Artificial neural network

Parcours Progis

Etudes, Medias, communication, Marketing

Bahareh Afshinpour

01.07.2026

Université Grenoble Alpes

Sciences Po Grenoble UGA

PROGIS
ETUDES, MEDIAS, COMMUNICATION, MARKETING

Suggested Reading

Perceptrons and MLPs and Back-Propagation training algorithm

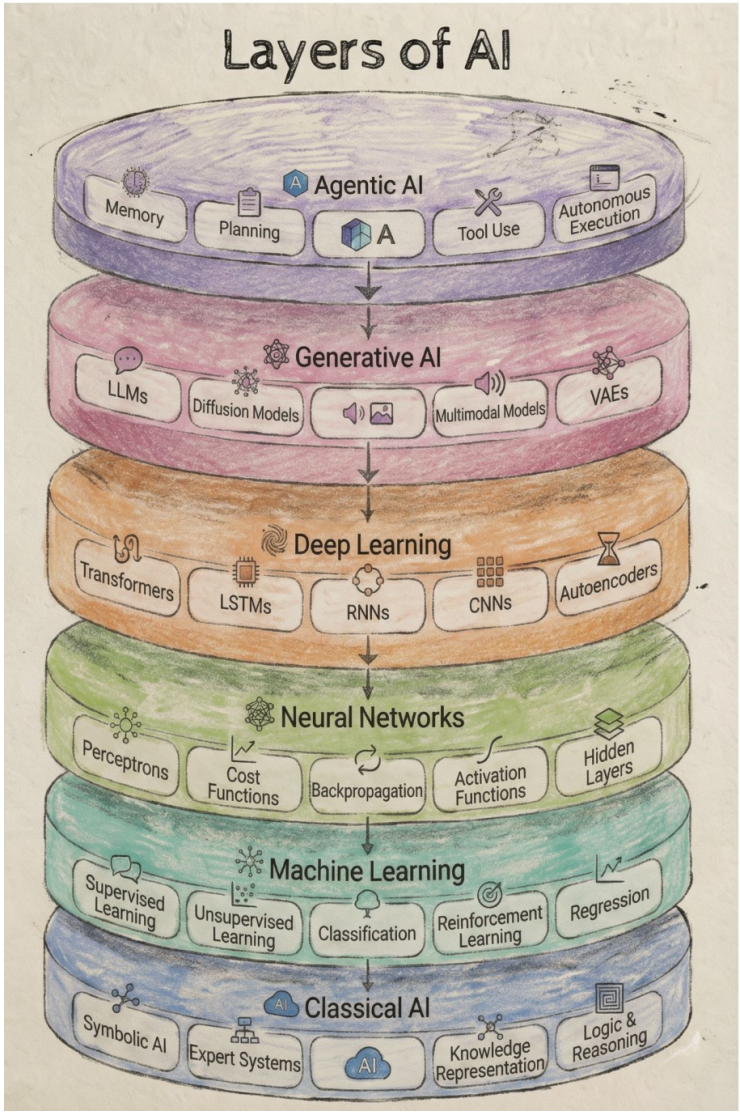
- Haykin: Sections 3.5 - 3.9 , 4.1, 4.2, 4.3, 4.4, 4.5, 4.6

- Bishop(NNs for PR): Sections 4.1, 4.8

- Hwang, Yoon. 2019. "Conducting Customer Analytics with Python." In Hands-On Data Science for Marketing. Chapter 11, 580–581

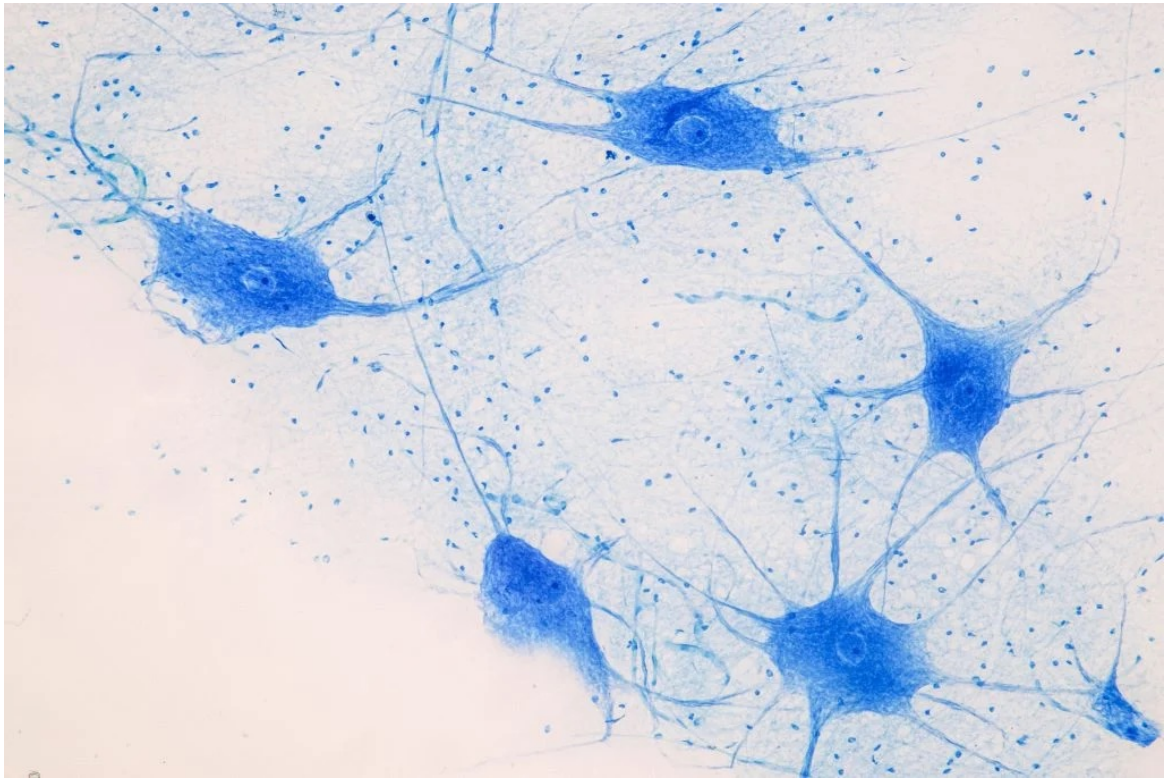
-https://www.youtube.com/watch?v=aircAruvnKk&list=PLZHQObOWTQDNU6R1_67000Dx_ZCJB-3pi&index=1

-https://xnought.github.io/backprop-explainer/


Layers of AI
Agentic AI
Memory
Planning
A
Tool Use
Autonomous Execution
Generative AI
LLMs
Diffusion Models
Multimodal Models
VAEs
Deep Learning
Transformers
LSTMs
RNNs
CNNs
Autoencoders
Neural Networks
Perceptrons
Cost Functions
Backpropagation
Activation Functions
Hidden Layers
Machine Learning
Supervised Learning
Unsupervised Learning
Classification
Reinforcement Learning
Regression
Classical AI
Symbolic AI
Expert Systems
AI
Knowledge Representation
Logic & Reasoning

PROGIS
ETUDES, MEDIAS, COMMUNICATION, MARKETING

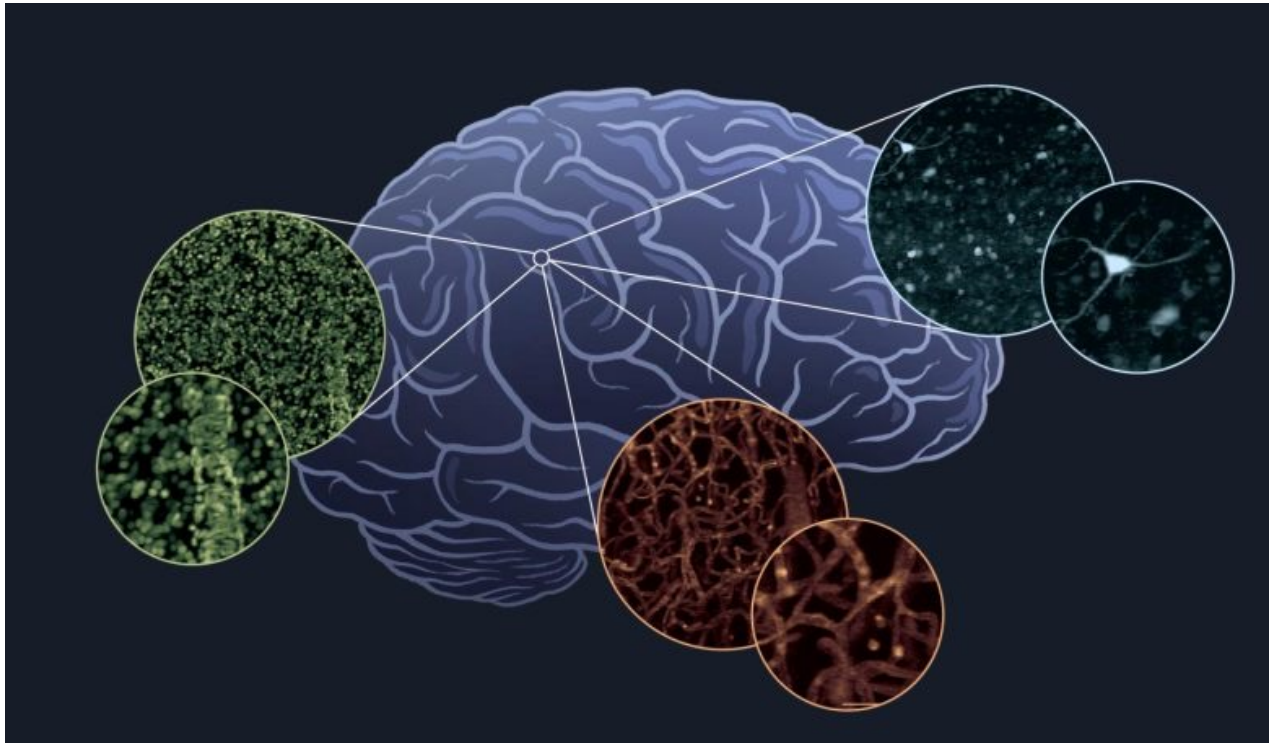
Neuron



PROGIS
ETUDES, MEDIAS, COMMUNICATION, MARKETING

Neuron

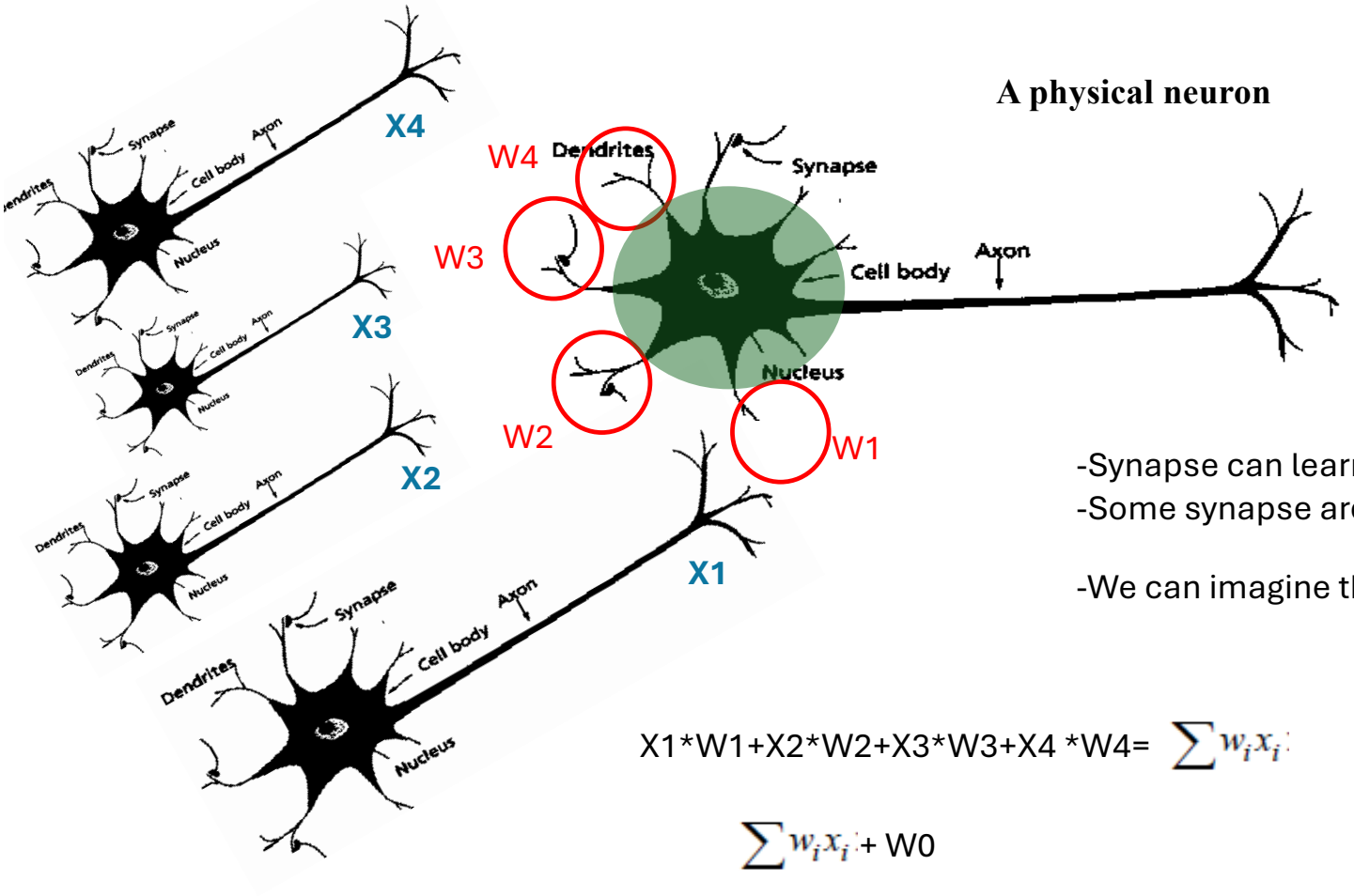
Artificial neural networks also have neurons that are linked to each other in various layers of the networks. These neurons are known as nodes.



PROGIS
ETUDES, MEDIAS, COMMUNICATION, MARKETING

Neuron By itself it's not that strong,but when you have lots of neurons together, they work together to do magic.

