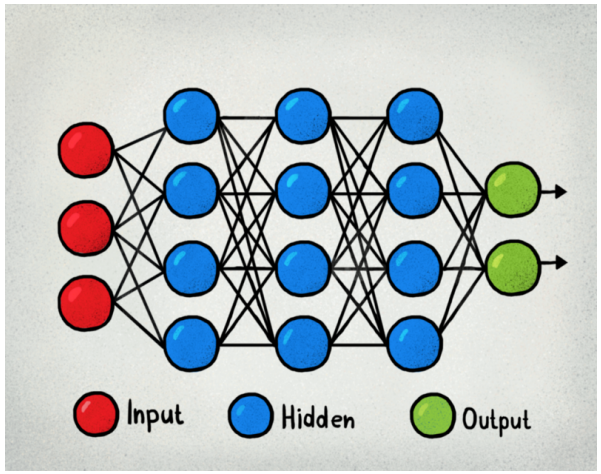
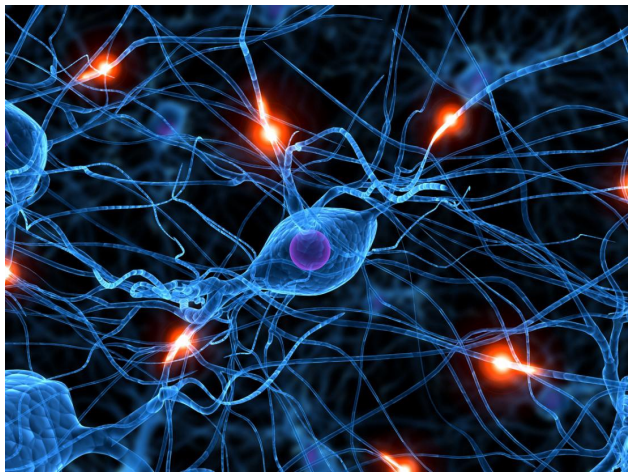
CSE 3521: Neural Networks



OR



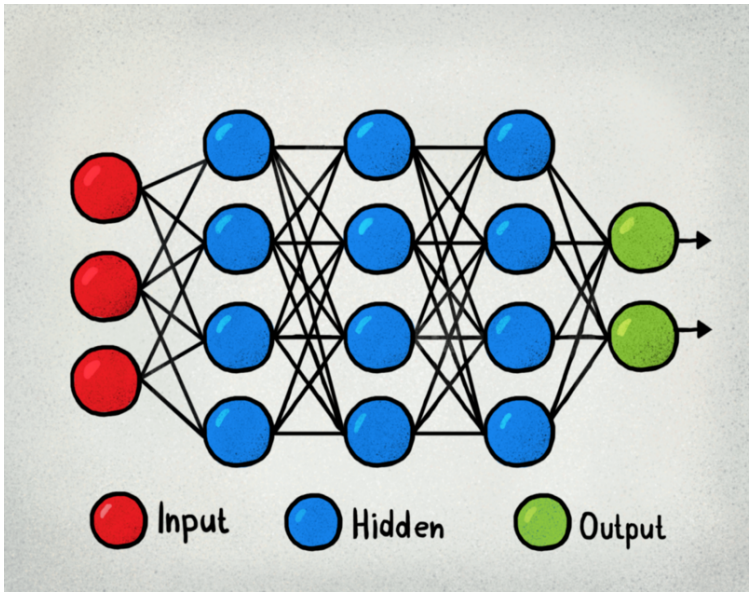
THE OHIO STATE UNIVERSITY

Today

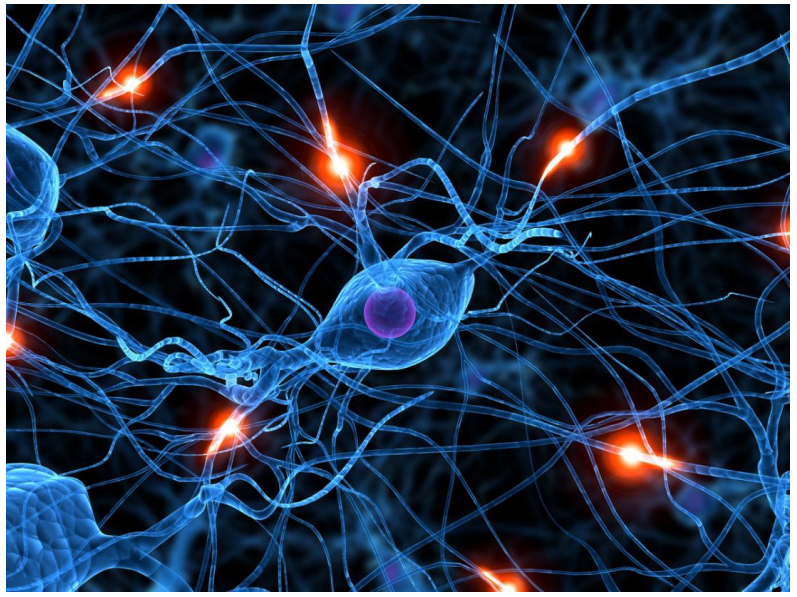
- Overview
  - Feedforward neural network
  - Recurrent neural network
- Perceptron
- Multi-layer Perceptron (MLP)

Overview of Neural Networks

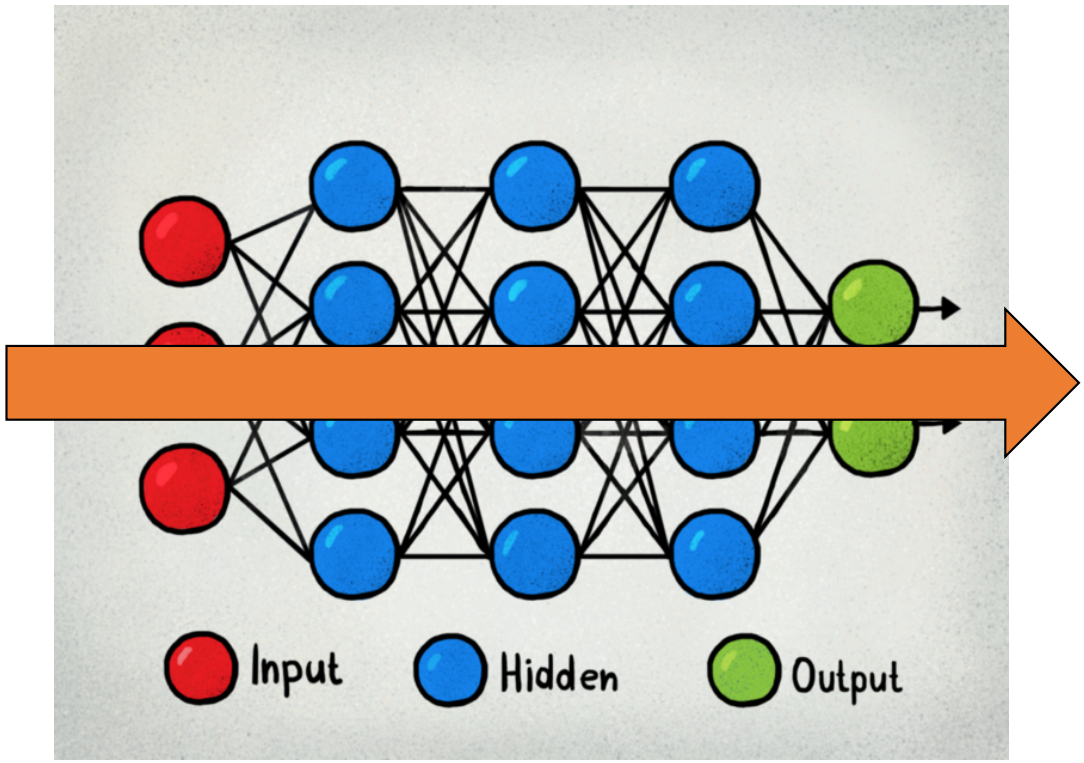
- Which one?



