Artificial Neural Networks

by

Dr. Vera DAMERJIAN PIETERS

DSTI
Data ScienceTech Institute

Plan :

- Introduction to Artificial Intelligence
- What are ANNs ?
- The Artificial Neuron
- ANN Layers
- The Perceptron
- Weights and Biases
- Activation Functions
- Forward Propagation

Plan :

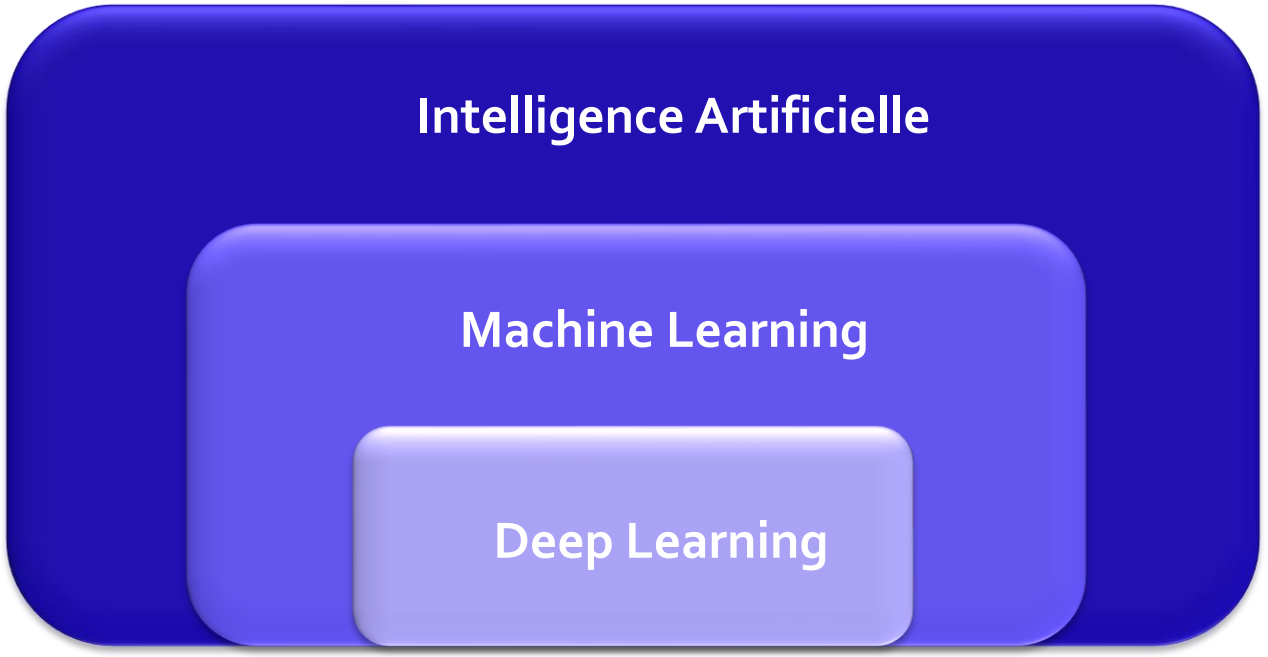
- Classification and Regression
- Cost Functions
- Optimizers
- Backpropagation
- The Learning Mechanism

Plan :

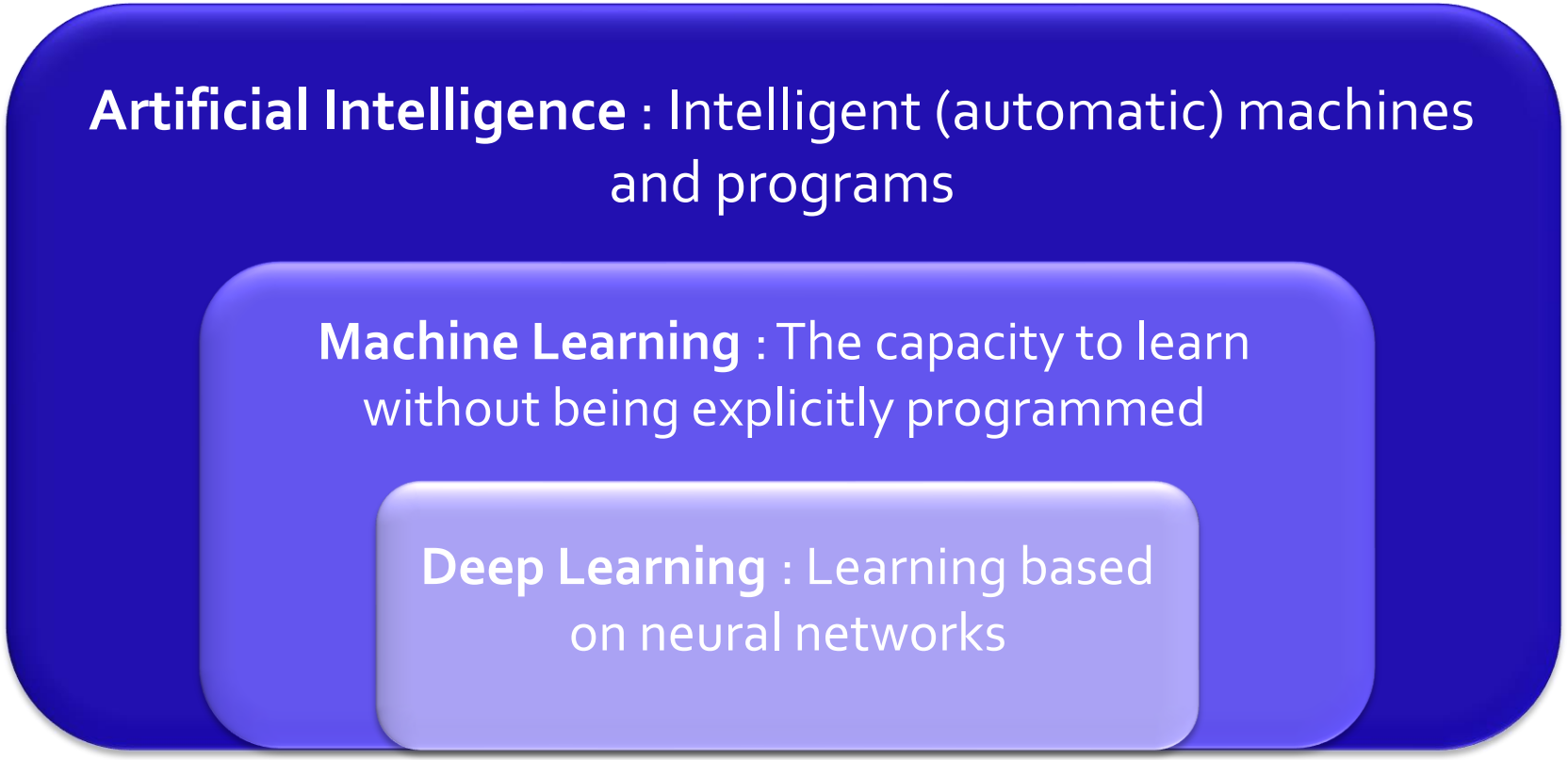
- Underfitting
- Overfitting
- Learning, Validation and Testing
- Regularisation
- Practice with Keras and TensorFlow

Introduction to Artificial Intelligence

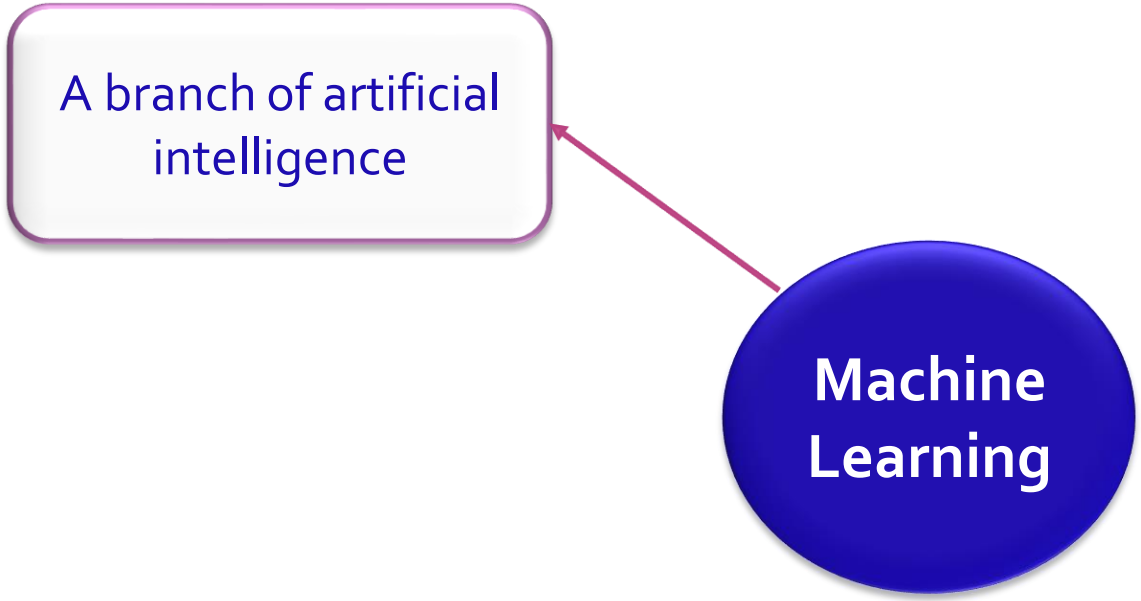
Introduction to Artificial Intelligence



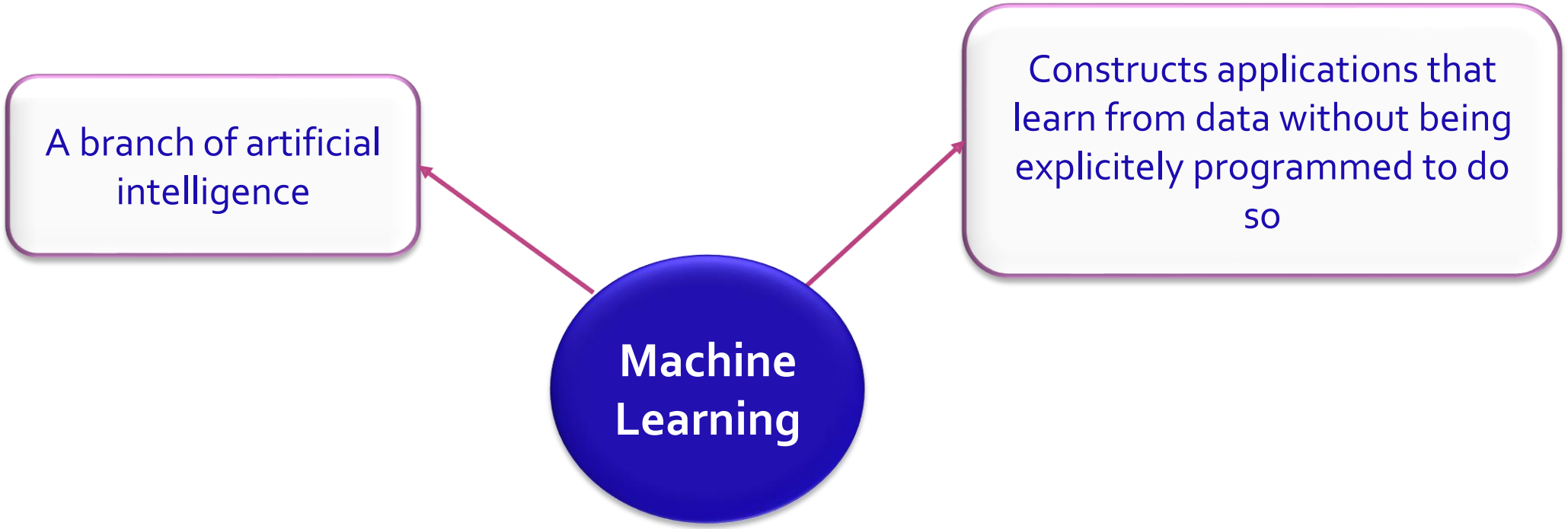
Introduction to Artificial Intelligence



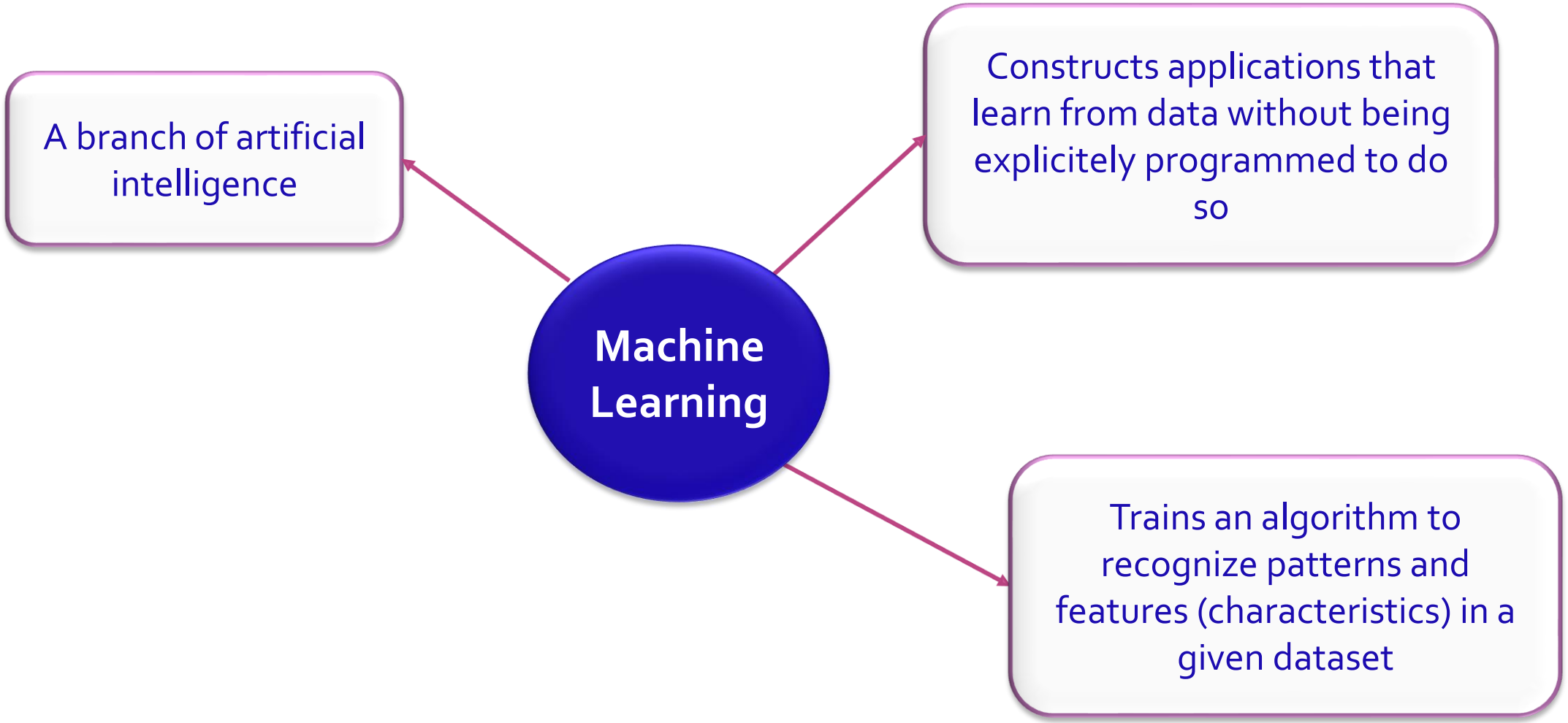
Introduction to Artificial Intelligence



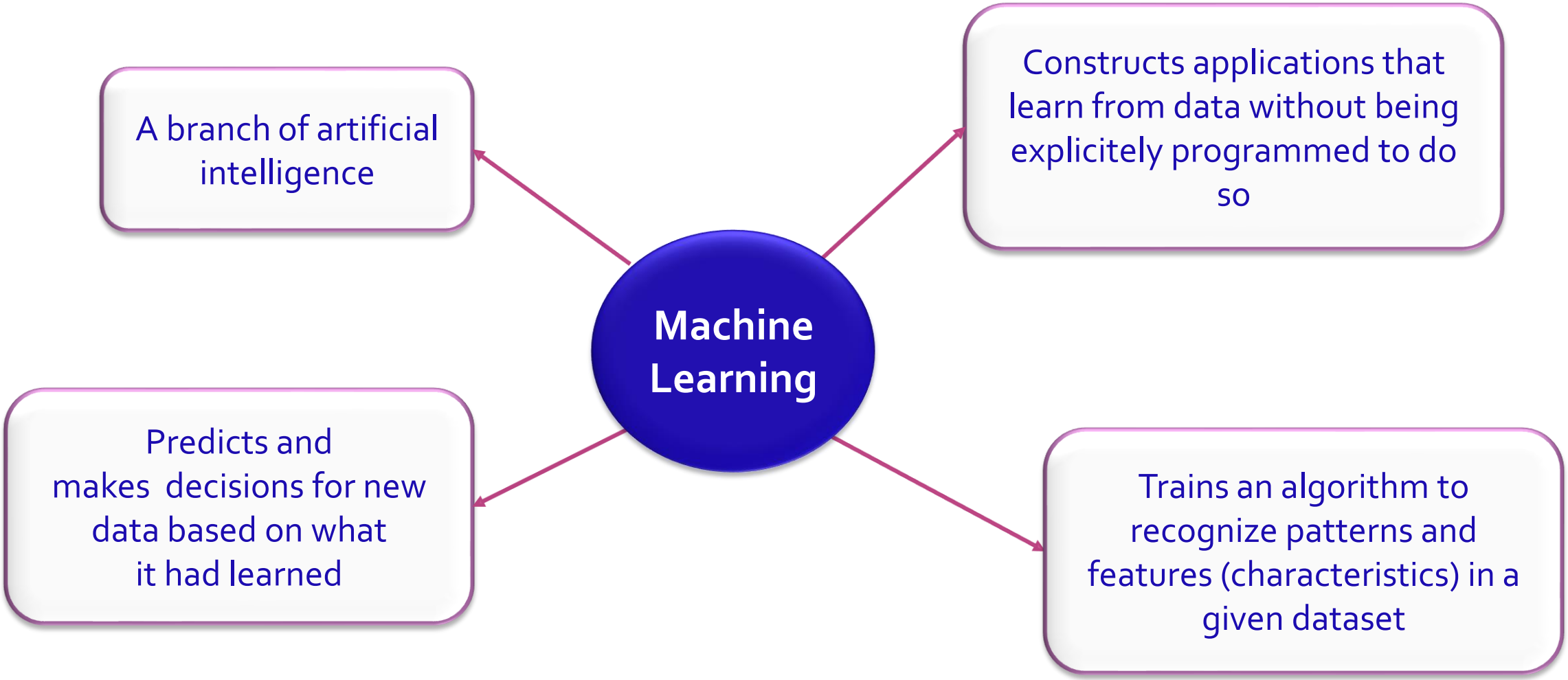
Introduction to Artificial Intelligence



Introduction to Artificial Intelligence



Introduction to Artificial Intelligence



Introduction to Artificial Intelligence

Machine Learning applications :

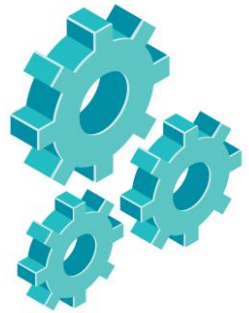
- Vocal assistance (Siri, Alexa, etc.)
- Recommendation systems (advertisements, movies, music, …)
- Medical image analysis and disease detection (tumors)
- Spam detection
- ⋮

Introduction to Artificial Intelligence

How does Machine Learning work ?

