LECTURER: NGHIA DUONG-TRUNG

DATA SCIENCE

TOPIC OUTLINE

UNIT 6

SELECTED ARTIFICIAL INTELLIGENCE TECHNIQUES

STUDY GOALS

On completion of this unit, you will have learned …

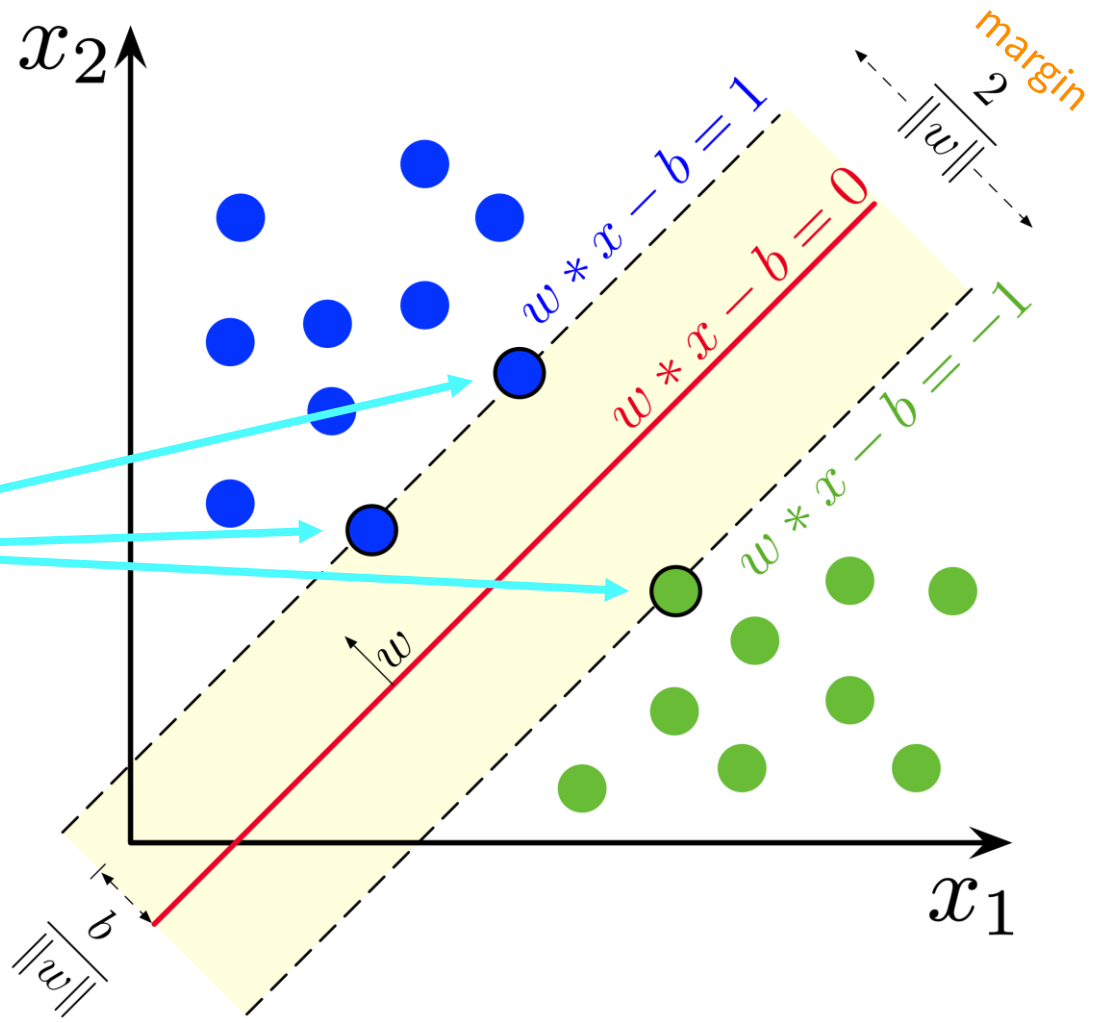
- data classification by support vector machines.
- the feedforward neural network structure.
- the back propagation algorithm in neural networks. how to develop an artificial neural networks prediction model.
- recurrent networks and reinforcement learning.
- basics about genetic algorithms, fuzzy logic, and Naïve Bayes classification.

EXPLAIN SIMPLY

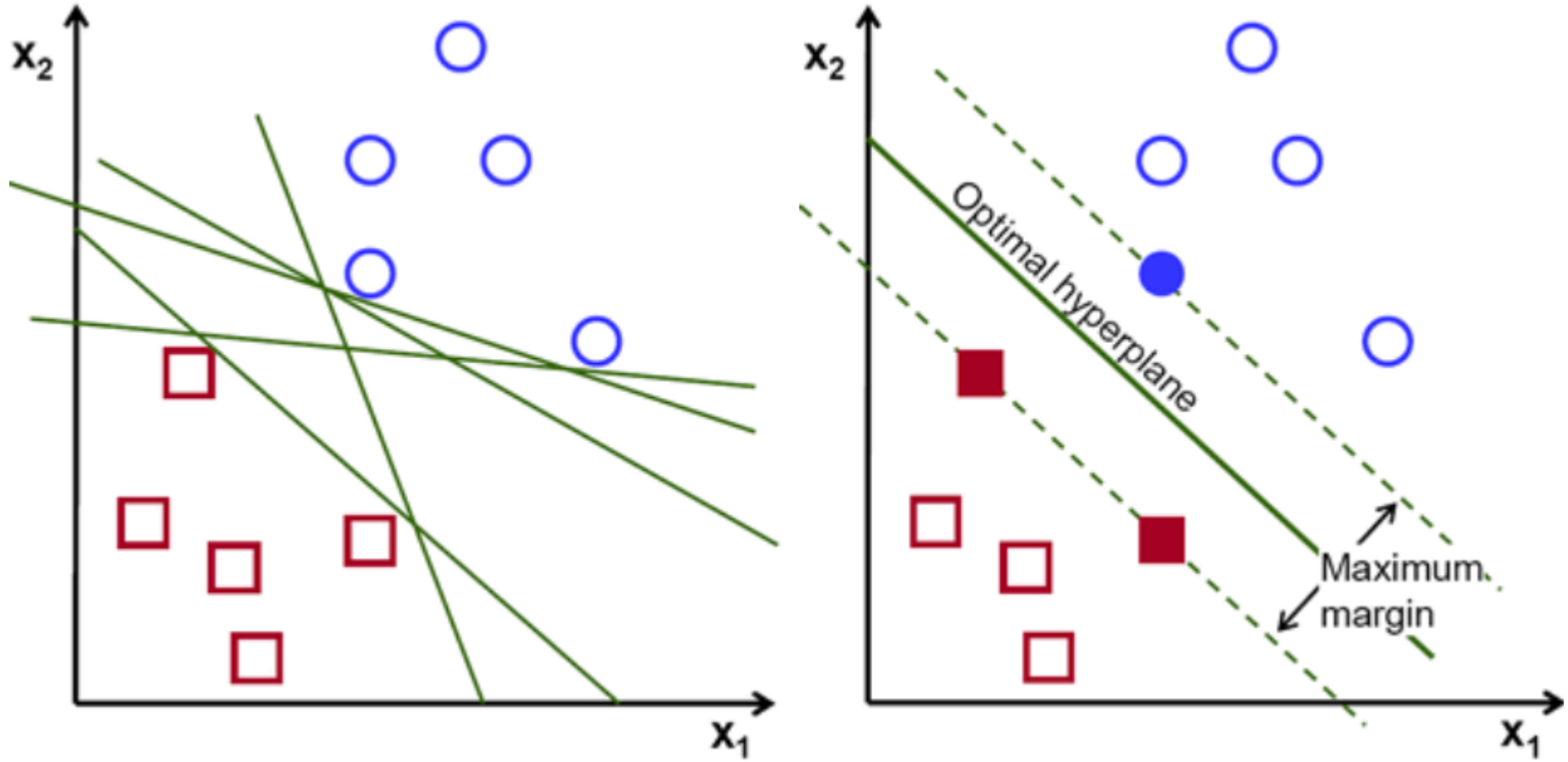
1. Explain the concept of Support Vector Machines (SVM) and the usage of the kernel tricks.
2. Name two activation functions. Can you draw them?
3. Describe the usage of Gradient Descent in Neural Networks.