OR



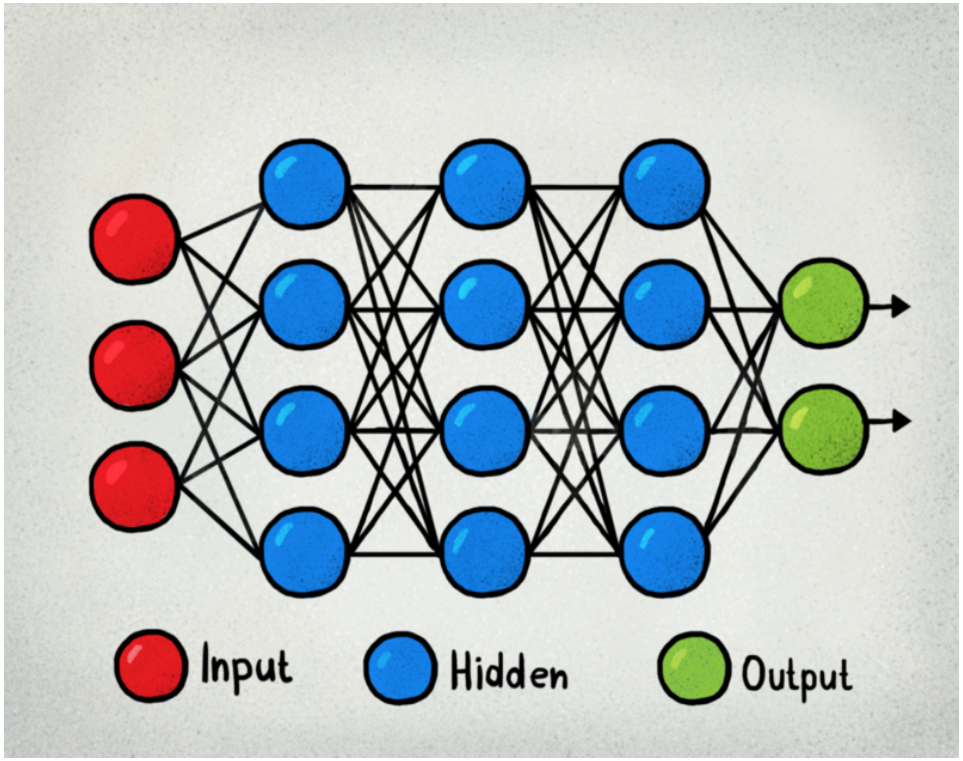
Overview of Neural Networks



Feed-forward
neural network

Overview of Neural Networks

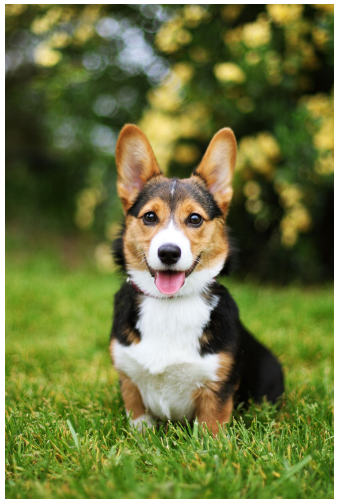
Image

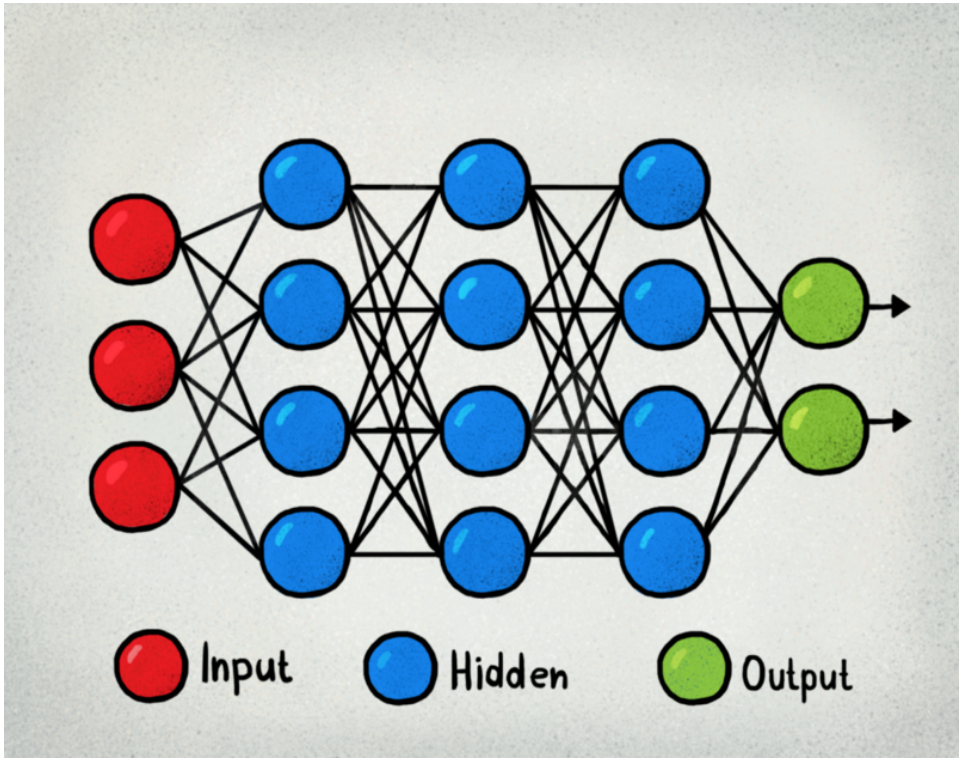


Cat score

Dog score

Overview of Neural Networks

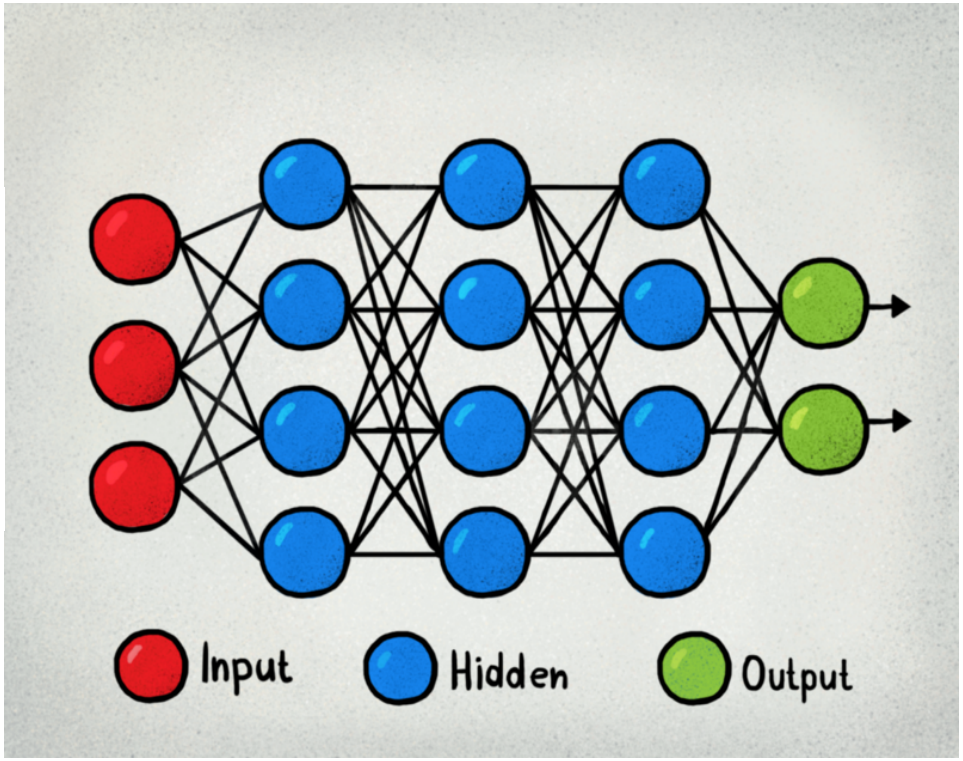




Cat score 0.1

Dog score **0.9**

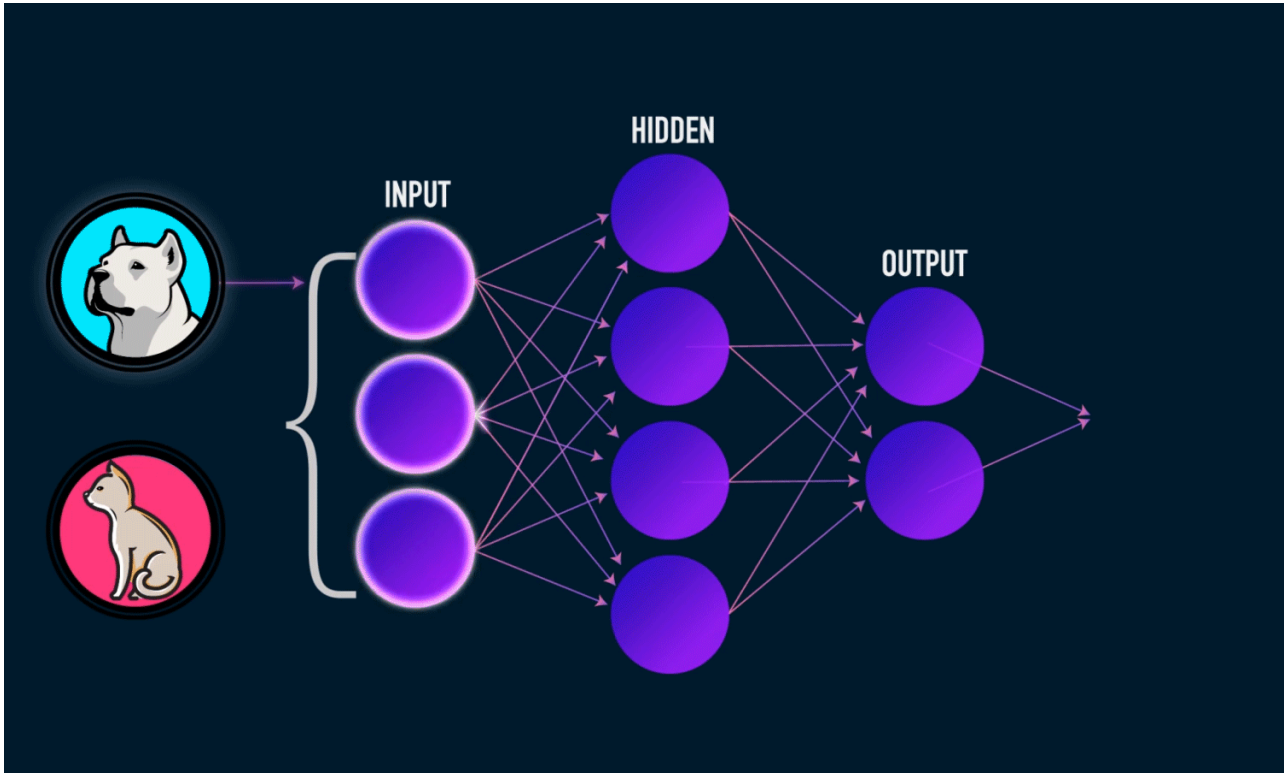
Overview of Neural Networks



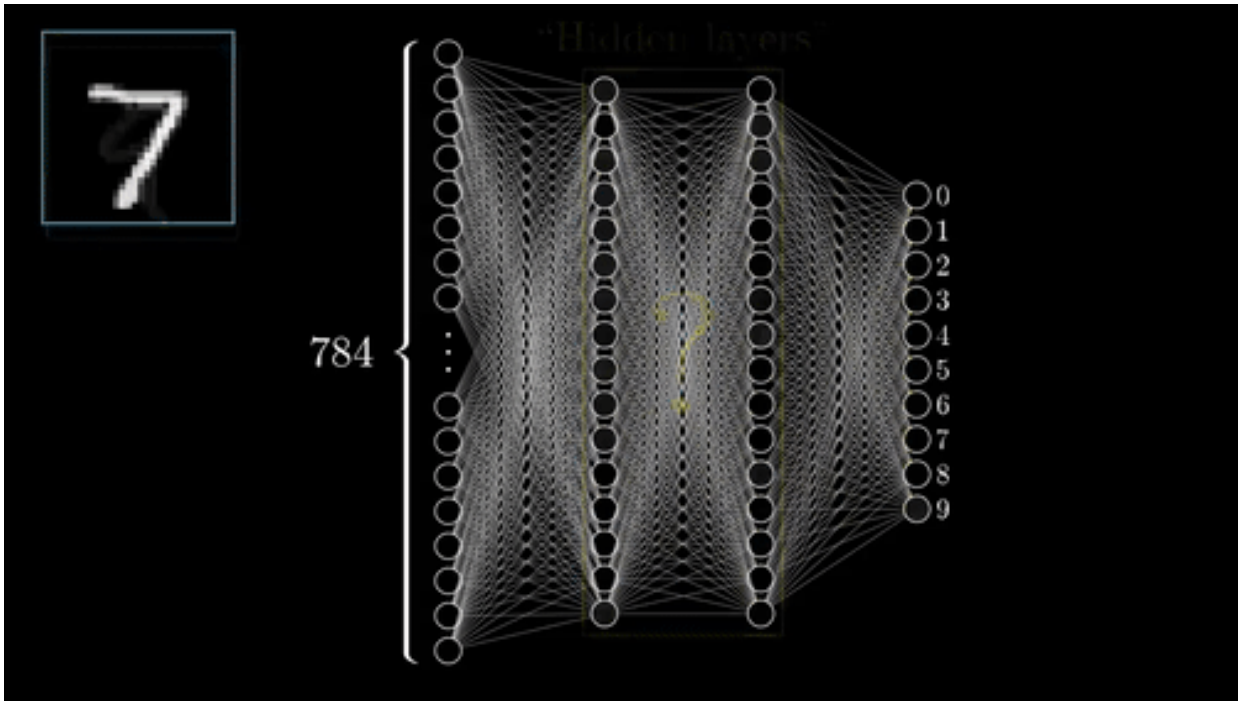
Cat score **0.9**

Dog score 0.1

Overview of Neural Networks