PROGIS
ETUDES, MEDIAS, COMMUNICATION, MARKETING



A physical neuron

-Synapse can learn during the time
-Some synapse are week

-We can imagine that we have **Wi** for each synapse

X1*W1+X2*W2+X3*W3+X4 *W4= $\sum w_i x_i$

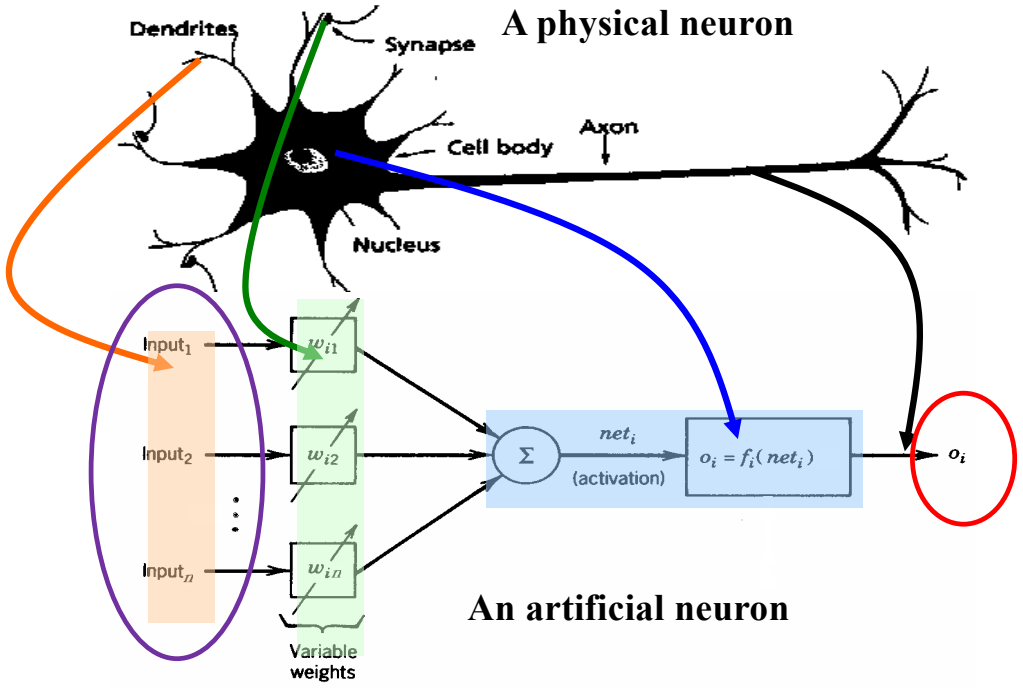
$\sum w_i x_i$ + W0

PROGIS
ETUDES, MEDIAS, COMMUNICATION, MARKETING

Neuron and Perceptron

- Inspiration from the brain
- we're going to model the brain
- Cognitive scientists and neuroscientists whose aim is to understand the functioning of the brain
- Models of the natural neural networks in the brain and make simulation studies.

How to recreate a neuron?

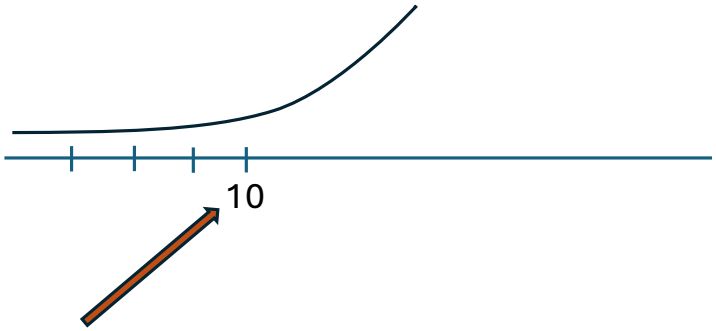


8

PROGIS
ETUDES, MEDIAS, COMMUNICATION, MARKETING

Activation

- For one specific moment, neuron starts to fire. (It start to create an electric pulse in real life.)
- If sum of the signals that go to the center, is less than for example 10, the neuron is off. After 10, it start to fire and have an output.

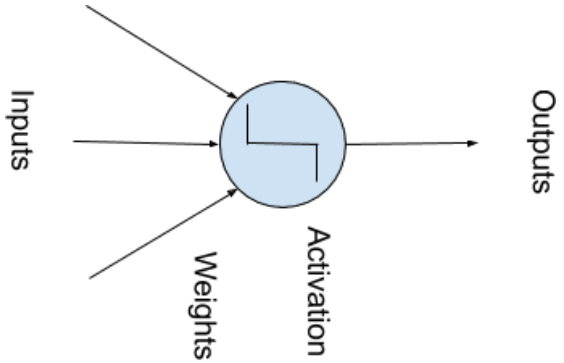


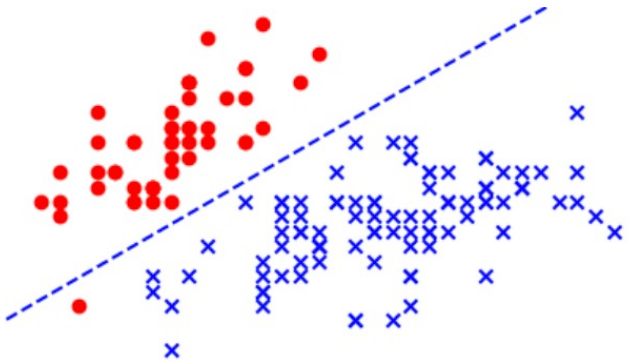
Fire
(activation function)

PROGIS
ETUDES, MEDIAS, COMMUNICATION, MARKETING

Perceptron

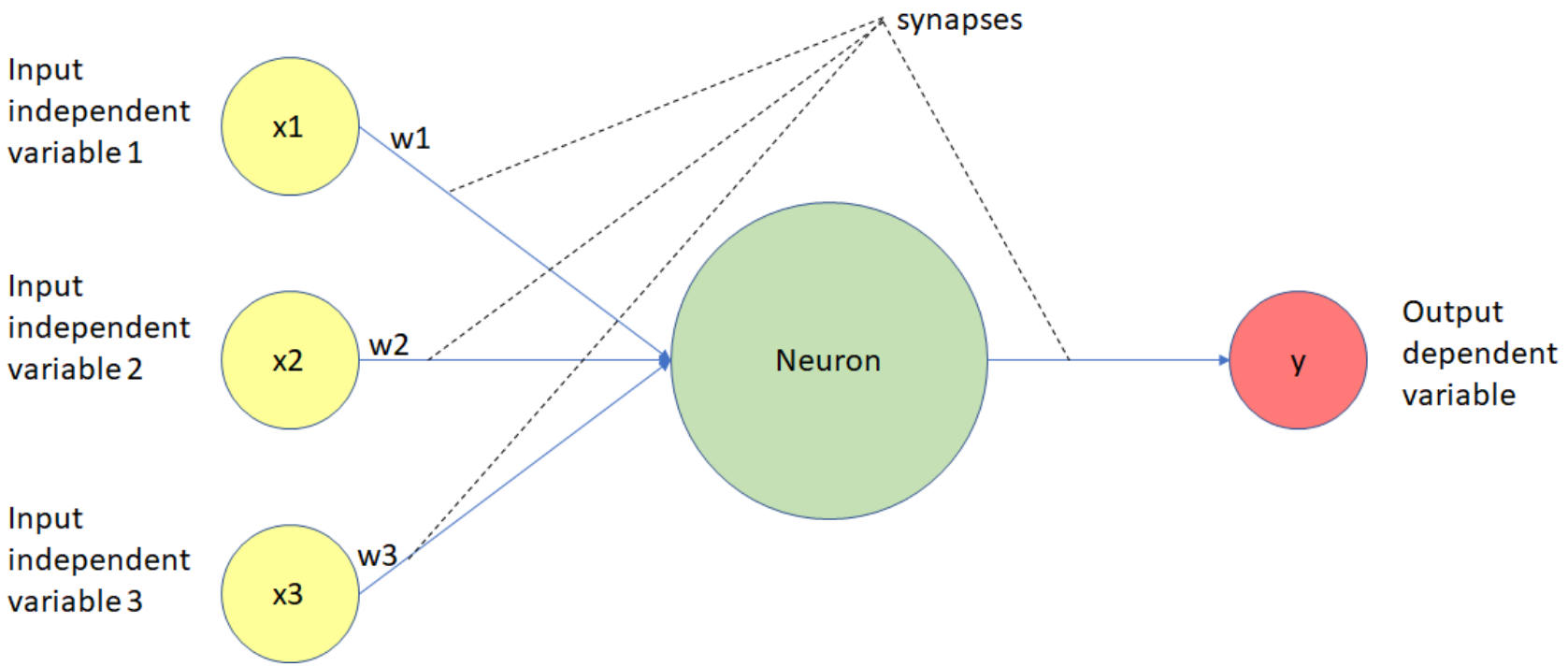
- The building blocks for neural networks are artificial neurons.
- These are simple computational units that have weighted input signals and produce an output signal using an activation function.
- A simple perceptron is used for the linear classification.





PROGIS
ETUDES, MEDIAS, COMMUNICATION, MARKETING

Perceptron



https://www.freecodecamp.org/news/deep-learning-neural-networks-explained-in-plain-english/

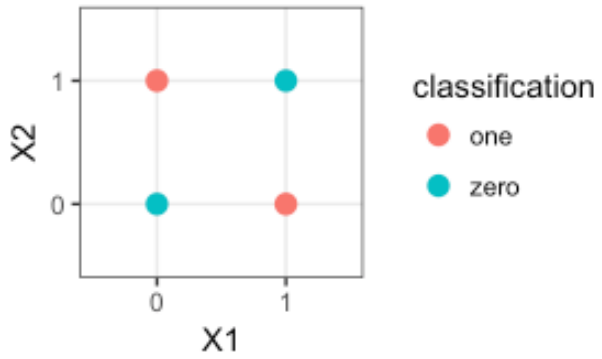
PROGIS
ETUDES, MEDIAS, COMMUNICATION, MARKETING

Historical Note and problem

There was great interest in Perceptrons in the '50s and '60s

- EXOR problem
  regions must be linearly separable.

you can not draw a straight line to separate the red point from the others.



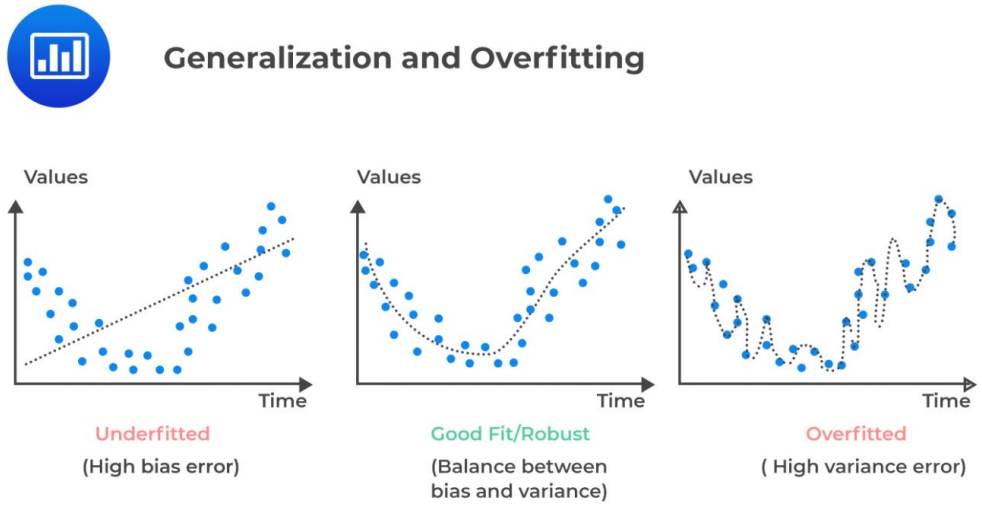
PROGIS
ETUDES, MEDIAS, COMMUNICATION, MARKETING

Why not just use regression or decision trees?

Linear Regression	Simple	Linear relationships only
Logistic Regression	Classification	Linear boundary
Decision Trees	Easy to interpret	Can overfit
Neural Networks	Learn complex patterns	Harder to interpret

Overfitting

Overfitting is a problem that arises when the machine learning algorithm fits the training data too well, making it unable to predict well using new data.



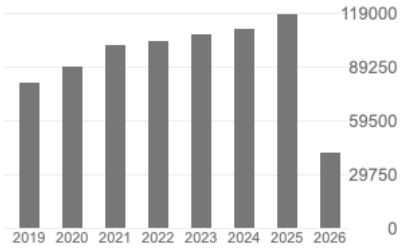
Example:
A model memorizes previous customers but performs badly on new customers.

PROGIS
ETUDES, MEDIAS, COMMUNICATION, MARKETING

Historical Note

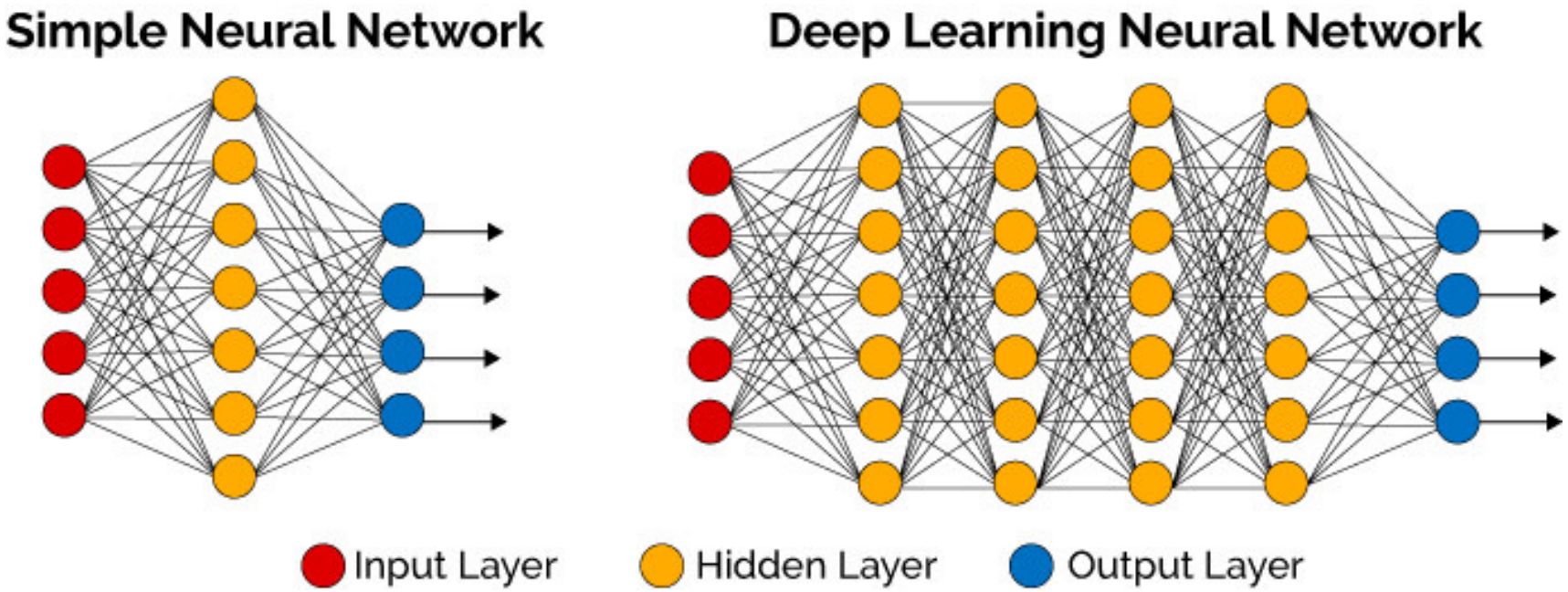
- In 2006, Geoffrey Hinton published a series of articles showing how they could efficiently train a neural network with **multiple layers**.