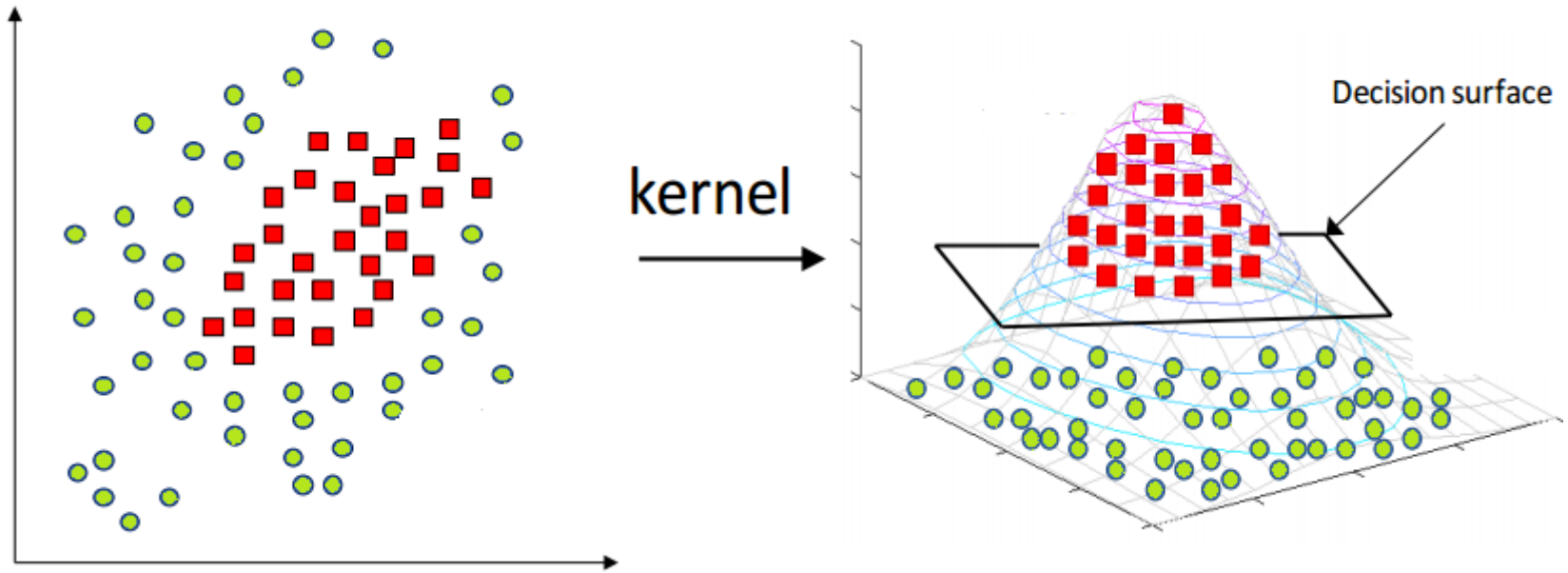
SUPPORT VECTOR MACHINES

- model used for regression & classification tasks
- identify hyperplane in data space that maximizes the margin between support vectors
- apply kernel trick for nonlinearly separable datasets



Source of the graphic: Larhmam, 2018, https://commons.wikimedia.org/wiki/File:SVM_margin.png, CC BY-SA 4.0.


x2
x1
x2
Optimal hyperplane
Maximum
margin
x1


kernel
Decision surface

- It is a binary linear classification whose decision boundary is explicitly constructed to minimize generalization error
- Very powerful and versatile machine learning model
- Suited for the classification of complex structures
- If observations are linearly separable, it fits the “decision boundary”
- Decision boundary – defined by the largest margin between the closest points for each class, called Maximum Margin Hyperplane (MMH)

- Feature scaling
  - **StandardScaler**
  - **MinMaxScaler**
  - **RobustScaler**
  - Codes: Session6_codes\01

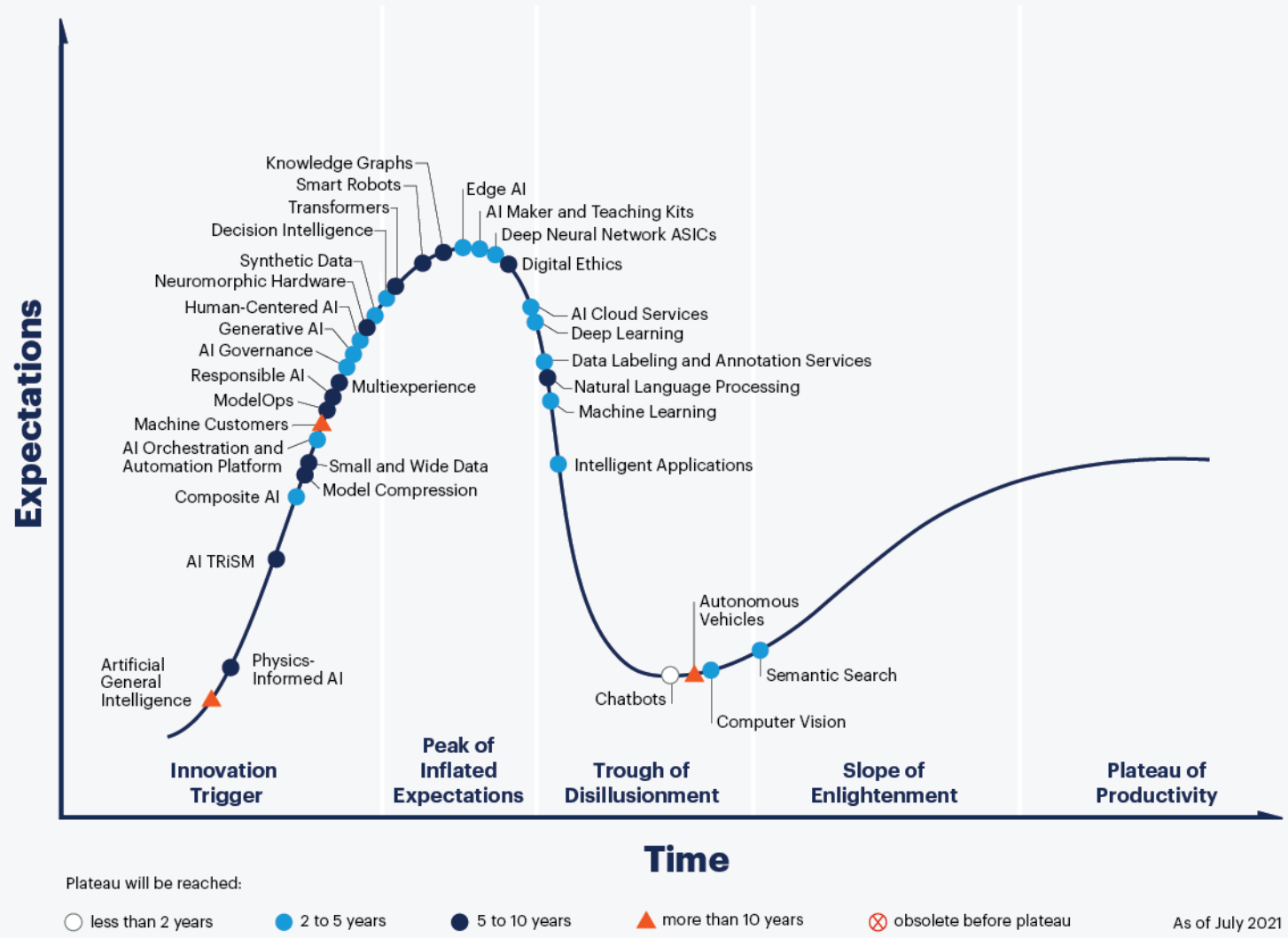
Further references:

- https://www.youtube.com/watch?v=_YPScrckx28

NEURAL NETWORKS

- Neural networks are what most people think about, when they think about AI
- They are modelled loosely after the human brain:
  - Very powerful
  - Very easy to execute
  - Very hyped
- Both supervised and unsupervised learning algorithm

Hype Cycle for Artificial Intelligence, 2021



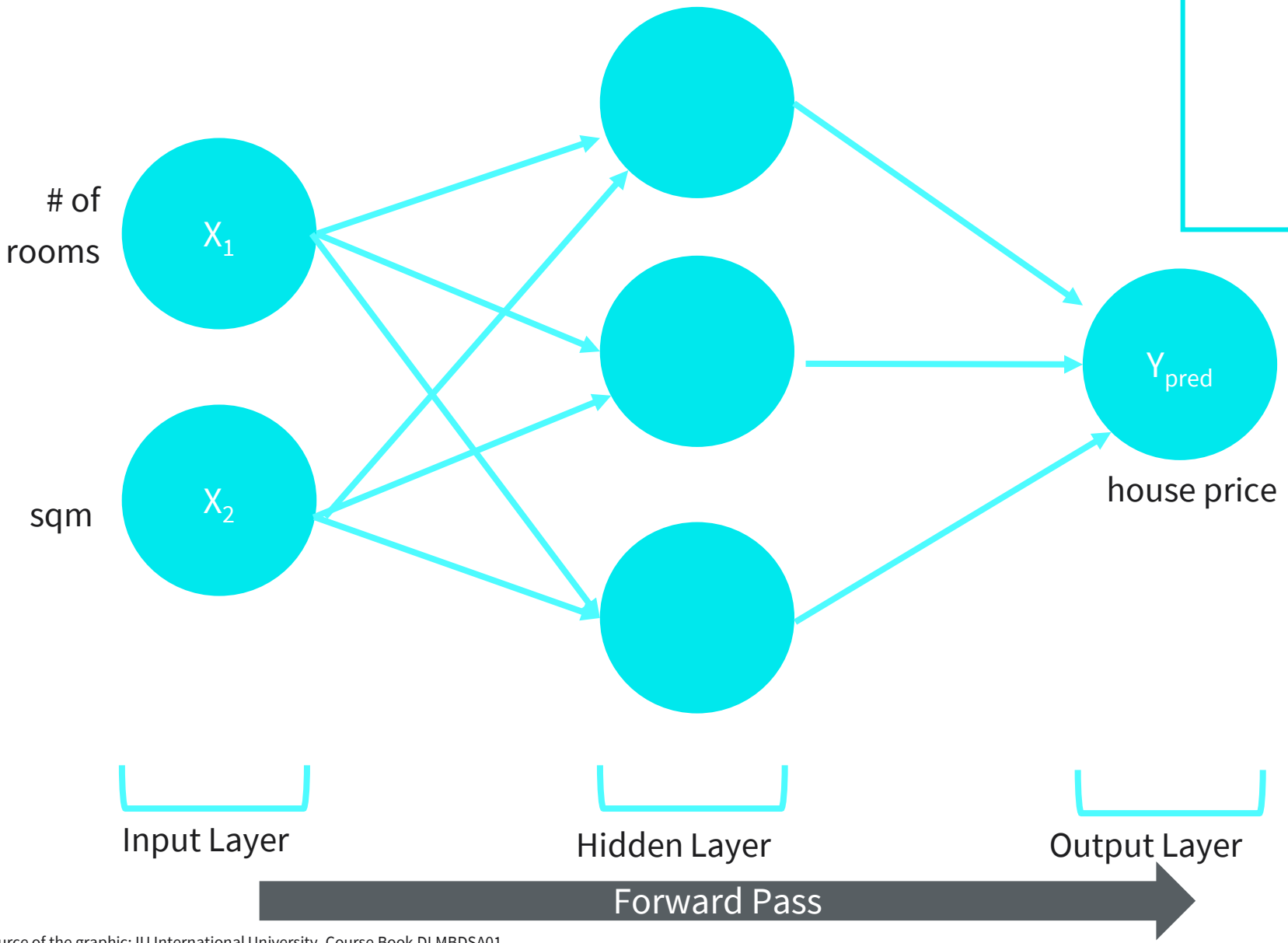
gartner.com

Source: Gartner

Gartner

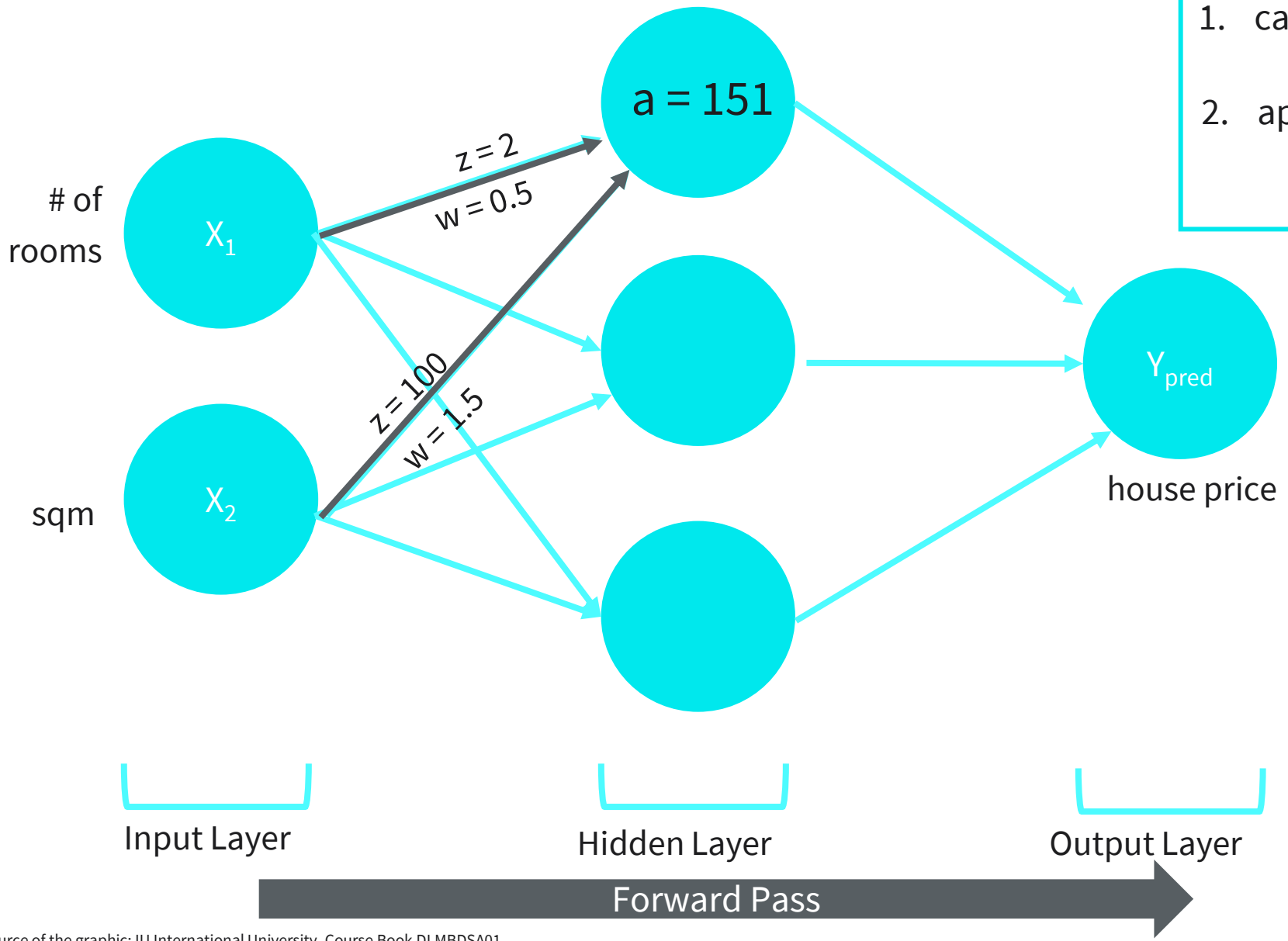
FEEDFORWARD NEURAL NETWORKS

Forward Pass



Source of the graphic: IU International University, Course Book DLMBDSA01.

