

Neural Network

Karina Hu

What is deep learning?
What is it used for?

Neural Network

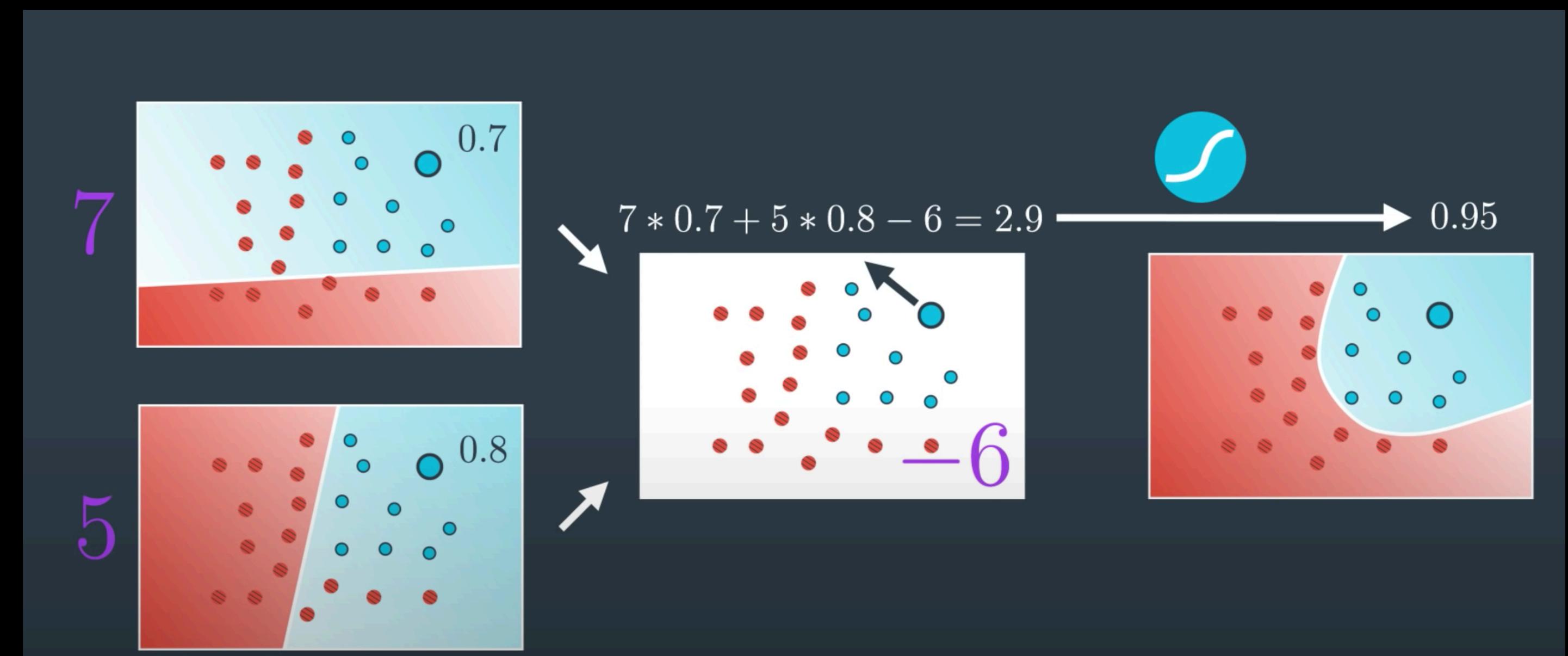
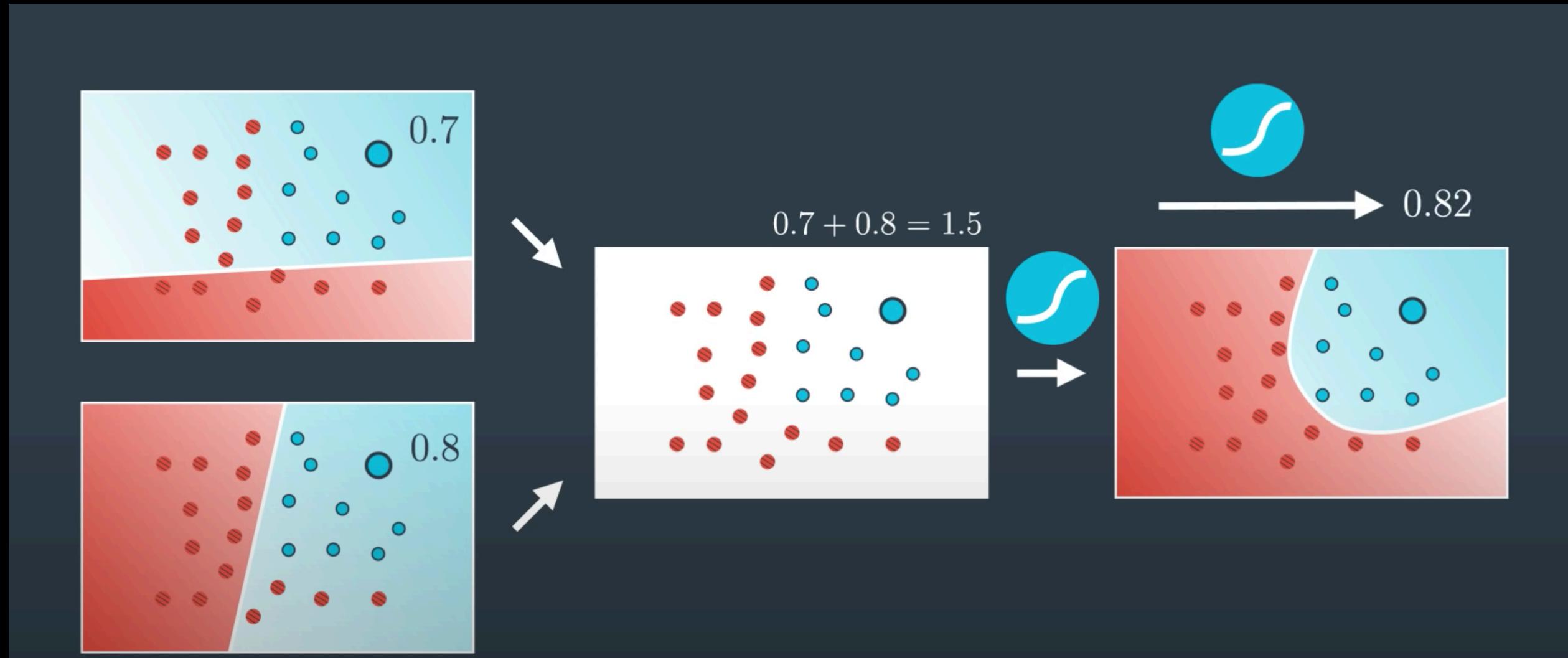
- Neural networks are an essential computational tools for language processing
- A modern neural network is a network of small computing units, each of which takes a vector of input values and produces a single output value.
- The use of modern neural nets is often called “deep learning”, because modern networks are often deep (have many layers)