Cited by		VIEW ALL
	All	Since 2021
Citations	1045317	582649
h-index	190	130
i10-index	532	385



PROGIS
ETUDES, MEDIAS, COMMUNICATION, MARKETING

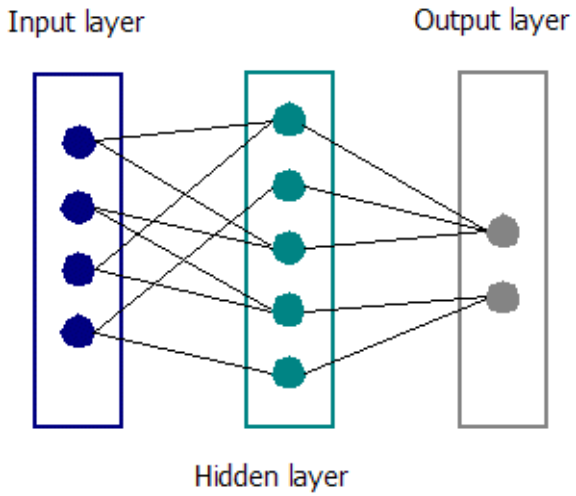
Multi Layer Neural Network (MLP)



https://thedatascientist.com/what-deep-learning-is-and-isnt/

PROGIS
ETUDES, MEDIAS, COMMUNICATION, MARKETING

Hidden layers



Here, the graphic shows a tri-layered neural network with a four-dimensional input layer and a two-dimensional output layer. The hidden layer is five-dimensional.

We separate it even further and we have not just one hidden layer, we have lots and lots and lots of hidden layers.

A deep neural network is simply a neural network with many layers.

PROGIS
ETUDES, MEDIAS, COMMUNICATION, MARKETING

- ANNs has evolved into a broad family of techniques that have advanced the state of the art across multiple domains.
- The simplest types have one or more static components, including number of units, number of layers, unit weights, and topology.
- **MLP model:**
  - As a neural network model that has at least one or more hidden layers of nodes, including one layer for the input and another layer for the output.
- **Deep learning:**
  - This is a type of ANN model where the number of layers between the input and the output layers is large.

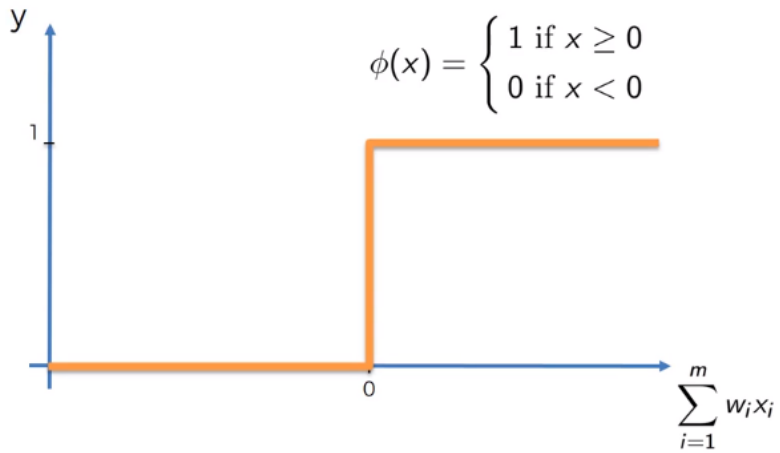
PROGIS
ETUDES, MEDIAS, COMMUNICATION, MARKETING

Activation Function

- An activation function is an internal state of a neuron that converts an input signal to an output signal.
- There are more different types of activation functions.

1- Threshold function:

- if the value is less than zero then the function is zero, but if the value is more than zero the function is one. It is basically a yes/no type of function.



https://www.andreaperlato.com/aipost/the-activation-function/

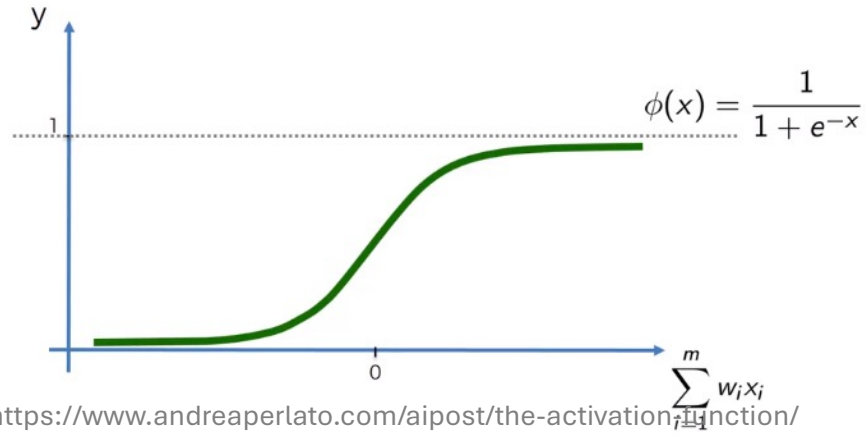
Without the activation function,
the output values from the neurons can range between
(- infinity) to (+infinity).

PROGIS
ETUDES, MEDIAS, COMMUNICATION, MARKETING

Activation Function

2- Sigmoid Function

- The value x in its formula is the value of the sum of the weighed.
- It is very similar to the logistic regression.
- It has a gradual progression and anything below zero is just zero, and above zero it approximates to one.
- In fact, outputs of sigmoid functions are interpretable as **probabilities**. (digit between0 and 1)
- It is especially useful in the output layer, especially when we are trying to predict probabilities.



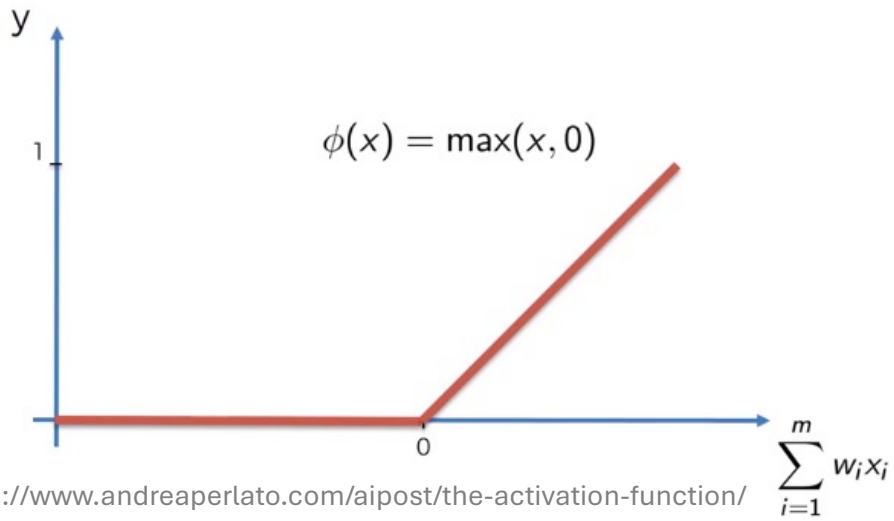
https://www.andreaperlato.com/aipost/the-activation-function/

What is the probability that the input data, belongs to the for example the first class?

PROGIS
ETUDES, MEDIAS, COMMUNICATION, MARKETING

Activation Function

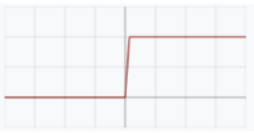
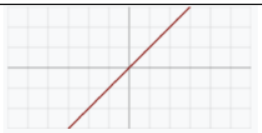
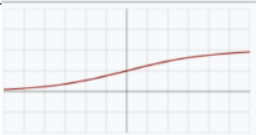
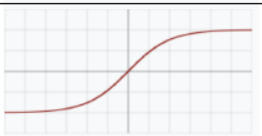
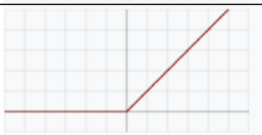
- **3-Rectifier Function**
  - is one of the most popular functions for ANN, so when it goes all the way to zero it is zero, and from there it is **gradually progresses** as the input value increase as well.
  - It is used in almost all the convolutional neural networks(CNN) or deep learning.



https://www.andreaperlato.com/aipost/the-activation-function/

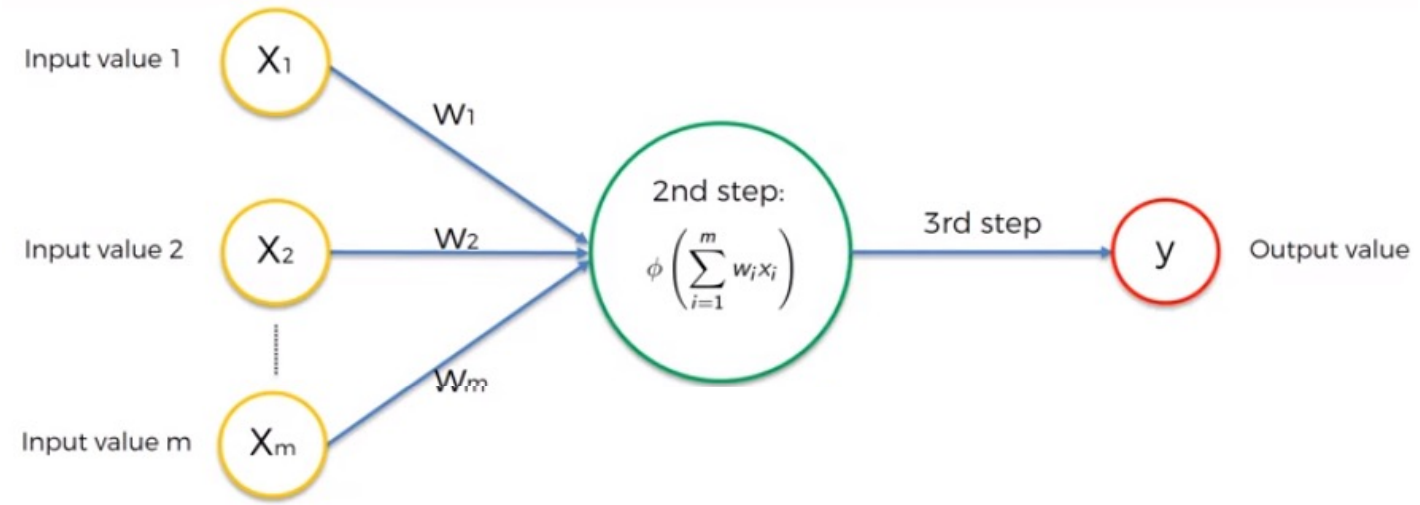
PROGIS
ETUDES, MEDIAS, COMMUNICATION, MARKETING

Activation Function (in French)

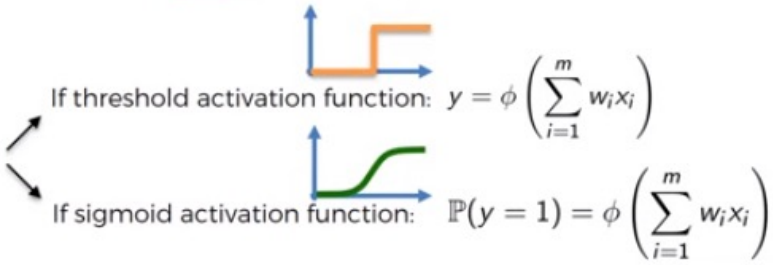
Nom	Définition	Graphe
Heaviside	$\bar{H}(z) = \begin{cases} 1, & \text{si } z \geqslant 0 \\ 0, & \text{sinon.} \end{cases}$	
Linéaire	$\bar{H}(z) = z$	
Sigmoïde	$\bar{H}(z) = \frac{1}{1+e^{-z}}$	
Tanh	$\bar{H}(z) = \frac{e^{z}-e^{-z}}{e^{z}+e^{-z}}$	
Relu	$\bar{H}(z) = \begin{cases} z, & \text{si } z \geqslant 0 \\ 0, & \text{sinon.} \end{cases}$	

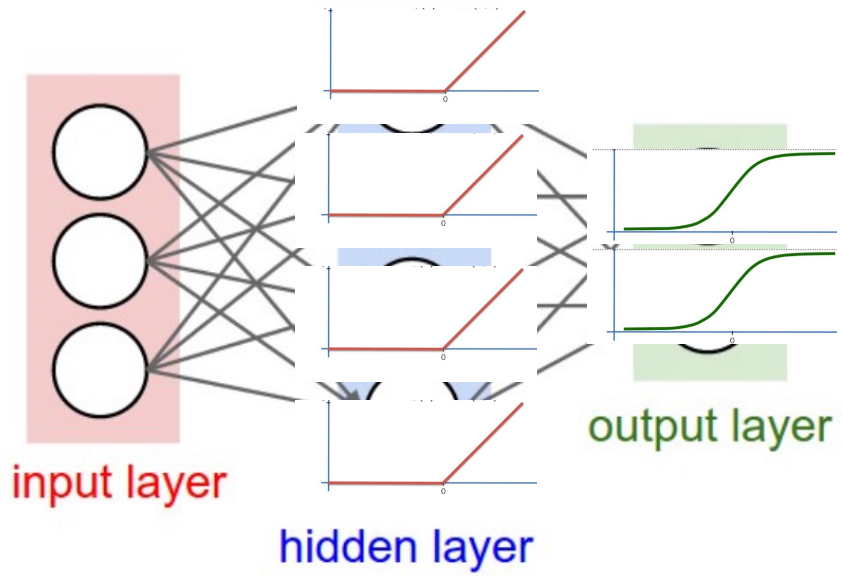
PROGIS
ETUDES, MEDIAS, COMMUNICATION, MARKETING

Question



Assume we only need two labels at the end (0 and 1). Which activation function should we use?





What is commonly used is to apply the **rectifier** activation function for the hidden layer and then the signals are passed on to the output layer where the **sigmoid** activation function is used, and that will be the final output that predict the probability.

PROGIS
ETUDES, MEDIAS, COMMUNICATION, MARKETING

Example to comprehend activation functions

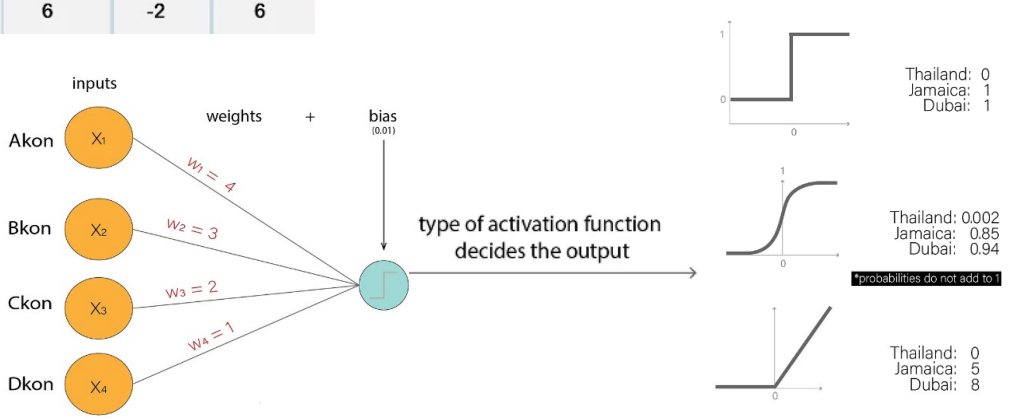
- Four friends are choosing to take a vacation together.
- Their list of possibilities has been reduced to three place. Jamaica, Dubai, and Thailand.
- They individually choose to vote for the places on a scale of -10 to 10 in order to choose the final destination.

