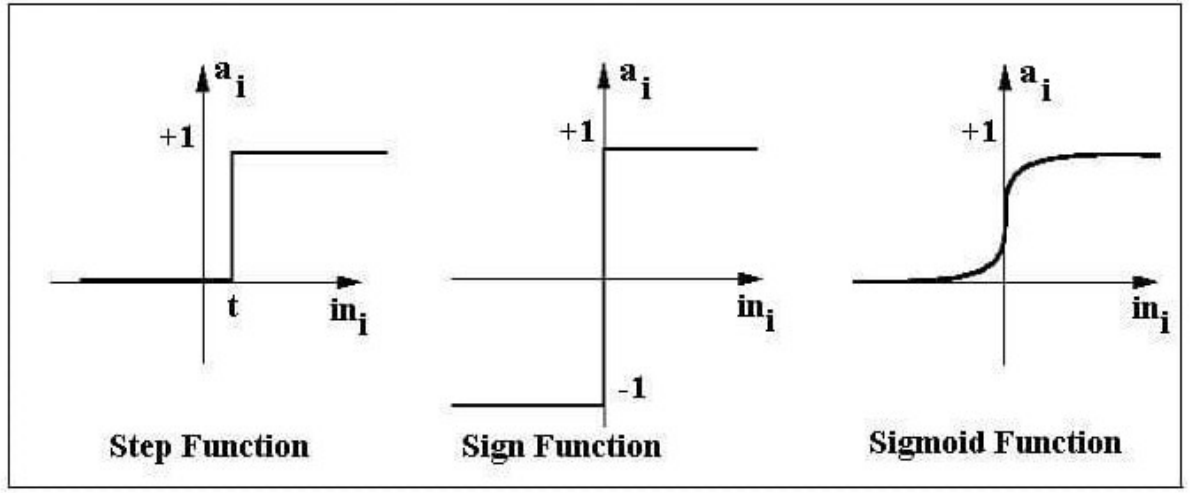
- ANN communicate signals (input numbers) through weights and activation functions that activate neurons. Using a training algorithm, these networks adjust those weights to solve a problem.
- Each node applies a mathematical transformation to the data it receives; it then passes its result to the other nodes in its path. Each connection between nodes represents a different parameter to the model.
- A perceptron is a single neuron model that was an originator for neural networks.

Structure of Neuron/Perceptron[mcculloch pitt model]



Cont…

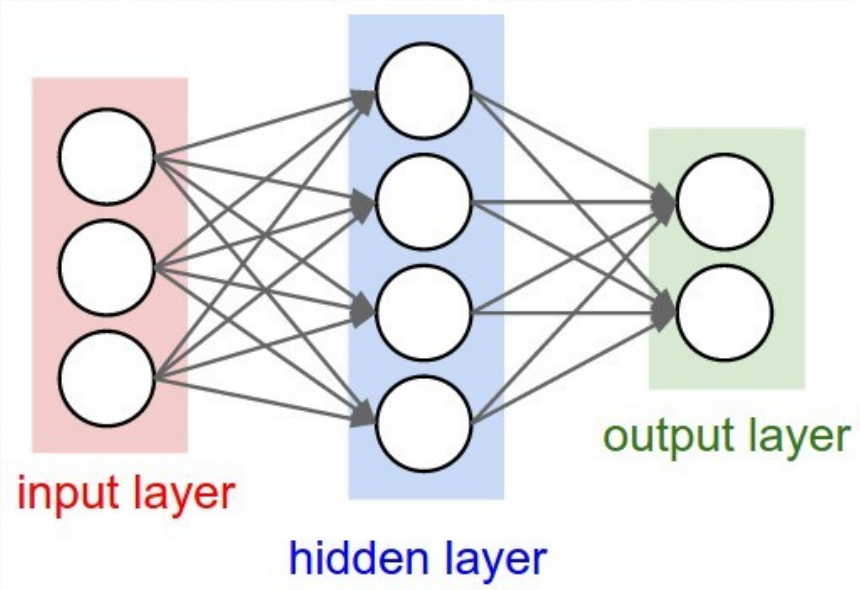
- **Activation function/Transfer function-** Activation function decides, whether a neuron should be activated or not by calculating weighted sum and further adding bias with it. In summary, it is used to determine the oid or linear.