Understanding the neural network

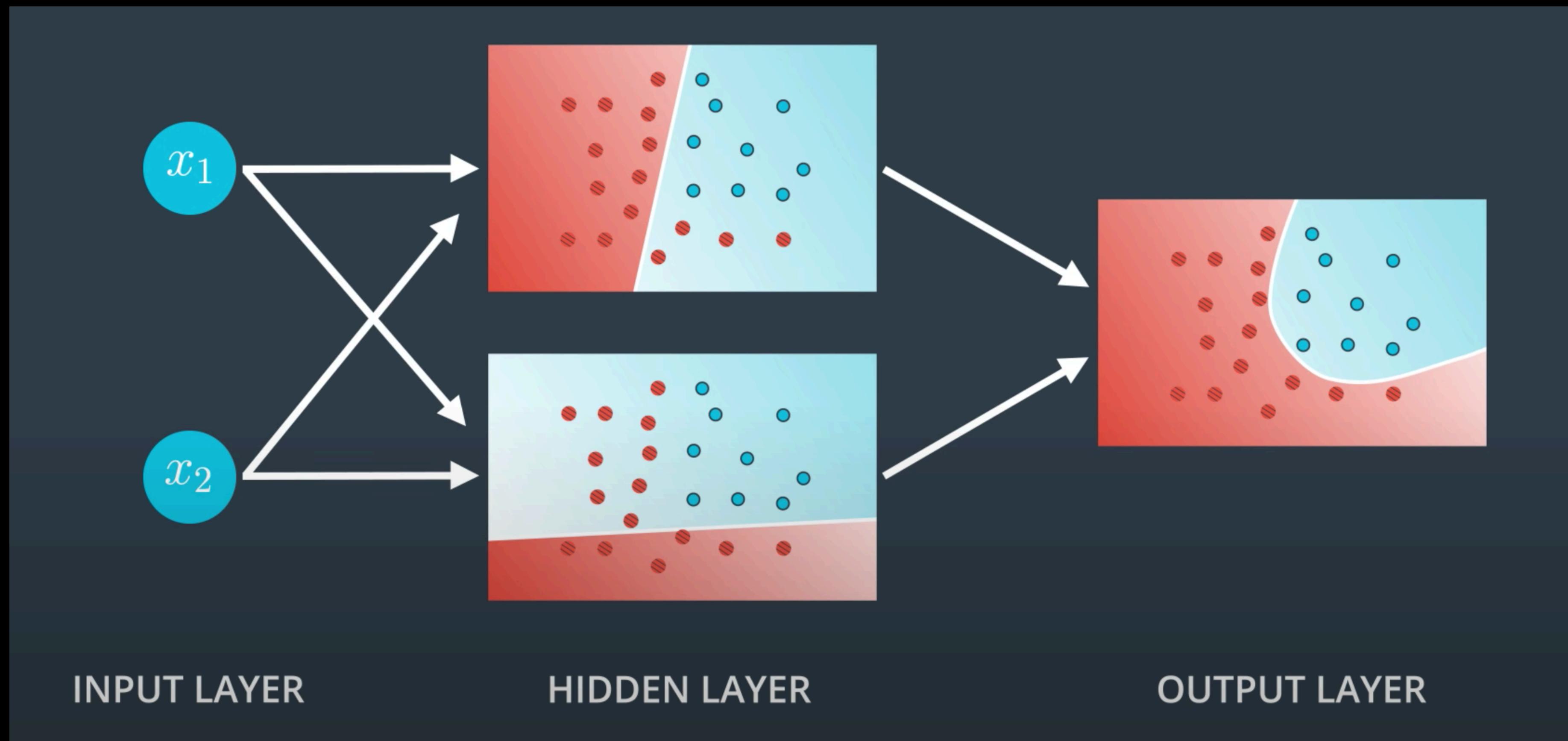
- Neuron: the basic processing unit of a neural network
- Bias terms : a neural units is taking a weighted sum of its inputs, with one additional term in the sum. A bias allows you to shift the activation function to the left or right. ($\text{output} = \text{sum}(\text{weights} * \text{input}) + \text{bias}$)
- Activation functions: sigmoid, Tanh, Rectified Linear Unit (ReLU)