	Akon	Bkon	Ckon	Dkon
Thailand	-9	10	3	-2
Jamaica	-4	5	-1	8
Dubai	-3	6	-2	6

$$Y = W_1X_1 + W_2X_2 + W_3X_3 + W_4X_4 + C \text{ (bias)}$$

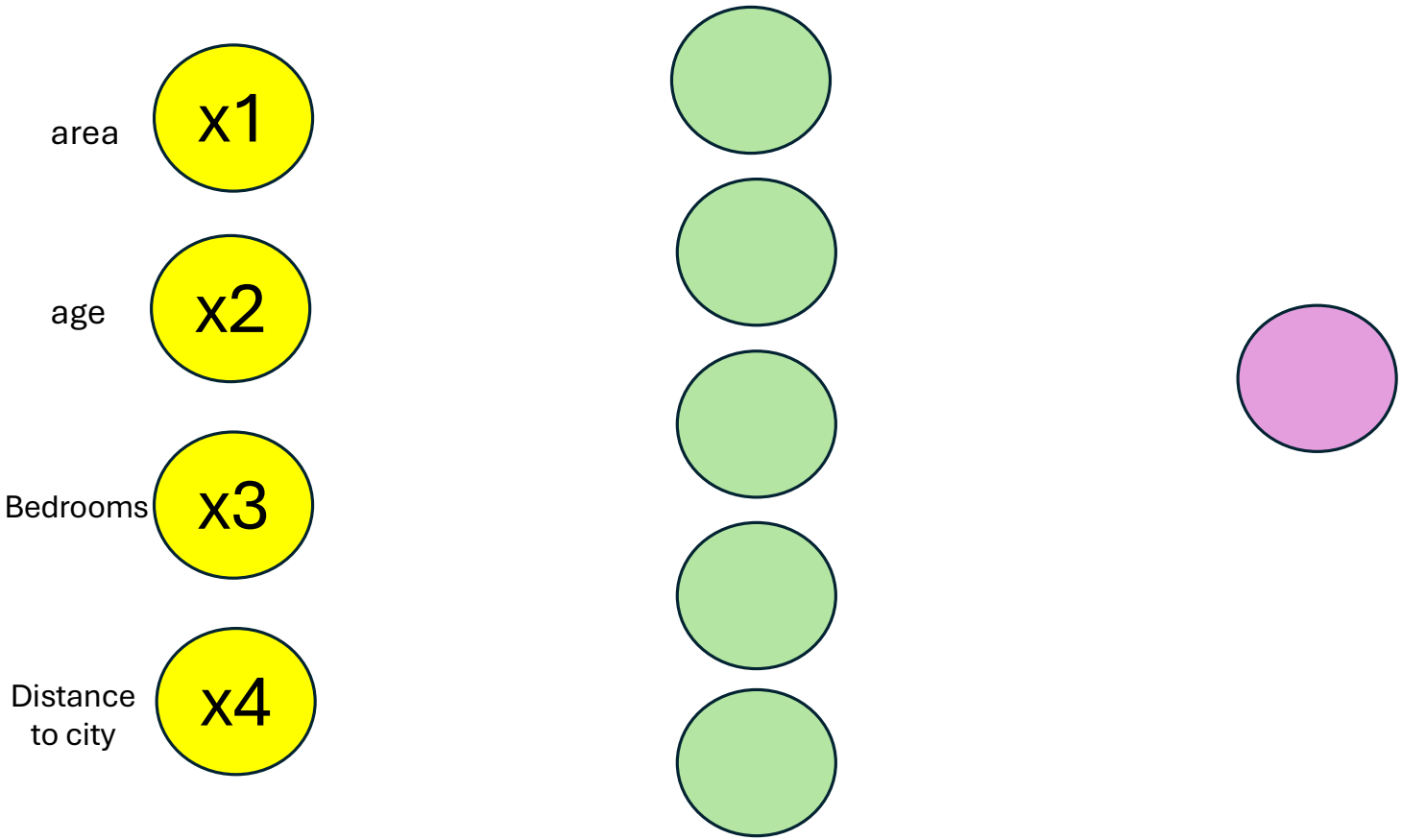
Thailand	(4x-9) + (3x10) + (2x3) + (1x-2)	=	- 2
Jamaica	(4x-4) + (3x5) + (2x-1) + (1x8)	=	5
Dubai	(4x-3) + (3x6) + (2x-2) + (1x6)	=	8

- assign initial random weights to each variable.
- The neuron calculates the weighted sum of its inputs and provides and output value.



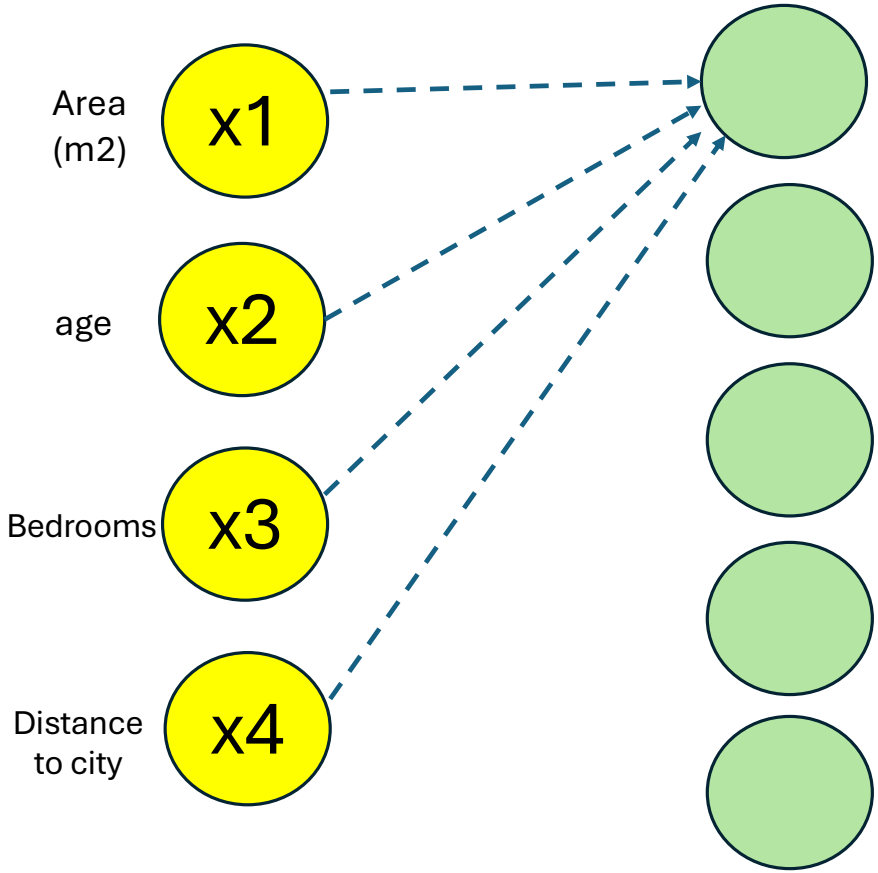
PROGIS
ETUDES, MEDIAS, COMMUNICATION, MARKETING

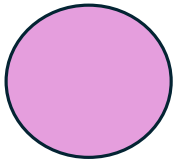
How do NNs work?



PROGIS
ETUDES, MEDIAS, COMMUNICATION, MARKETING

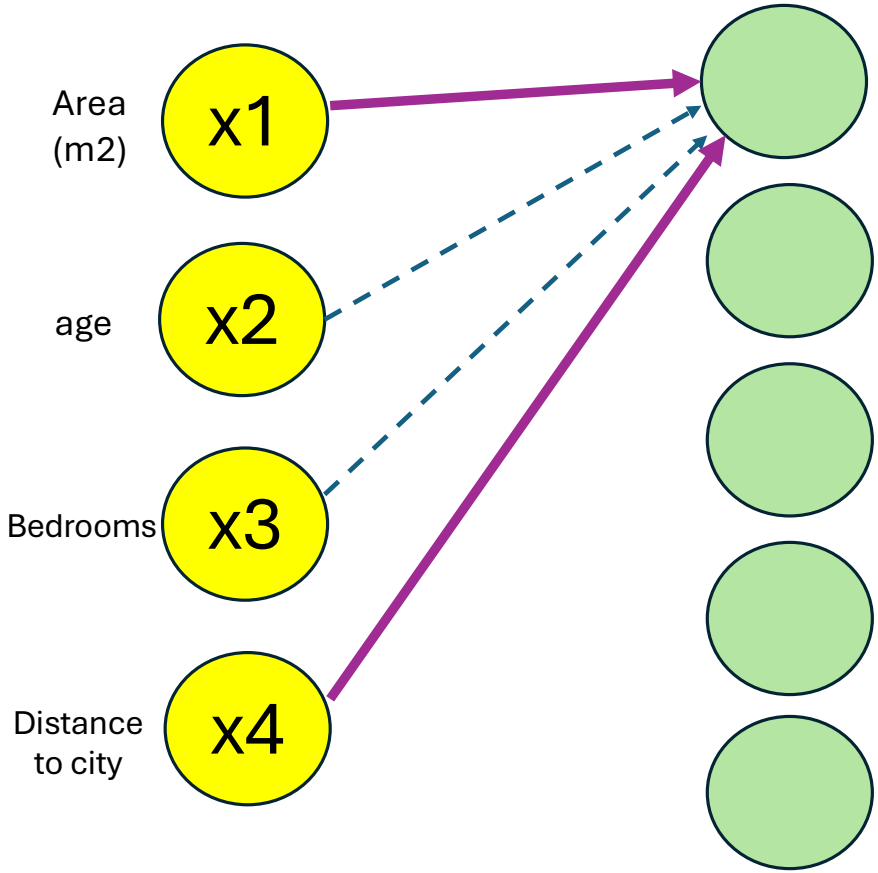
How do NNs work?

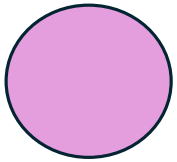




PROGIS
ETUDES, MEDIAS, COMMUNICATION, MARKETING

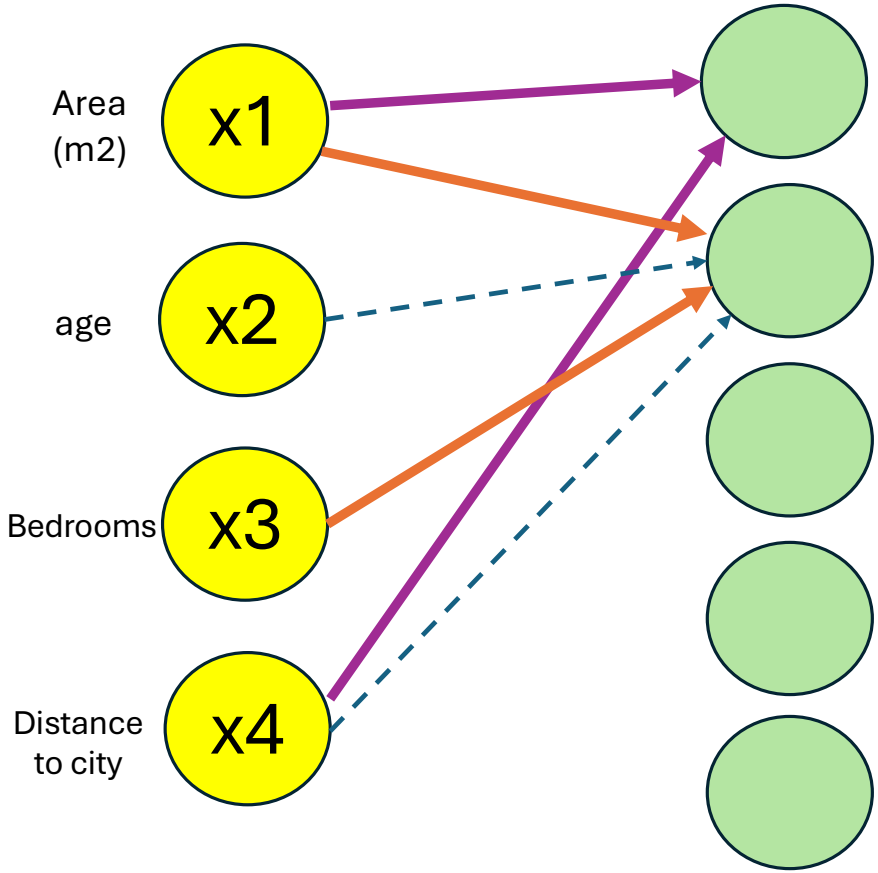
How do NNs work?

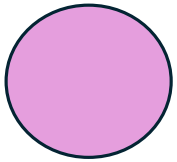




PROGIS
ETUDES, MEDIAS, COMMUNICATION, MARKETING

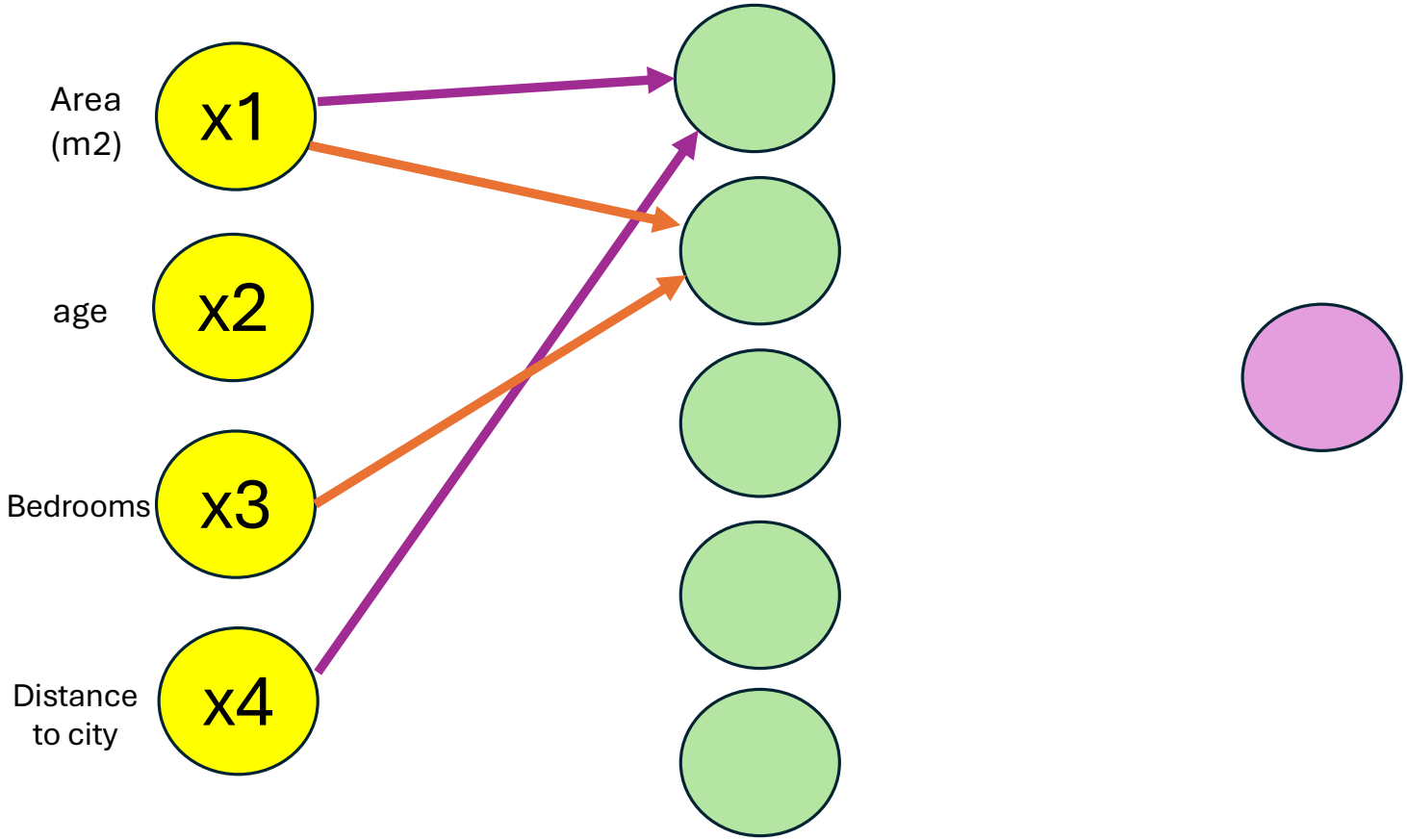
How do NNs work?