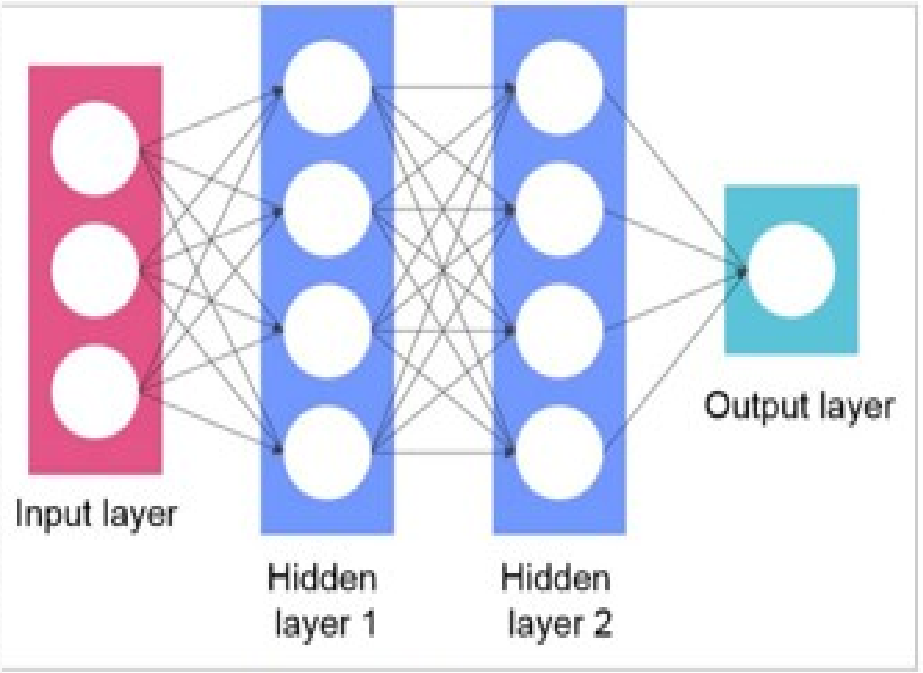


Cont…

- A Perceptron can be viewed as an algorithm for supervised learning of binary classifiers. This algorithm enables neurons to learn and processes elements in the training set one at a time.
- There are two types of Perceptrons: Single layer and Multilayer.
  - *Single layer* Perceptrons can learn only linearly separable patterns.
  - *Multilayer* Perceptrons or feedforward neural networks with two or more layers have the greater processing power.
- **Perceptron Learning Rule** states that the algorithm would automatically learn the optimal weight coefficients. The input features are then multiplied with these weights to determine if a neuron fires or not.