Neural network architecture

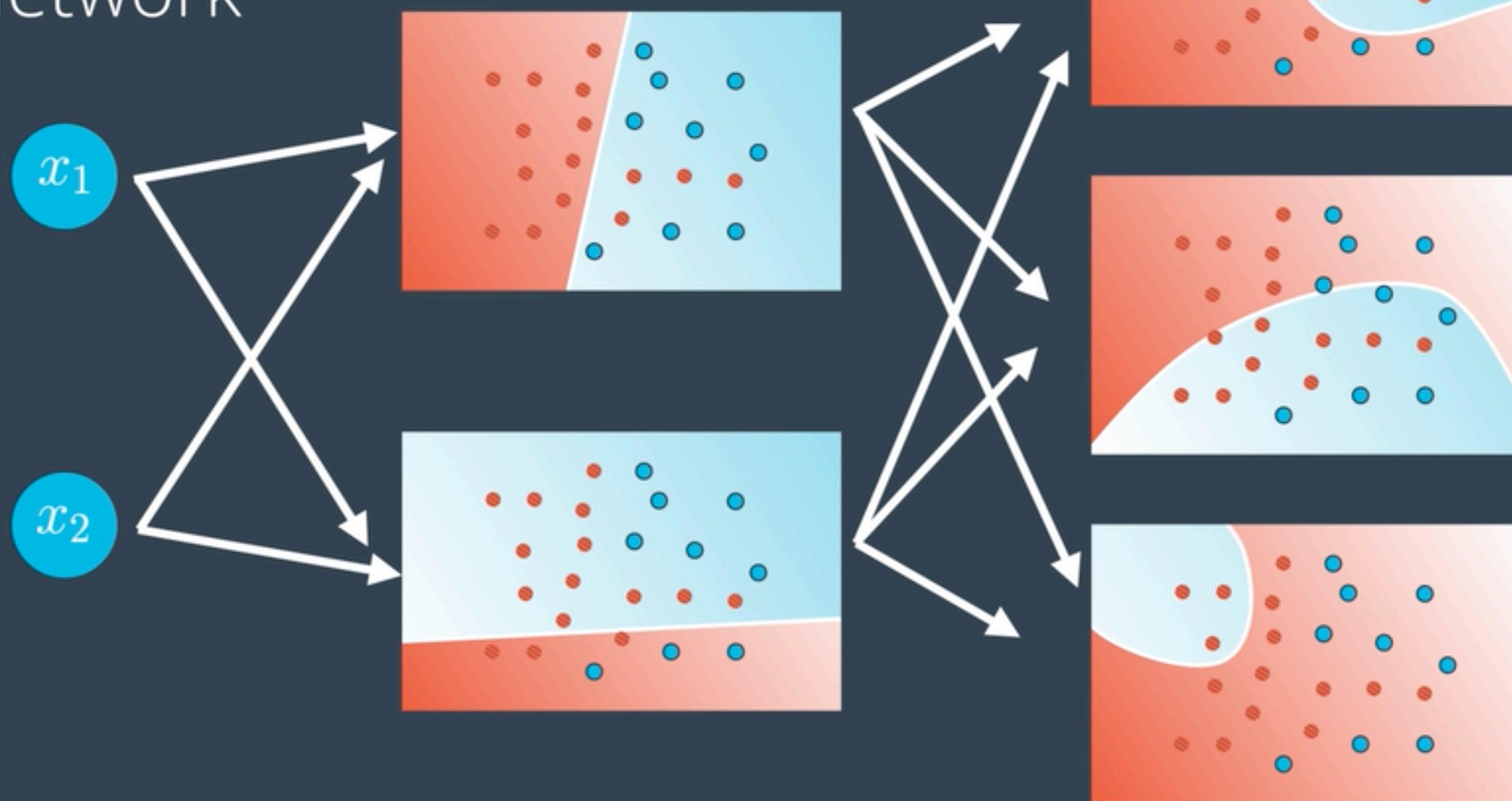
Put building blocks together and build great neural networks