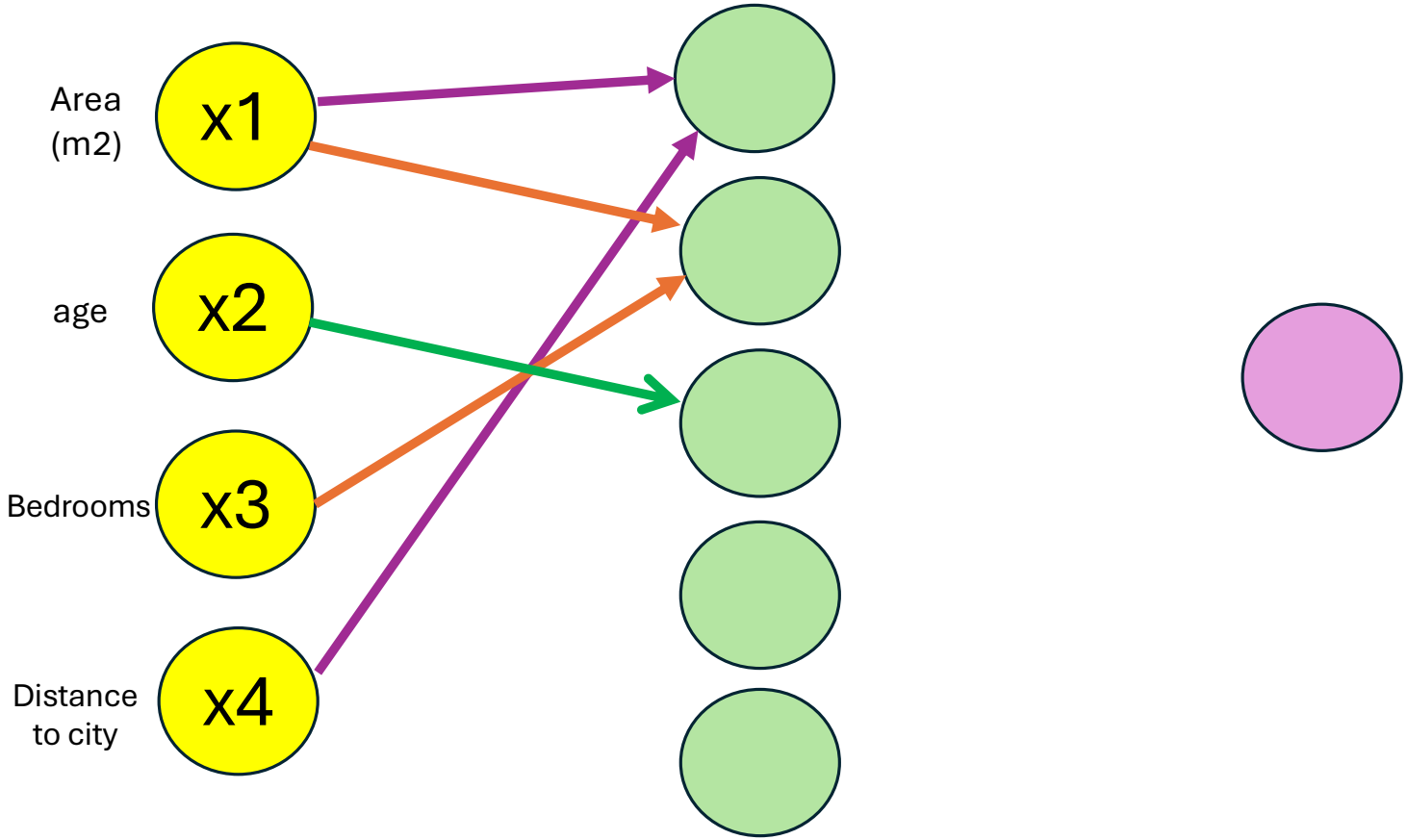
PROGIS
ETUDES, MEDIAS, COMMUNICATION, MARKETING

How do NNs work?



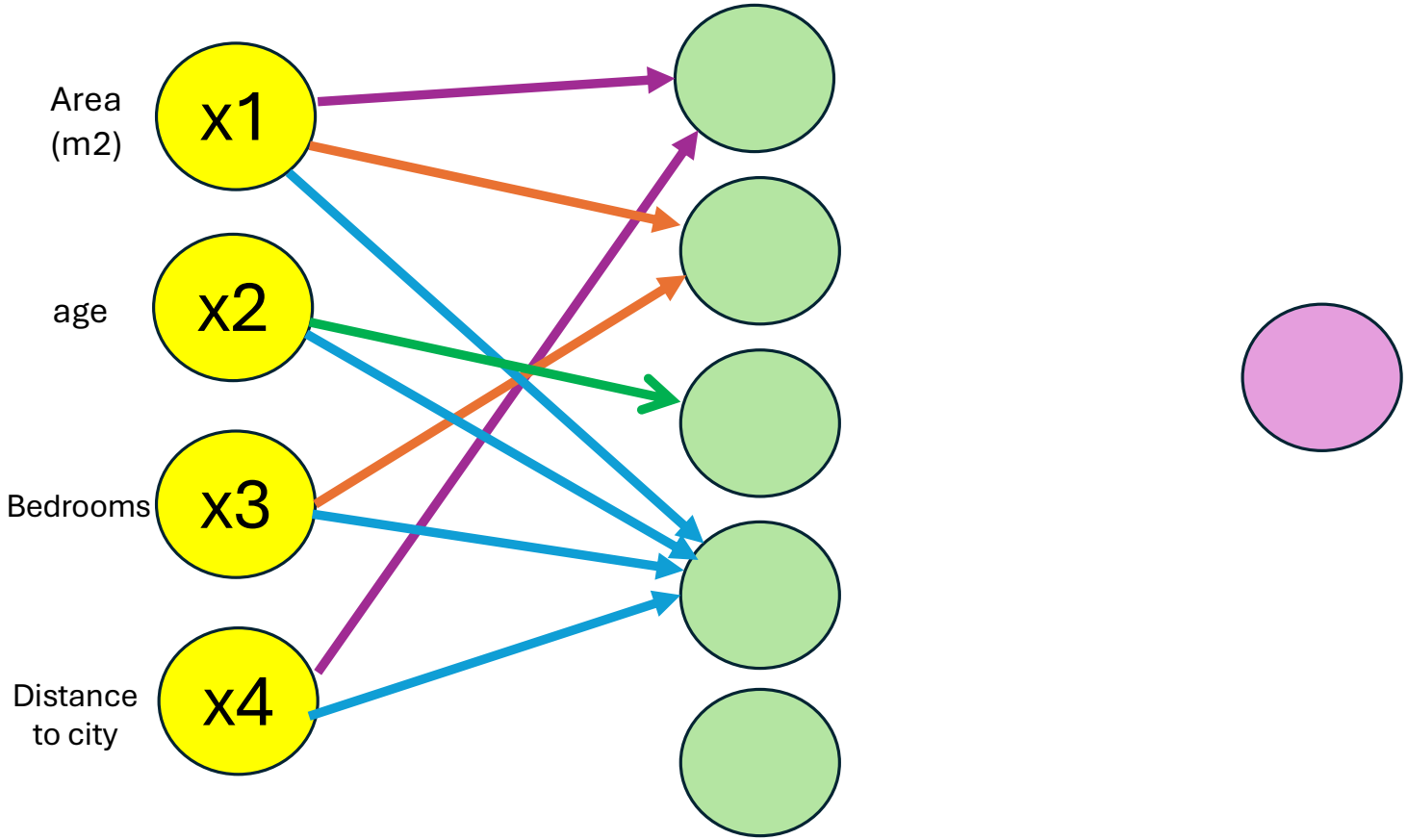
PROGIS
ETUDES, MEDIAS, COMMUNICATION, MARKETING

How do NNs work?



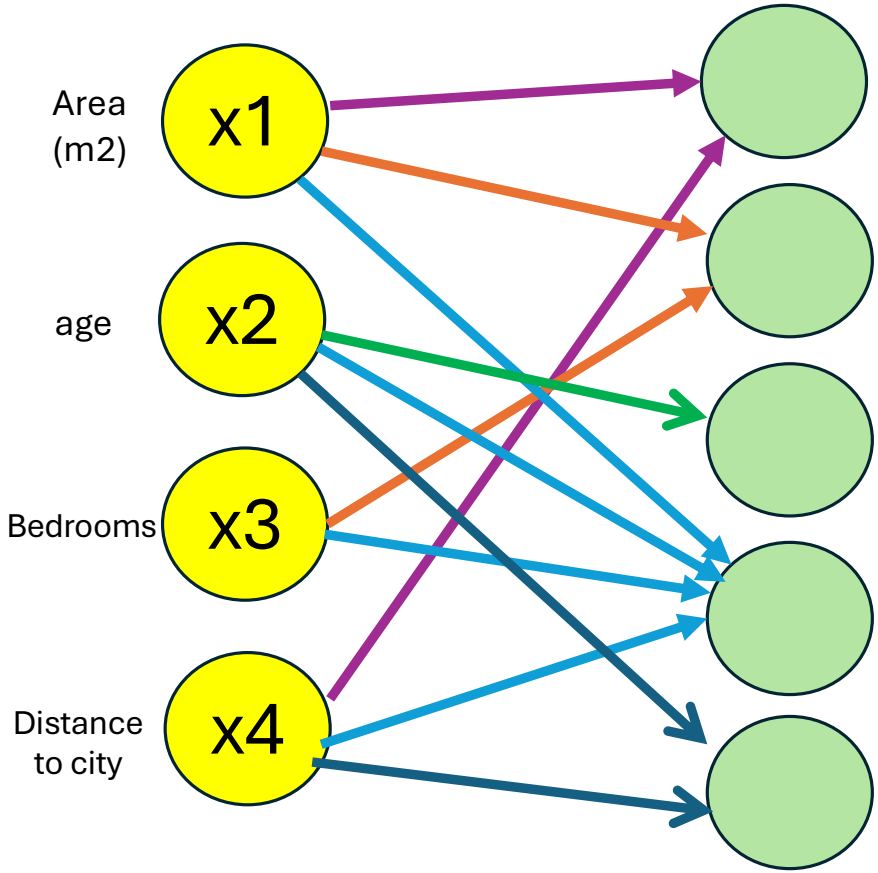
PROGIS
ETUDES, MEDIAS, COMMUNICATION, MARKETING

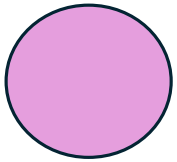
How do NNs work?



PROGIS
ETUDES, MEDIAS, COMMUNICATION, MARKETING

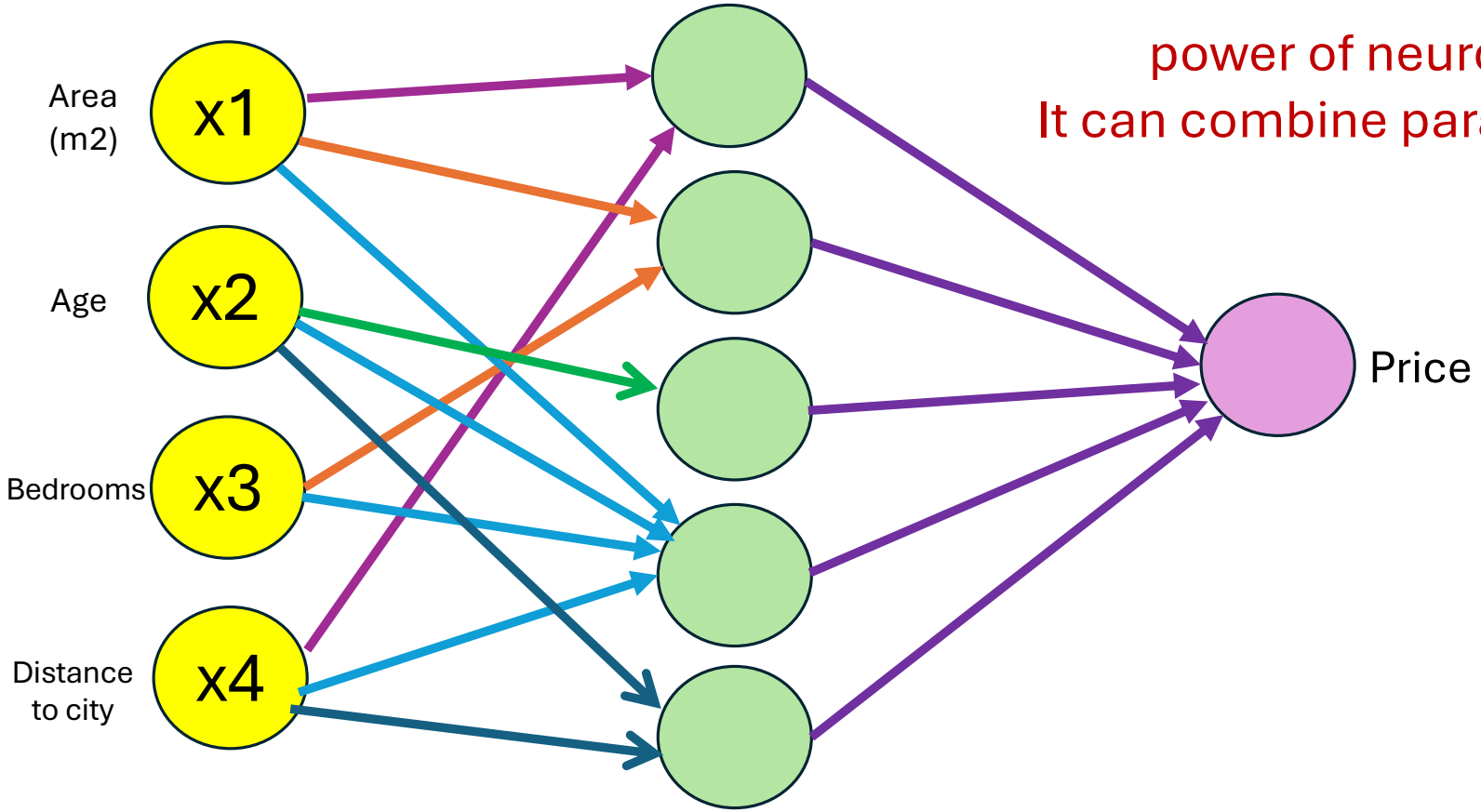
How do NNs work?





PROGIS
ETUDES, MEDIAS, COMMUNICATION, MARKETING

How do NNs work?