Single Layer Vs Multilayer NN





Comparison of a Biological Network and Artificial Neuron

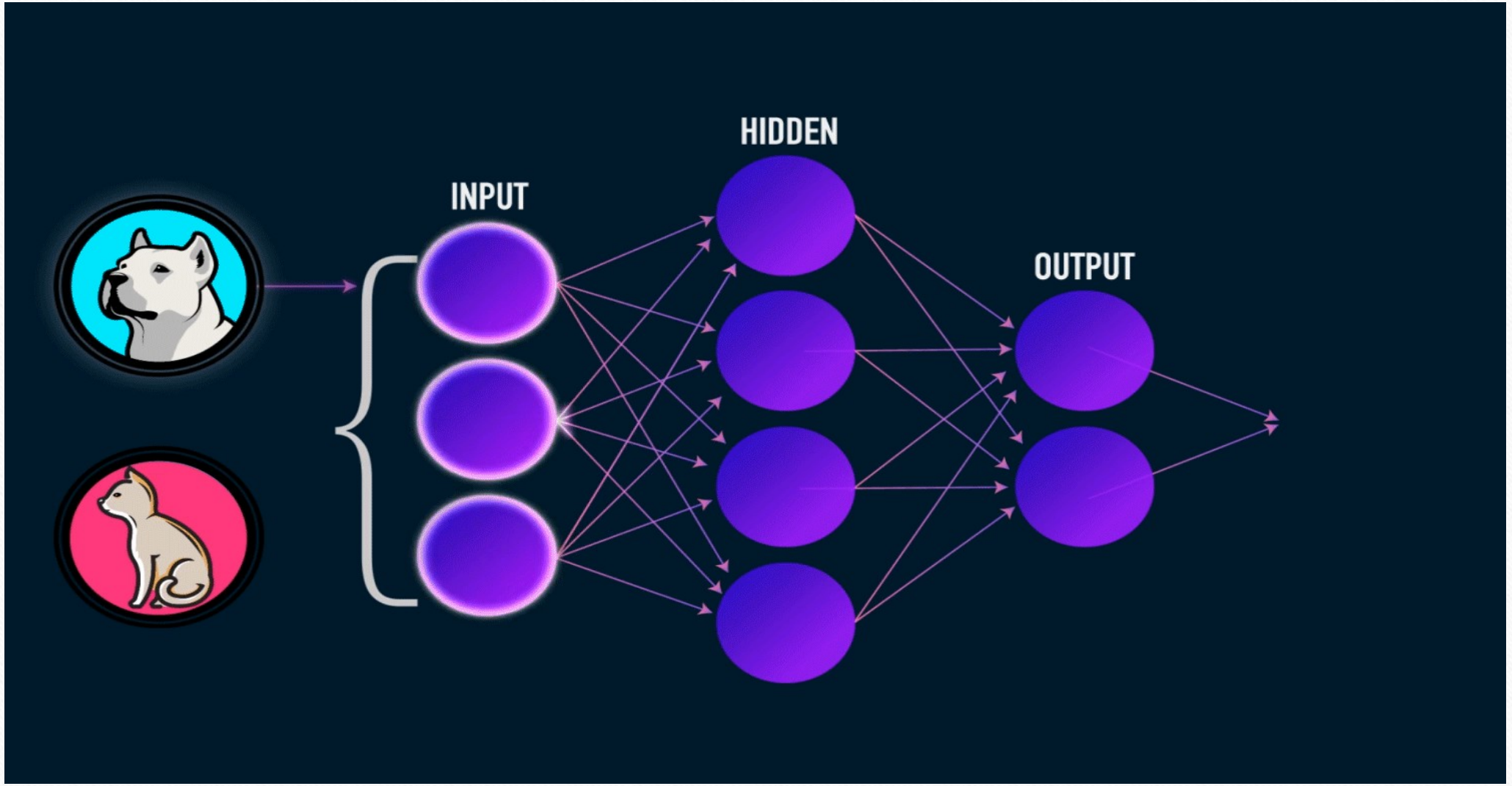
Biological Neuron vs. Artificial Neuron

The biological neuron is analogous to artificial neurons in the following terms:

Biological Neuron	Artificial Neuron
Cell Nucleus (Soma)	Node
Dendrites	Input
Synapse	Weights or interconnections
Axon	Output

General structure of a ANN

- A neural network is composed of three or more layers: **an input layer**, one or many **hidden layers**, and an **output layer**.
- Data is imported through the input layer. Then, the data is modified in the hidden and output layers based on the weights that are applied to their nodes. The typical neural network can consist of thousands or even millions of simple processing nodes that are densely interconnected.
- Each link is associated with a weight .ANN are capable of learning what takes place by altering these weights.