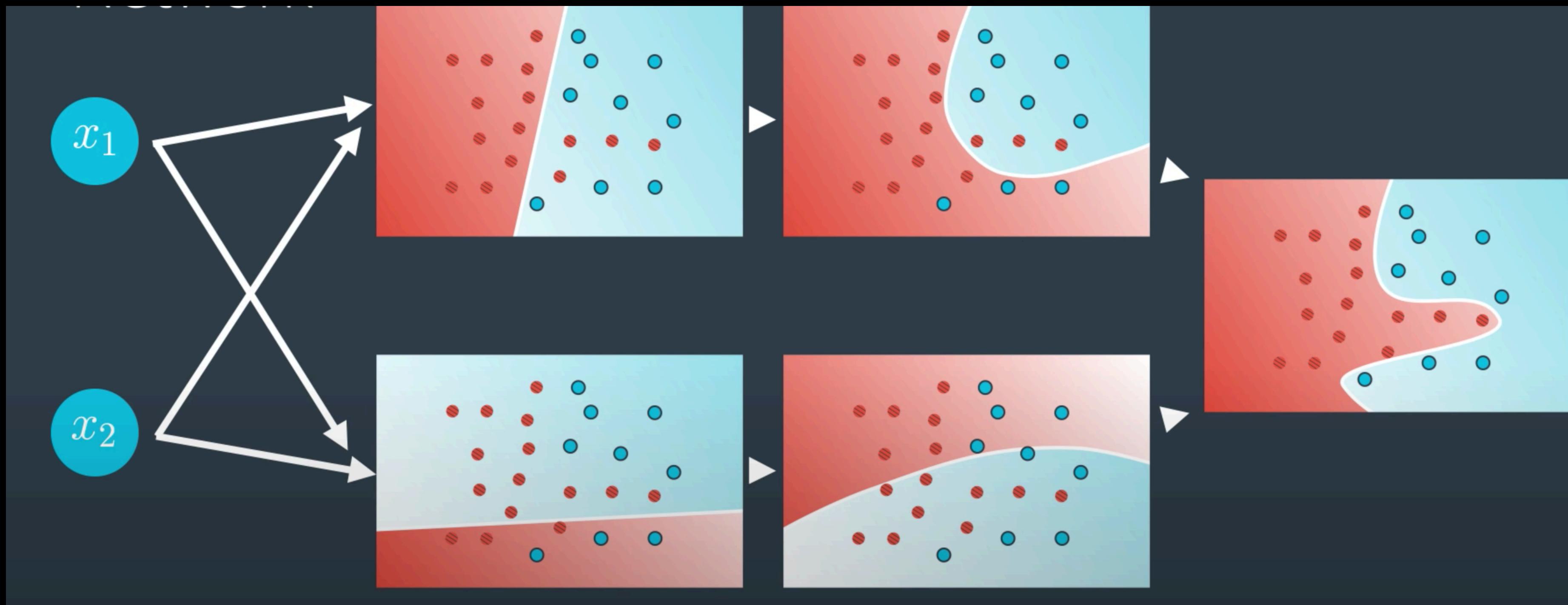
Deep neural network



Deep Neural Network



Deep neural network

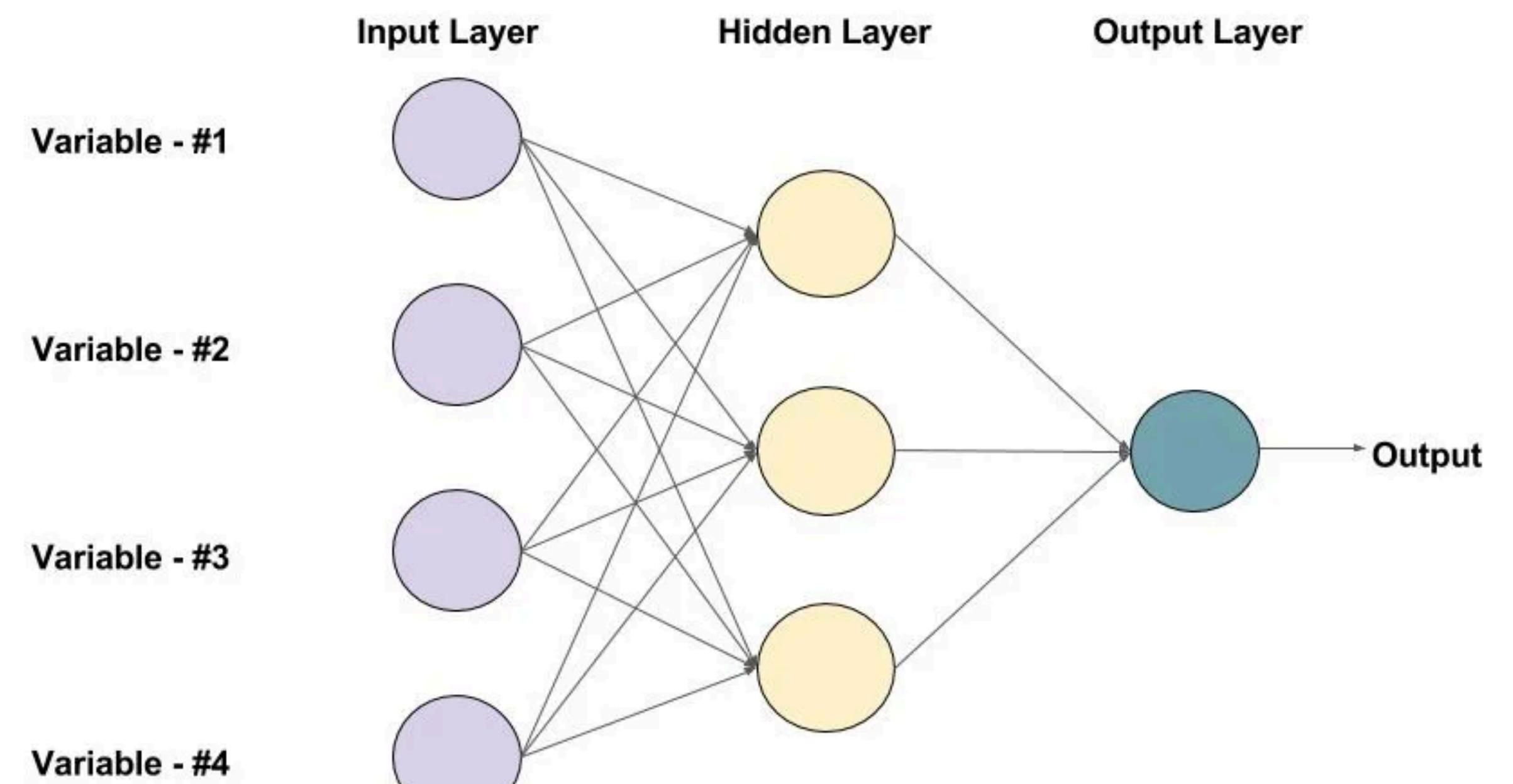


Feed-forward neural network

- A feed-forward network is a multilayer network in which the units are connected with no cycles; the outputs from units in each layer are passed to units in the next higher layer, and no outputs are passed back to lower layers.