FEEDFORWARD NEURAL NETWORKS



Forward Pass

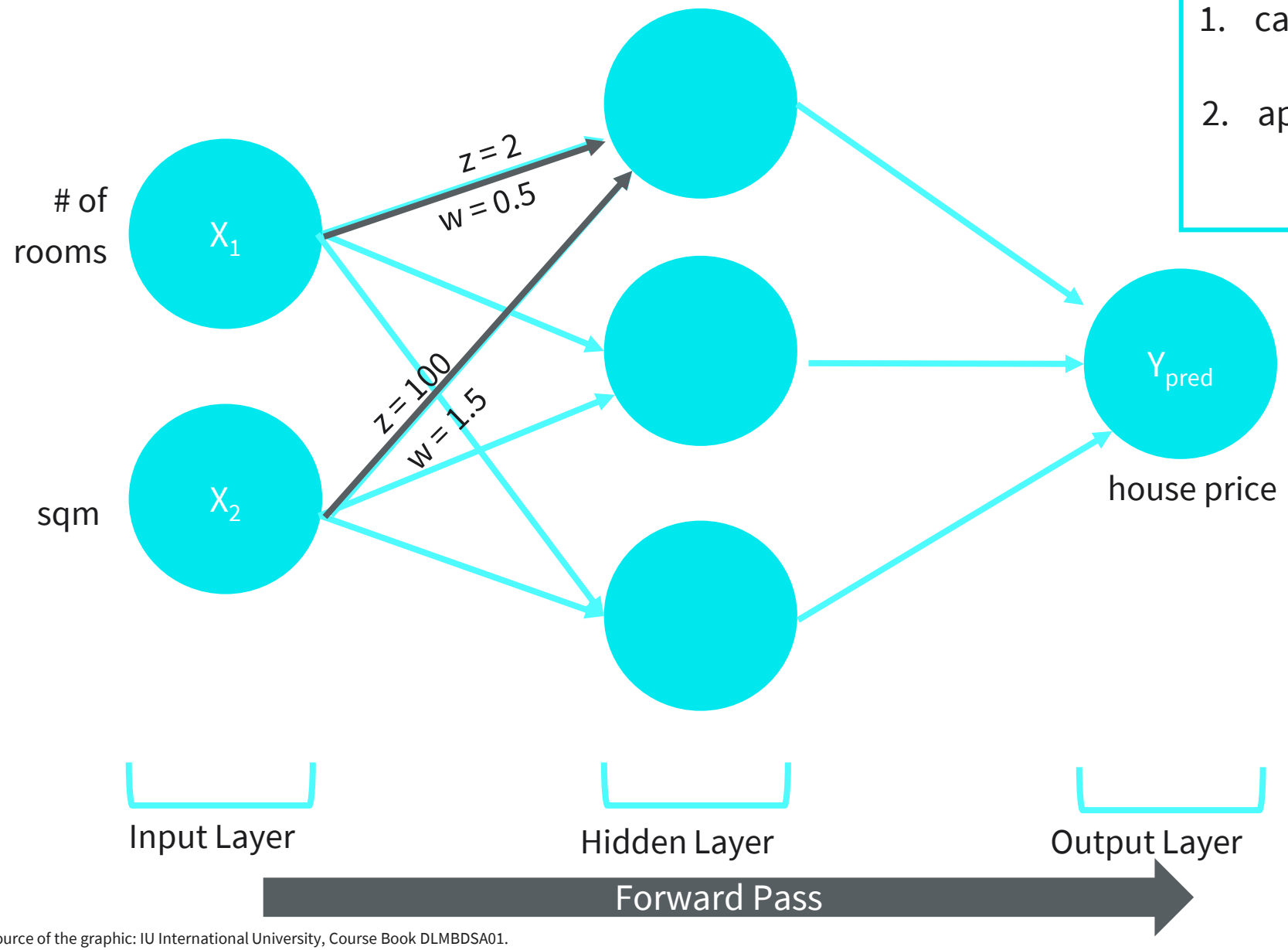
1. calculate weighted sum a from X_1 & X_2

   $a = 2 \cdot 0.5 + 100 \cdot 1.5 = 151$

2. apply Activation Function f(a) to get z

Source of the graphic: IU International University, Course Book DLMBDSA01.