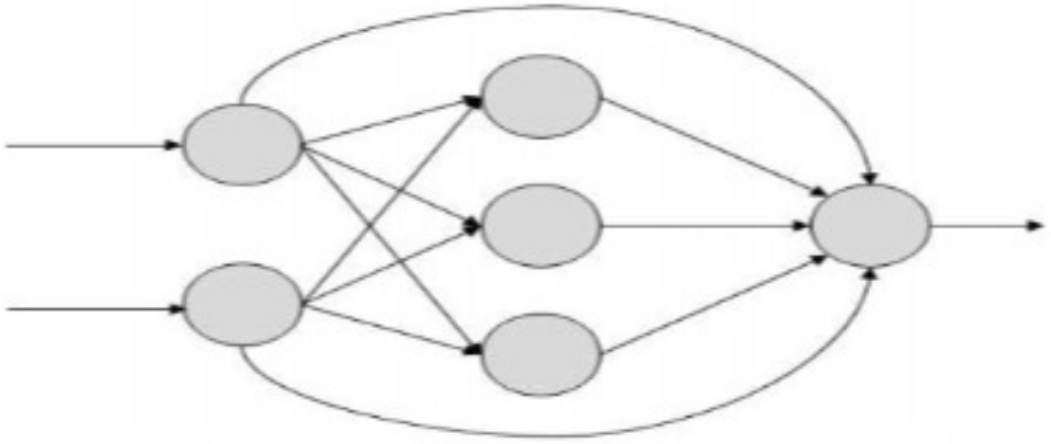

HIDDEN
INPUT
OUTPUT

Types of ANN

- 2 main types exist as follows:
  - a) Feed forward ANN[multilayer perceptron]
  - b) Feed Back ANN[recurrent ANNs]

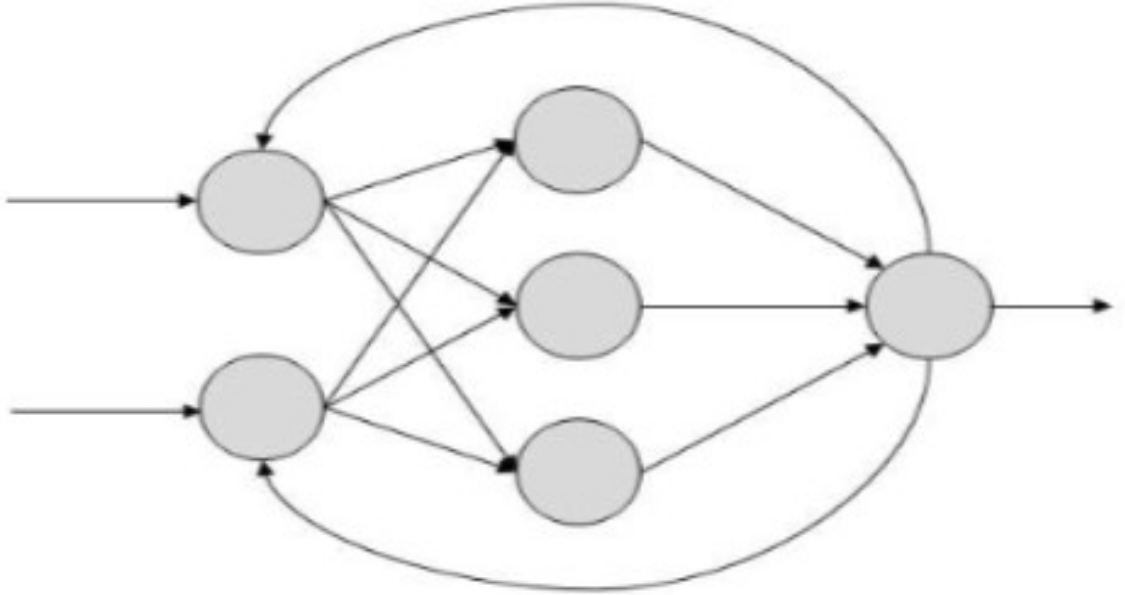
Feed forward ANNs

- Information flow is unidirectional
- A unit send information to the next unit from which it does not receive any informat
- It has no feedback loops
- It has fixed input and ou
- Used in pattern recognit
- and classification



Feed back ANNs

- A unit sends information to the next unit from which it does can also receive
- Have feedback loops



Back propagation[Technique for training ANNs]

- **Back propagation** is an algorithm for training neural networks that have many layers. If you give an example you want the network to solve, it changes the network weights until it reaches desired output . It works in two phases:
  - **First phase:** The propagation of inputs through a neural network to the final layer (called feed-forward).
  - **Second phase:** The algorithm computes an error. An error value is then calculated by using the wanted output and the actual output for each output neuron in the network. The error value is propagated backward through the weights of the network (adjusting the weights) beginning with the output neurons through the hidden layer and to the input layer (as a function of the contribution of the error).