Introduction to Artificial Intelligence

How does Machine Learning work ?



1. Database preparation for training/learning

Introduction to Artificial Intelligence

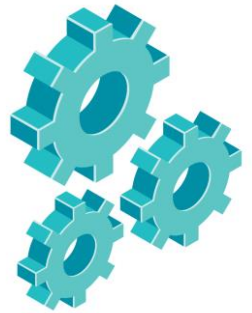
How does Machine Learning work ?

1. Database preparation for training/learning

2. Machine Learning algorithm selection

Introduction to Artificial Intelligence

How does Machine Learning work ?



1. Database preparation for training/learning

2. Machine Learning algorithm selection

3. Algorithm training (learning) to obtain the model

Introduction to Artificial Intelligence

How does Machine Learning work ?

1. Database preparation for training/learning

2. Machine Learning algorithm selection

3. Algorithm training (learning) to obtain the model

4. Using the trained model

Introduction to Artificial Intelligence

How does Machine Learning work ?

1. Database preparation for training/learning

Training database :

Data that we use to teach our model

in order to solve a certain problem

Labeled :

The classes or categories of our data

are specified by humans

Introduction to Artificial Intelligence

How does Machine Learning work ?

1

Database preparation for training/learning

Training database :

Data that we use to teach our model

in order to solve a certain problem

Labeled :

The classes or categories of our data

are specified by humans

Non-Labeled :

The model extracts the features and

defines the categories by itself

Introduction to Artificial Intelligence

How does Machine Learning work ?

1. Database preparation for training/learning

➡ Labeled data - Example 1 :

Input	Labels (classes)
	dog
	dog
	cat
	cat

Classes to learn are defined by humans

Introduction to Artificial Intelligence

How does Machine Learning work ?

1. Database preparation for training/learning

➡ Labeled data - Example 2 :

Age	Smoker	Nodule size (mm)	Location	Label
70	1	5	Top left	malignant
40	0	8	Top left	benign
45	1	2	Bottom middle	benign
65	1	4	Bottom right	benign

Classes to learn are defined by humans

Introduction to Artificial Intelligence

How does Machine Learning work ?

1 Database preparation for training/learning

➡ Labeled data - Example 3 :

Area	Number of rooms	Distance from train station	city	Price (label)
32	3	5	Amiens	60 000
25	1	8	Lille	70 000
80	4	10	Versailles	600 000
55	3	10	Nice	450 000

Classes to learn are defined by humans

Introduction to Artificial Intelligence

How does Machine Learning work ?

1. Database preparation for training/learning

➡ Unlabeled data - Example 1 :

The model extracts the characteristics then defines the classes itself, with no human intervention

Input



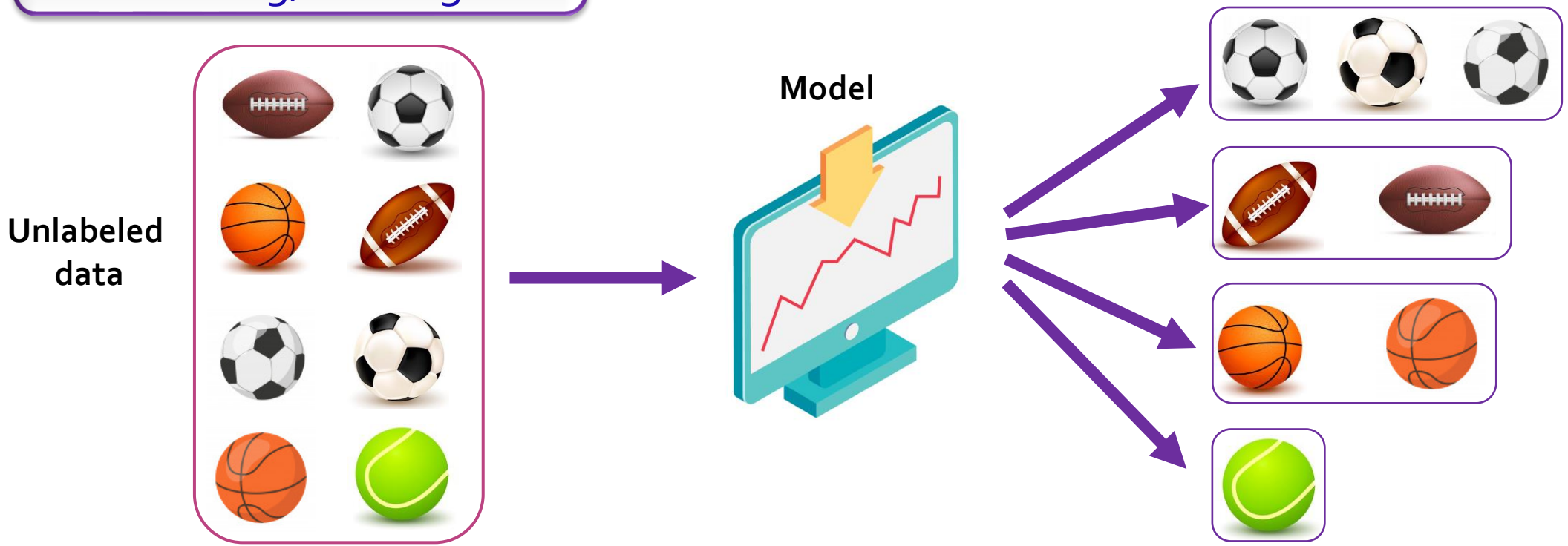
Introduction to Artificial Intelligence

How does Machine Learning work ?

1

Database preparation for training/learning

Unlabeled data - Example 2 :



Introduction to Artificial Intelligence

How does Machine Learning work ?

1. Database preparation for training/learning

- Learning database
  - balanced (unbiased)
  - random
  - No duplicates

Introduction to Artificial Intelligence

How does Machine Learning work ?

1. Database preparation for training/learning

Database :

- Learning database
- Validation (testing) database

Introduction to Artificial Intelligence

How does Machine Learning work ?

2

Machine Learning algorithm selection

Machine Learning types :

- Unsupervised Learning (Self-Supervised Learning)
- Semi-Supervised Learning
- Supervised Learning
- Reinforcement Learning

Introduction to Artificial Intelligence

How does Machine Learning work ?

2

Machine Learning algorithm selection

Supervised Learning

- Labeled database
- Human supervision and guidance
- Detection of pre-determined information

Introduction to Artificial Intelligence

How does Machine Learning work ?

2

Machine Learning algorithm selection

Supervised Learning

- Labeled database
- Human supervision and guidance
- Detection of pre-determined information

Good prediction quality

Introduction to Artificial Intelligence

How does Machine Learning work ?

2

Machine Learning algorithm selection

Supervised Learning

- Labeled database
- Human supervision and guidance
- Detection of pre-determined information

Good prediction quality

Dataset labeling is done by hand :
➔ *Slow and expensive procedure*

A great quantity of information is needed (millions sometimes)

Introduction to Artificial Intelligence

How does Machine Learning work ?

2

Machine Learning algorithm selection

Unsupervised Learning
(Self-Supervised Learning)

- Unlabeled database
- No human supervision
- Detection of information that has not been specified previously

Introduction to Artificial Intelligence

How does Machine Learning work ?

2

Machine Learning algorithm selection

Unsupervised Learning (Self-Supervised Learning)

- Unlabeled database
- No human supervision
- Detection of information that has not been specified previously

Saving money, time and energy

Introduction to Artificial Intelligence

How does Machine Learning work ?

2

Machine Learning algorithm selection

Unsupervised Learning (Self-Supervised Learning)

- Unlabeled database
- No human supervision
- Detection of information that has not been specified previously

Saving money, time and energy

More complex algorithms

Less predictable

Introduction to Artificial Intelligence

How does Machine Learning work ?

2

Machine Learning algorithm selection

Semi-supervised Learning

- Labeled + unlabeled data
- Labeled data size is smaller

Introduction to Artificial Intelligence

How does Machine Learning work ?

2

Machine Learning algorithm selection

Semi-supervised Learning

- Labeled + unlabeled data
- Labeled data size is smaller

Useful when data is sparse, difficult to obtain or rare

Introduction to Artificial Intelligence

How does Machine Learning work ?

2

Machine Learning algorithm selection

Semi-supervised Learning

- Labeled + unlabeled data
- Labeled data size is smaller

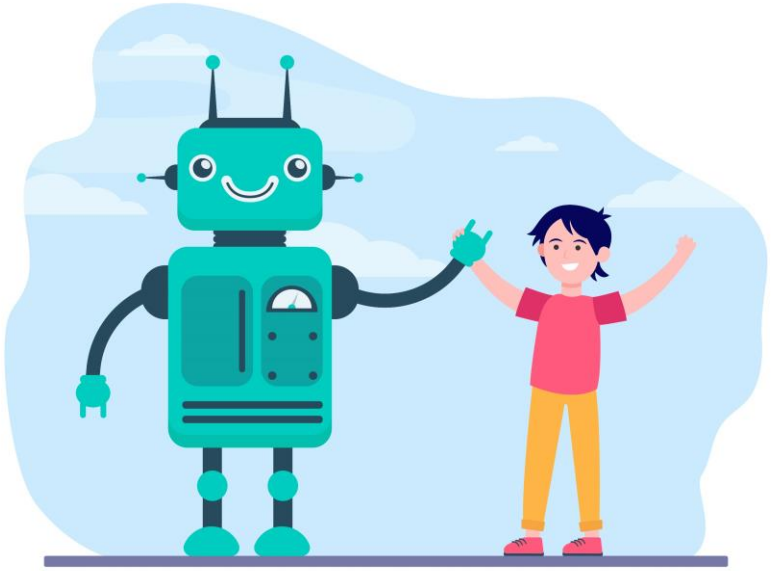
Useful when data is sparse, difficult to obtain or rare

Lower accuracy

Unstable

Introduction to Artificial Intelligence

How does Machine Learning work ?



2

Machine Learning algorithm selection

Reinforcement Learning

- The system makes a **decision** with respect to its present state and receives a positive or a negative **reward** in response to its decision
- The system generates an optimal **strategy** to maximise the obtained rewards
- **Iterative** procedure
- Used in robotic training

Introduction to Artificial Intelligence

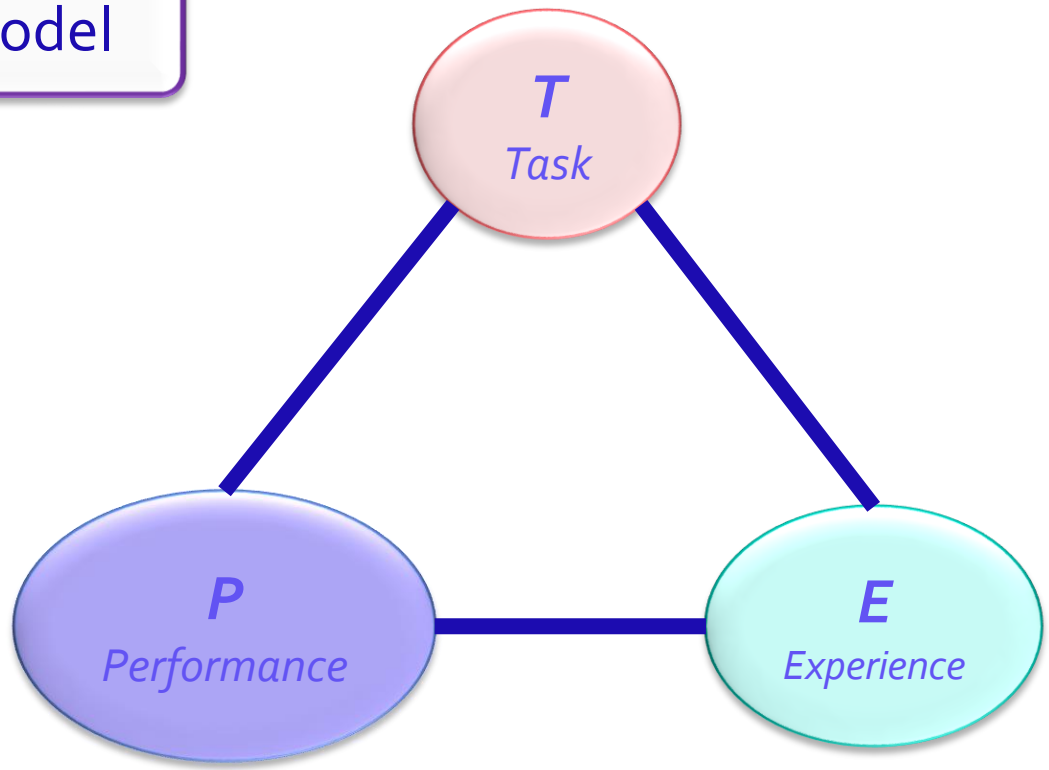
How does Machine Learning work ?

3

Algorithm training (learning) to obtain the model

*« **Machine learning** is the study of computer **algorithms** that allow computer programs to automatically improve through experience. »*

Tom Mitchell
Machine Learning, 1997

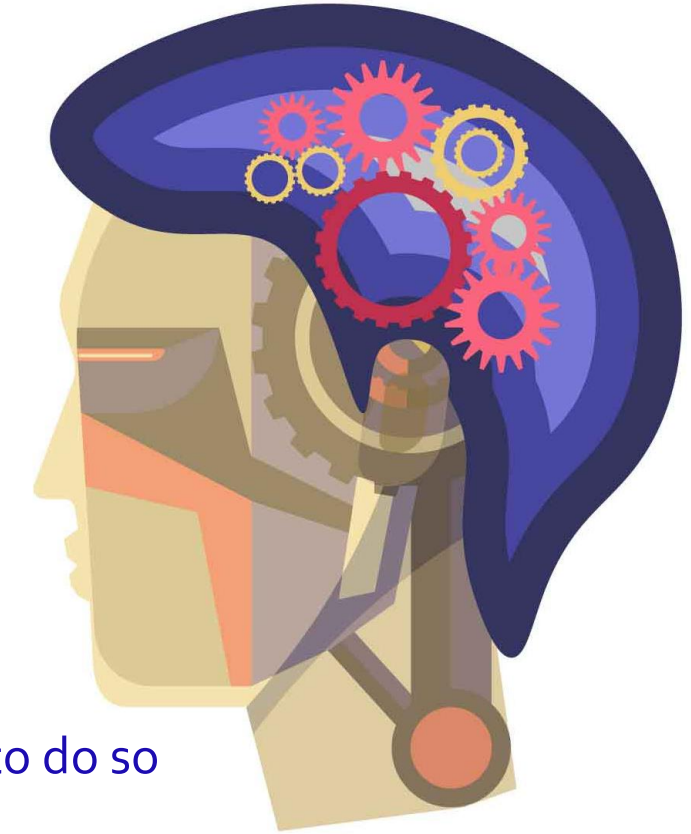


Introduction to Artificial Intelligence

How does Machine Learning work ?

3

Algorithm training (learning) to obtain the model



Machine Learning creates an algorithm that :

→ Learns with time and experience

→ Has the capacity to learn without being explicitely programmed to do so

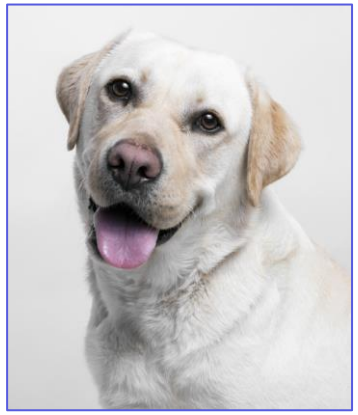
Introduction to Artificial Intelligence

How does Machine Learning work ?

4

Using the trained model

Input



Model of cat/dog classification

Trained Model

Output

« Dog »

Introduction to Artificial Intelligence

How does Machine Learning work ?