FEEDFORWARD NEURAL NETWORKS



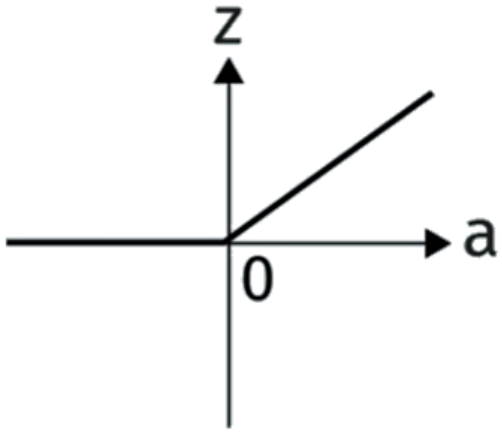
Forward Pass

1. calculate weighted sum a from X_1 & X_2

   $a = 2 \cdot 0.5 + 100 \cdot 1.5 = 151$

2. apply Activation Function f(a) to get z

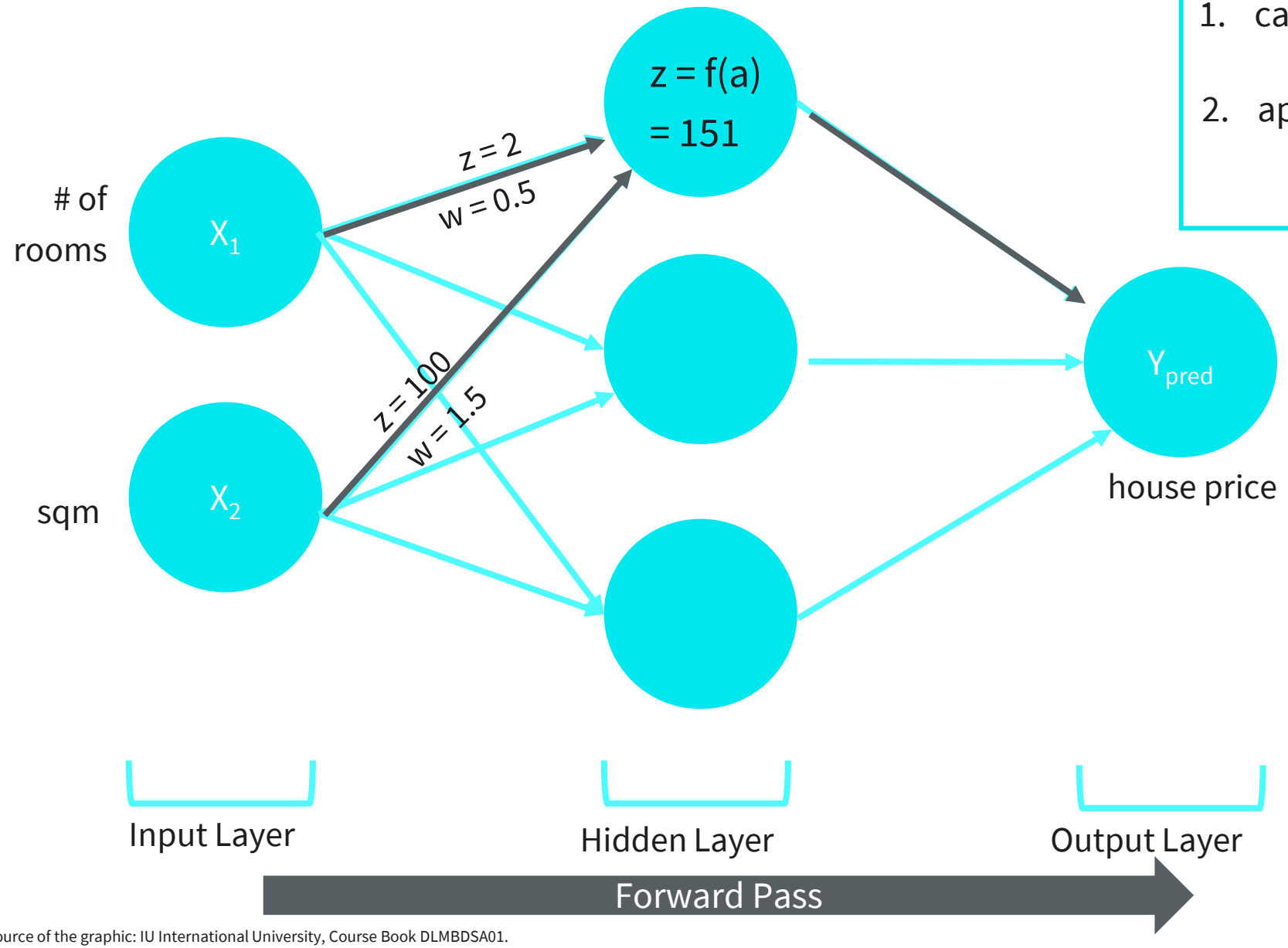
Rectified Linear Unit (ReLU)



$$z = \max(a, 0)$$

Source of the graphic: IU International University, Course Book DLMBDSA01.

FEEDFORWARD NEURAL NETWORKS



Forward Pass

1. calculate weighted sum a from X_1 & X_2

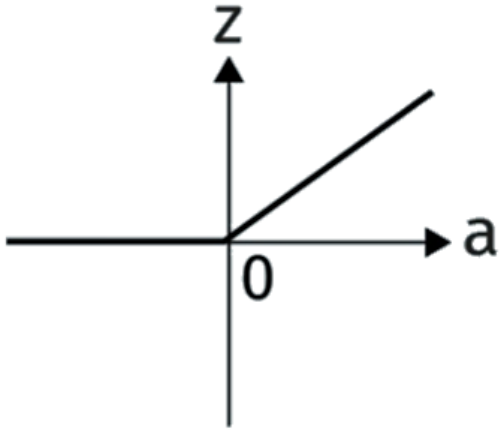
   $a = 2 \cdot 0.5 + 100 \cdot 1.5 = 151$

2. apply Activation Function f(a) to get z

   $z = f(151) = 151$

... repeat until Output Neuron

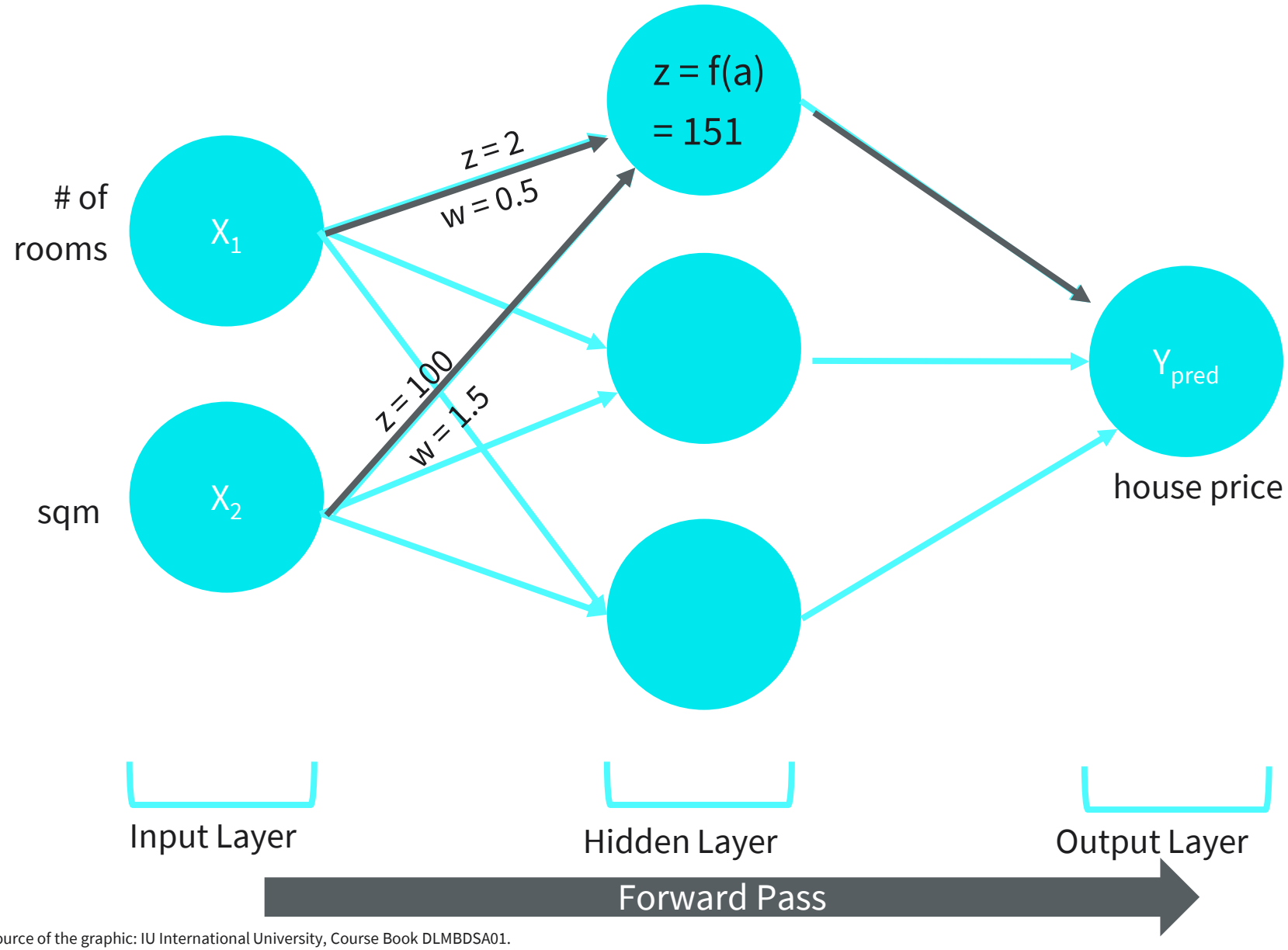
Rectified Linear Unit (ReLU)