Feed-forward network

The computation proceeds iteratively from one layer of units to the next

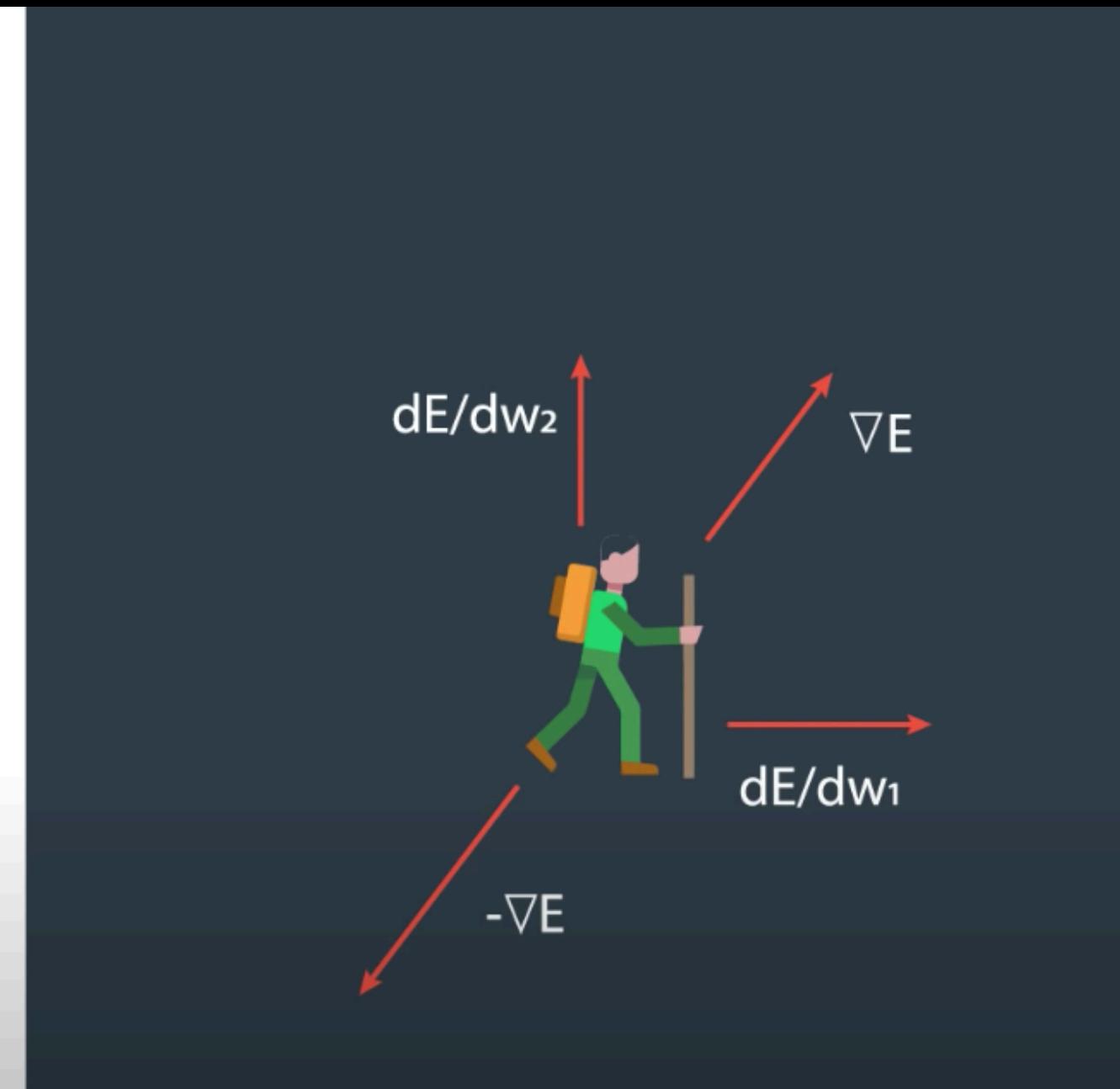