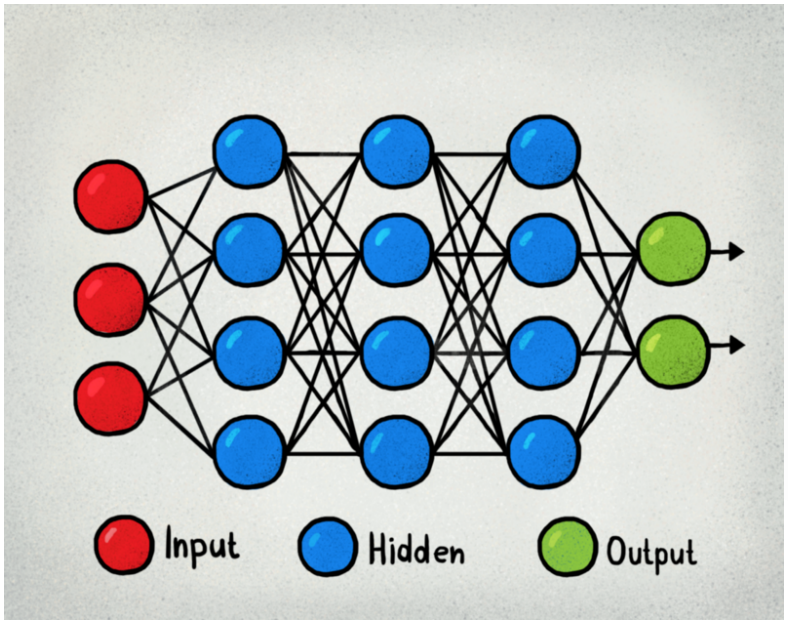


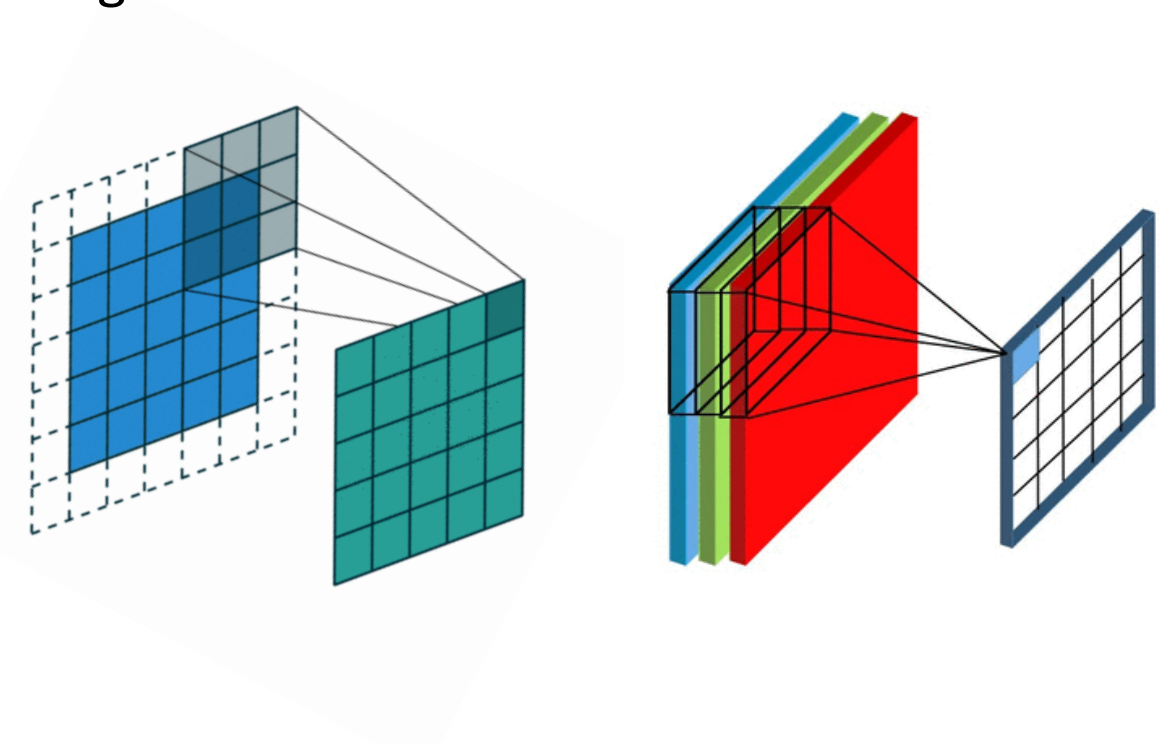
Overview of Neural Networks



Convolutional Neural Networks

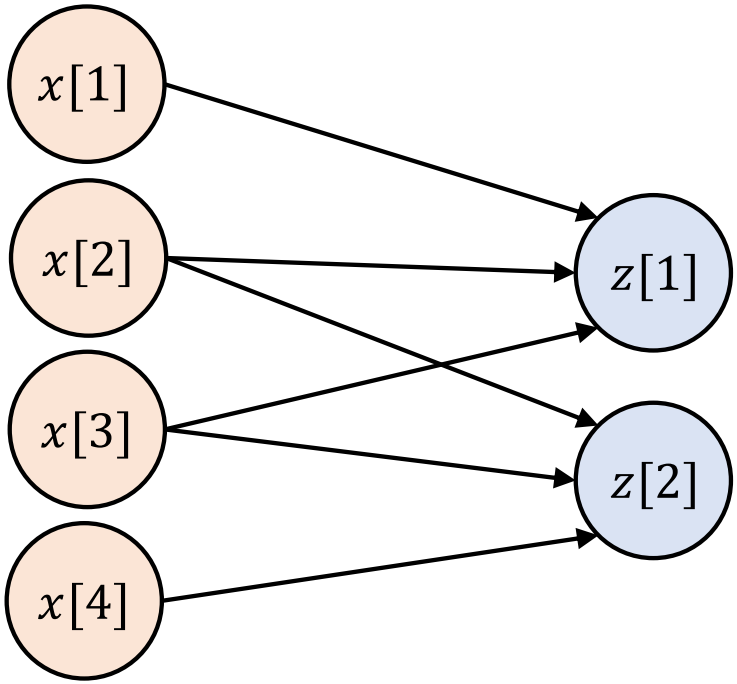
- A current node is not directly affected by “all nodes in the previous layer”
- The network “weights” on the edges can be “re-used”

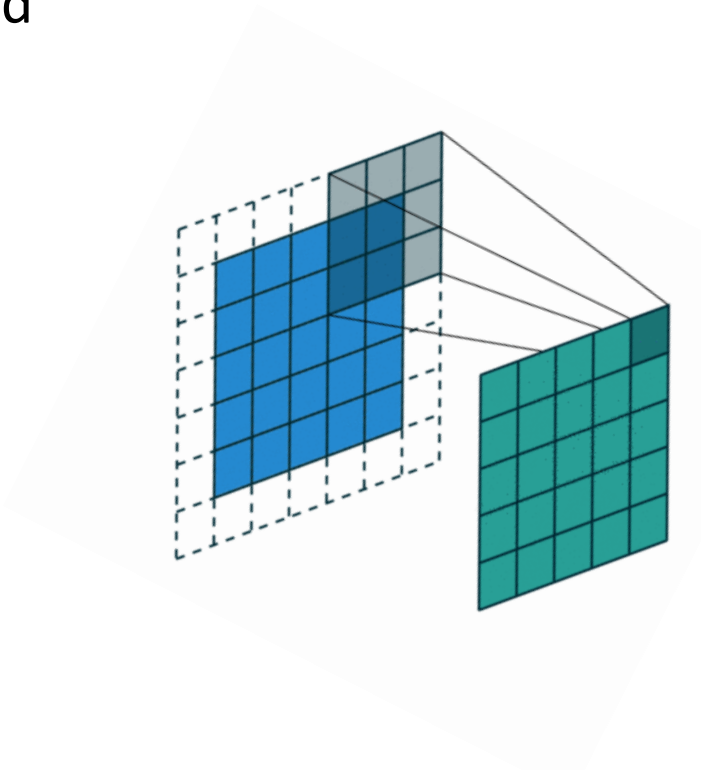




Convolutional Neural Networks

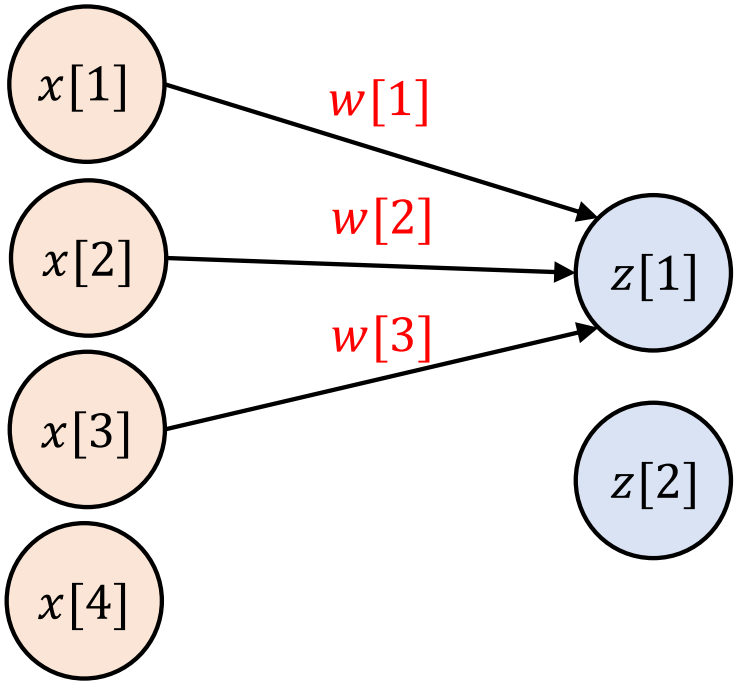
- A current node is not directly affected by "all nodes in the previous layer"
- The network "weights" on the edges can be "re-used"