$$z = \max(a, 0)$$

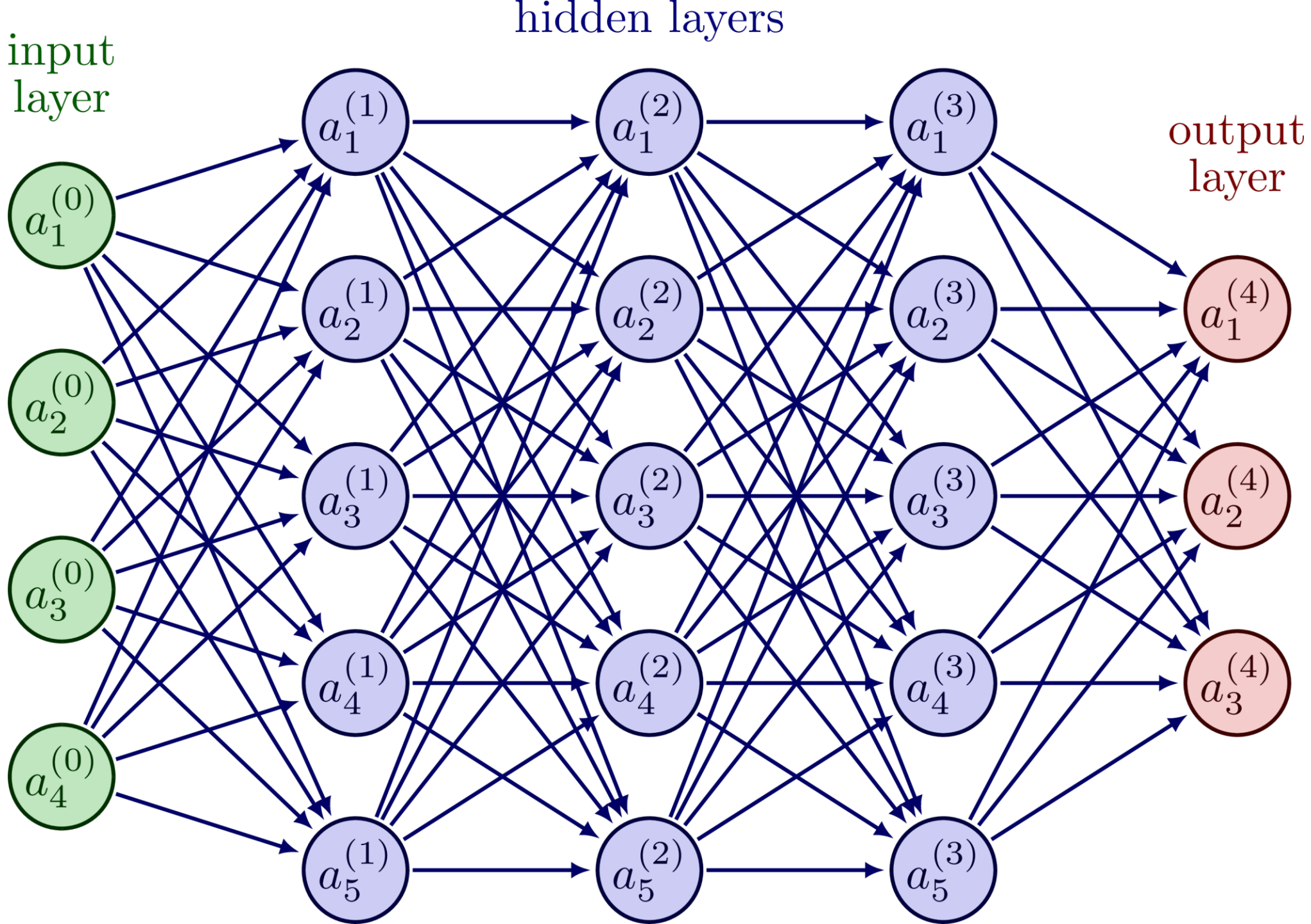
Source of the graphic: IU International University, Course Book DLMBDSA01.

FEEDFORWARD NEURAL NETWORKS



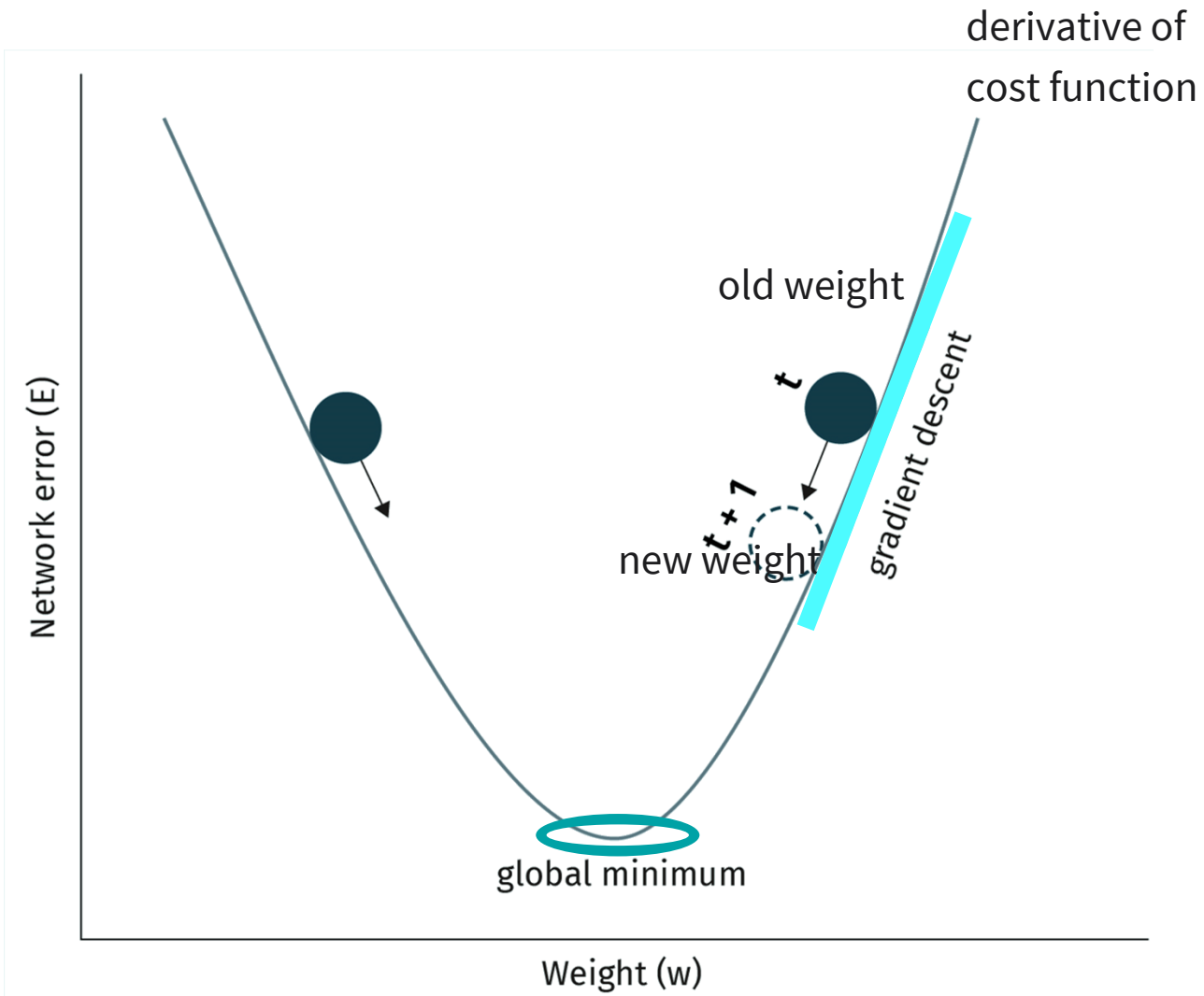
… how do we define network weights w?
→ backpropagation

Source of the graphic: IU International University, Course Book DLMBDSA01.


hidden layers
input layer
output layer
$a_1^{(0)}$
$a_2^{(0)}$
$a_3^{(0)}$
$a_4^{(0)}$
$a_1^{(1)}$
$a_2^{(1)}$
$a_3^{(1)}$
$a_4^{(1)}$
$a_5^{(1)}$
$a_1^{(2)}$
$a_2^{(2)}$
$a_3^{(2)}$
$a_4^{(2)}$
$a_5^{(2)}$
$a_1^{(3)}$
$a_2^{(3)}$
$a_3^{(3)}$
$a_4^{(3)}$
$a_5^{(3)}$
$a_1^{(4)}$
$a_2^{(4)}$
$a_3^{(4)}$