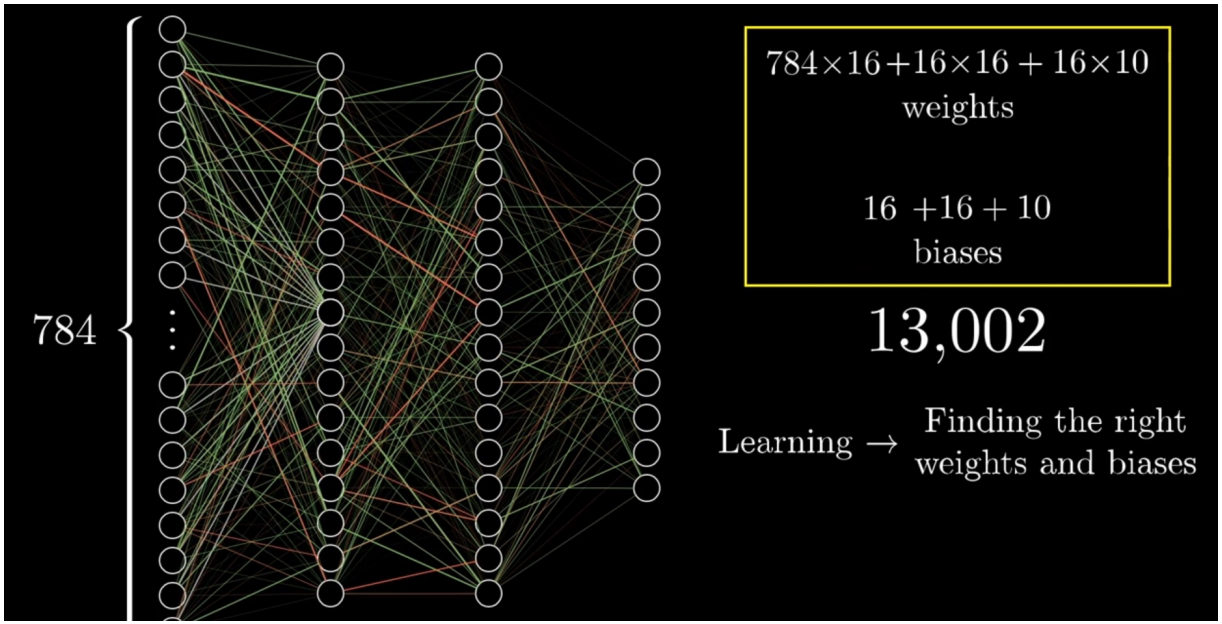


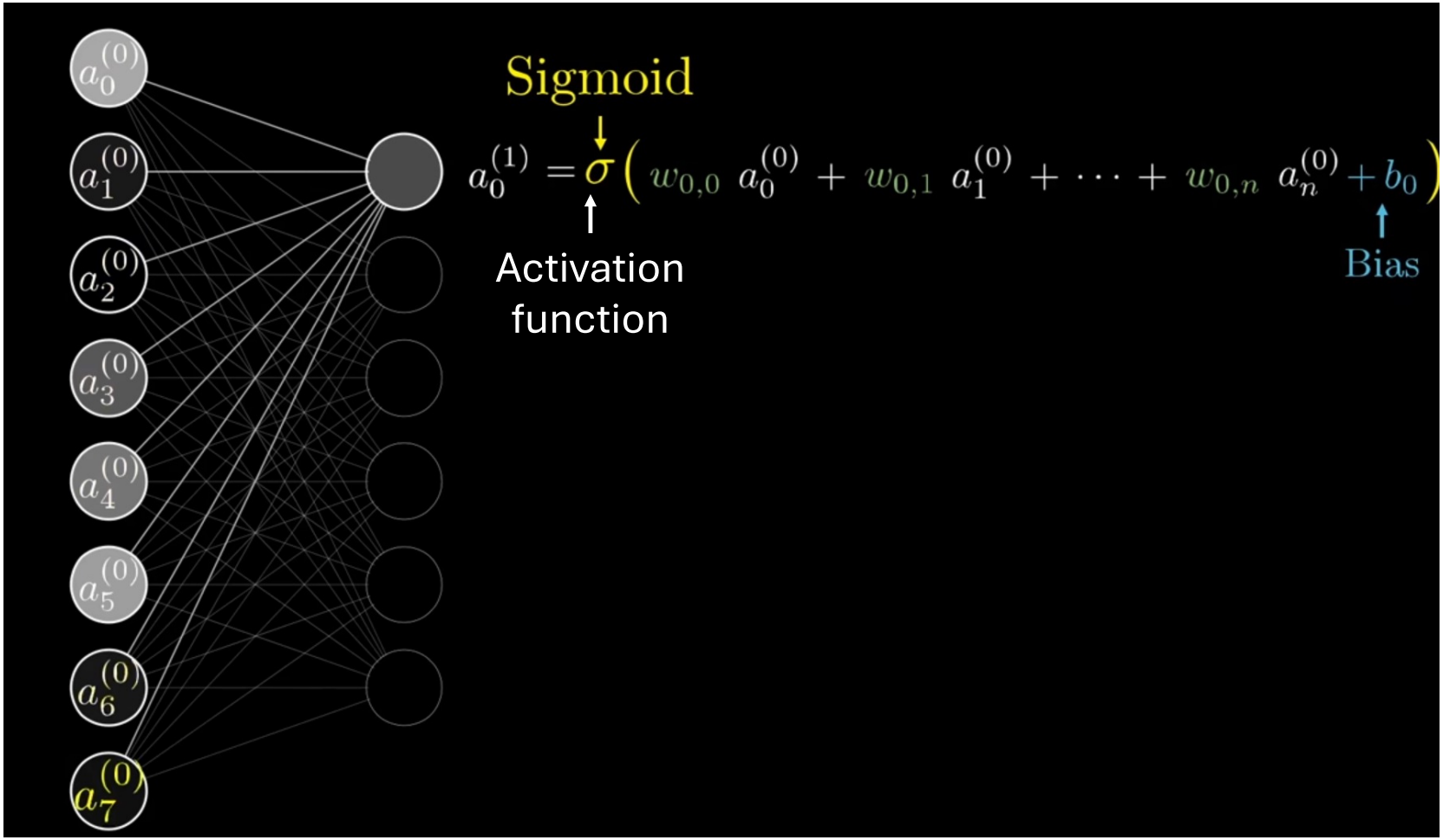
power of neuron .
It can combine parameters

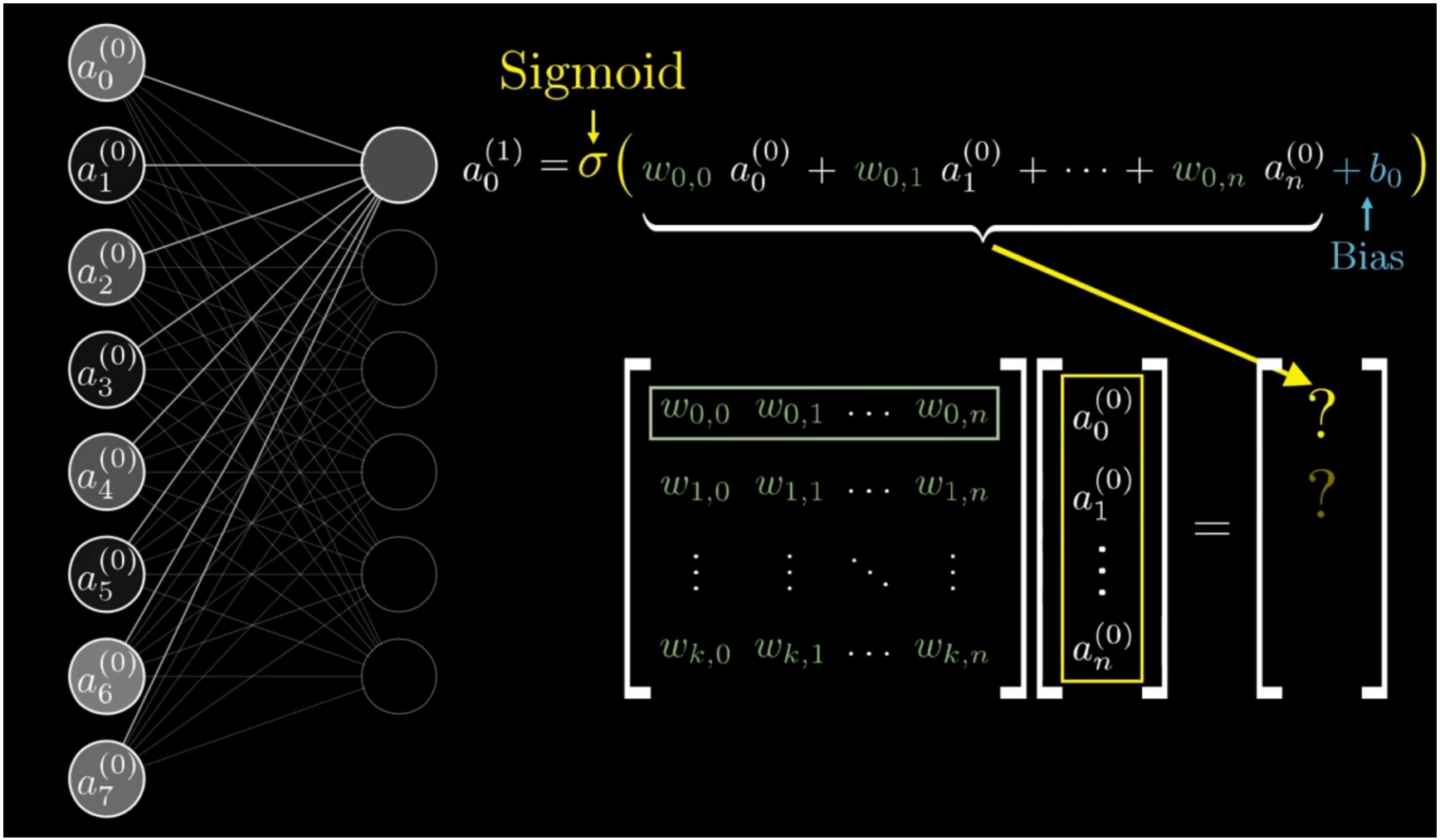
PROGIS
ETUDES, MEDIAS, COMMUNICATION, MARKETING

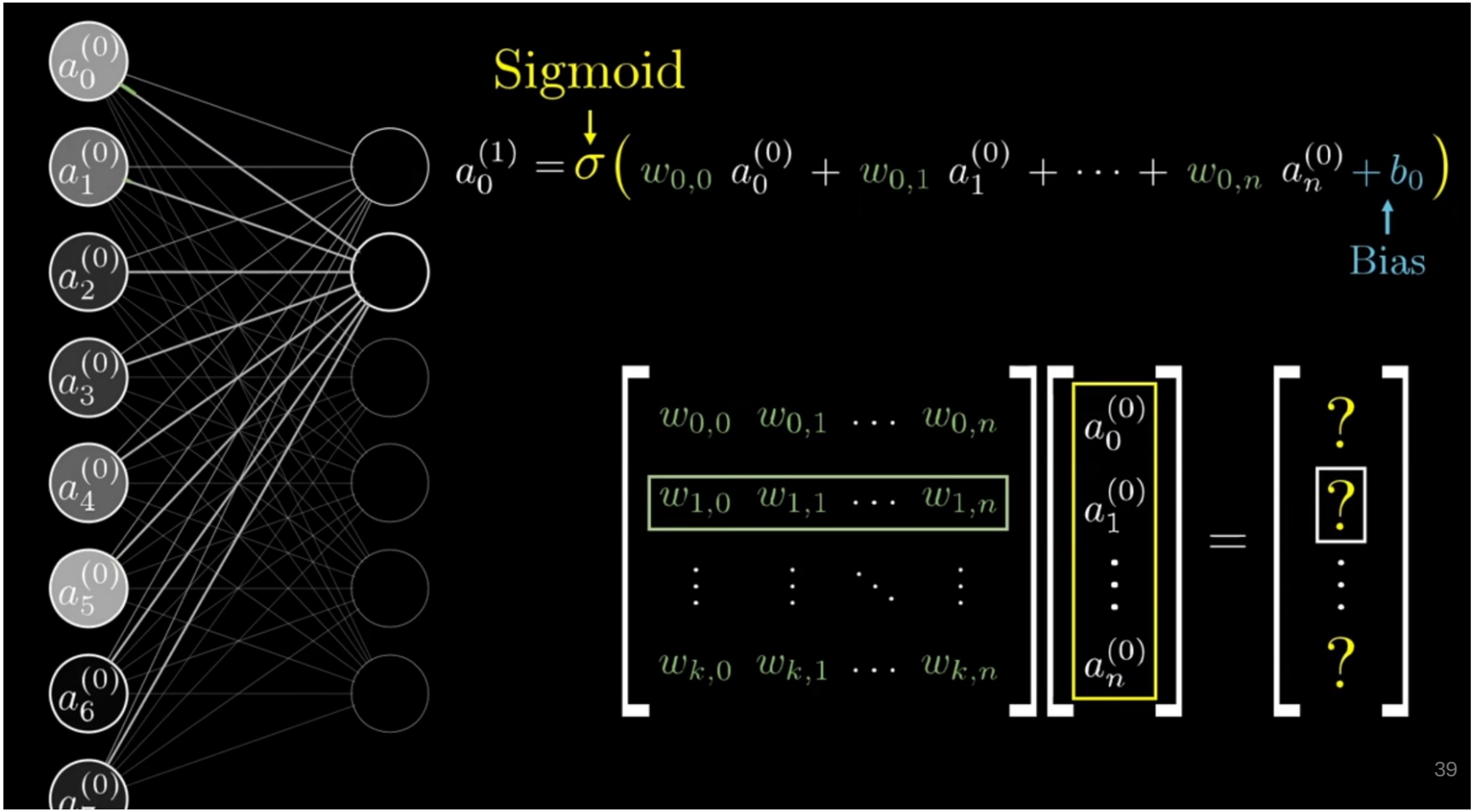
Bias

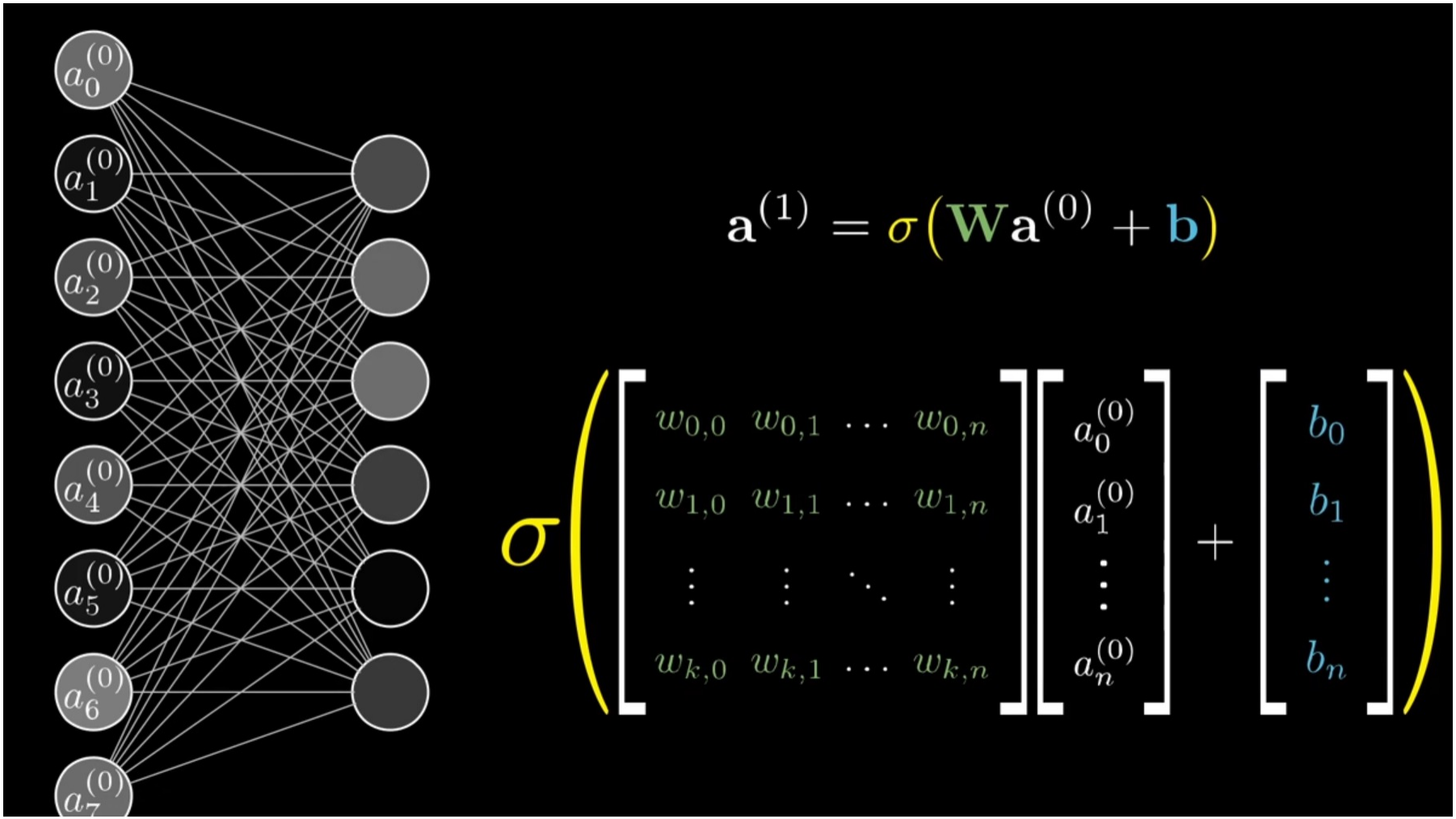
- How high the weighted sum needs to be, before the neuron starts getting meaningfully active.
- Each neuron in hidden layer has one bias