Learning rules used for ANNs

- Error correction/ trial and error rules
- Boltzmann rule
- Hebbian rule
- Competitive learning rule

Applications of ANN

- Automotive guidance[ALVINN]
- Speech recognision
- Time series prediction
- Transportation[vehicle scheduling]
- Medicine[cancer diagnosis]
- Image recognition[CNN as used for processing pixel data]

Advantages of ANNs

- Can adopt to unknown conditions/ Uncertainty
- Can be use to model complex functions[e.g human mind functioning]

Disadvantages of ANNs

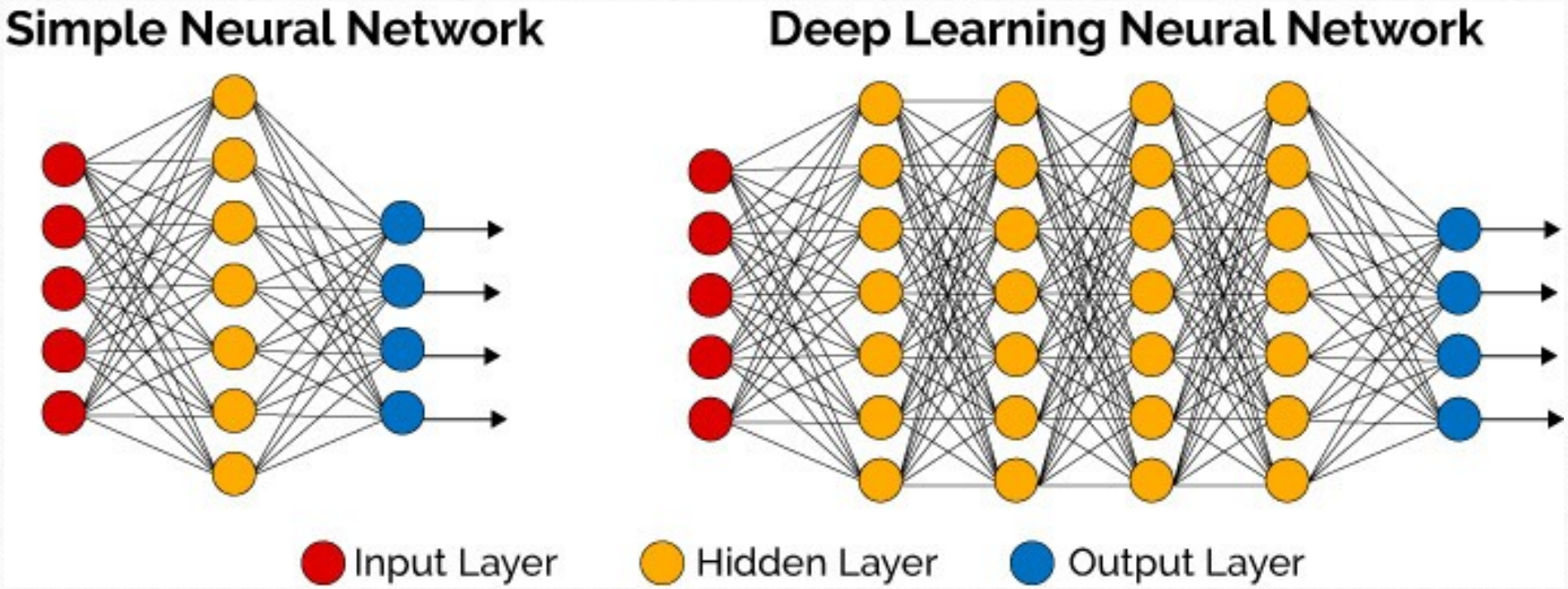
- They become very complex for large problems
- They are not good for preciseness/not exact

Deep Learning

- Deep learning is a machine learning technique that uses neural networks to learn.
- It is Similar to a traditional neural network, but it has many **more hidden layers.**
- The **more** complex the problem, the **more** hidden layers there are in the model.