GRADIENT DESCENT

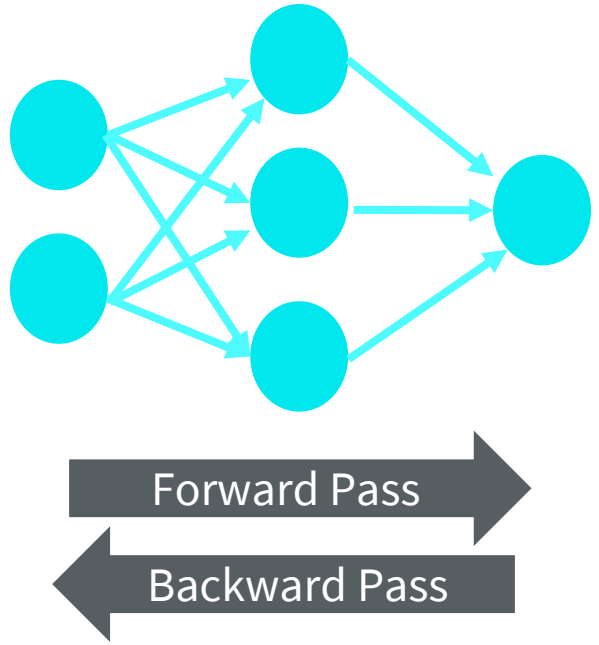
- Algorithm for finding a local minimum of a differentiable function
- in ML: **find weights that minimize the error function**
- calculate the gradient of the error function with respect to network weights



Source of the graphic: IU International University, Course Book DLMBDSA01.

BACKPROPAGATION ALGORITHM

1. Randomly initialize weights
2. Calculate output of every neuron
3. Calculate the error for 2.
4. Update the weights with GD



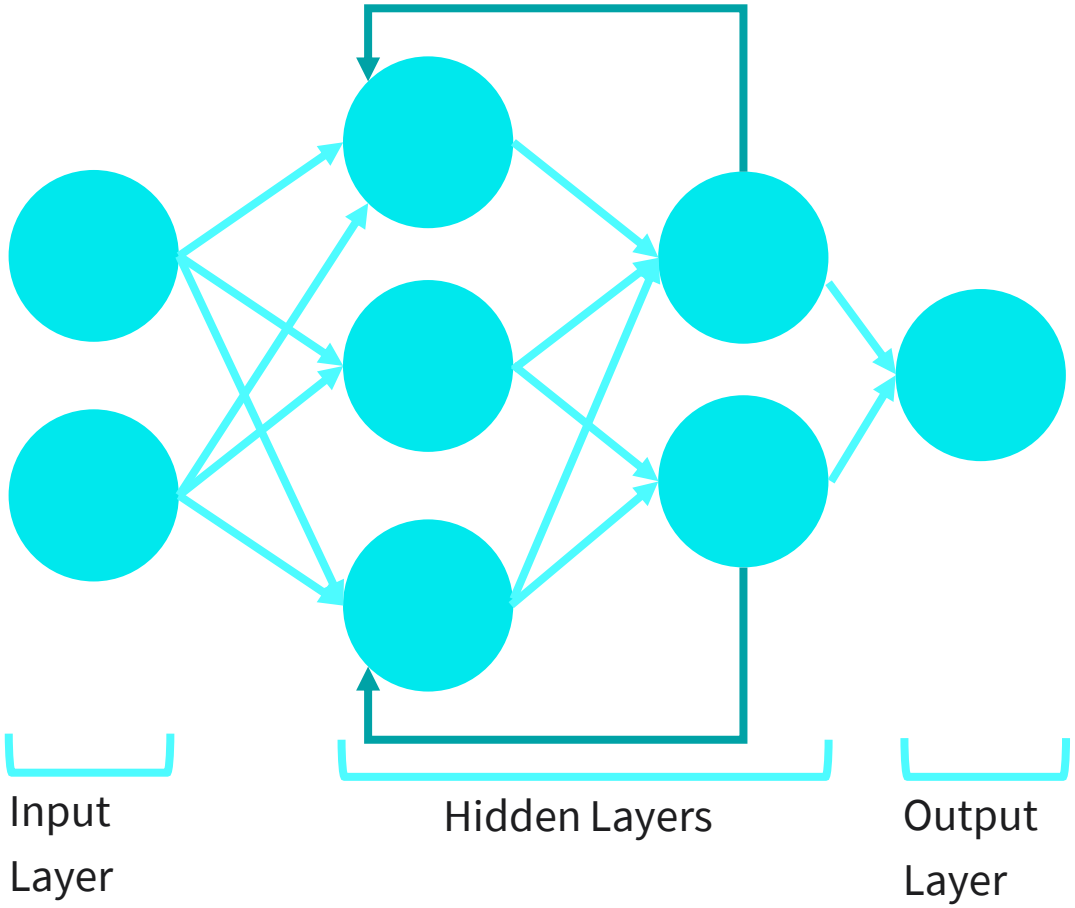
$$w_{new} = w_{old} - \eta \left(\frac{\partial \text{ Error}}{\partial w_{old}}\right)$$

Derivative of Error with respect to weights

learning rate

5. Start new forward pass with updated weights
6. Repeat steps 2-4 until no improvement in Error achieved

Feedforward Neural Networks



Recurrent Neural Networks

- Allow connections to previous layers
- Memory cells to retain information in deeper neural networks

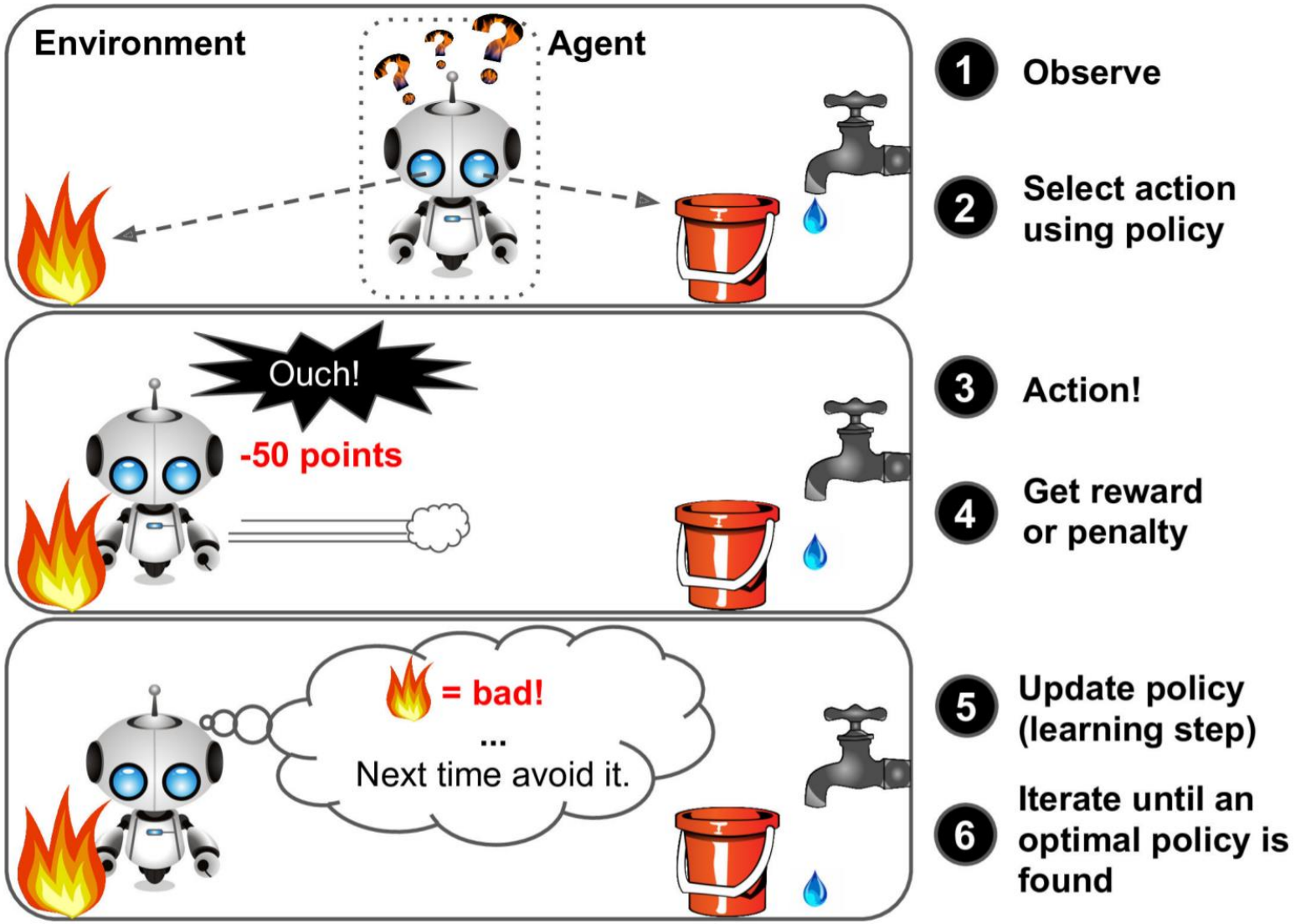
Source of the graphic: IU International University, Course Book DLMBDSA01.