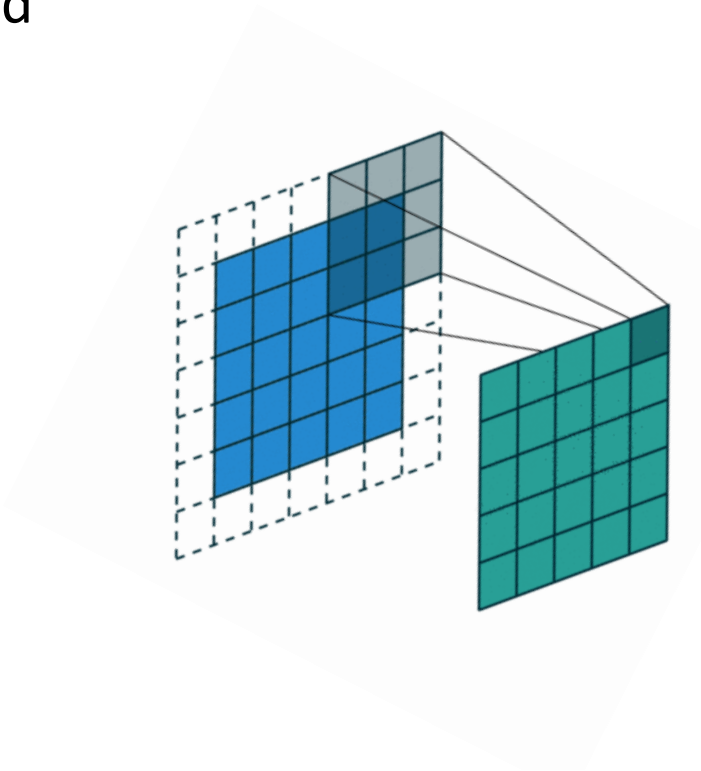




Convolutional Neural Networks

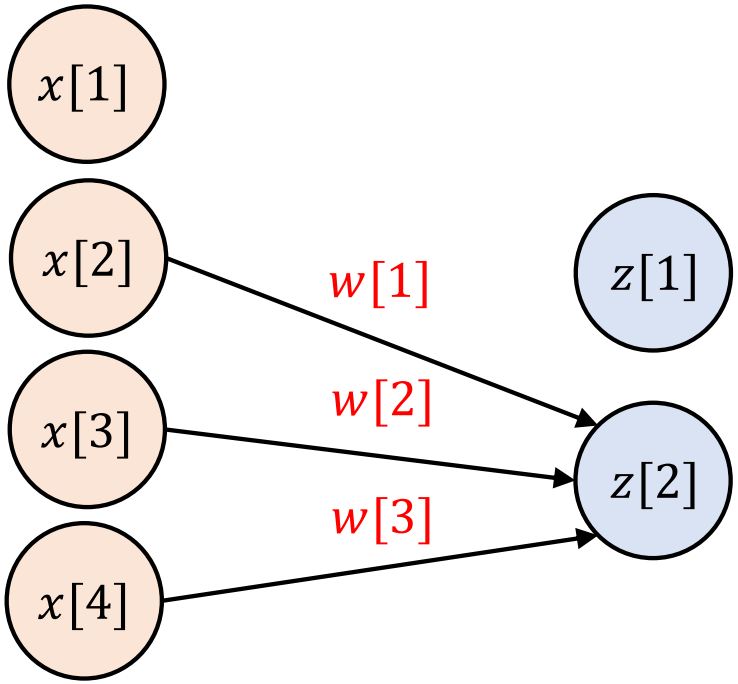
- A current node is not directly affected by "all nodes in the previous layer"
- The network "weights" on the edges can be "re-used"

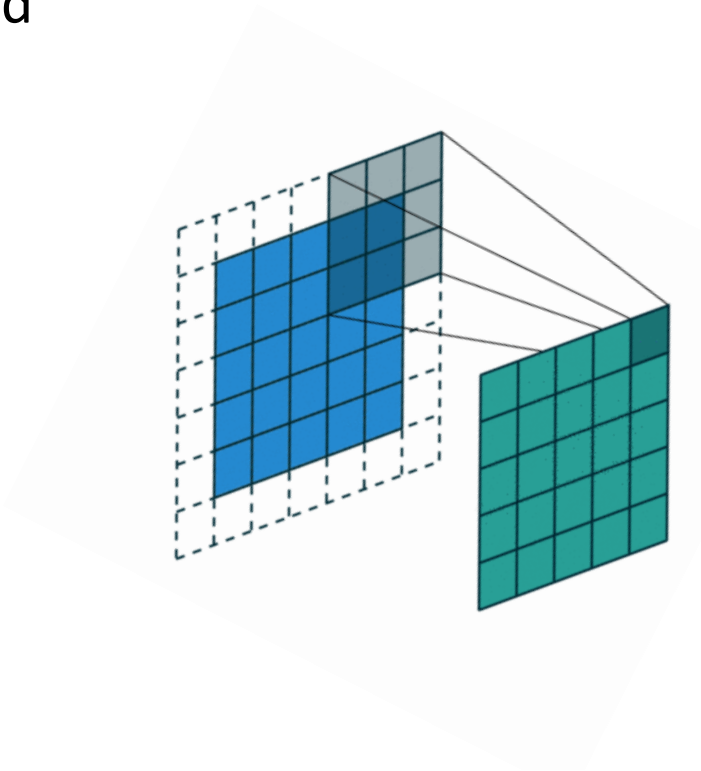




Convolutional Neural Networks

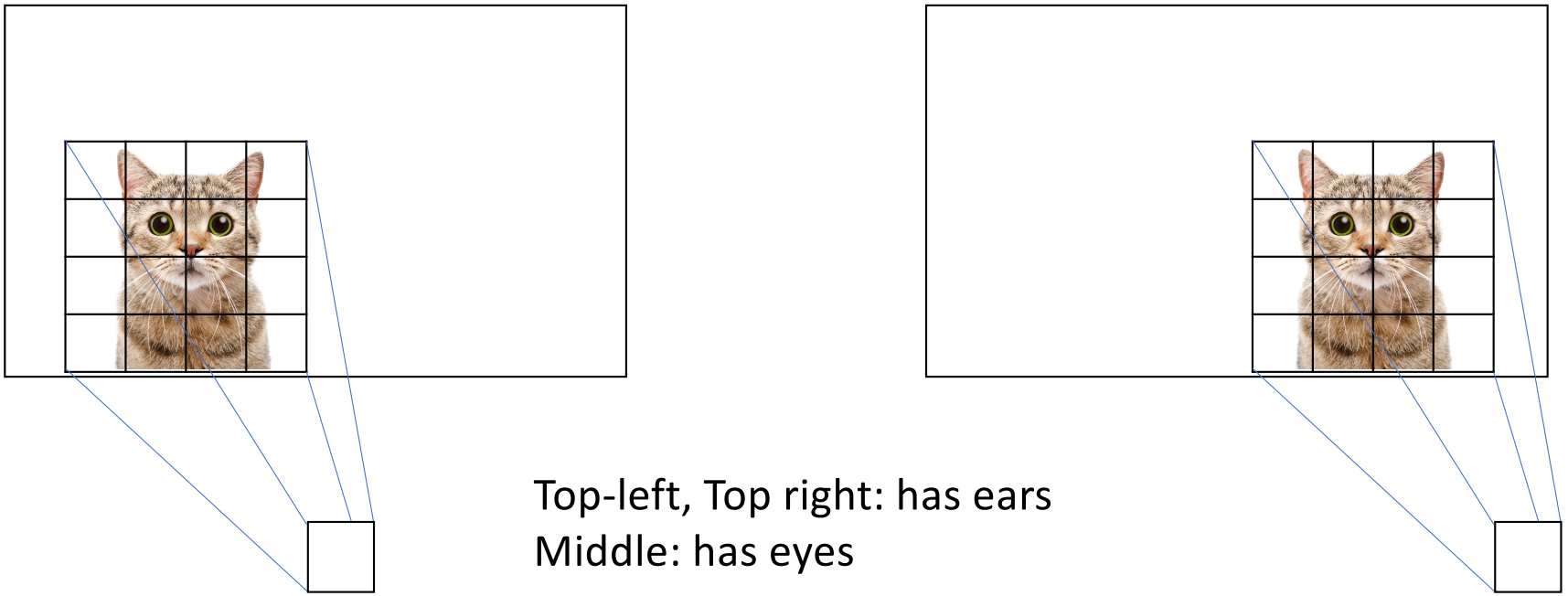
- A current node is not directly affected by "all nodes in the previous layer"
- The network "weights" on the edges can be "re-used"





Convolutional Neural Networks

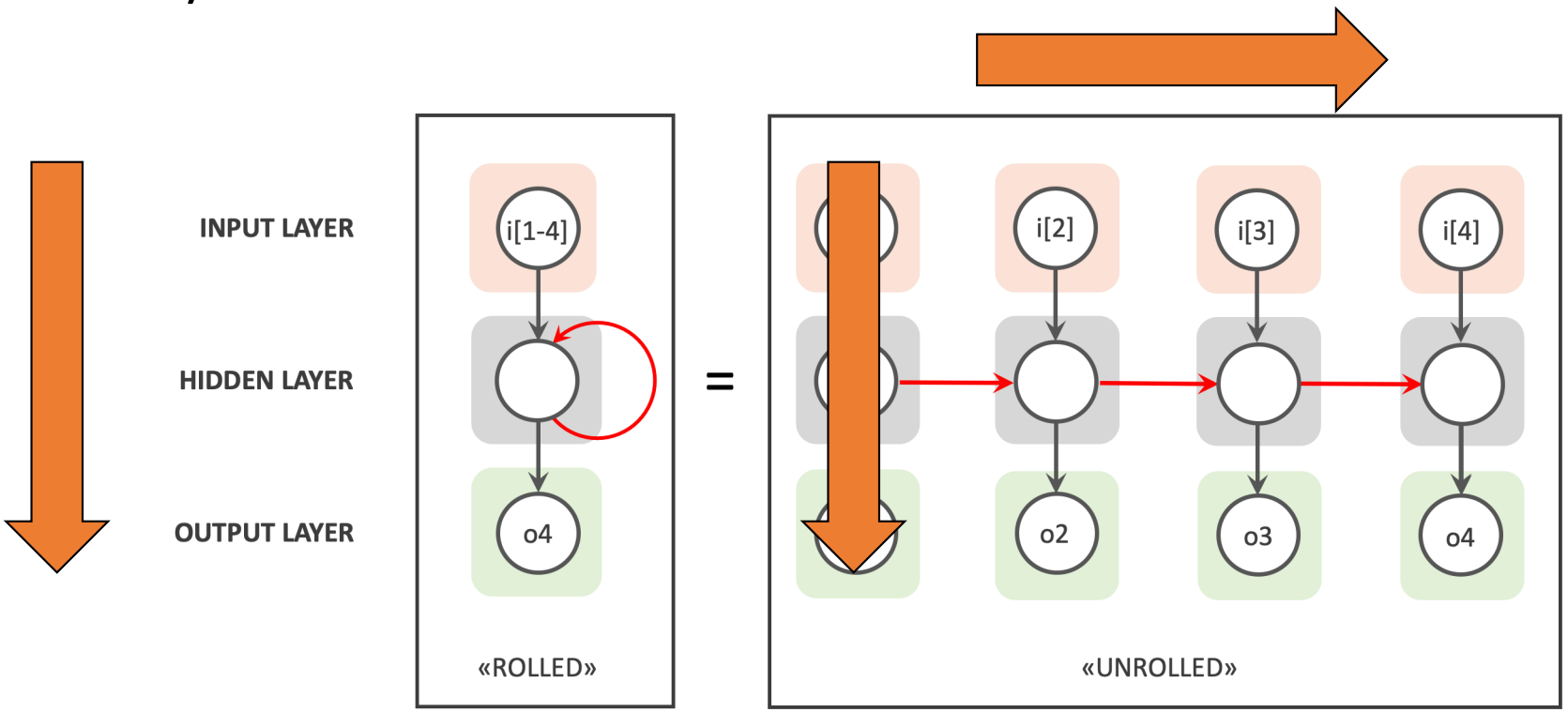
- “Local patterns” can show up at different image locations