4

Using the trained model

Input

Area = 43

Number of rooms = 3

Distance from train station = 7

City = Strasbourg

Regression model
(house price predictor)

Trained Model

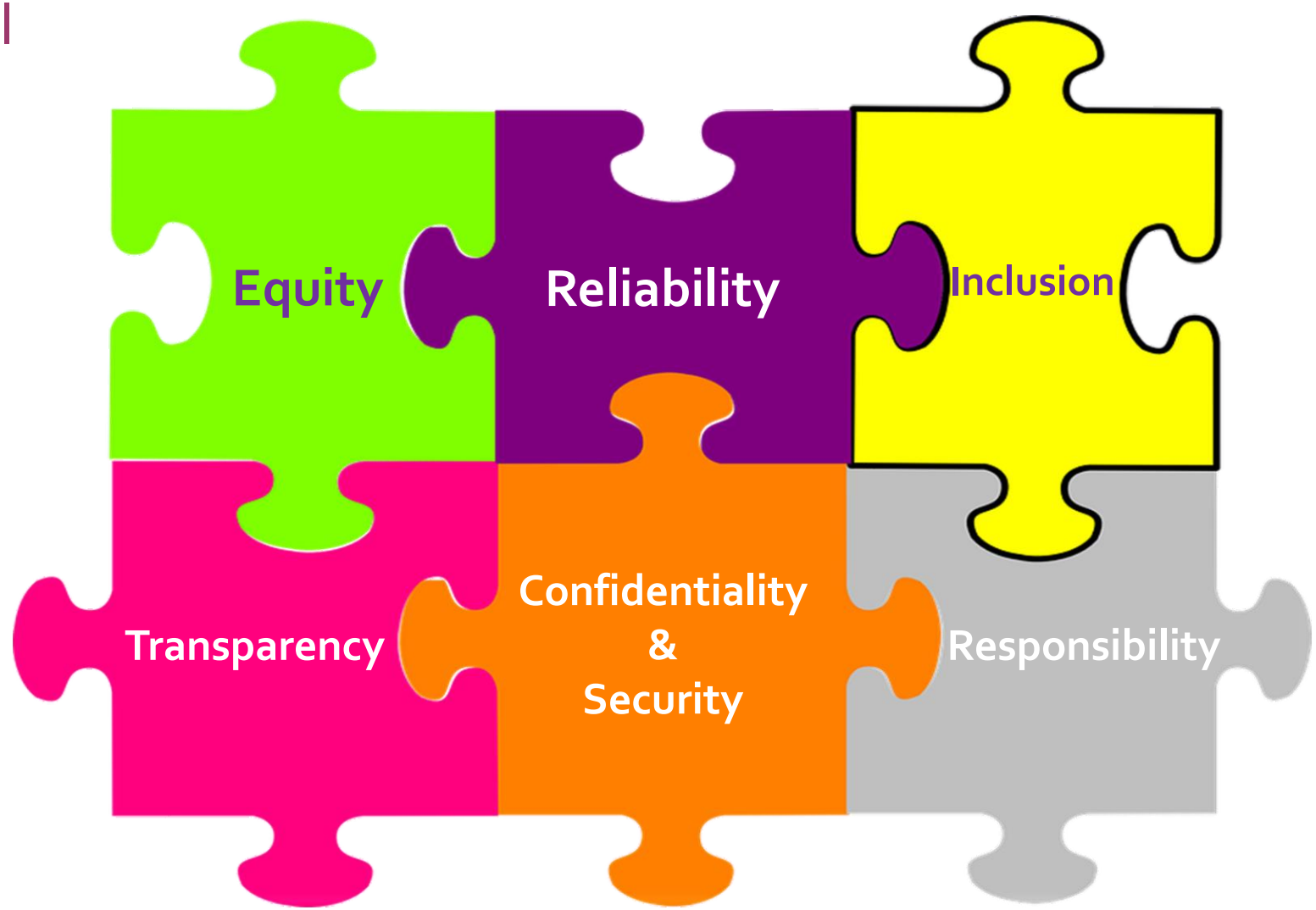
Output

210 000 €

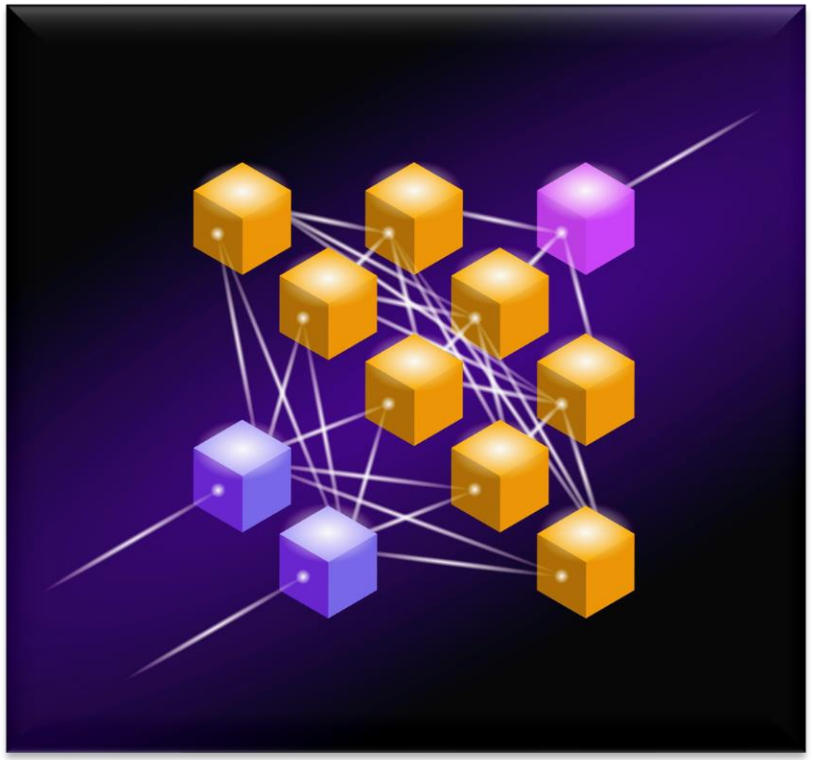
Introduction to Artificial Intelligence

Ethical and Responsible AI

En ethical AI system is one that has:



What are Artificial Neural Networks ?

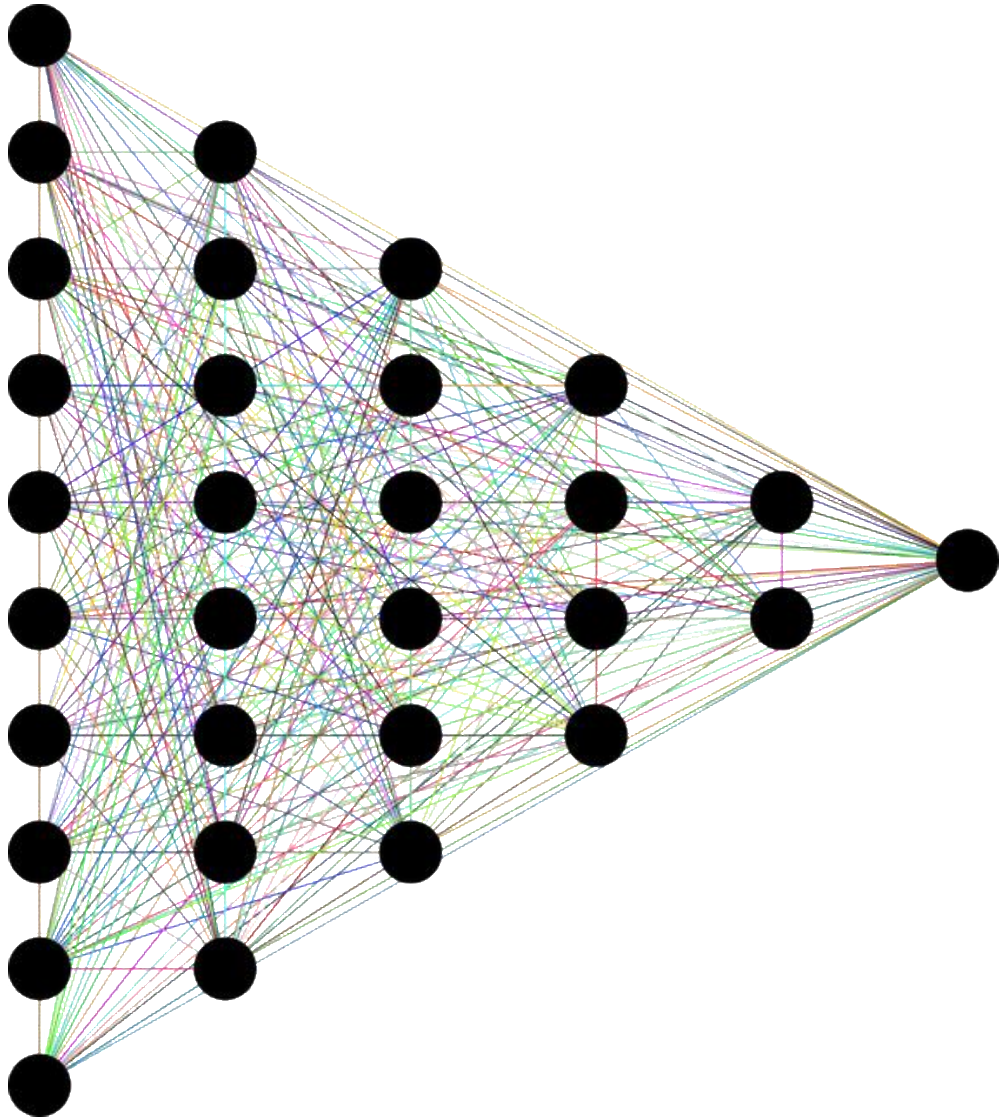


What are Artificial Neural Networks ?

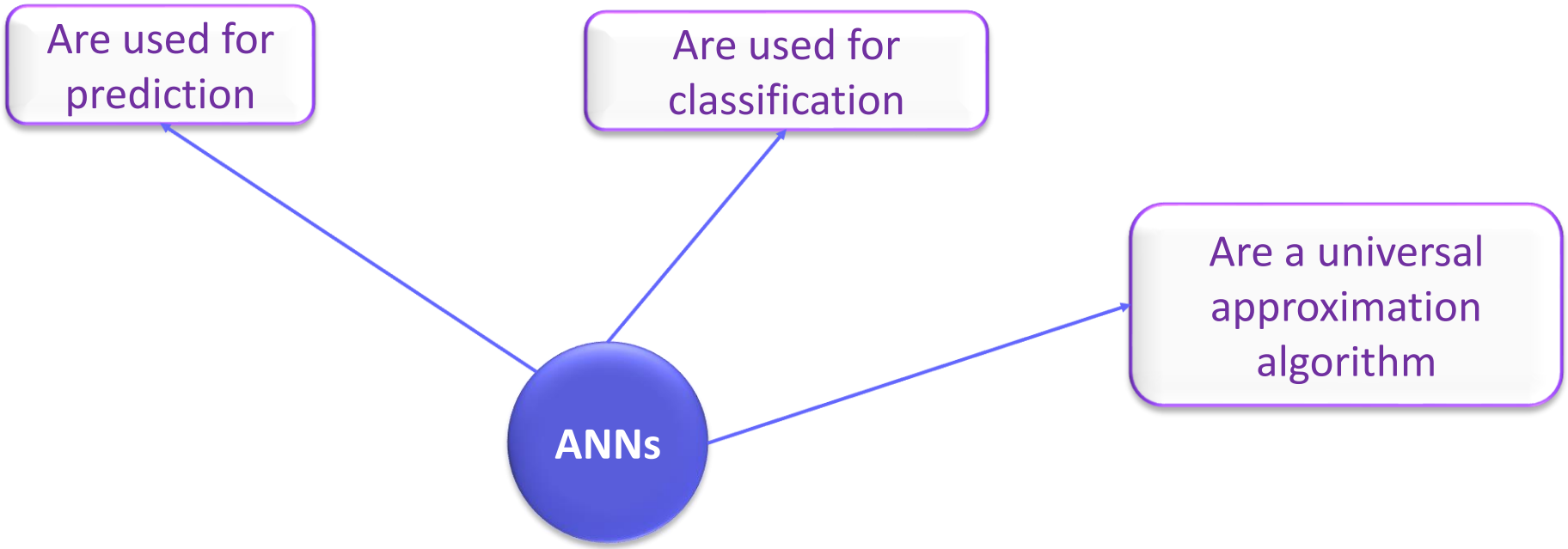
Artificial Neural Networks (**ANNs**), or

simply **Neural Networks**, are :

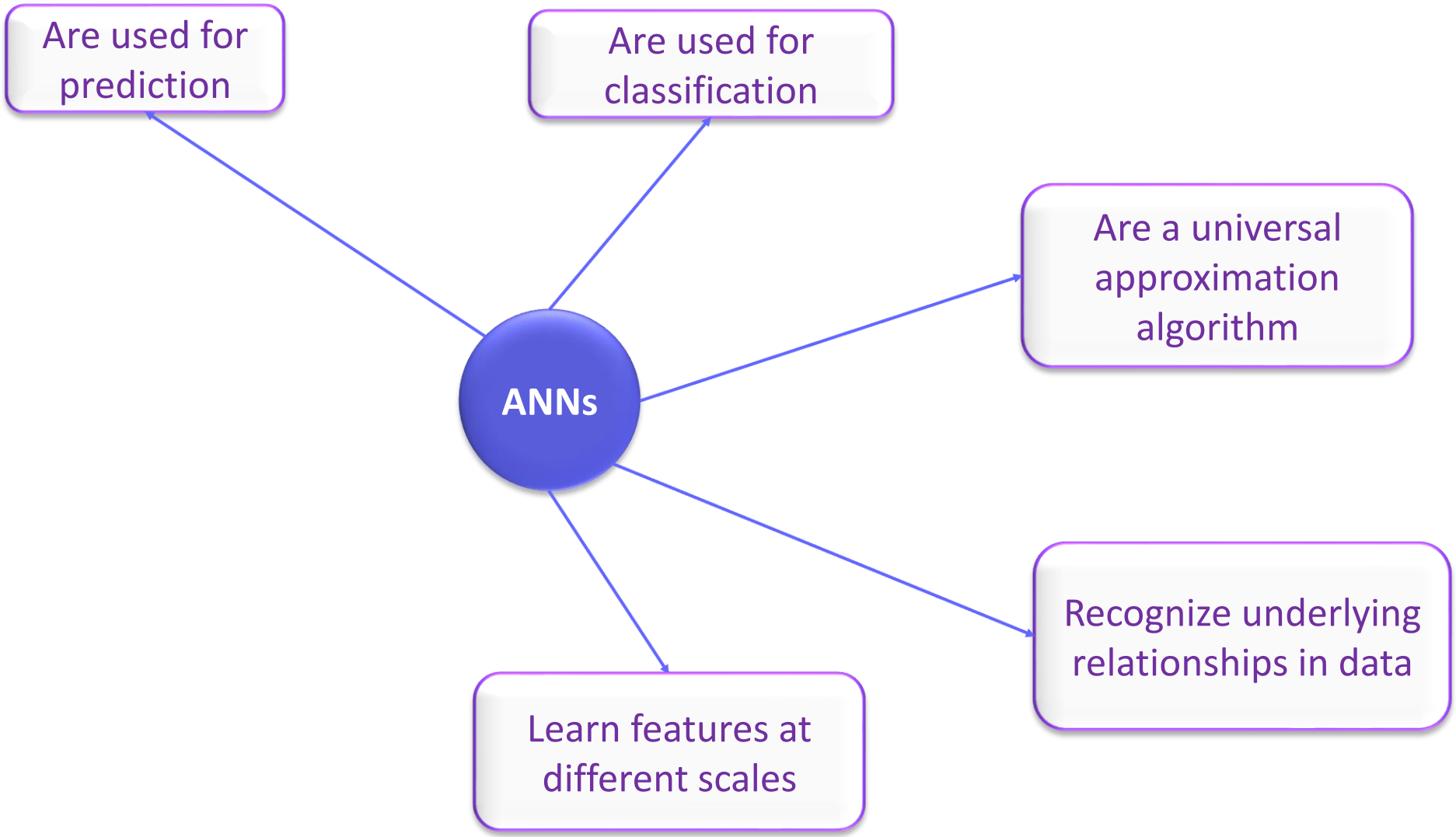
- A subset of Machine Learning
- At the core of Deep Learning



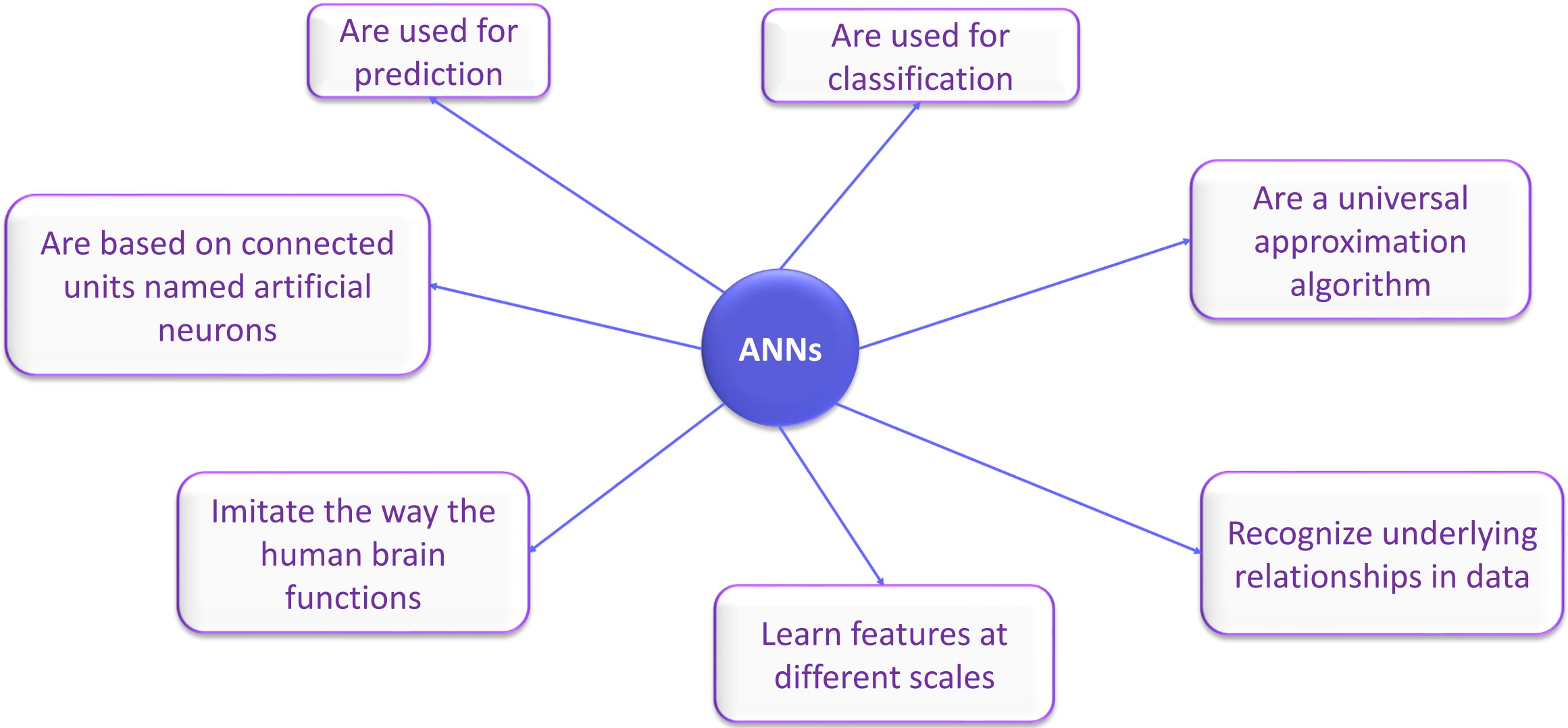
What are Artificial Neural Networks ?