Sigmoid
$a_0^{(1)} = \sigma\left(w_{0,0}\ a_0^{(0)} + w_{0,1}\ a_1^{(0)} + \cdots + w_{0,n}\ a_n^{(0)} + b_0 \right)$
Activation function
Bias

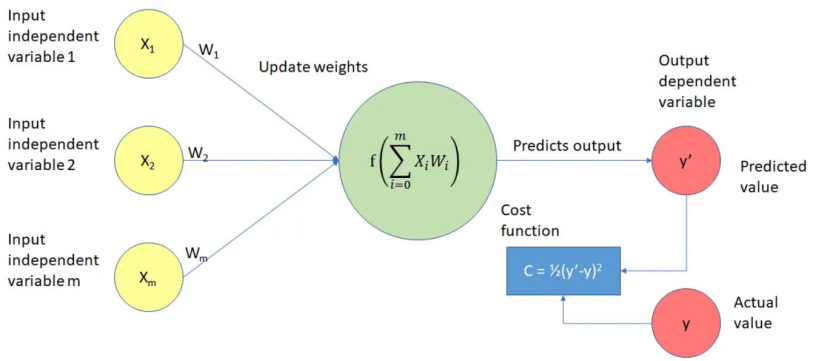

Sigmoid
$a_0^{(1)} = \sigma\left(w_{0,0}\, a_0^{(0)} + w_{0,1}\, a_1^{(0)} + \cdots + w_{0,n}\, a_n^{(0)} + b_0 \right)$
Bias
$a_0^{(0)}$
$a_1^{(0)}$
$a_2^{(0)}$
$a_3^{(0)}$
$a_4^{(0)}$
$a_5^{(0)}$
$a_6^{(0)}$
$a_7^{(0)}$
$\begin{bmatrix} w_{0,0} & w_{0,1} & \cdots & w_{0,n} \\ w_{1,0} & w_{1,1} & \cdots & w_{1,n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{k,0} & w_{k,1} & \cdots & w_{k,n} \end{bmatrix} \begin{bmatrix} a_0^{(0)} \\ a_1^{(0)} \\ \vdots \\ a_n^{(0)} \end{bmatrix} = \begin{bmatrix} ? \\ ? \end{bmatrix}$


Sigmoid
$a_0^{(1)} = \sigma\left(w_{0,0}\ a_0^{(0)} + w_{0,1}\ a_1^{(0)} + \cdots + w_{0,n}\ a_n^{(0)} + b_0 \right)$
Bias
$\begin{bmatrix} w_{0,0} & w_{0,1} & \cdots & w_{0,n} \\ w_{1,0} & w_{1,1} & \cdots & w_{1,n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{k,0} & w_{k,1} & \cdots & w_{k,n} \end{bmatrix} \begin{bmatrix} a_0^{(0)} \\ a_1^{(0)} \\ \vdots \\ a_n^{(0)} \end{bmatrix} = \begin{bmatrix} ? \\ ? \\ \vdots \\ ? \end{bmatrix}$


$a_0^{(0)}$
$a_1^{(0)}$
$a_2^{(0)}$
$a_3^{(0)}$
$a_4^{(0)}$
$a_5^{(0)}$
$a_6^{(0)}$
$a_7^{(0)}$
$\mathbf{a}^{(1)} = \sigma(\mathbf{W}\mathbf{a}^{(0)} + \mathbf{b})$
$\sigma\left(\begin{bmatrix} w_{0,0} & w_{0,1} & \cdots & w_{0,n} \\ w_{1,0} & w_{1,1} & \cdots & w_{1,n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{k,0} & w_{k,1} & \cdots & w_{k,n} \end{bmatrix}\begin{bmatrix} a_0^{(0)} \\ a_1^{(0)} \\ \vdots \\ a_n^{(0)} \end{bmatrix} + \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_n \end{bmatrix}\right)$

Cost function

- We initializeall of weights and biases totally randomly.
- So, the output layer is just looks like a mess.
- Now we need to define a cost function.
- We are trying to optimize the value of cost function.



https://medium.com/@zeeshanmulla/cost-activation-loss-function-neural-network-deep-learning-what-are-these-91167825a4de

PROGIS
ETUDES, MEDIAS, COMMUNICATION, MARKETING

Cost function

- To learn, we need to compare the **output value** to the **actual value** that we want the neural network to get.
- We are going to calculate a function called **cost function**,
  - It is calculated as one half of the squared difference between the actual value and the output value.

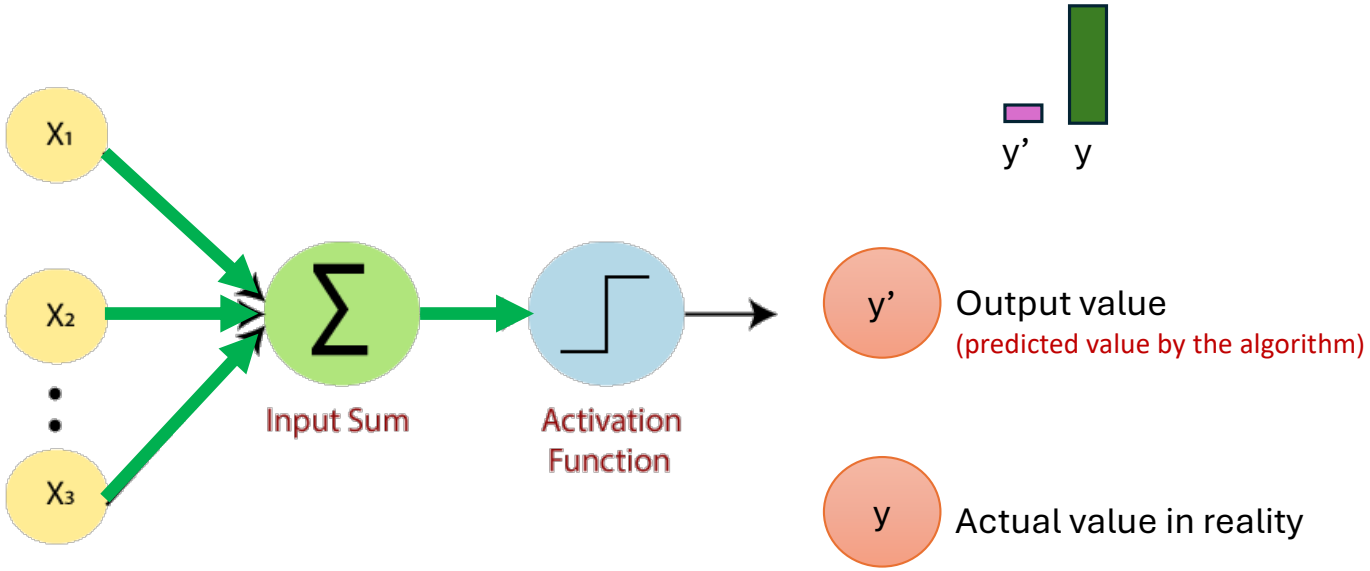
$$C = \frac{1}{2}(y'-y)^2$$

  - There are many diffrent function (here we consider error)
  - **Our goal is to minimize the cost function** (the lower the cost function the closer the y hat is to y)

Single layer feedforward neural network (a perceptron)

Cost function is a function that **measures the performance of a Machine Learning model** for given data.

Cost Function quantifies the error between predicted values and expected values and **presents it in the form of a single real number**.



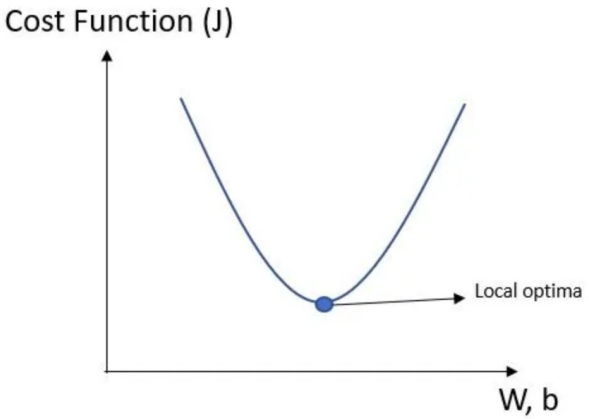
Cost function $C = \frac{1}{2}(y'-y)^2$

the lower the cost function, the closer the y hat is to y.

As mentioned, the goal of an artificial neural network is to minimize the value of the cost function.

The cost function is minimized when your algorithm's predicted value is as close to the actual value as possible. Said differently, the goal of a neural network is to minimize the error it makes in its predictions!

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$



https://medium.com/@zeeshanmulla/cost-activation-loss-function-neural-network-deep-learning-what-are-these-91167825a4de

- What the cost function is telling is:

  What is the error that you have in your prediction

- Just telling the computer what job it is doing and the result is not good is not very helpful.

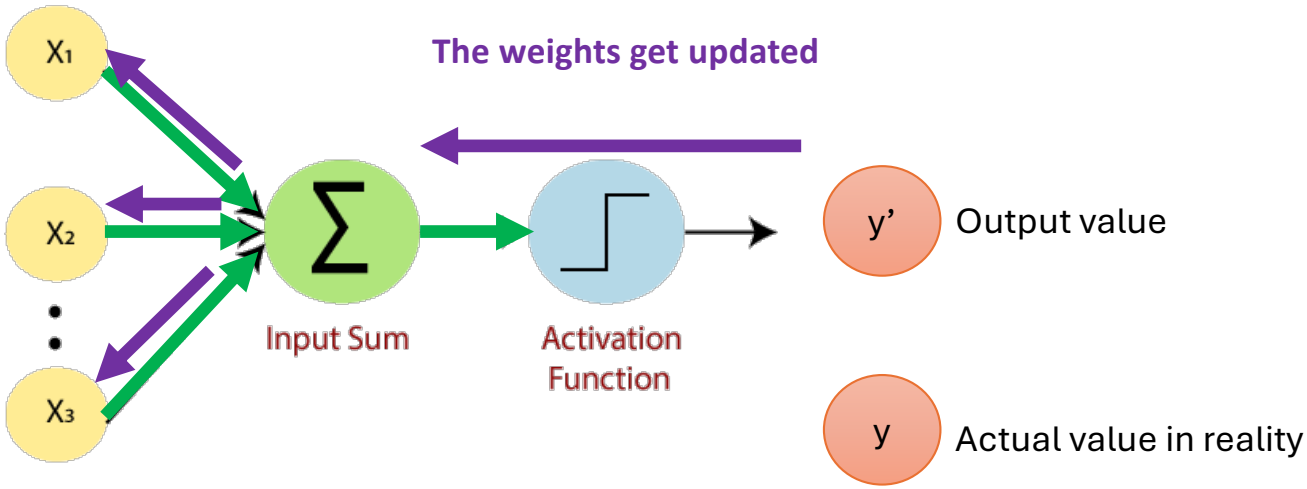
- To make things better, we need to tell it how to change those weights and biases.

- The algorithm for computing gradient efficiently which is effectively the heart of how a neural network learns is called **back propagation**.

PROGIS
ETUDES, MEDIAS, COMMUNICATION, MARKETING

Single layer feedforward neural network

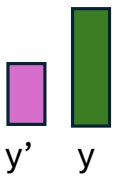
It is necessarily to know here, we're dealing with just the **one row**.