- Further references:
  - https://www.youtube.com/watch?v=GvQwE2OhL8I
  - https://www.analyticsvidhya.com/blog/2021/05/beginners-guide-to-artificial-neural-network/
  - https://playground.tensorflow.org/
  - Gradient descent: https://www.youtube.com/watch?v=sDv4f4s2SB8

- Codes: Session6_codes\02

REINFORCEMENT LEARNING

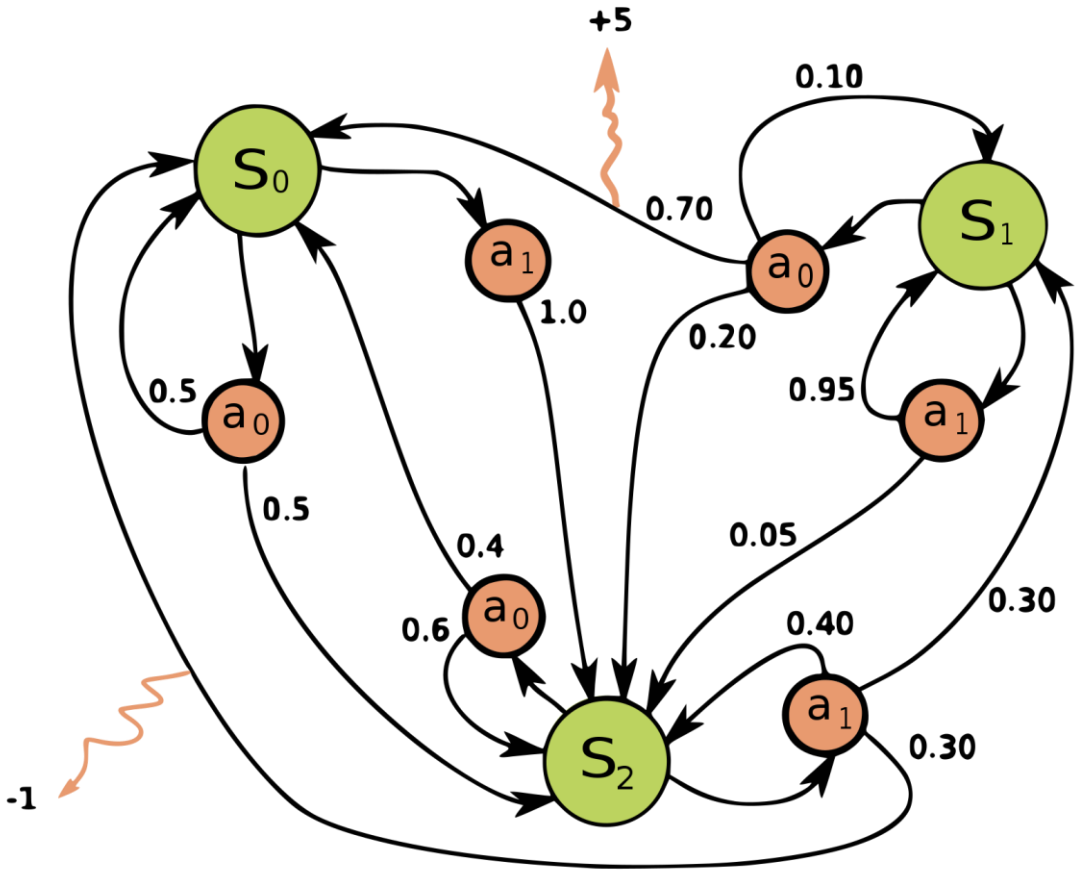
- Algorithm learns a policy how to act in a given environment through trial-and-error actions
- goal: maximize the reward for the agent



Source of graphics: IU International University, Course Book DLMBDSA01; Géron, 2019, p. 15.

MARKOV DECISION PROCESS

- framework to solve reinforcement learning problems
- set of states $\{s_0, s_1, s_2\}$
- set of actions to take a path $\{a_0, a_1 \ldots\}$
- set of rewards {+10,+40,-50}
- policy for selected path $\{s_0 \rightarrow s_1 \rightarrow s_2\}$



Source of the graphic: Waldo Alvarez, 2017, https://commons.wikimedia.org/wiki/File:Markov_Decision_Process.svg, CC BY-SA 4.0.

You have learned …

- data classification by support vector machines.
- the feedforward neural network structure.
- the back propagation algorithm in neural networks. how to develop an artificial neural networks prediction model.
- recurrent networks and reinforcement learning.
- basics about genetic algorithms, fuzzy logic, and Naïve Bayes classification.

SESSION 6

TRANSFER TASK

TRANSFER TASK

1. Discuss the parameter of ***learning rate η*** in the context of Gradient Descent.
2. How does it influence the process?
3. Can you foresee challenges in choosing the adequate learning rate?

TRANSFER TASK
PRESENTATION OF THE RESULTS

Please present your results.

The results will be discussed in plenary.

1. The Naïve Bayes approach assumes that the independent variables are…

   a) random variables.
   b) orthogonal variables.
   c) normalized variables.
   d) structured data variables.

2. A memory cell is a concept which exists in …

   a) feedforward networks.
   b) recurrent networks.
   c) reinforcement learning.
   d) support vector machines.

3. The Kernel trick is employed in support vector machines to…

   a) maximize the margin between the two classes.
   b) minimize the classification error.
   c) deal with nonlinearly separable dataset.
   d) define the set of support vectors.

How did you like the course?

LIST OF SOURCES

Alvarez, W. (2017). *Markov Decision Process* [Image]. https://commons.wikimedia.org/wiki/File:Markov_Decision_Process.svg, CC BY-SA 4.0.
Géron, A. (2019). *Hands-on machine learning with scikit-learn, keras, and tensorflow : Concepts, tools, and techniques to build intelligent systems*. O'Reilly Media, Incorporated.
Larhmam (2018). *SVM-Margin* [Image]. https://commons.wikimedia.org/wiki/File:SVM_margin.png, CC BY-SA 4.0.
Jordon, J. (2018). *Setting the learning rate of your neural network*. [Image]. https://www.jeremyjordan.me/nn-learning-rate/