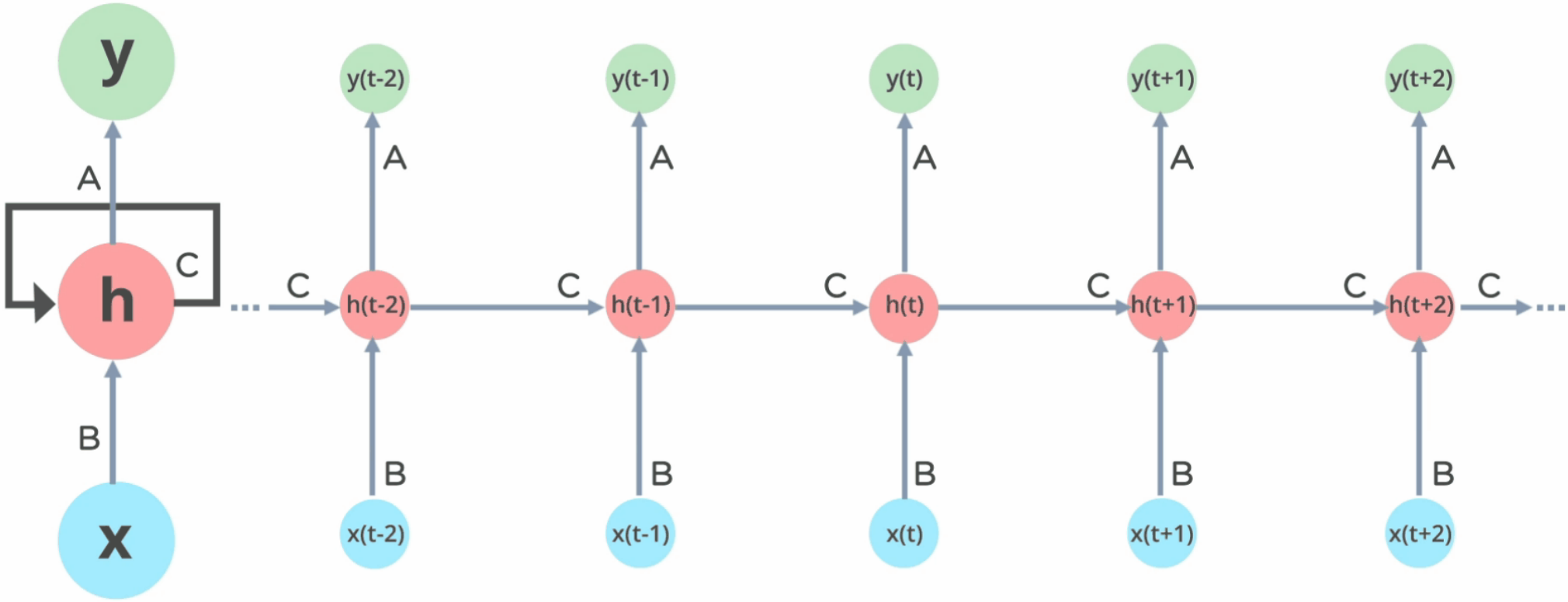
Top-left, Top right: has ears
Middle: has eyes

Recurrent Neural Network

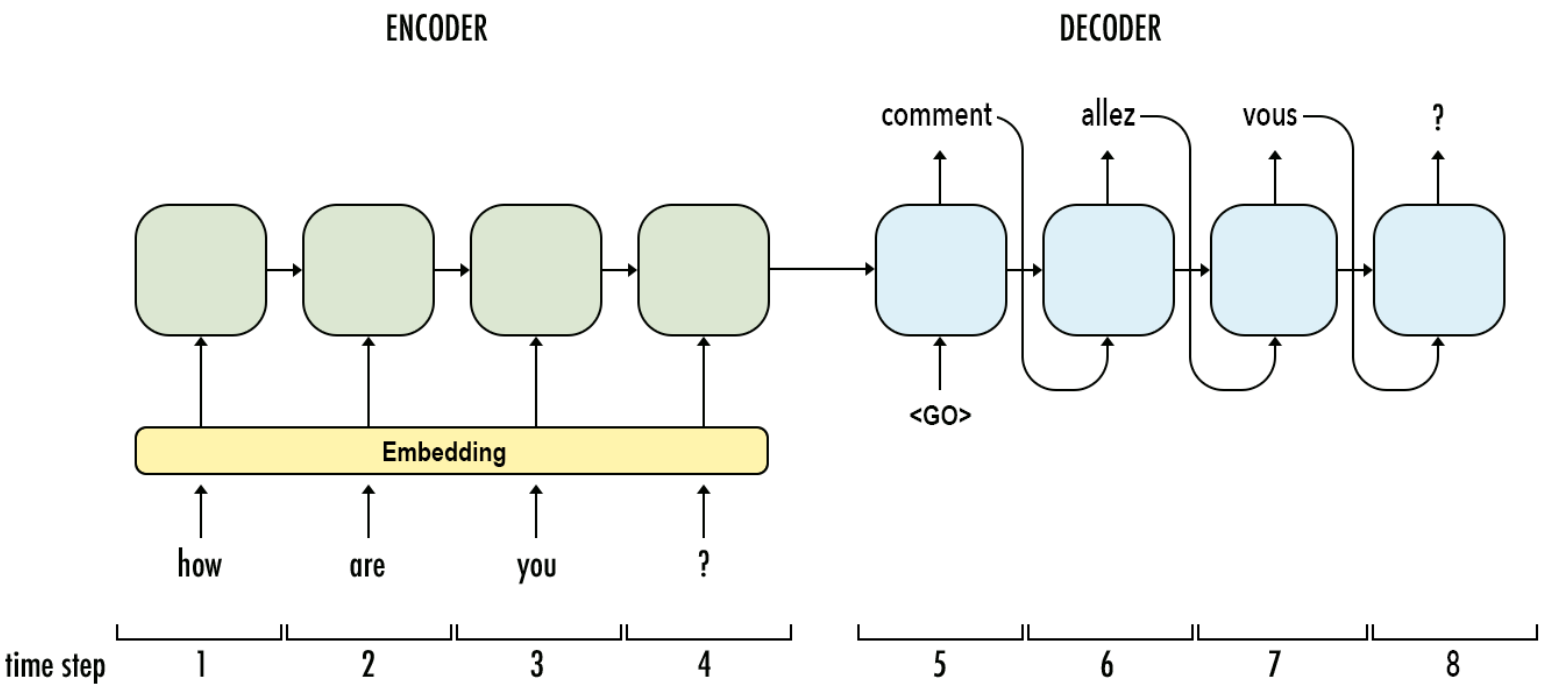
- Other styles: Recurrent neural network



Recurrent Neural Network



Recurrent Neural Network



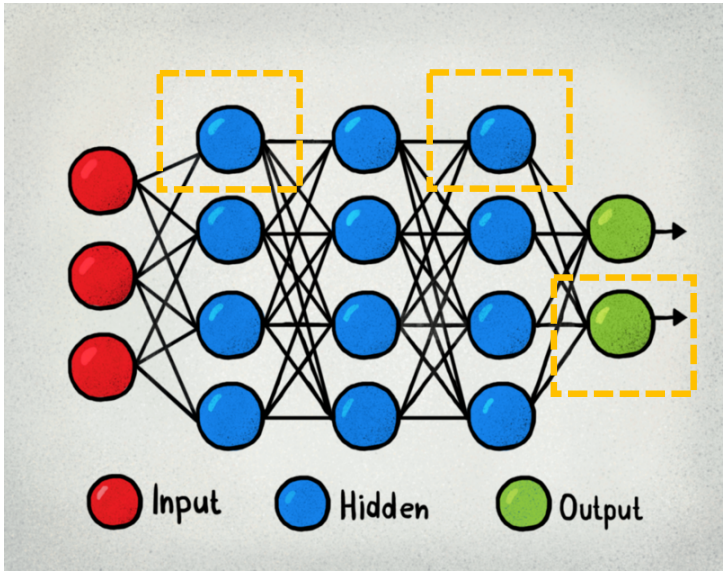
Today

- Overview
  - Feedforward neural network
  - Recurrent neural network
- Perceptron
- Multi-layer Perceptron (MLP)