What are Artificial Neural Networks ?



What are Artificial Neural Networks ?



What are Artificial Neural Networks ?

A brief history

1943

McCulloch & Pitts model a basic NN using electric circuits

1957

Rosenblatt starts his work on the Perceptron

1962

Widrow & Hoff develop the Least Mean Squared learning rule

1979

Fukushima uses the very first convolutional NN

1982

Hopfield creates bi-directional networks

1983

Hinton & Sejnowski invent Boltzmann Machines, unsupervised NN

1985

Rumelhart, Hinton & Williams : backpropagation can be used in NNs & the introduction of RNNs

What are Artificial Neural Networks ?

A brief history

1997

Hochreiter & Schmidhuber develop the long short-term memory (LSTM)

1998

LeCun publishes : Stochastic gradient descent with backpropagation are very successful in deep learning

2000

Bengio's work kickstarts Natural Language Processing

2009

Fei-Fei Li creates the ImageNet database used in image recognition research

2011

Krizhevsky, Sutskever & Hinton use a CNN with ReLU and dropout regularization

2014

Goodfellow & Bengio publish the paper 'Generative Adversarial Nets', a revolution in computer vision

2017

Google's AI teams propose transformers based on attention mechanisms

What are Artificial Neural Networks ?

ACM A. M. Turing Award Laureates

Fathers of the Deep Learning Revolution :

Yoshua Bengio

Geoffrey Hinton

Yann LeCun

Source : https://awards.acm.org/about/2018-turing

What are Artificial Neural Networks ?

ACM A. M. Turing Award Laureates

Fathers of the Deep Learning Revolution :

Yoshua Bengio

Geoffrey Hinton

Yann LeCun

The A.M. Turing Award :

- *is ACM's most prestigious award.*
- *rewards major contributions of long-term interest to computing.*
- *was created in 1966*

* ACM = Association for Computing Machinery

Source : https://awards.acm.org/about/2018-turing

What are Artificial Neural Networks ?

Andrew Ng is **Founder & CEO of Landing AI, Founder of deeplearning.ai, Co-Chairman and Co-Founder of Coursera, and is currently an Adjunct Professor at Stanford University**. He was also Chief Scientist at Baidu Inc., and Founder & Lead for the Google Brain Project.

https://hai.stanford.edu › people › andrew-ng

Andrew Ng | Stanford HAI

The Artificial Neuron

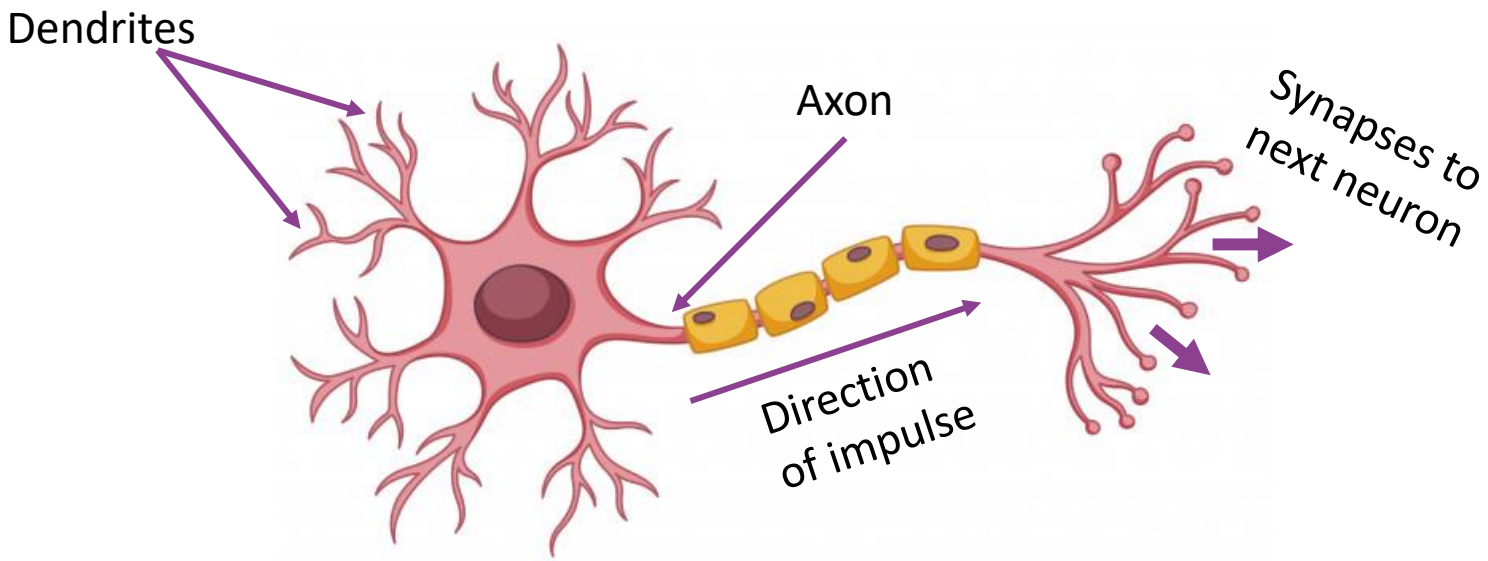
The Artificial Neuron

The purpose of neural networks :

imitation of the biological brain

Recreate the basic unit of our nervous system :

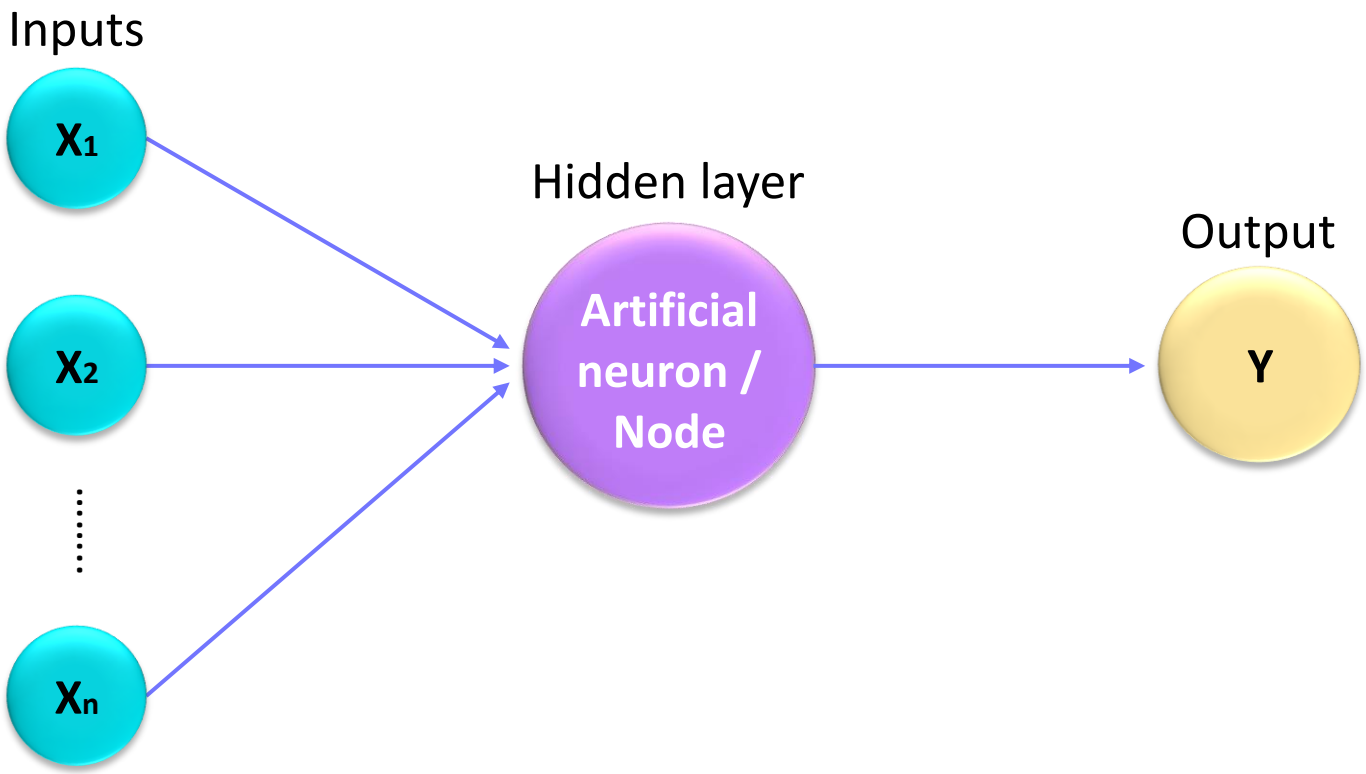
the **neuron**



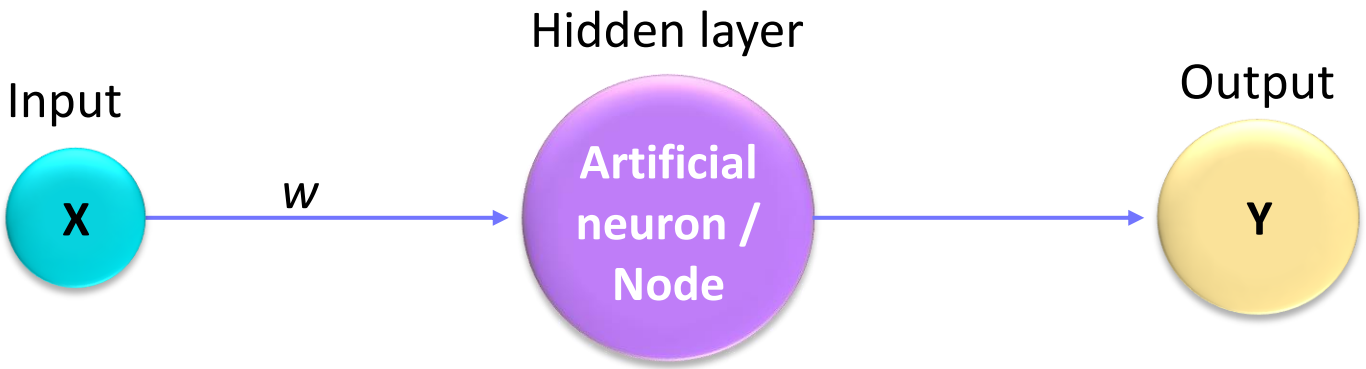
The Artificial Neuron

The artificial neuron :

- *The building block of neural networks*
- *Computational unit that takes inputs and yields outputs*