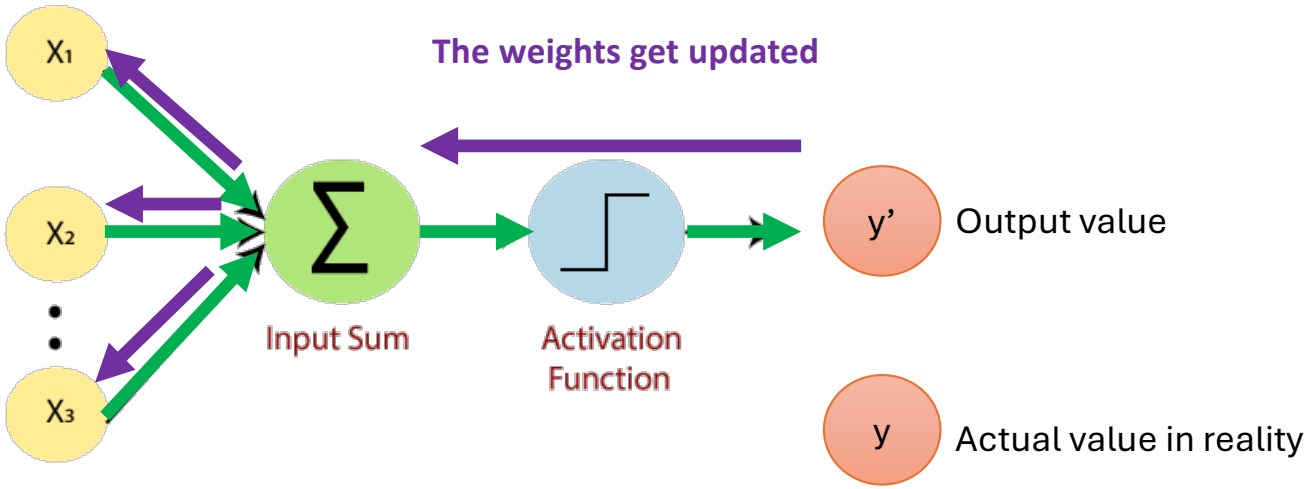


Cost function $(Y_i - \hat{Y}_i)^2$

area	#bedroom	#bathroom	basement	price
7420	4	2	no	13300000

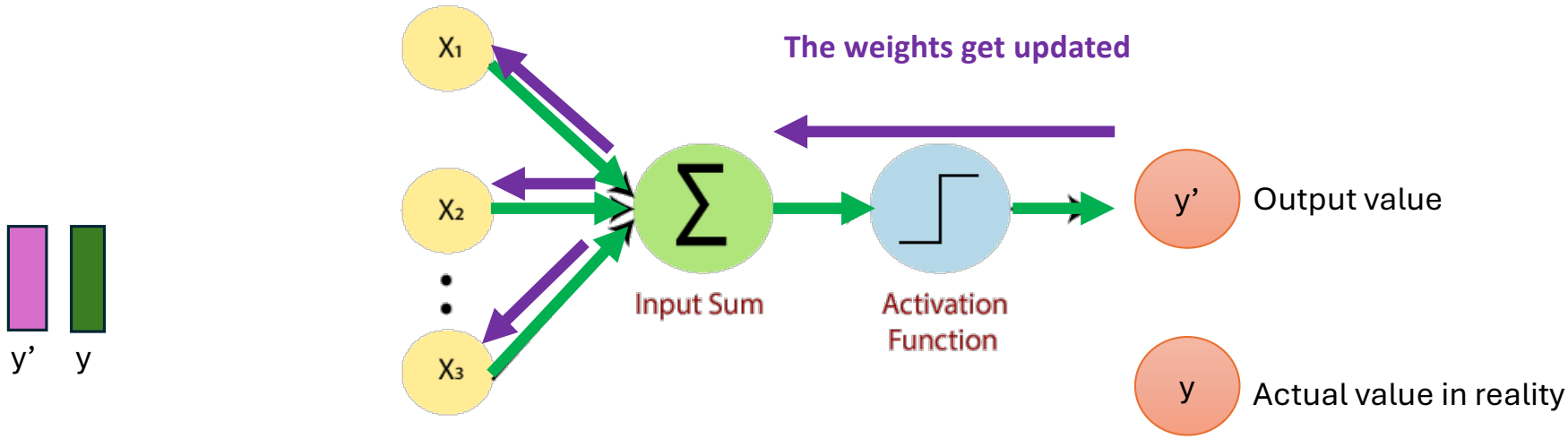
Single layer feedforward neural network





Our output value, is changing, so our cost function is chqnging $(Y_i - \hat{Y}_i)^2$
We back again, the weights get adjusted again

Single layer feedforward neural network

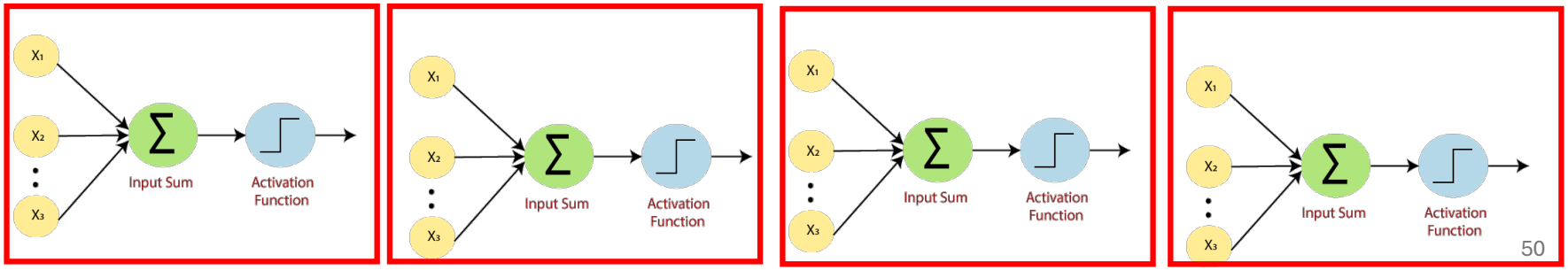


We feed the information back to the neural network again, the weights get adjusted again
We try to minimize cost function

$$(Y_i - \hat{Y}_i)^2$$

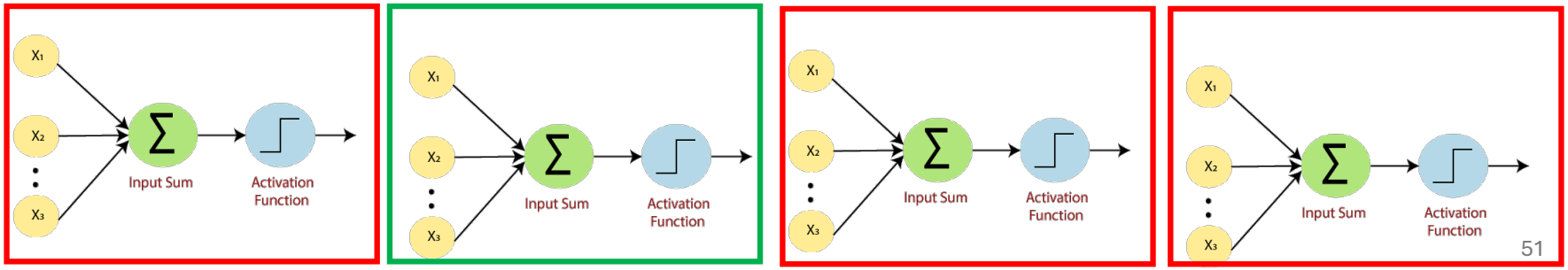
- One epoch: when we go through a whole data set and we train our neural network in all of these rows.

area	#bedroom	#bathroom	basement	price
7420	4	2	no	13300000
8960	4	2	no	12250000
9960	4	3	yes	12450000
7420	4	1	yes	11410000



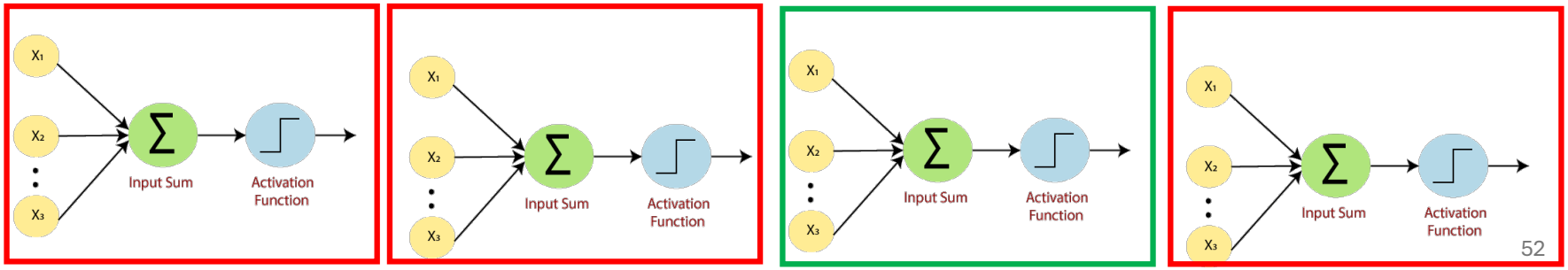
- One epoch: when we go through a whole data set and we train our neural network in all of these rows.

area	#bedroom	#bathroom	basement	price
7420	4	2	no	13300000
8960	4	2	no	12250000
9960	4	3	yes	12450000
7420	4	1	yes	11410000



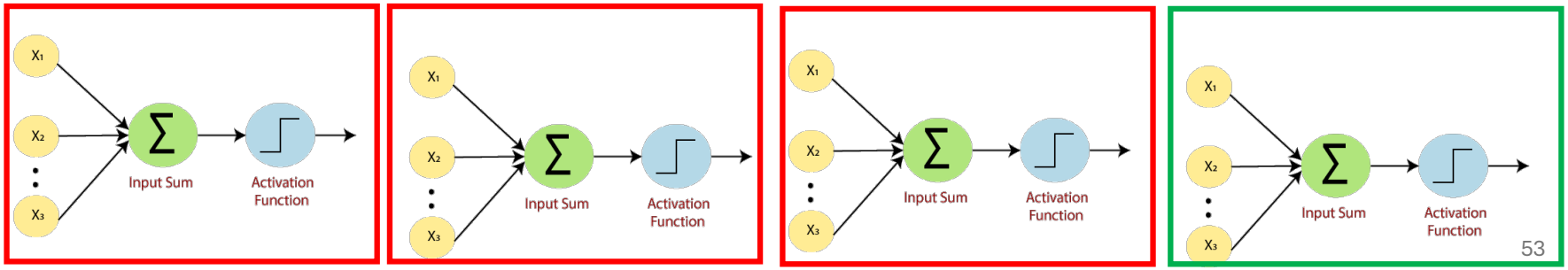
- One epoch: when we go through a whole data set and we train our neural network in all of these rows.

area	#bedroom	#bathroom	basement	price
7420	4	2	no	13300000
8960	4	2	no	12250000
9960	4	3	yes	12450000
7420	4	1	yes	11410000



- One epoch: when we go through a whole data set and we train our neural network in all of these rows.

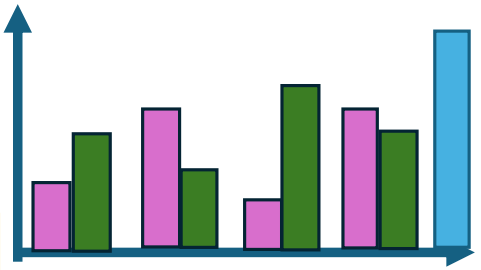
area	#bedroom	#bathroom	basement	price
7420	4	2	no	13300000
8960	4	2	no	12250000
9960	4	3	yes	12450000
7420	4	1	yes	11410000

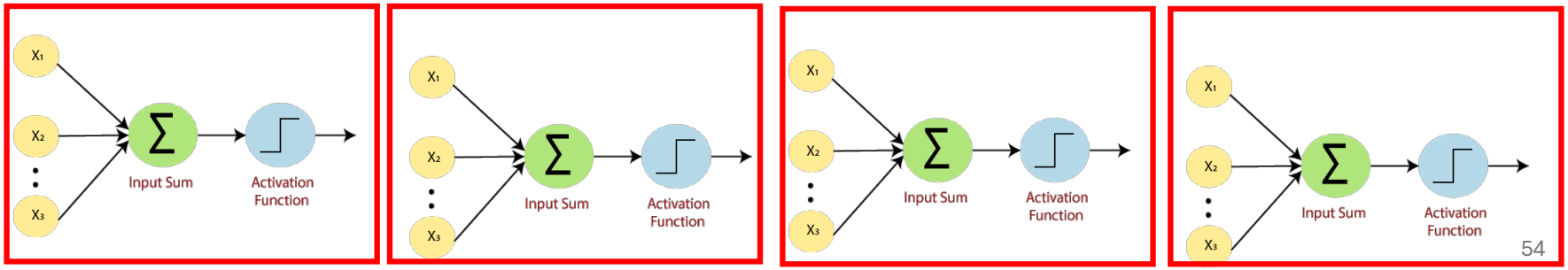


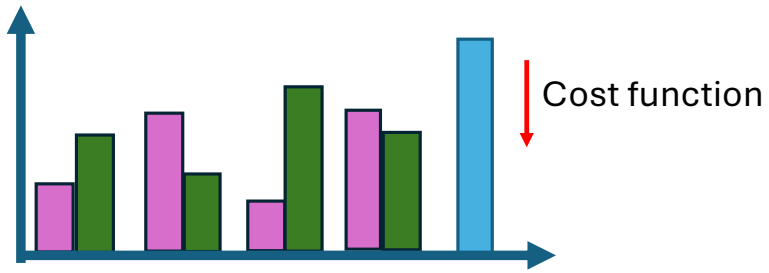
- One epoch: when we go through a whole data set and we train our neural network in all of these rows.

$$\frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$

area	#bedroom	#bathroom	basement	price
7420	4	2	no	13300000
8960	4	2	no	12250000
9960	4	3	yes	12450000
7420	4	1	yes	11410000







- The goal here that is minimize the cost function
- We go to the second epoch and adjust the weights
- This whole process is called **backpropagation**.
- Backpropagation is the learning mechanism that allows the Multilayer Perceptron to iteratively adjust the weights in the network, with the goal of minimizing the cost function.
- With propagation we are able to adjust all of the weights at the same time.

PROGIS
ETUDES, MEDIAS, COMMUNICATION, MARKETING

Number of Epochs

- Too few epochs vs too many epochs

Epochs	Result
Too few	Underfitting
Good number	Best performance
Too many	Overfitting

PROGIS
ETUDES, MEDIAS, COMMUNICATION, MARKETING

Neural Networks in Marketing

- EXAMPLE:

- **Customer churn prediction:** Will a customer leave?
  - Inputs: Age , Purchases , Visits , Complaints
  - Output: Leave / Stay

- **Sentiment analysis**
  - Input: Customer review
  - Output: Positive , Neutral , Negative

- **Product recommendation**
  - Input: Previous purchases , Browsing history
  - Output: Recommended products

PROGIS
ETUDES, MEDIAS, COMMUNICATION, MARKETING

Evaluation Metrics for Classification

- Example: Predicting High Social Media Engagement

Imagine we have **100 social media posts**.

In reality:

45 posts had **High Engagement** (class = 1)
55 posts had **Low Engagement** (class = 0)

Our neural network tries to identify which posts will have High Engagement.

- After prediction, we obtain:

	Actually High Engagement	Actually Low Engagement
Predicted High Engagement	40	10
Predicted Low Engagement	5	45

PROGIS
ETUDES, MEDIAS, COMMUNICATION, MARKETING

Evaluation Metrics for Classification

	Actually High Engagement	Actually Low Engagement
Predicted High Engagement	40 True positive	10
Predicted Low Engagement	5	45

A post receives many likes, comments, and shares.
The model predicted High Engagement.
The model was right.

PROGIS
ETUDES, MEDIAS, COMMUNICATION, MARKETING

Evaluation Metrics for Classification

	Actually High Engagement	Actually Low Engagement
Predicted High Engagement	40	10
Predicted Low Engagement	5	45

False positive

Marketing team decides to promote a post because the model says it will perform well.
Unfortunately, the post performs poorly.

The model predicted:
"This post will have High Engagement."
But in reality:
The post had Low Engagement.
Wrong prediction

PROGIS
ETUDES, MEDIAS, COMMUNICATION, MARKETING

Evaluation Metrics for Classification

	Actually High Engagement	Actually Low Engagement
Predicted High Engagement	40	10
Predicted Low Engagement	5	45

False Negative (FN)

The model predicted:
"This post will have Low Engagement."
But in reality:
The post had High Engagement.
Wrong prediction

PROGIS
ETUDES, MEDIAS, COMMUNICATION, MARKETING

Evaluation Metrics for Classification

	Actually High Engagement	Actually Low Engagement
Predicted High Engagement	40	10
Predicted Low Engagement	5	45

True Negative (TN)

The model predicted:
"This post will have Low Engagement."
And in reality:
The post indeed had Low Engagement.
Correct prediction

PROGIS
ETUDES, MEDIAS, COMMUNICATION, MARKETING

Precision = TP / (TP + FP)

- The model predicted High Engagement for:
- 40+10=50 posts

	Actually High Engagement	Actually Low Engagement
Predicted High Engagement	40	10

Among those 50 posts: 40 were actually successful, 10 were not
Precision = 40 / 50 = 0.80

When the model says "this post will be successful", it is correct 80% of the time.

Modern AI Connection