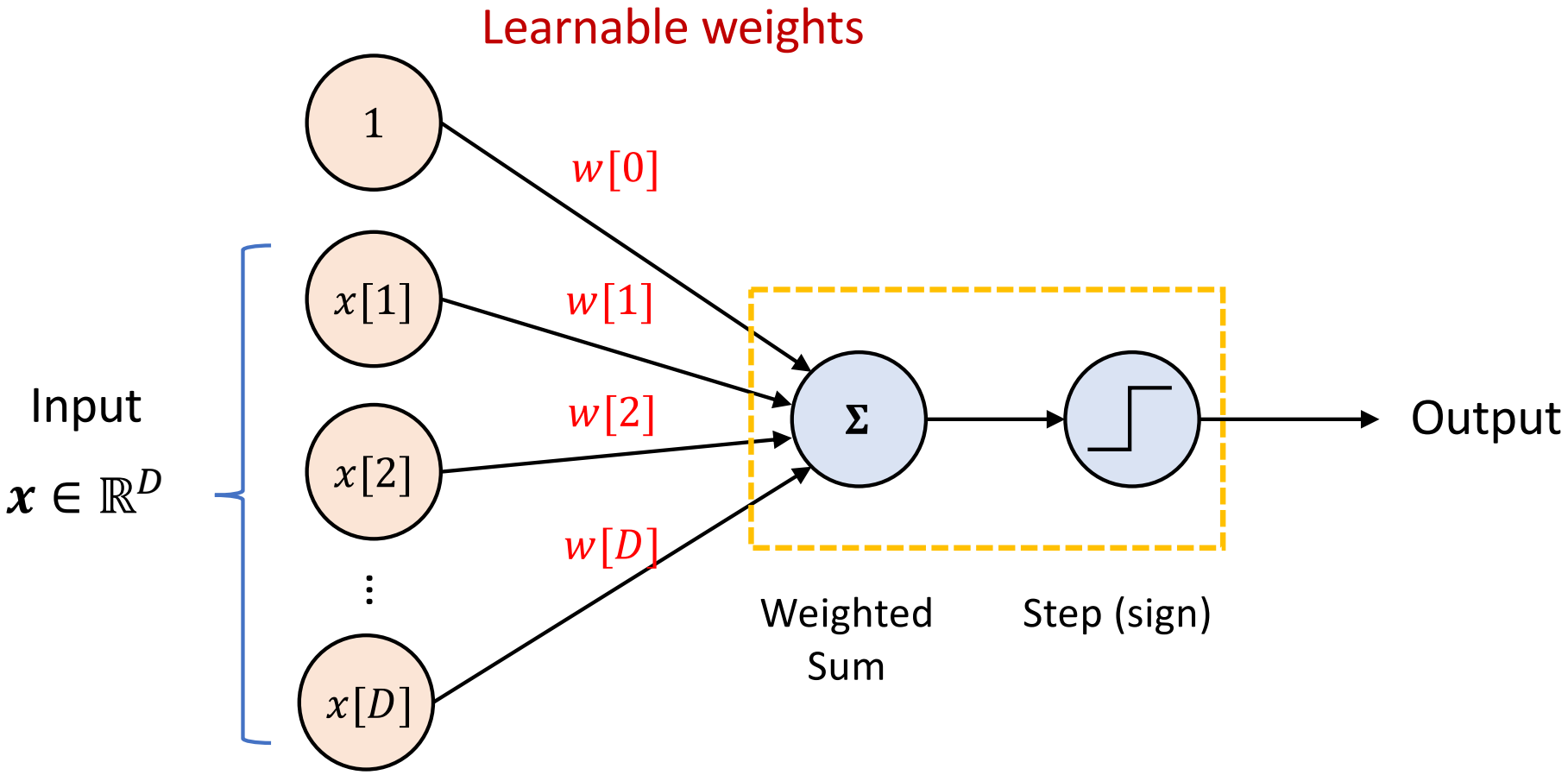
“Artificial” Neuron

- A basic “computational unit” of neural networks

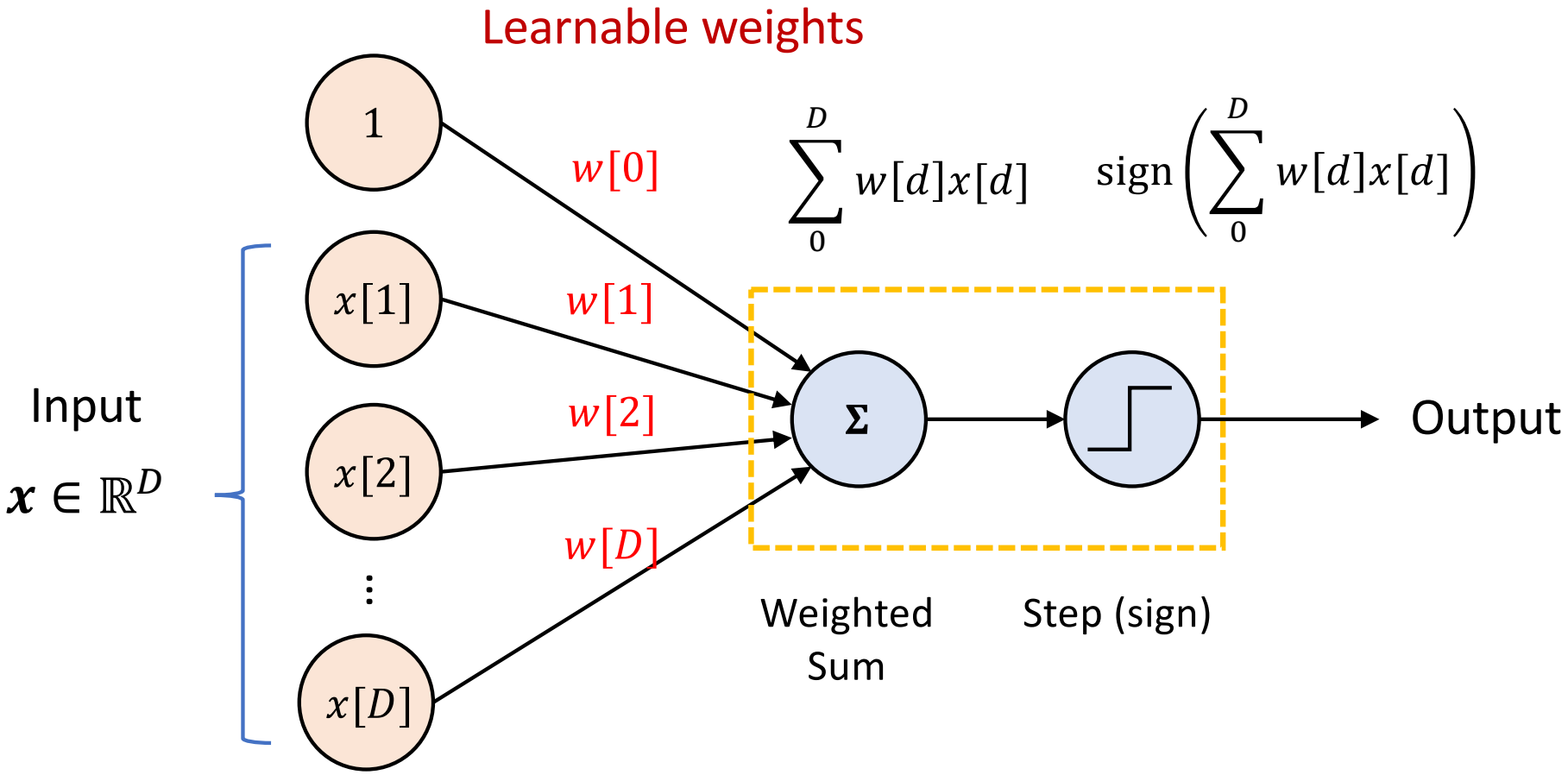
x $z^{(1)}$ $z^{(2)}$ $z^{(3)}$ y



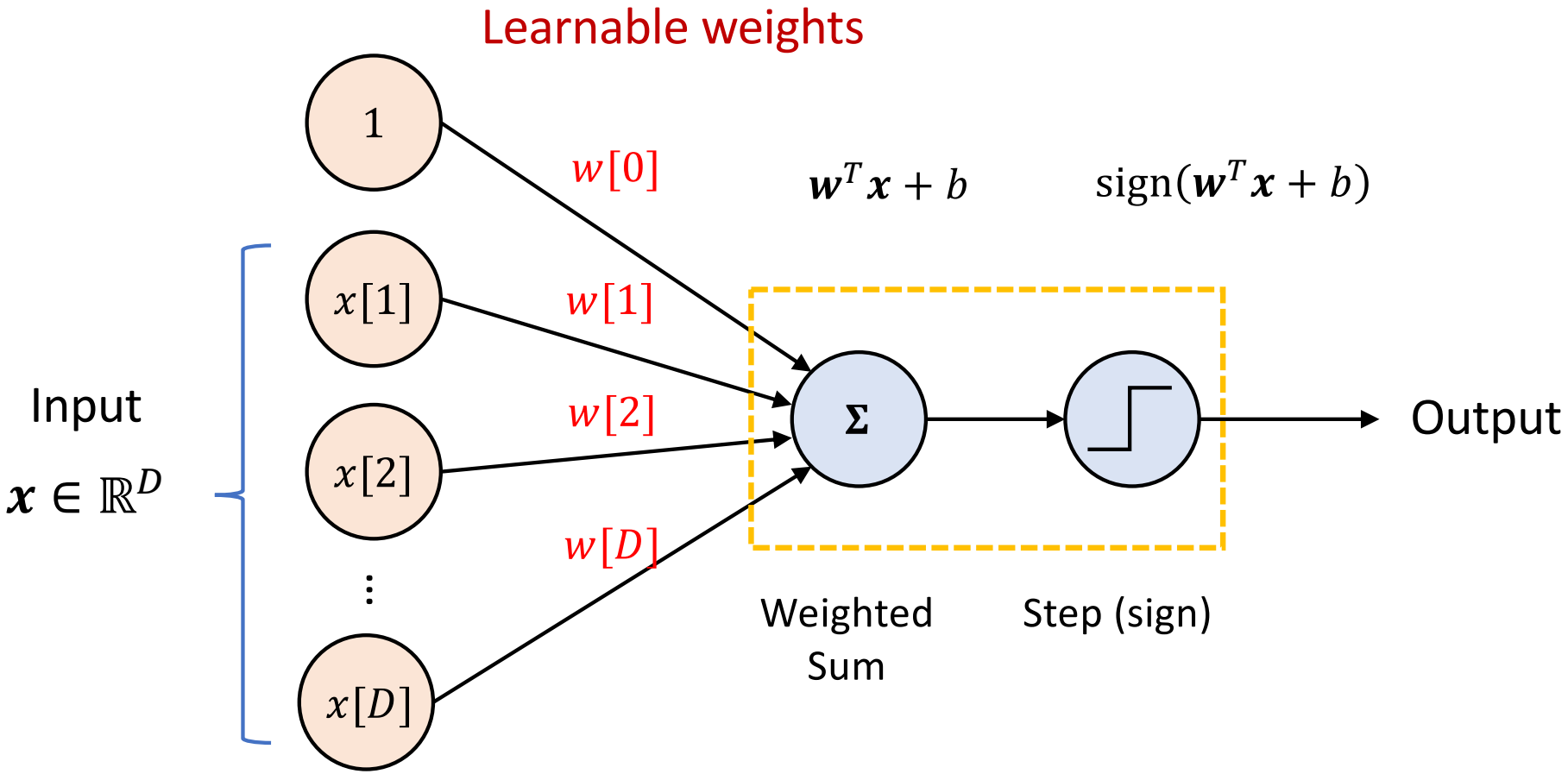
Perceptron (single neural for classification)



Perceptron (single neural for classification)

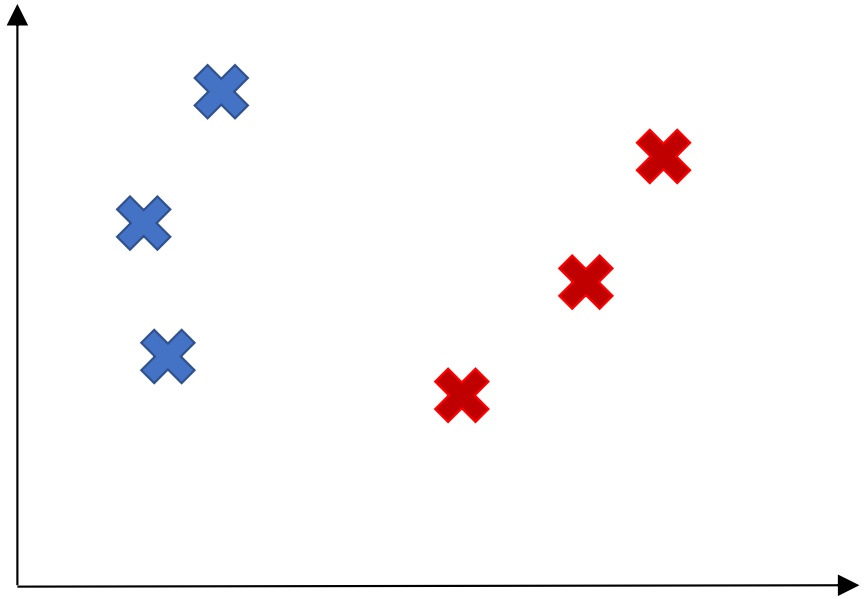


Perceptron (single neural for classification)