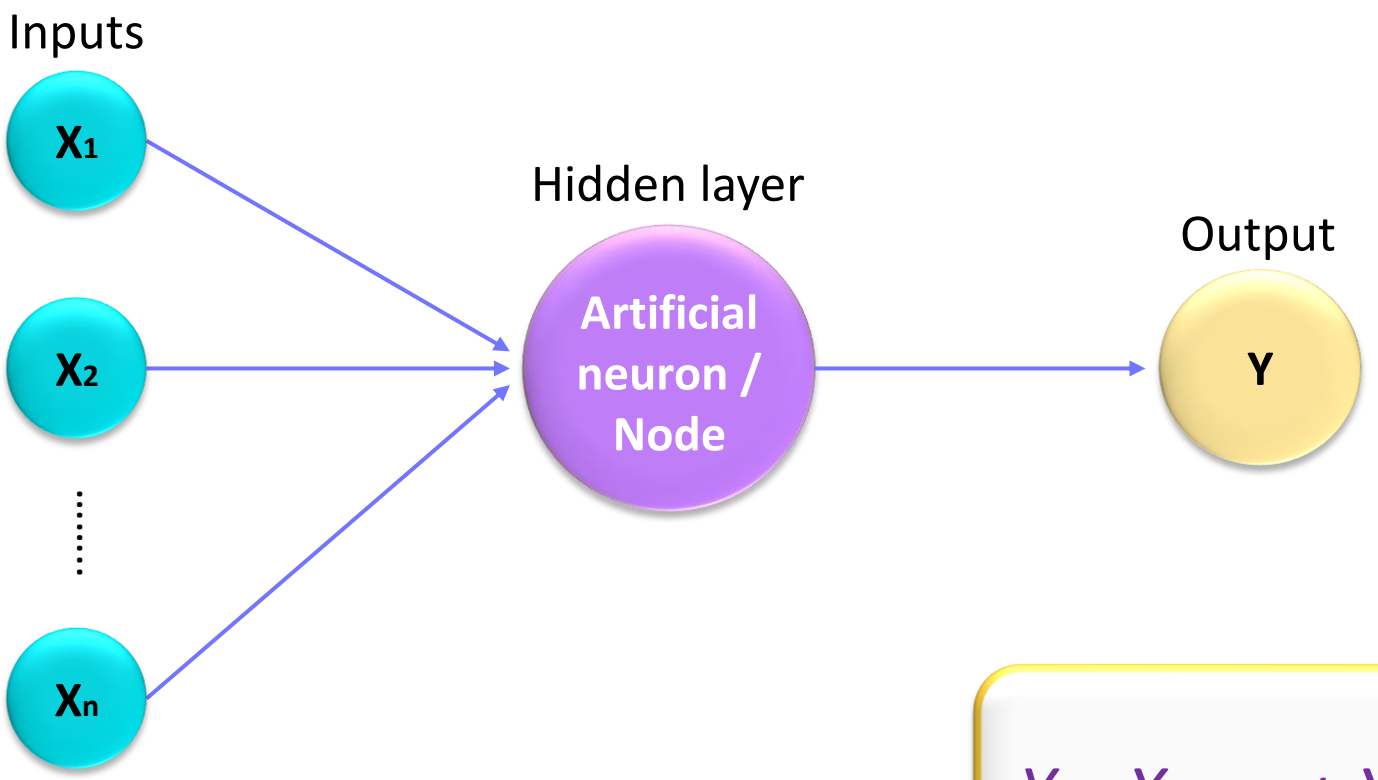
The Artificial Neuron



$$Y = Input \times weight$$

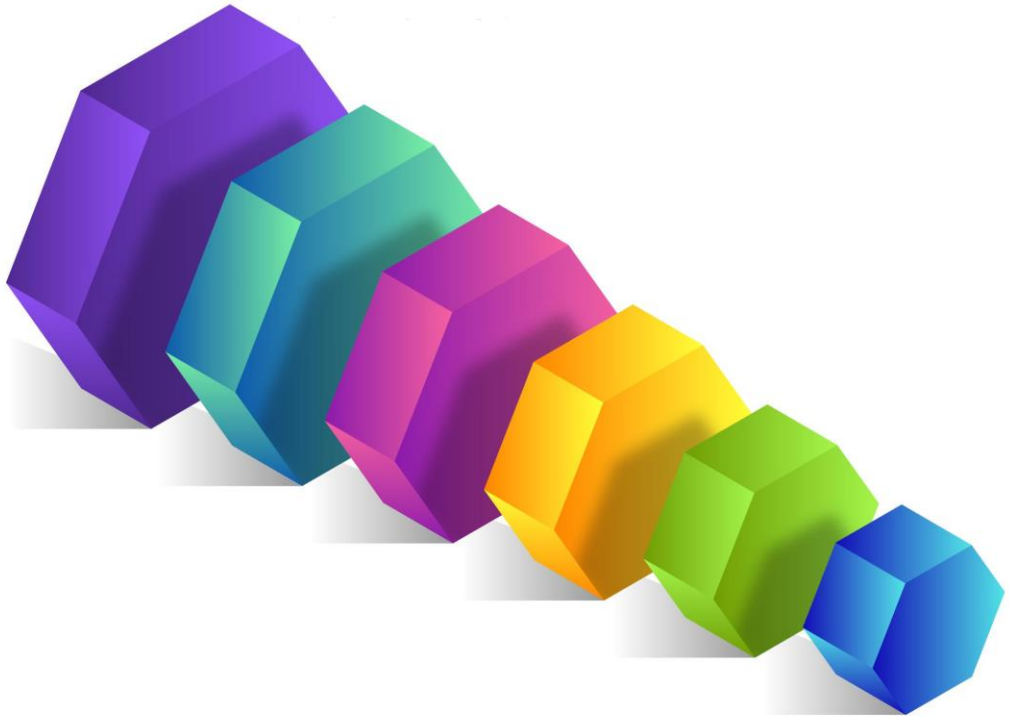
$$Y = X \times w$$

The Artificial Neuron



$$Y = X_1 w_1 + X_2 w_2 + \ldots + X_n w_n$$

ANN Layers

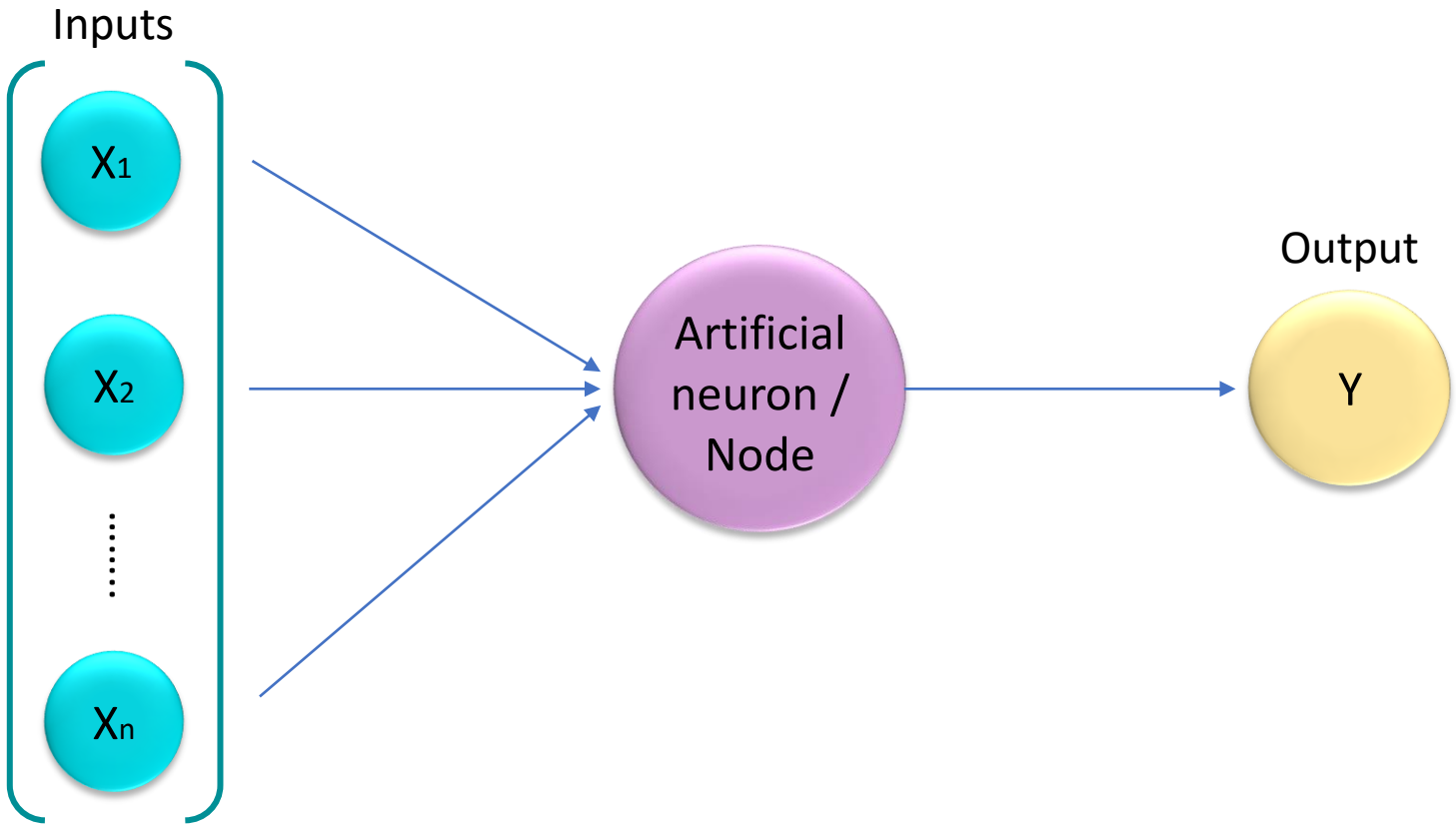


ANN Layers : Inputs

The inputs: Independent variables for a single observation

The input features (independent variables) can be :

- categorical
- numeric



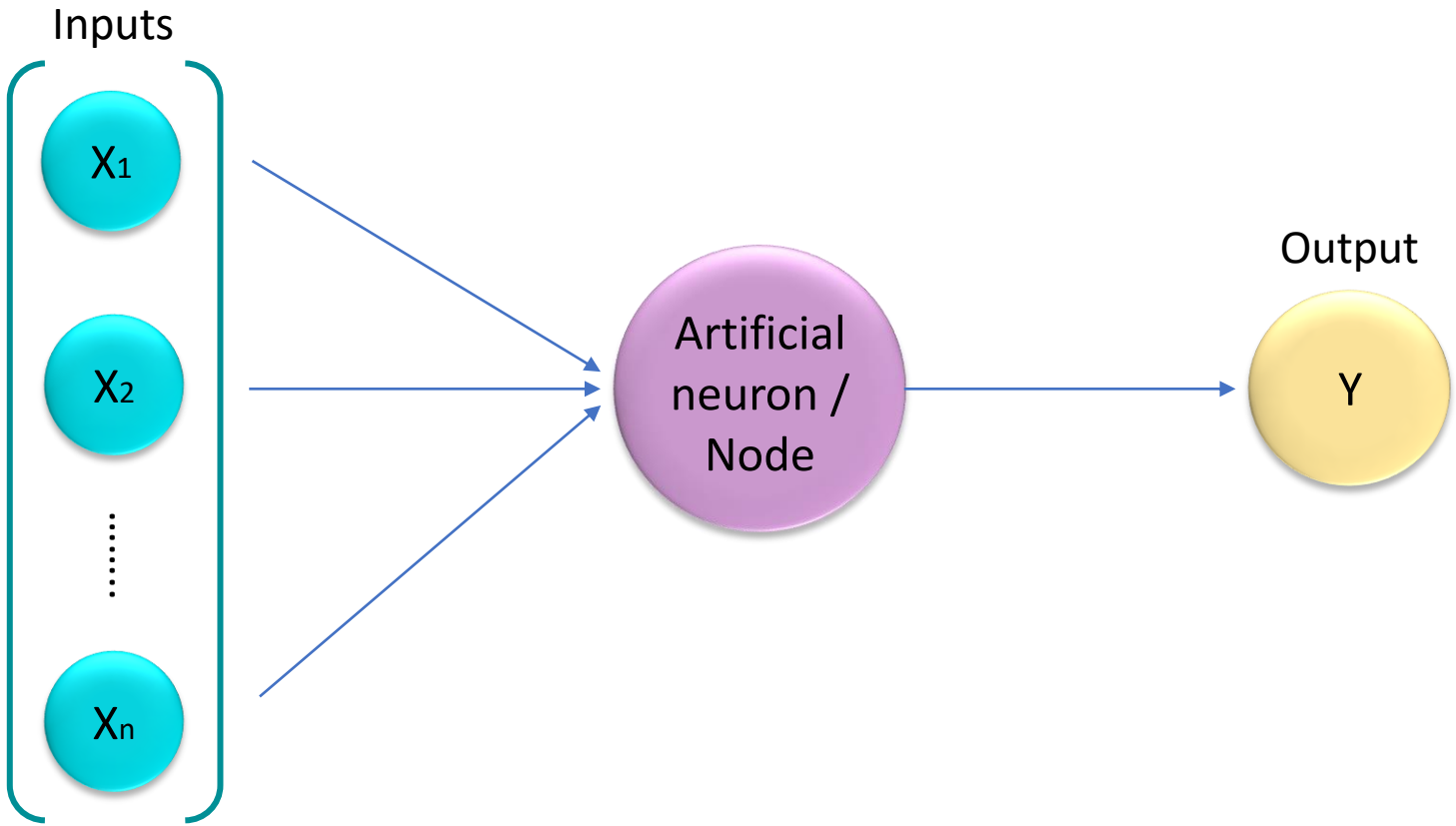
ANN Layers : Inputs

The inputs: Independent variables for a single observation

Categorical input features :

Inputs are labels (categories)

→ cats & dogs



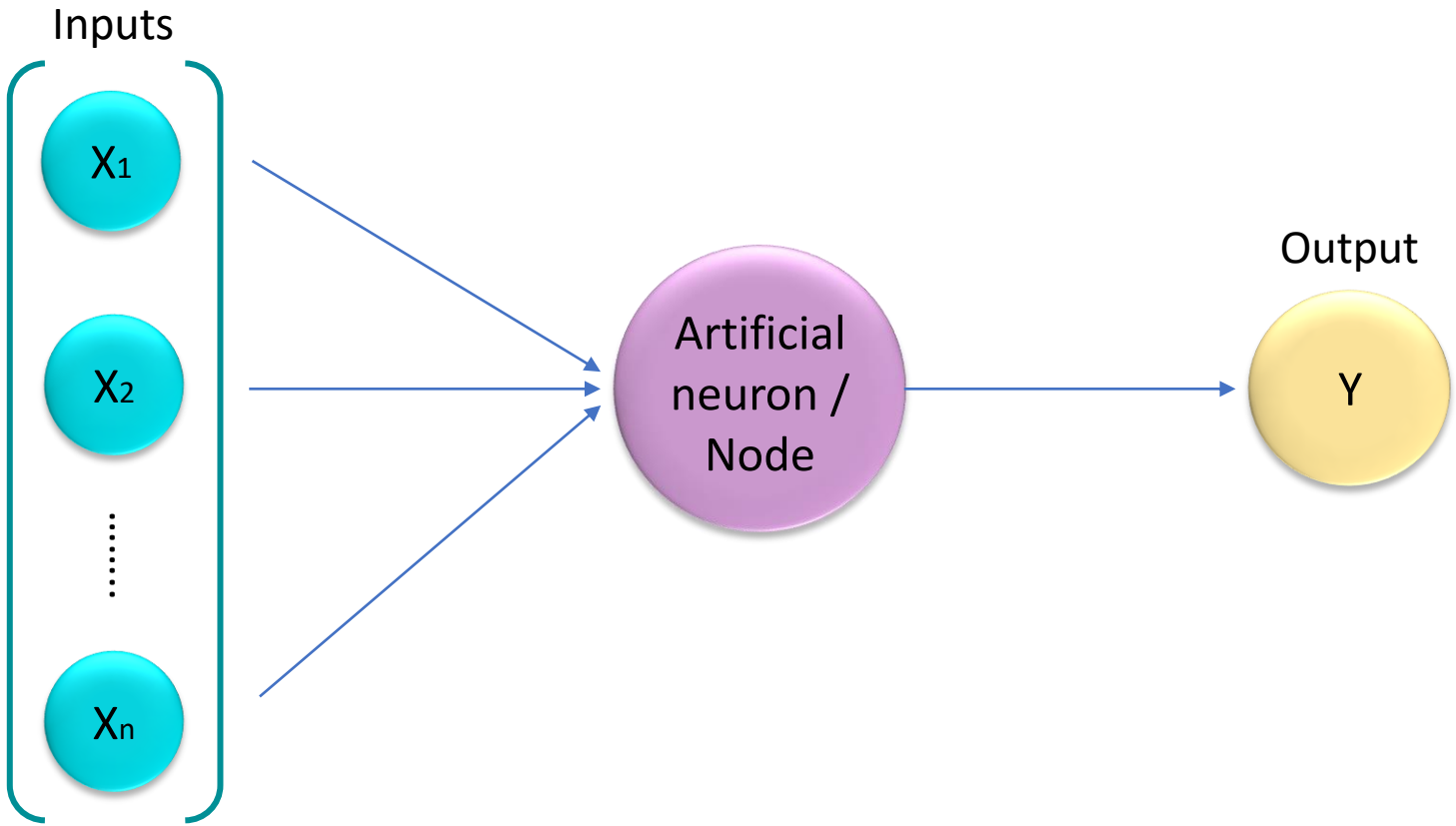
ANN Layers : Inputs

The inputs: Independent variables for a single observation

Numeric input features :

Inputs contain numerical values

→ area, weight, distance, etc.



ANN Layers : Inputs

The inputs: Independent variables for a single observation

Example :

If our *observation* is a *car* :

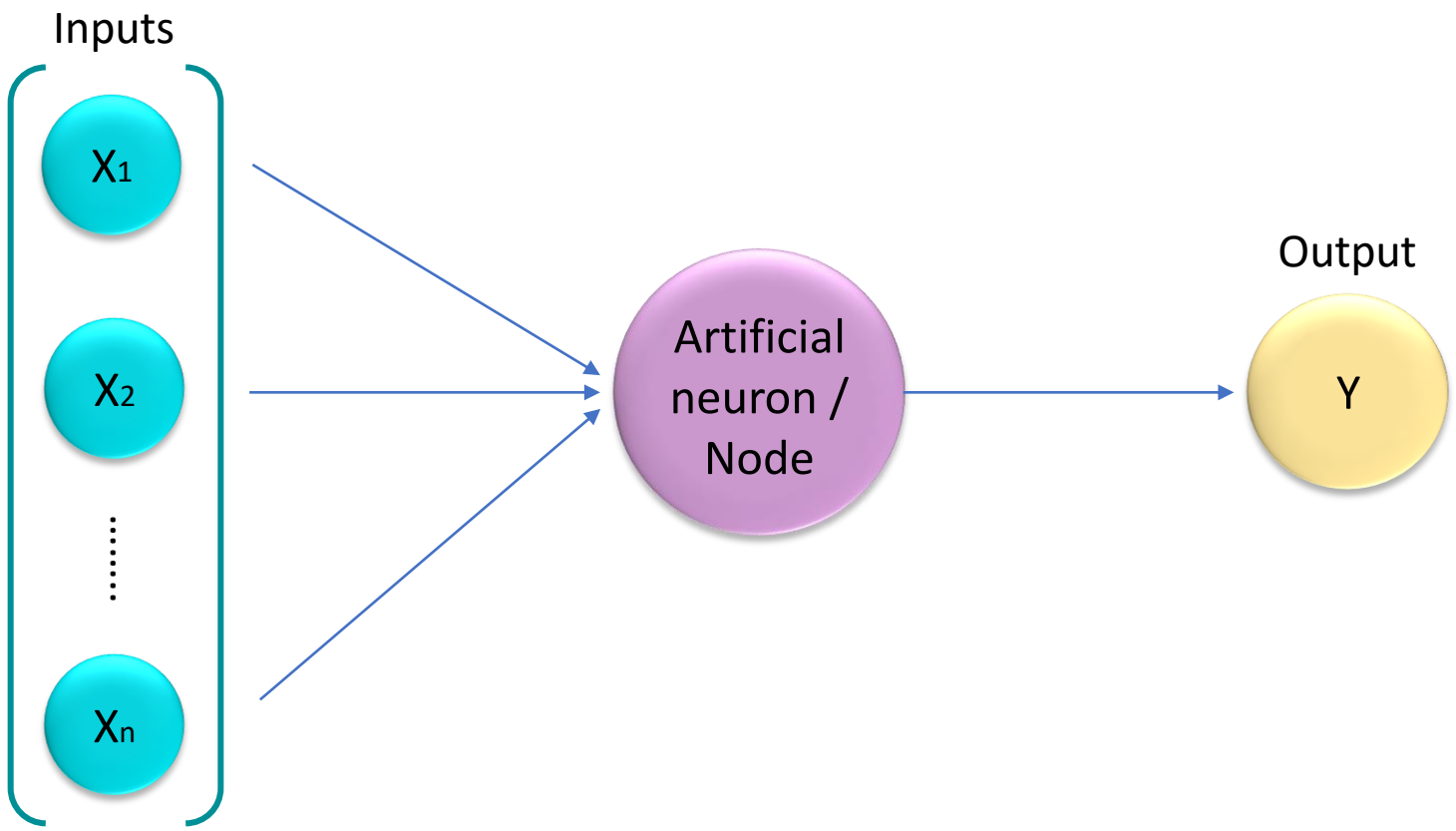
X1 : brand

X2 : manufacturing year

X3 : color

X4 : automatic/mechanical

⋮

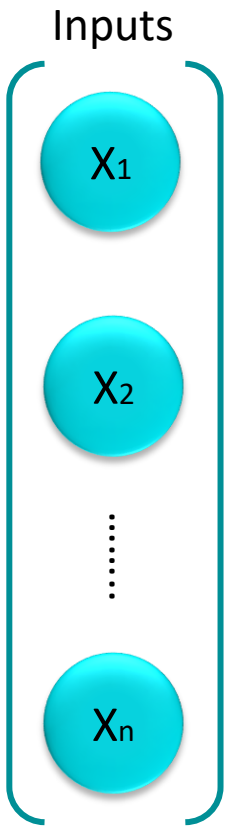


ANN Layers : Inputs

The inputs:

Data preparation :

To avoid false influence of the different input values and intervals on the trained model



1

Standardization :

1. Subtract the mean
2. Divide by the standard deviation

$Xi - mean \ / \ standard\_dev$

$==> Mean = 0 \ \& \ Variance = 1$

2

Normalization :

1. Subtract minimum value
2. Divide by the data range

$X_i - X_{min} \ / \ (X_{max} - X_{min})$

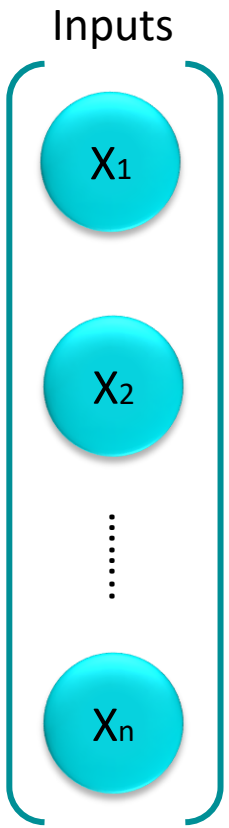
$==> 0 \leq X_{i\_new} \leq 1$

ANN Layers : Inputs

The inputs:

Data preparation :

To avoid false influence of the different input values and intervals on the trained model



1

Standardization :

Z-Score normalization

Mean = 0

Variance = 1

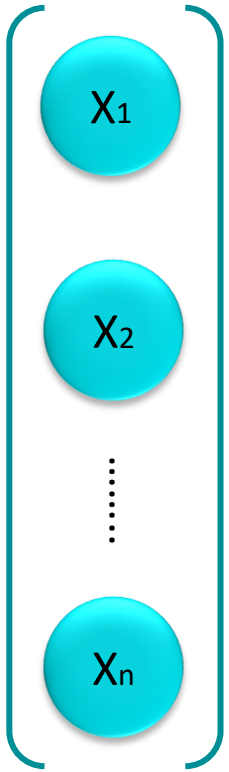
More robust to outliers (preserves them

Can be applied when the indepent variables follow normal (Gaussian) distributions with different means and standard deviations

ANN Layers : Inputs

The inputs:

Inputs



Data preparation :

To avoid false influence of the different input values and intervals on the trained model

2

Normalization :

Min-Max Scaling

1. Subtract minimum value
2. Divide by the data range

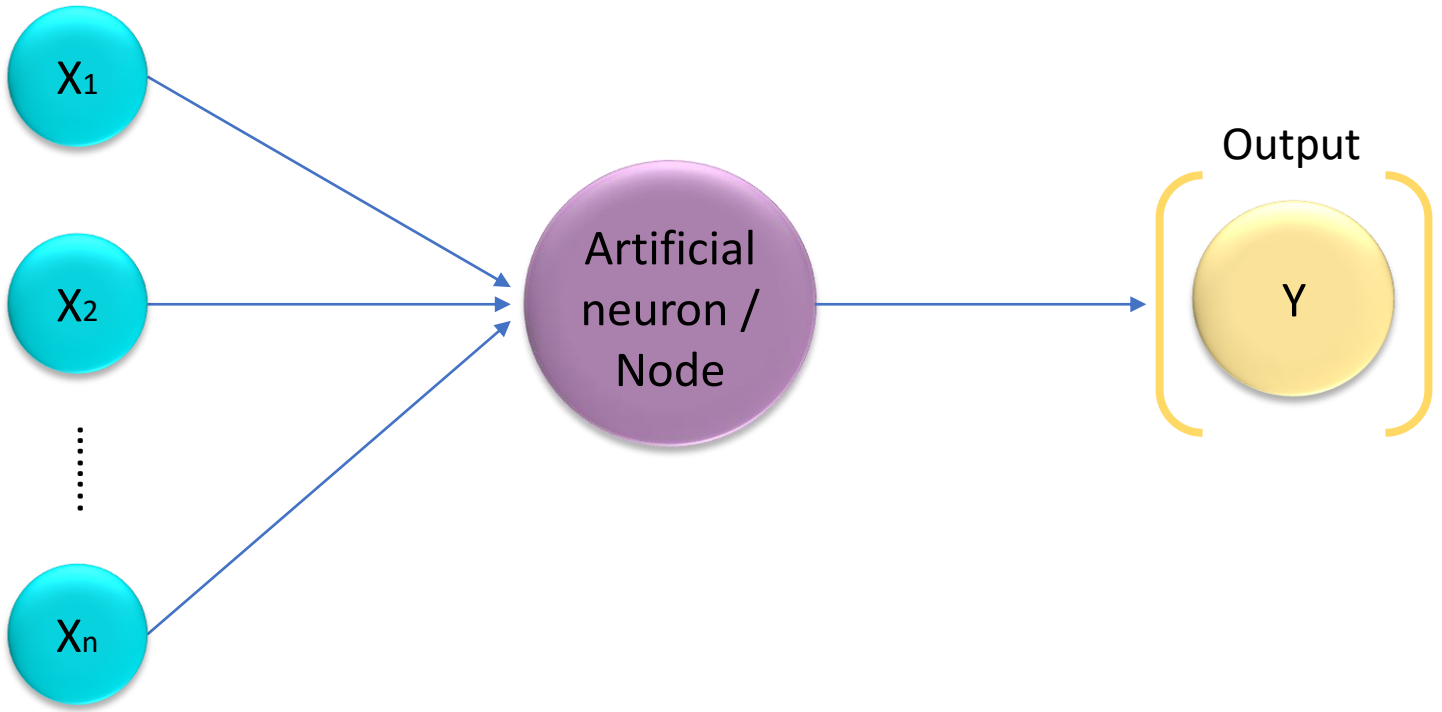
$$X_i - X_{min} \,/\, (X_{max} - X_{min})$$

$$==> 0 \leq X_{i\_new} \leq 1$$

Can be applied when the indepent variables vary in different scales and intervals

ANN Layers : Outputs

The output:



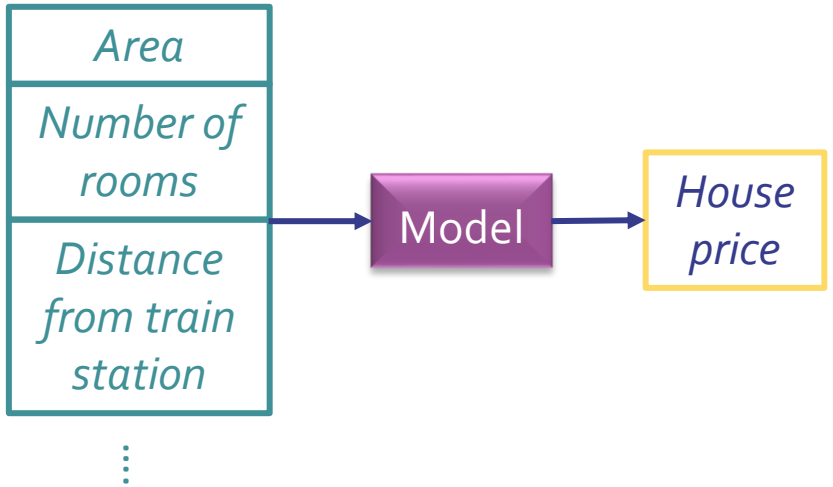
1. Continuous
2. Binary
3. Categorical

ANN Layers : Outputs

The output:

Continuous

Value prediction

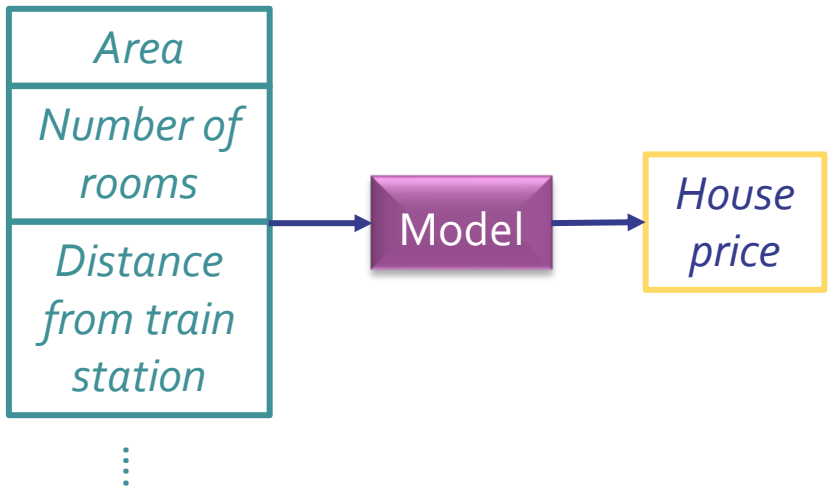


ANN Layers : Outputs

The output:

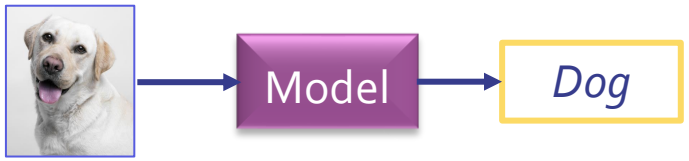
Continuous

Value prediction



Binary

Classification among 2 classes



(out of 2 possible classes (Dog & Cat)

ANN Layers : Outputs

The output:

Continuous

Value prediction

Area

Number of rooms

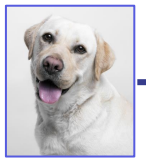
Distance from train station

⋮

Model → *House price*

Binary

Classification among 2 classes



Model → *Dog*

(out of 2 possible classes (Dog & Cat)

Categorical

Classification among several classes

Model → *Dog*

(out of more than 2 possible classes (Dog, Cat, Turtle, Elephant, Lion, etc.))

ANN Layers : Inputs & Outputs

Categorical Data Encoding : Most Machine Learning models do not accept text labels for inputs and outputs

➔ We have to encode those categorical labels to numerical values.

1. Ordinal Encoding : Every label is represented by an ordered number

Example : If we are classifying countries :

Country (Label)	Ordinal encoding
India	1
Japan	2
Madagascar	3
USA	4

ANN Layers : Inputs & Outputs

Categorical Data Encoding : Most Machine Learning models do not accept text labels for inputs and outputs

➔ We have to encode those categorical labels to numerical values.

1. Ordinal Encoding : Every label is represented by an ordered number

Example : If we are classifying countries :

Country (Label)	Ordinal encoding
India	1
Japan	2
Madagascar	3
USA	4

The system will probably establish false relations between these countries

➔ Example : It can consider India superior to Japan while there is no hierarchical relation between the countries !

ANN Layers : Inputs & Outputs

Categorical Data Encoding : Most Machine Learning models do not accept text labels for inputs and outputs

➔ We have to encode those categorical labels to numerical values.

2. One-Hot Encoding : Creates new binary variables based on the number of labels

→ N labels are represented by N bits, while only one bit = 1

Example : If we are classifying countries :

Country (Label)	One-Hot encoding
India	1 0 0
Japan	0 1 0
Madagascar	0 0 1

ANN Layers : Inputs & Outputs

Categorical Data Encoding : Most Machine Learning models do not accept text labels for inputs and outputs

➔ We have to encode those categorical labels to numerical values.

2. One-Hot Encoding : Creates new binary variables based on the number of labels

→ N labels are represented by N bits, while only one bit = 1

Example : If we are classifying countries :

Country (Label)	One-Hot encoding
India	1 0 0
Japan	0 1 0
Madagascar	0 0 1

If we have a large number of categories, the variable can grow to demand a considerable amount of memory storage

ANN Layers : Inputs & Outputs

Categorical Data Encoding :

So, which encoding should we use ?

Ordinal Encoding :

- When the labels are ordered (primary school, secondary school, university)
- The number of categories is very large

One-Hot Encoding :

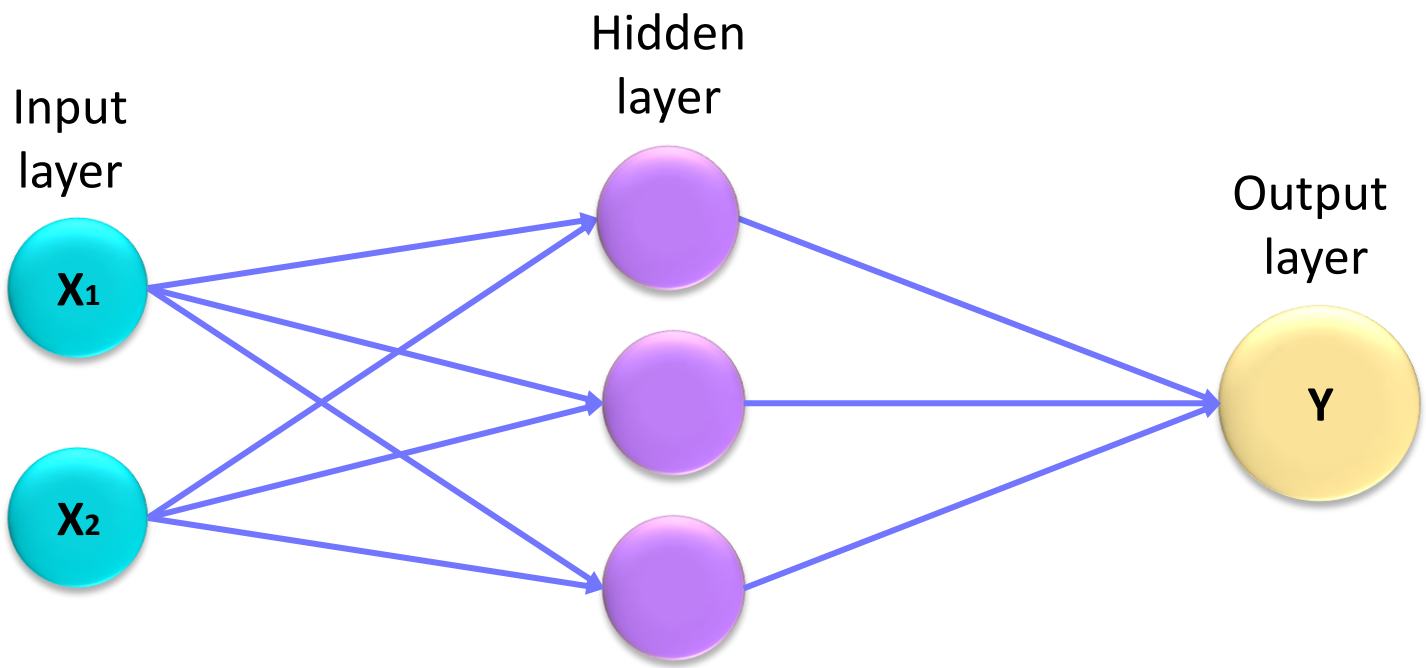
- When the labels are not ordered (countries, animals)
- The number of categories is not too large to cause memory problems

ANN Layers : Hidden Layers

The hidden layer is situated between inputs and outputs of a neural network

The hidden nodes have no direct connection with the output world

A collection of hidden nodes forms a « *hidden layer* ».