[Facebook is working on ‘deep learning’ neural networks to learn even more about your personal life. nobody has access to more raw information than Facebook.]

Cont..


Simple Neural Network
Deep Learning Neural Network
Input Layer
Hidden Layer
Output Layer

Cont....

- Deep learning has emerged now because of the following reasons:
  - The continuous increase in big data requires data processing scaling to analyze and use this data correctly.
  - Improvement in processing power and the usage of GPUs to train neural networks.
  - Advancement in algorithms like the rectified linear unit (ReLU) instead of the Sigmoid algorithm helps make gradient descent converge faster.

Applications of Deep learning

- **Speech recognition.** All major commercial speech recognition systems (like Microsoft Cortana, Alexa, Google Assistant, Apple Siri) are based on deep learning.
- **Pattern recognition.** Pattern recognition systems are already able to give more accurate results than the human eye in medical diagnosis.
- **Natural language processing.** Neural networks have been used to implement language models since the early 2000s. The invention of LSTM helped improve machine translation and language modeling.
- **Discovery of new drugs.** For example, the AtomNet neural network has been used to predict new biomolecules that can potentially cure diseases such as Ebola and multiple sclerosis.
- **Recommender systems.** Today, deep learning is being used to study user preferences across many domains. Netflix is one of the brightest examples in this field.

Applications of Deep learning

- **Multilayer perceptron (MLP):** Classification and regression, for example, a house price prediction.
- **Convolutional neural network (CNN):** For image processing like facial recognition.
- **Recurrent neural network (RNN):** For one-dimensional sequence input data. Like audio and languages.
- **Hybrid neural network:** Covering more complex neural networks, for example, autonomous cars.

How To Evaluate a Machine learning Model

- After you have successfully trained your model, you need a methodology to follow to evaluate your machine learning model performance .
- The objective of a good machine learning model is to generalize for any future data points.
- A classic mistake is to use the same sample data that is used in training to test a model, which produces a false perfect score. This is called **“overfitting”** (also referred as “high variance”). The problem with overfitting is that your model fails at predicting future unseen data.

How To Evaluate a Machine learning Model

- Another case that can cause overfitting is where you have unbalanced data. For example, assume that you are working on a data set for churn analysis. The customers who churned are actually 2% of your data set. Using this data set “as is” causes overfitting
- Overfitting can also occur if you are using too many features. Relatively, if the number of features is the same as or greater than the number of training samples, that can cause overfitting.

Cont..

- Solutions to overfitting problem
  - One of the solutions to overcome overfitting is to **increase the number of data set samples** that is used for training compared to features.
  - Another solution is to **manually decrease the number of features**, but that might result in removing useful information.
  - Another solution is to perform model selection by using **cross-validation**.

Cont....

- **Underfitting** (also referred to as “high bias”) occurs when a machine learning model cannot fit the training data or generalize to new data.
- A possible reason might be that the model is using a simple estimator. For example, you might be using a linear estimator.
- Another reason might be that you are not using enough features, so your estimator fails to capture the structure of the data

- Solutions to underfitting problem
- Using the right quadratic or higher degree polynomial estimator to develop your model like in “Right fit” graph instead of a simple linear estimator equation
- A possible solution would be to add more features and try a different estimator.

Cont....

- **Cross-validation (CV)** is a process to evaluate a model by dividing the data set once or several times in training and testing.
- **Hold-out method:** Randomly splits the data set into a training set and test set.
- **K-fold cross validation:** Splits data into K subsamples where each subsample gets a chance to be the validation set, and K-1 is the training set.
- **Leave one out cross validation (LOO-CV):** Similar to K-fold except that one subsample that contains one data point is held out, and the rest of data is used for training