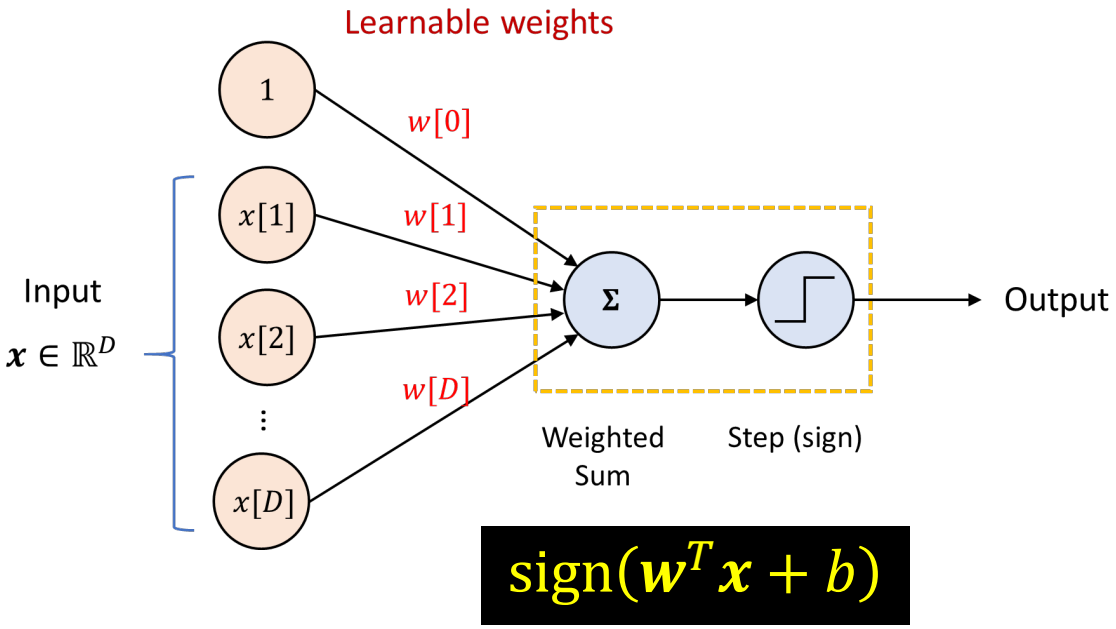
Perceptron (single neural for classification)

- Exercise:
  - Is perceptron a “linear classifier” that generates linear decision boundary?

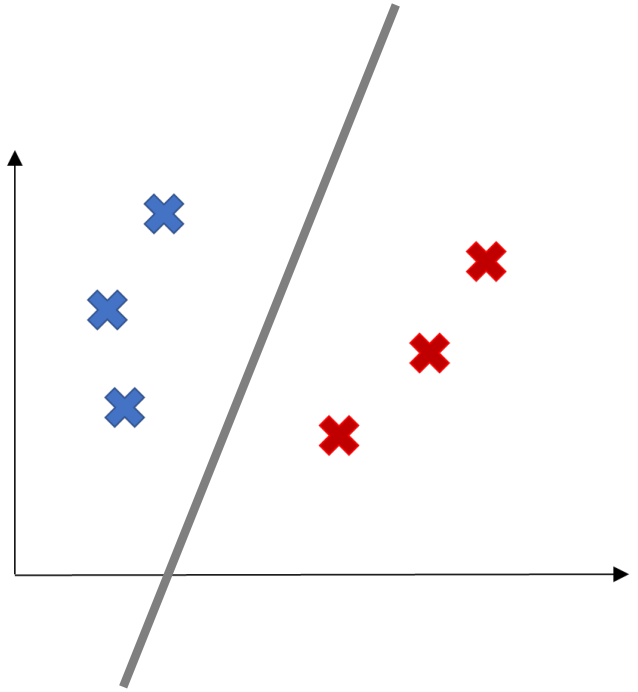


Perceptron Algorithm

- An algorithm to learn a perceptron for binary classification

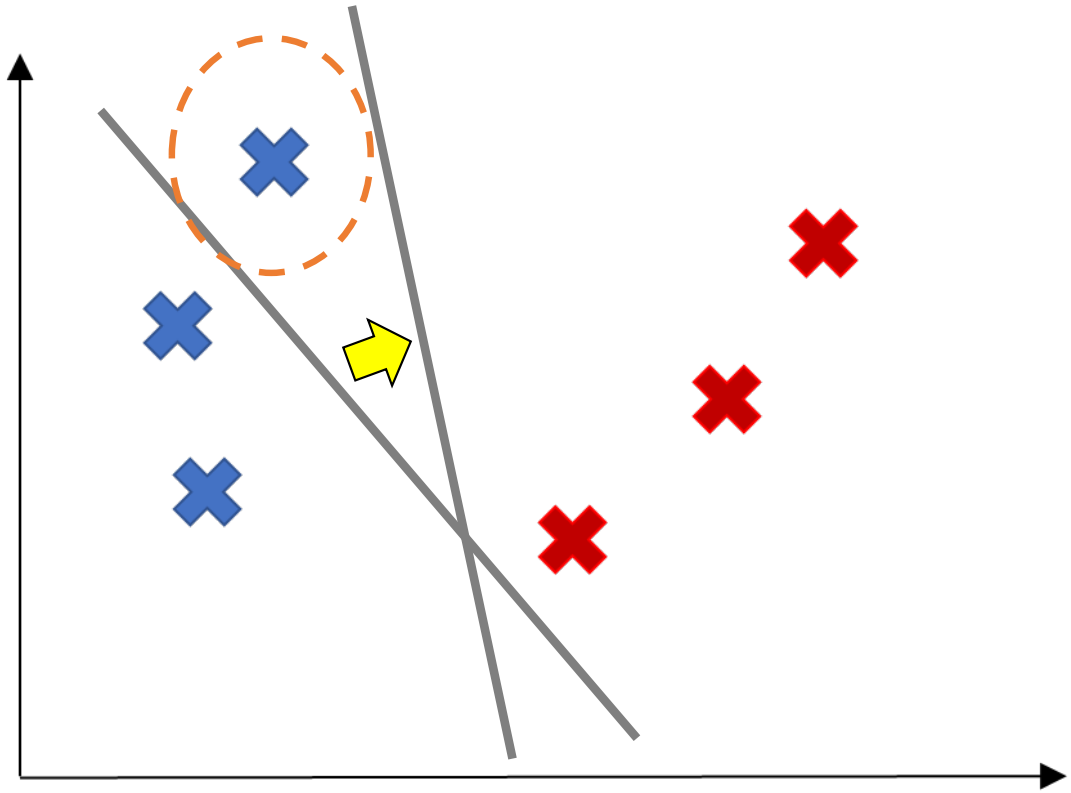


$$\text{sign}(\boldsymbol{w}^T\boldsymbol{x} + b)$$



Perceptron Algorithm

- Exercise:



Perceptron Algorithm

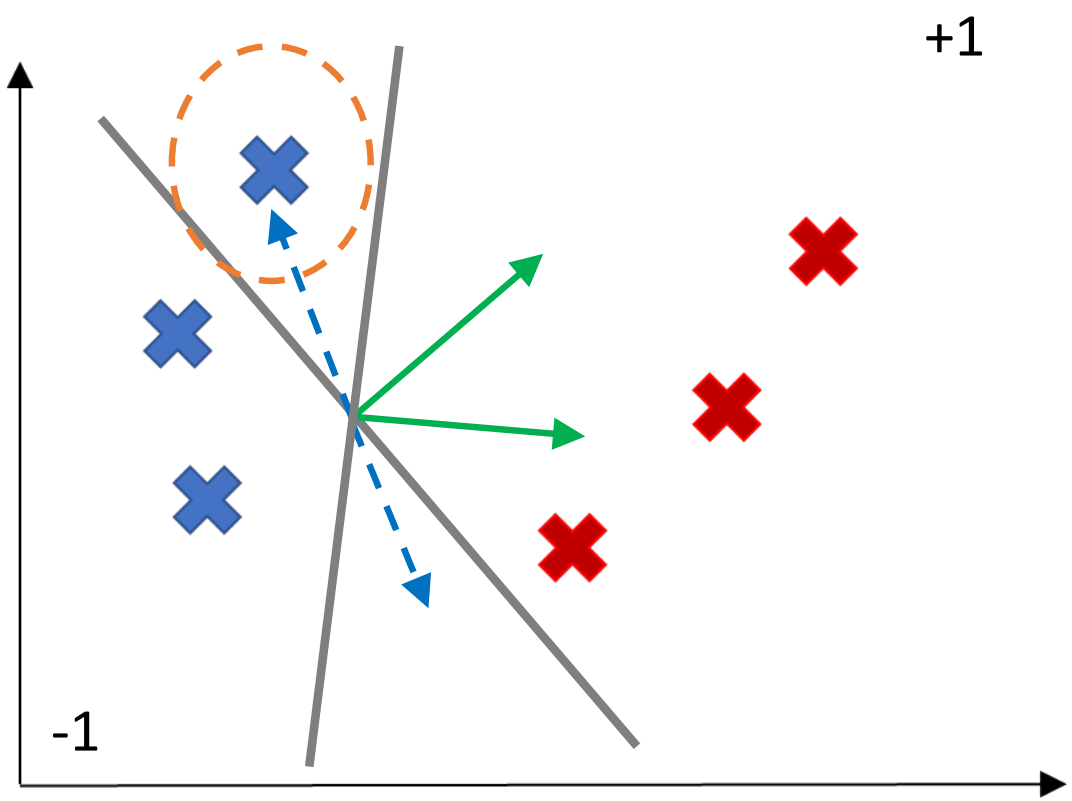
- Let $\boldsymbol{x} \in \mathbb{R}^D$ and $\boldsymbol{w} \in \mathbb{R}^D$ and ignore b, $y \in \{-1,1\}$

- Initialize weight vector $\boldsymbol{w} = \boldsymbol{0}$
- Loop for T iterations
  - Loop for all training examples $\boldsymbol{x}_n$ (random order!)
  - Predict $\hat{y}_n = \text{sign}(\boldsymbol{w}^T \boldsymbol{x}_n)$
  - If $\hat{y}_n \neq y_n$
    - Update: $\boldsymbol{w} \leftarrow \boldsymbol{w} + \eta(y_n \boldsymbol{x}_n)$

Perceptron Algorithm

- Exercise:

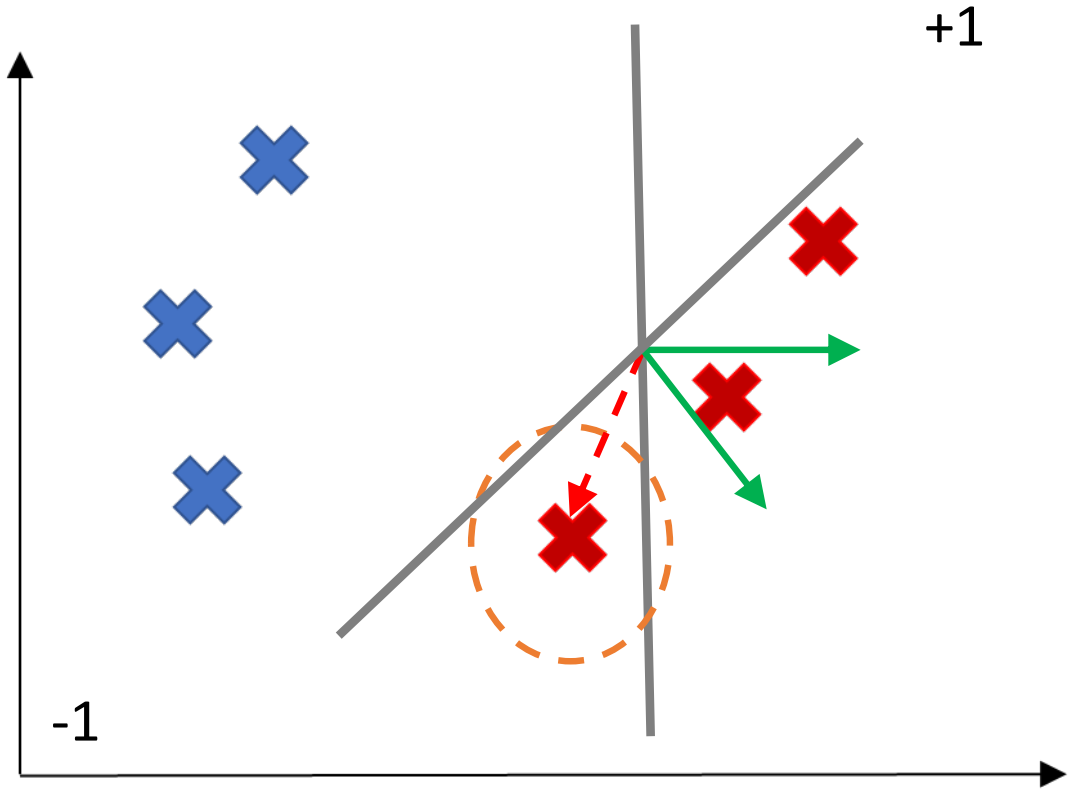
$$\boldsymbol{w} \leftarrow \boldsymbol{w} + \eta(y_n \boldsymbol{x}_n)$$



Perceptron Algorithm

- Example:

$$\boldsymbol{w} \leftarrow \boldsymbol{w} + \eta(y_n \boldsymbol{x}_n)$$



Perceptron Algorithm

- Why does it work?
  - If $\hat{y}_n = \text{sign}(\boldsymbol{w}^T \boldsymbol{x}_n) = -1$ but $y_n = 1$
  - $(\boldsymbol{w} + \eta(y_n \boldsymbol{x}_n))^T \boldsymbol{x}_n = \boldsymbol{w}^T \boldsymbol{x}_n + \eta y_n \boldsymbol{x}_n{}^T \boldsymbol{x}_n \geq \boldsymbol{w}^T \boldsymbol{x}_n$

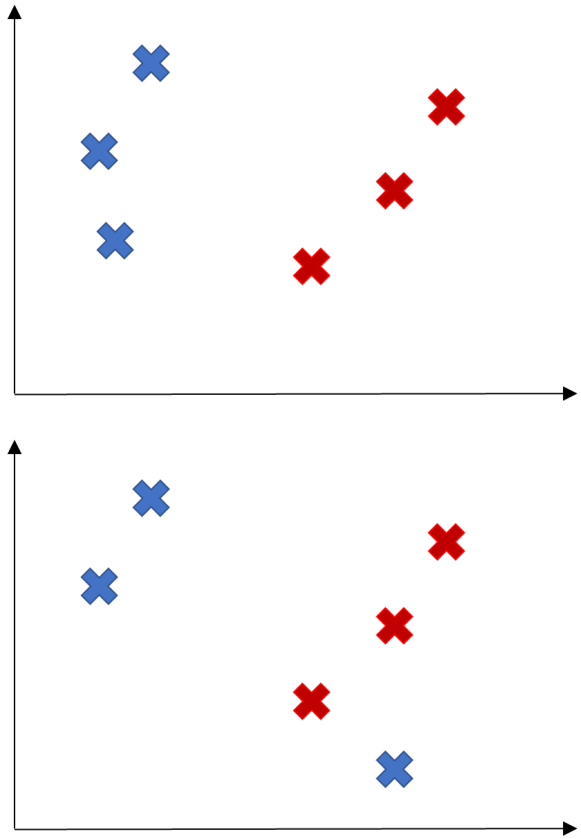
Perceptron Algorithm

- Let $\boldsymbol{x} \in \mathbb{R}^D$ and $\boldsymbol{w} \in \mathbb{R}^D$ and ignore b, $y \in \{-1,1\}$

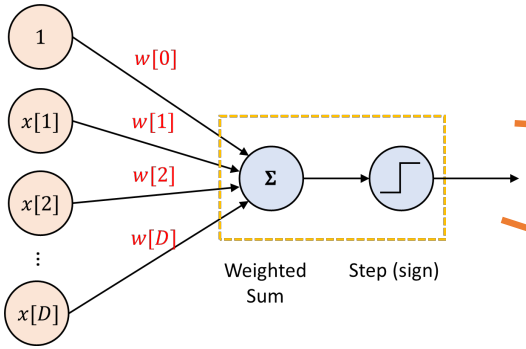
- Initialize weight vector $\boldsymbol{w} = \boldsymbol{0}$
- Online version: Get on training data $(\boldsymbol{x}, y)$
  - Predict $\hat{y} = \text{sign}(\boldsymbol{w}^T \boldsymbol{x})$
  - If $\hat{y} \neq y$
    - Update: $\boldsymbol{w} \leftarrow \boldsymbol{w} + \eta(y\boldsymbol{x})$

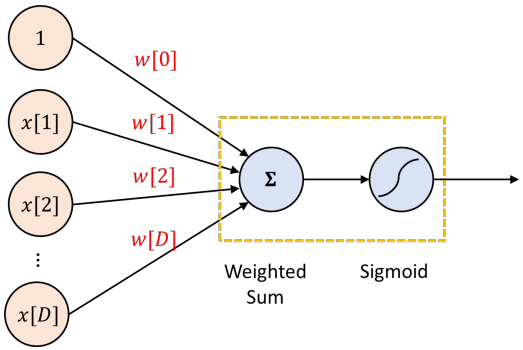
Perceptron Algorithm

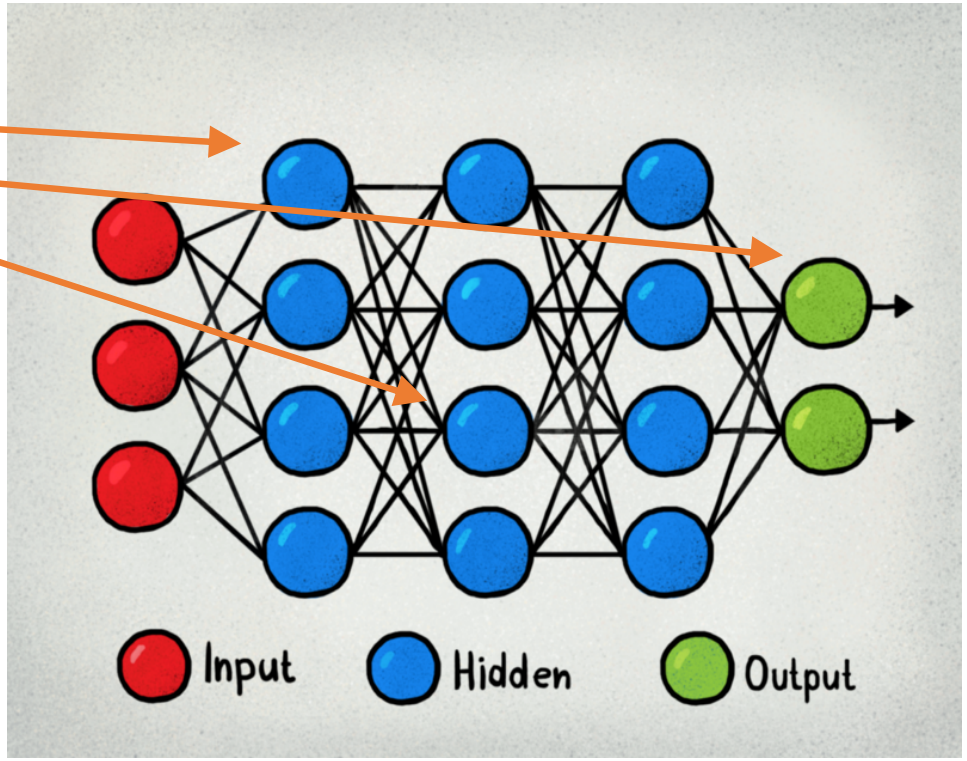
- Convergence:
  - If the data is linear separable
- Not convergence:
  - If the data is not linear separable



Multi-layer perceptron

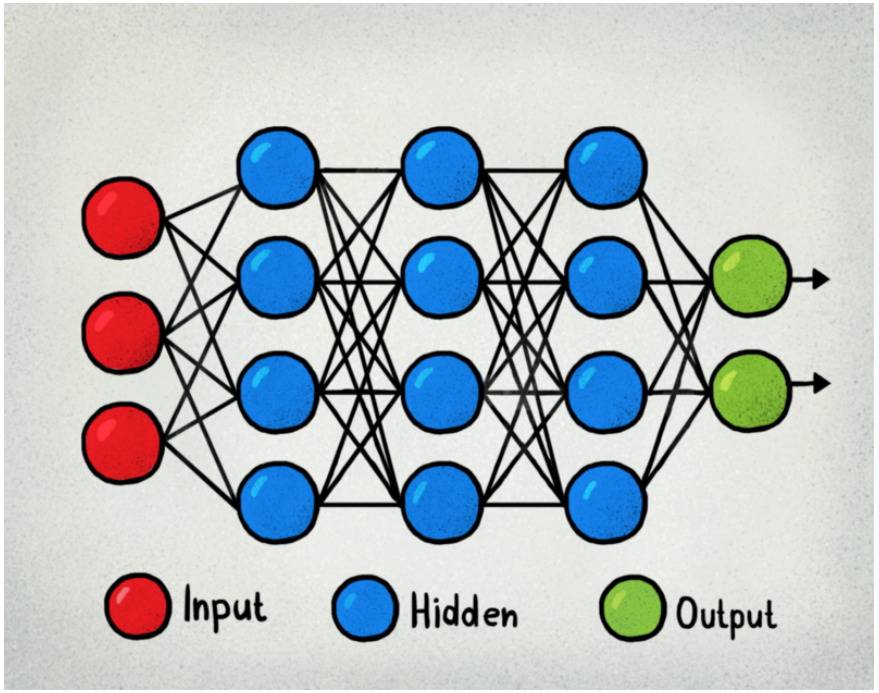






Multi-layer perceptron

- Learning Perceptron algorithm
  - Back-propagation
  - Mini-batch gradient descent



Multi-Layer Perceptron (MLP)