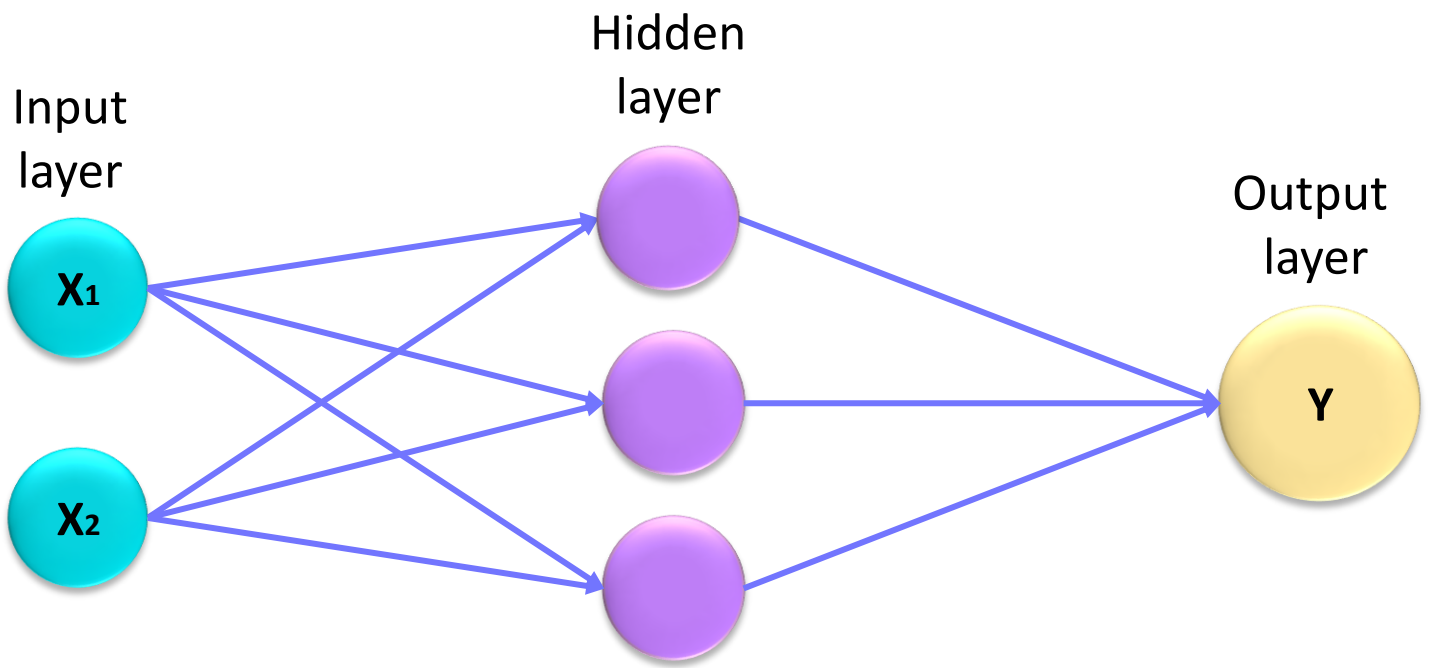
ANN Layers : Hidden Layers

The hidden layer is situated between inputs and outputs of a neural network

The output of one neuron is the input of another

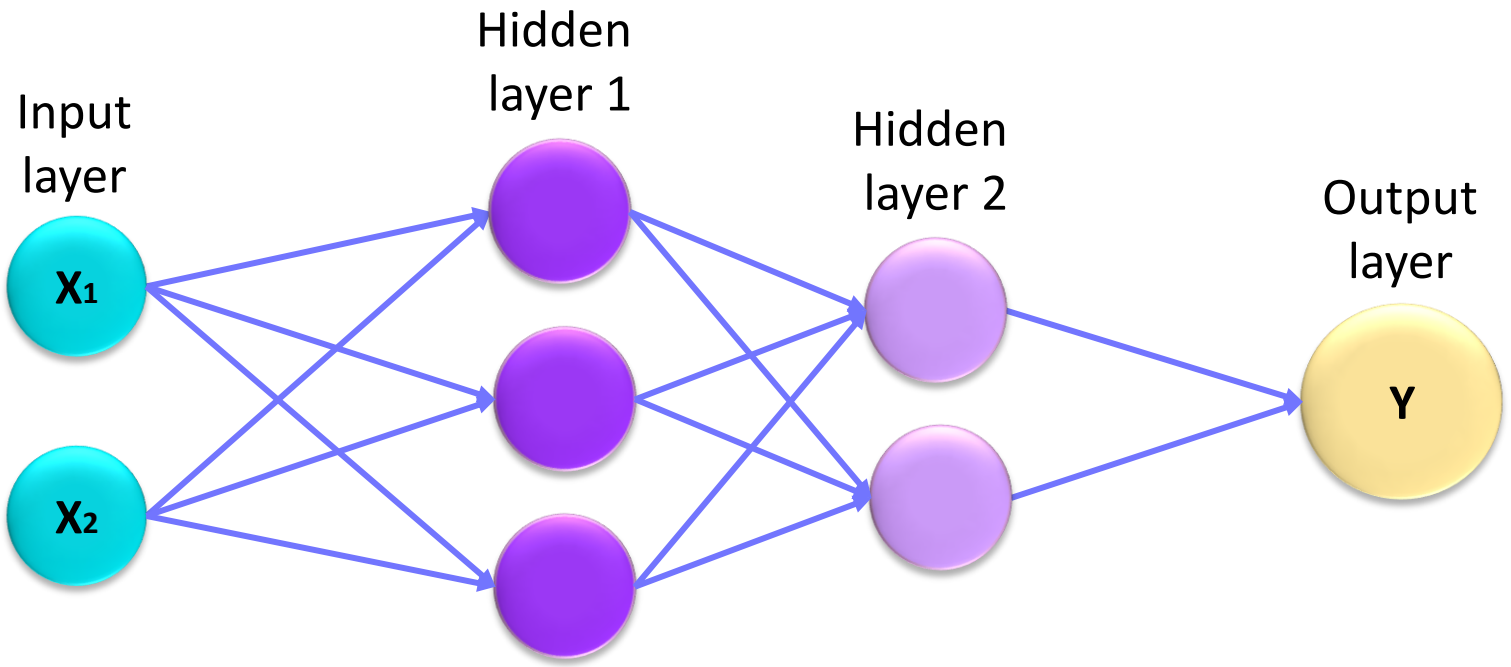
The hidden nodes have no direct connection with the output world

A collection of hidden nodes forms a « *hidden layer* ».



ANN Layers : Hidden Layers

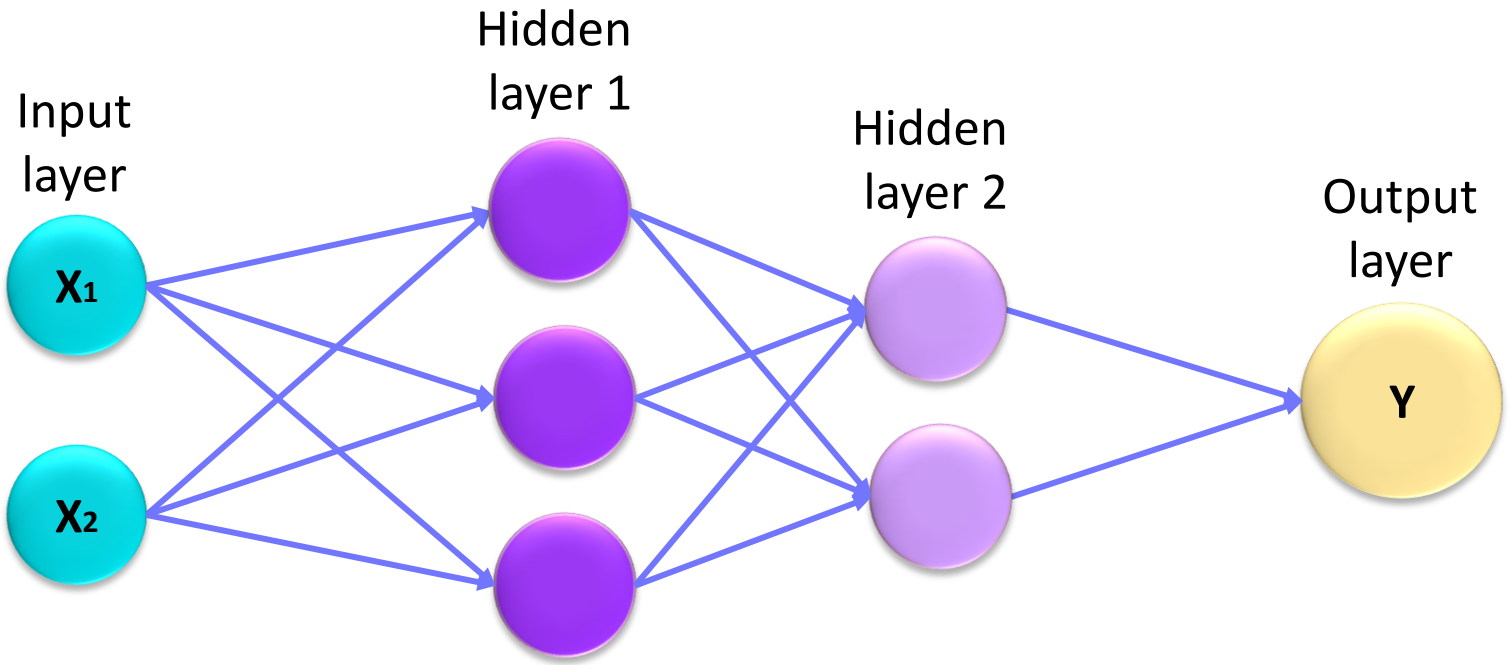
There usually are multiple hidden layers in a neural network



ANNs can have zero or multiple hidden layers

ANN Layers : Hidden Layers

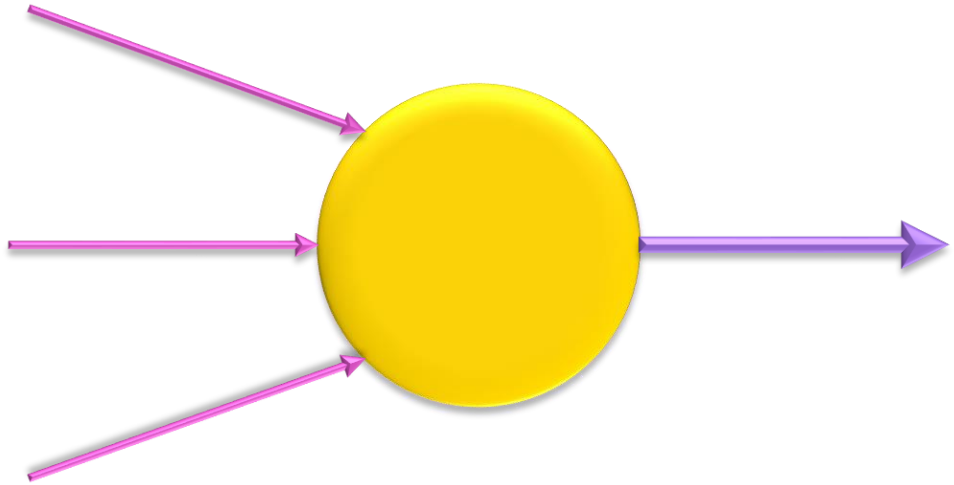
There usually are multiple hidden layers in a neural network



Traditional neural networks typically have fully-connected hidden layers

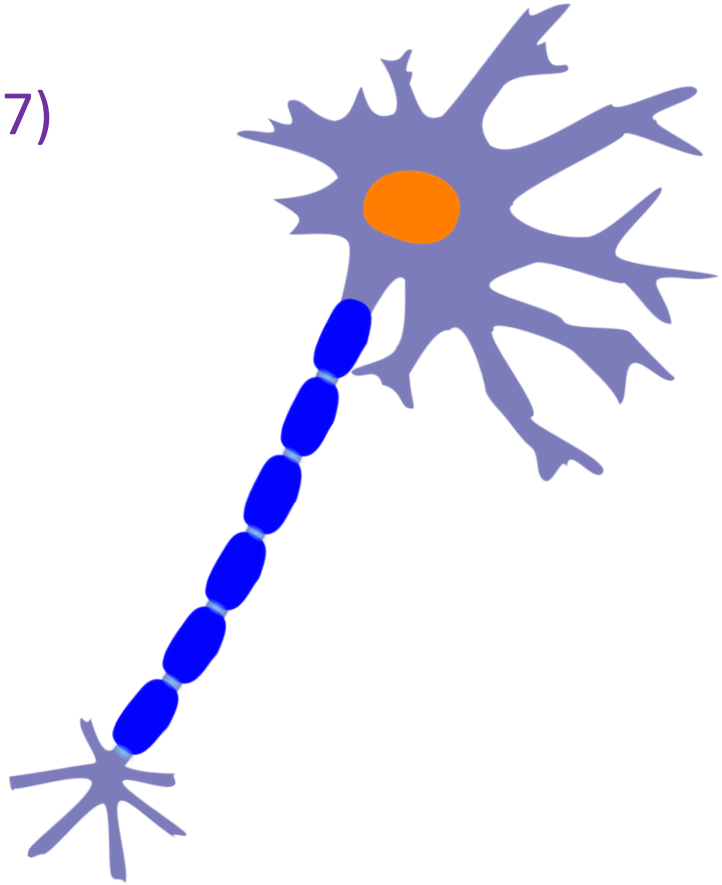
Fully-connected layer : every neuron is connected to all the neurons in the previous and the next layers

The Perceptron



The Perceptron

- The Perceptron model is proposed by Frank Rosenblatt (1957)



Single Layer Perceptron

Multi Layer Perceptron

- Modeled after the essential unit of the human brain – the neuron

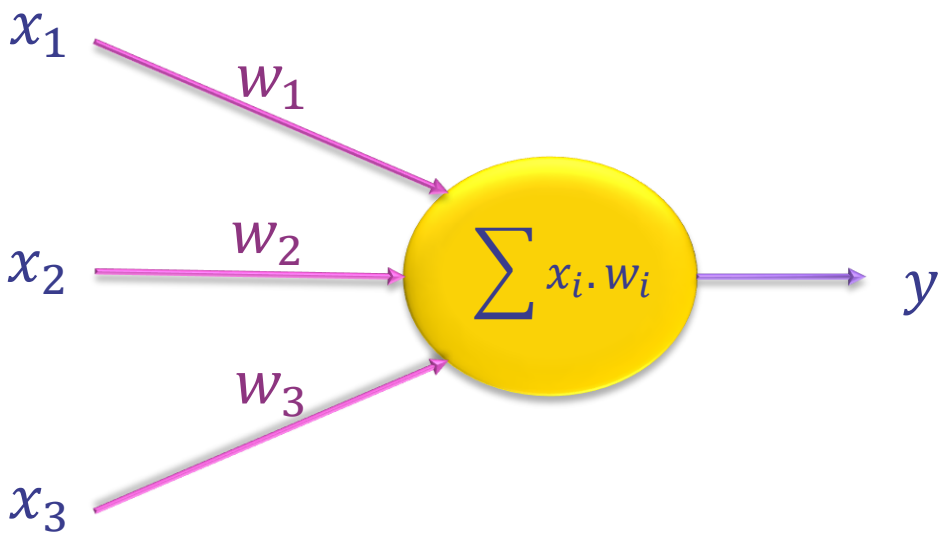
The Perceptron

1 Single Layer Perceptron

- SLP is the simplest type of ANN
- contains **no hidden layers**
- can only classify linearly separable classes (a straight line to separate classes)
- can only perform binary classification (2 classes)

The Perceptron

1 Single Layer Perceptron

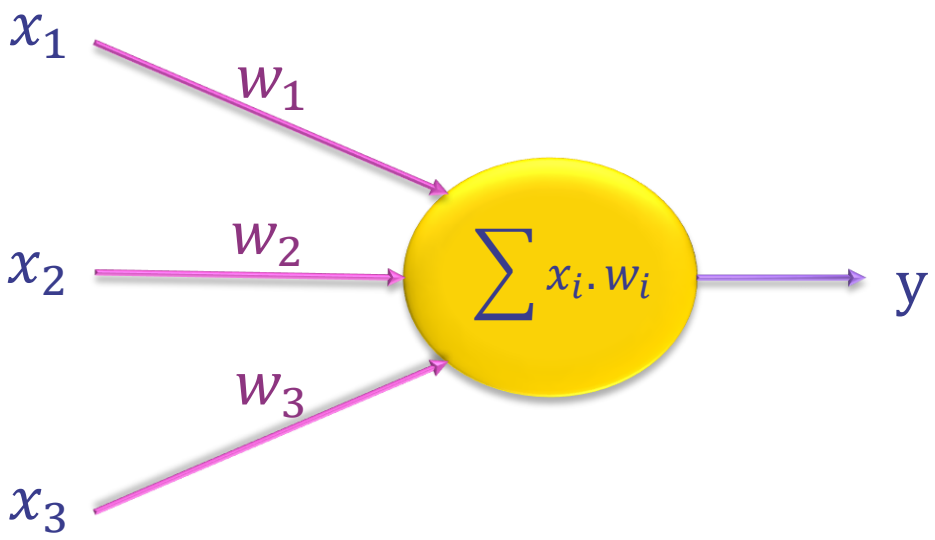


$$y = \begin{cases} 1 \; if \sum x_i . w_i \; > \; \theta \; (predetermined) \\ 0 \quad otherwise \end{cases}$$

The Perceptron

1

Single Layer Perceptron



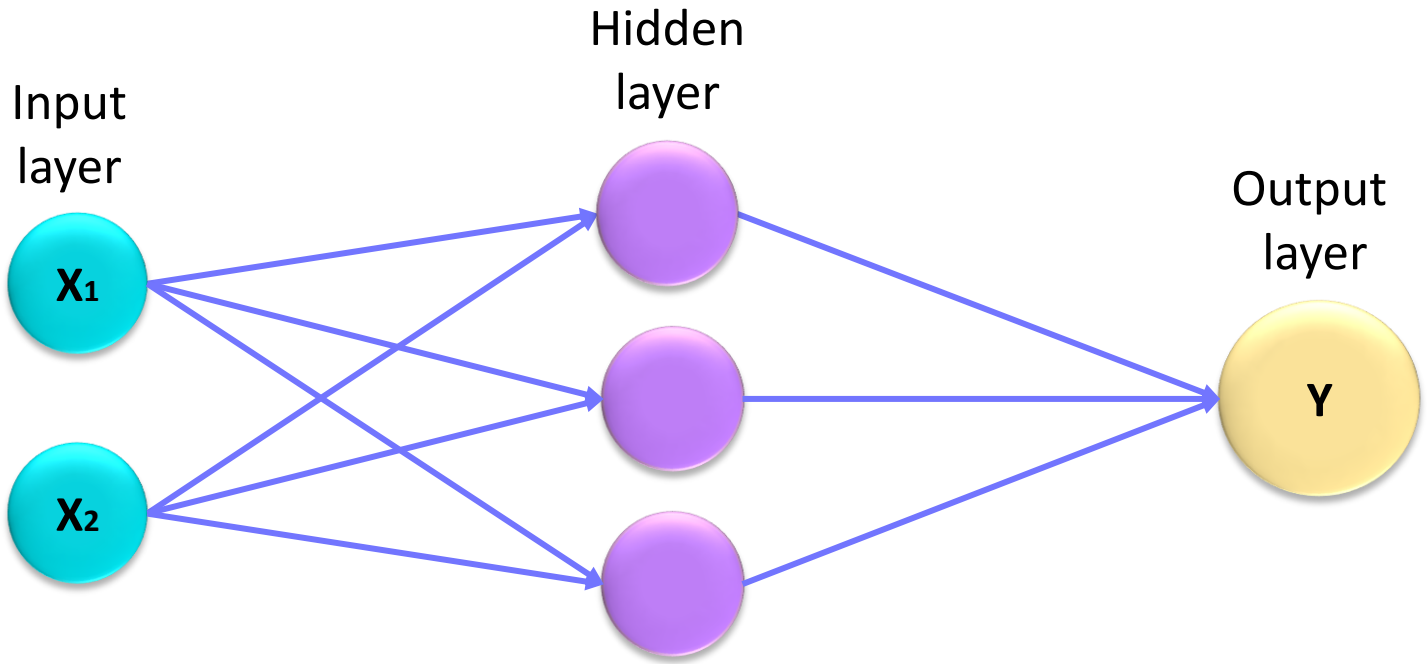
$$y = \begin{cases} 1 \; if \sum x_i . w_i \; > \; \theta \; (predetermined) \\ 0 \quad otherwise \end{cases}$$

- Has no hidden layers
- Each input has a corresponding weight to designate how important it is
- The output *y* provides a decision (one of two classes)

The Perceptron

2

Multi Layer Perceptron

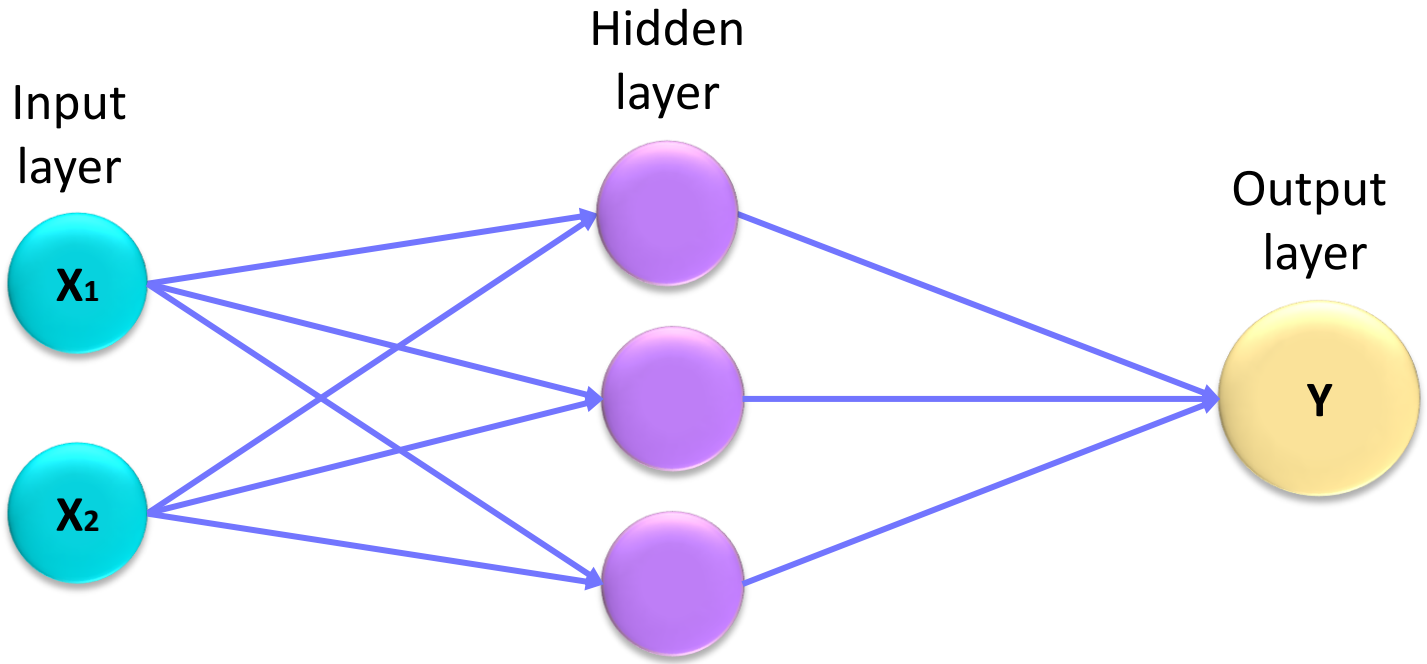


- Has **one or more** hidden layers
- Perceptrons team up with each other to solve complex problems
- Each layer can have multiple perceptrons, and there can be multiple layers

The Perceptron

2

Multi Layer Perceptron

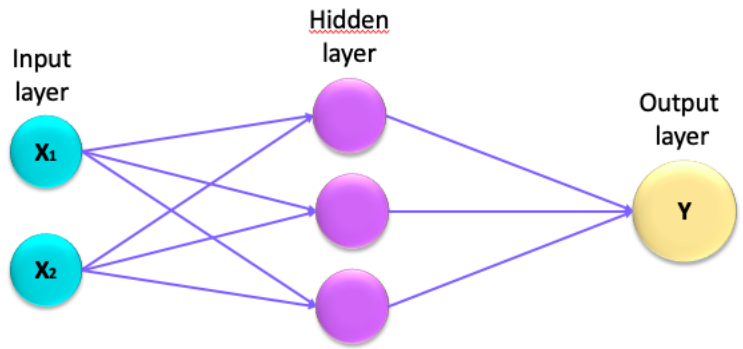


MLP is a class of feedforward ANNs

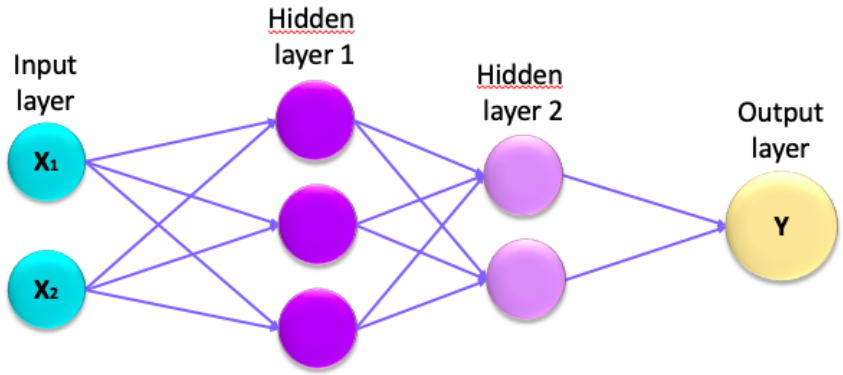
- Has one or more hidden layers
- Perceptrons team up with each other to solve complex problems (even non-linearly seperated ones)
- Each layer can have multiple perceptrons, and there can be multiple layers

The Perceptron

2 Multi Layer Perceptron



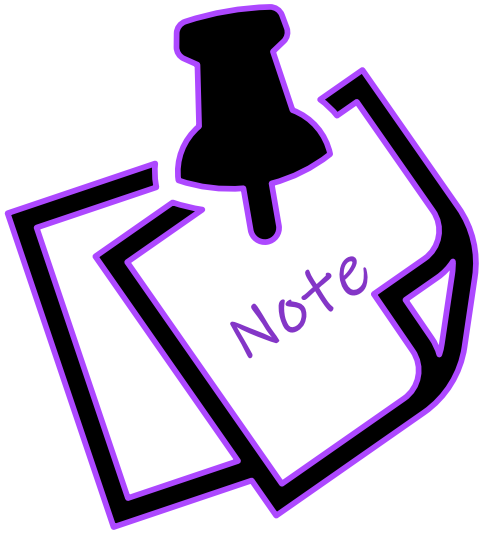
An ANN with one hidden layer is called a **shallow** or **non-deep neural network**



An ANN with more than one hidden layer is called **deep neural network**

The Perceptron

2 Multi Layer Perceptron



Difference between MLP and Neural Network :

MLP :

- Uses a step function for decision
- The decision is binary

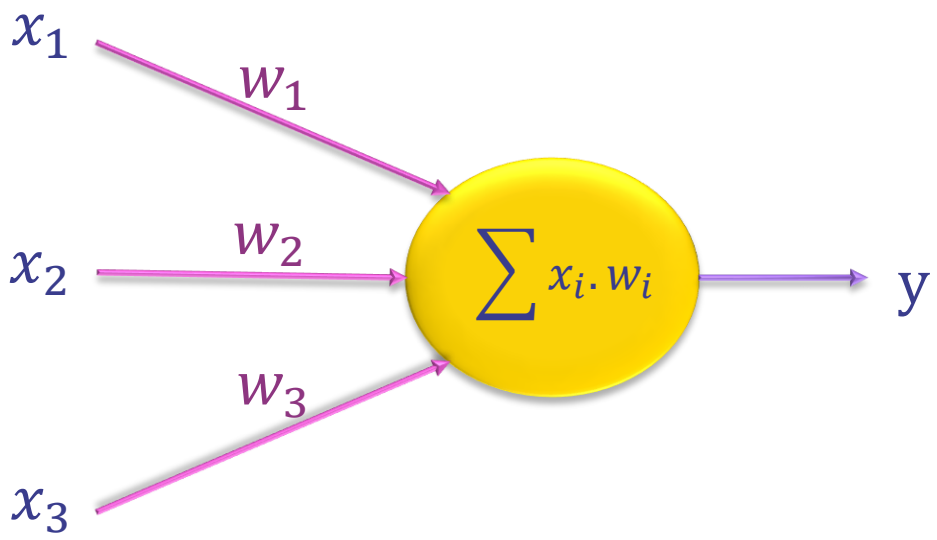
Neural Networks :

- Evolved from MLP
- Other activation functions can be used
- Outputs are real values, probability-based or classes

Weights & Biases

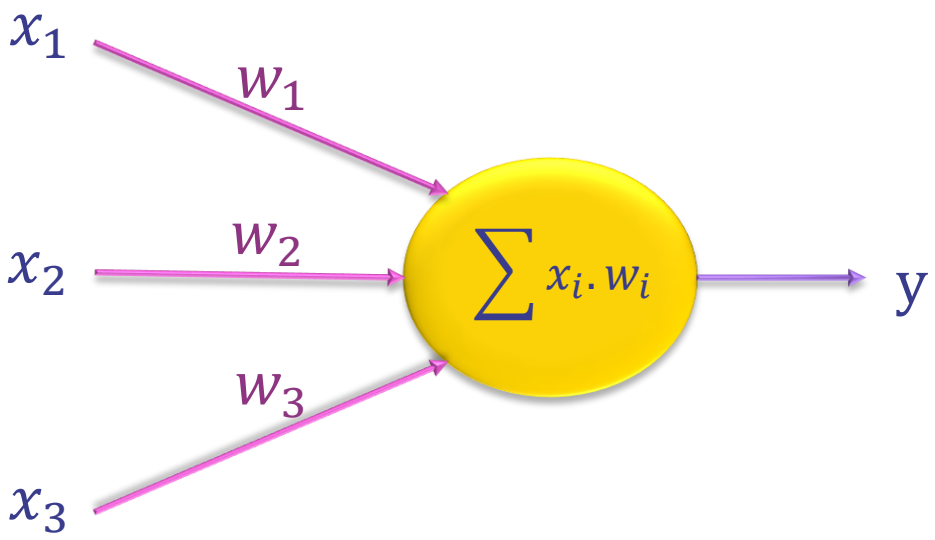
Weights and Biases

1 Weights



Weights and Biases

1 Weights

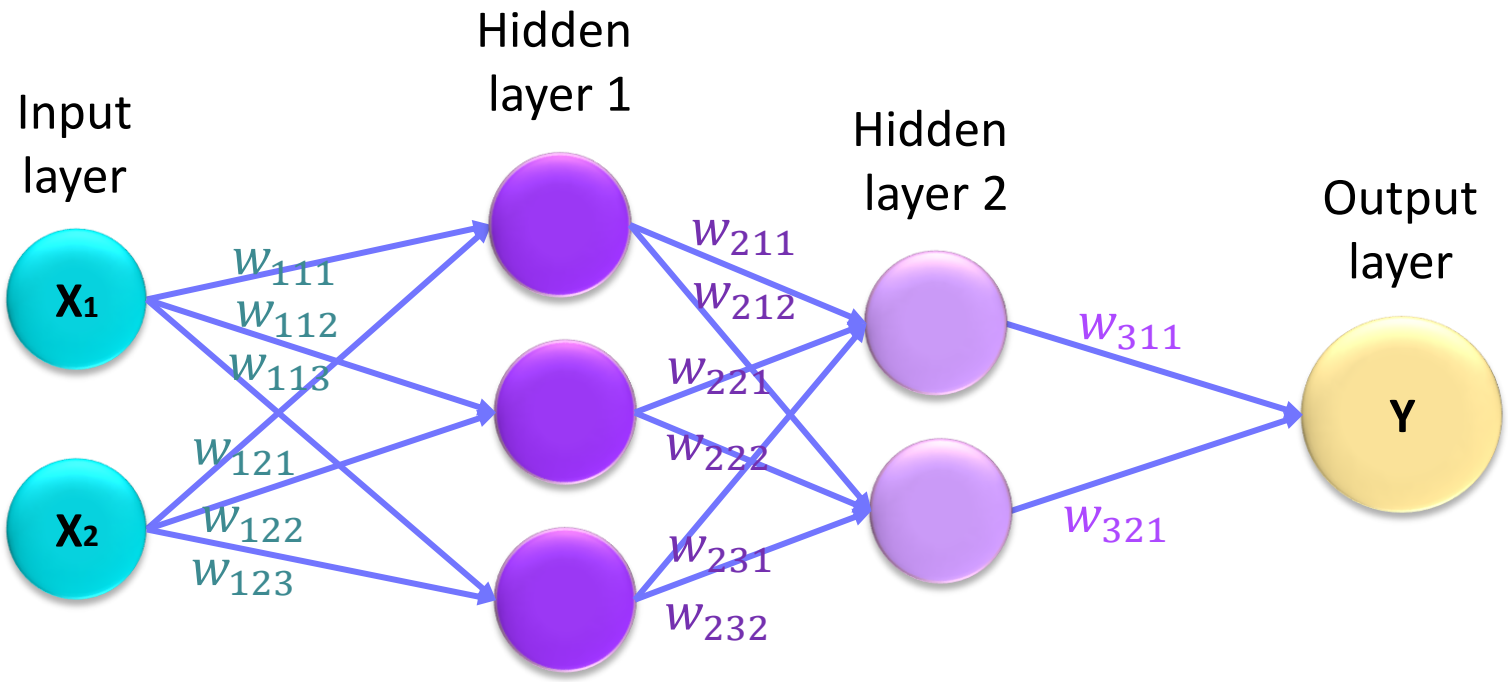


Weights determine the relative importance of every feature in the classification decision.

$$y = \sum x_i . w_i$$

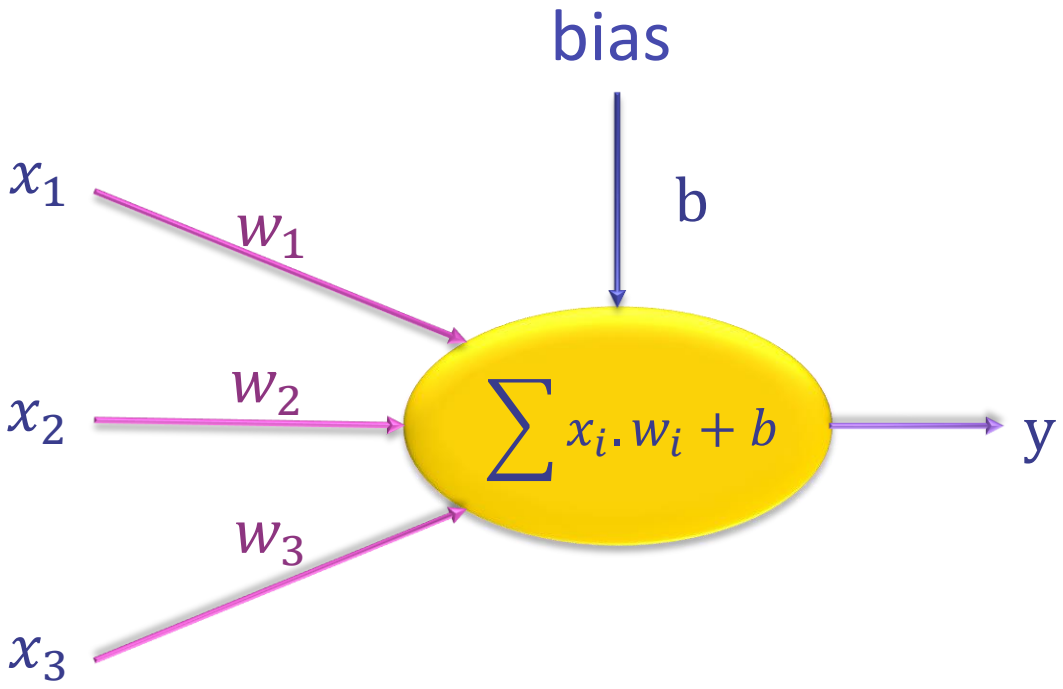
Weights and Biases

1 Weights



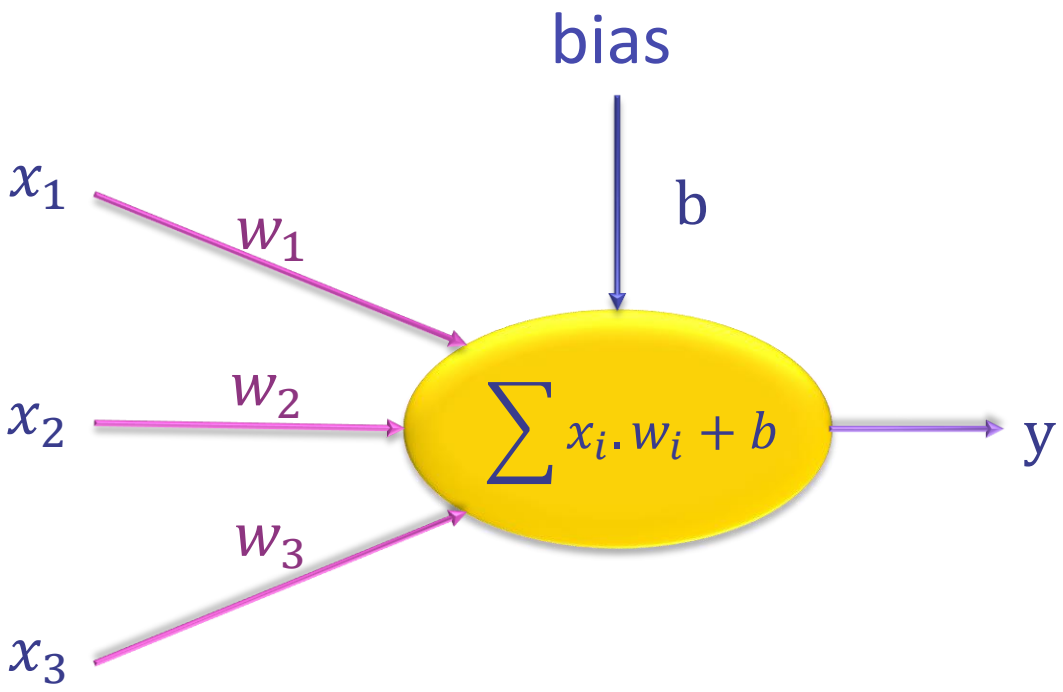
Weights and Biases

2 Biases



Weights and Biases

2 Biases



A bias b :

- is a constant that helps the model fit the given data best.
- behaves as the intercept of a linear equation.

$$y = \sum x_i . w_i + b$$

Weights and Biases

2 Biases

$$y = \sum x_i . w_i + b$$