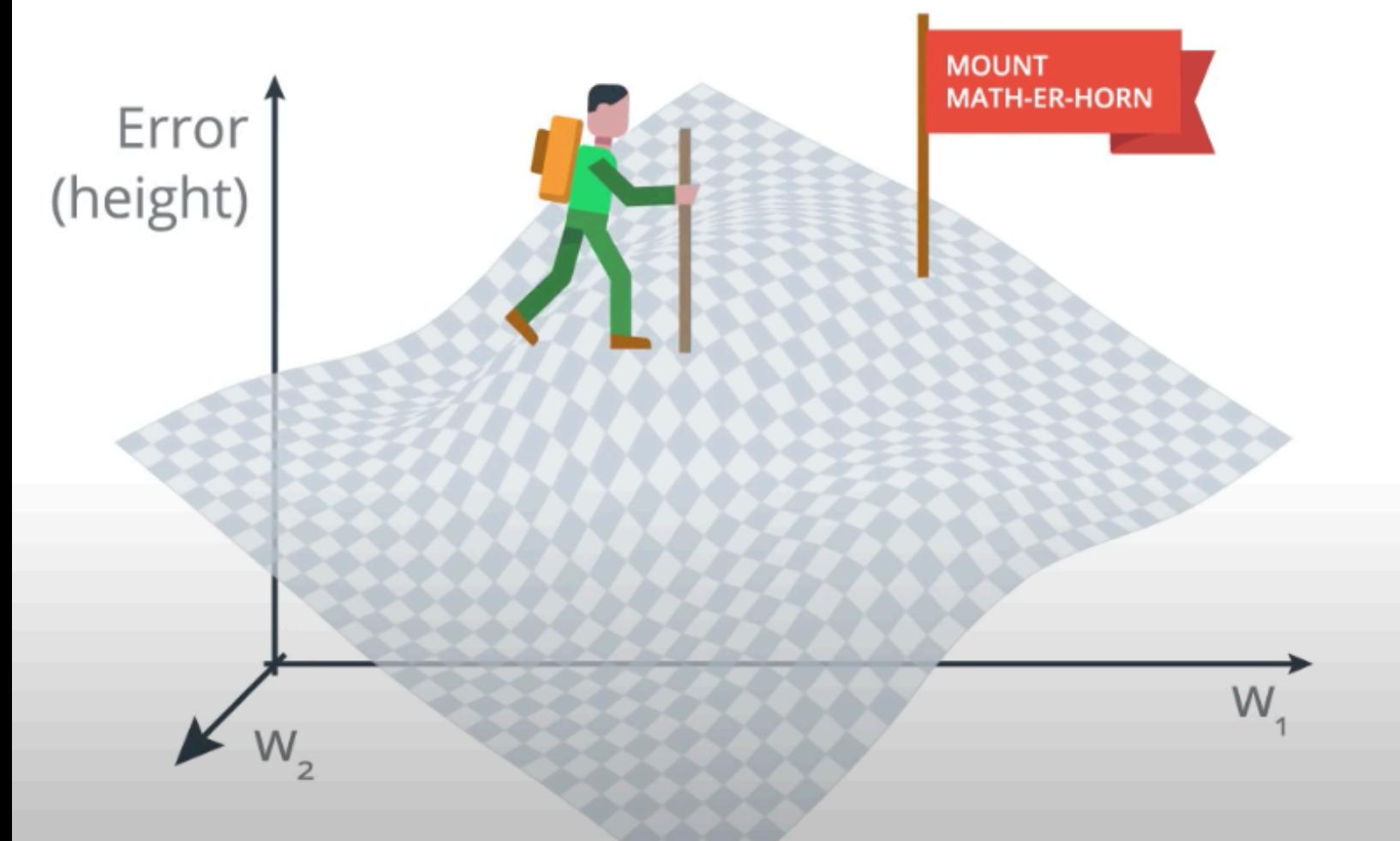


An example of a Feed-forward Neural Network with one hidden layer (with 3 neurons)

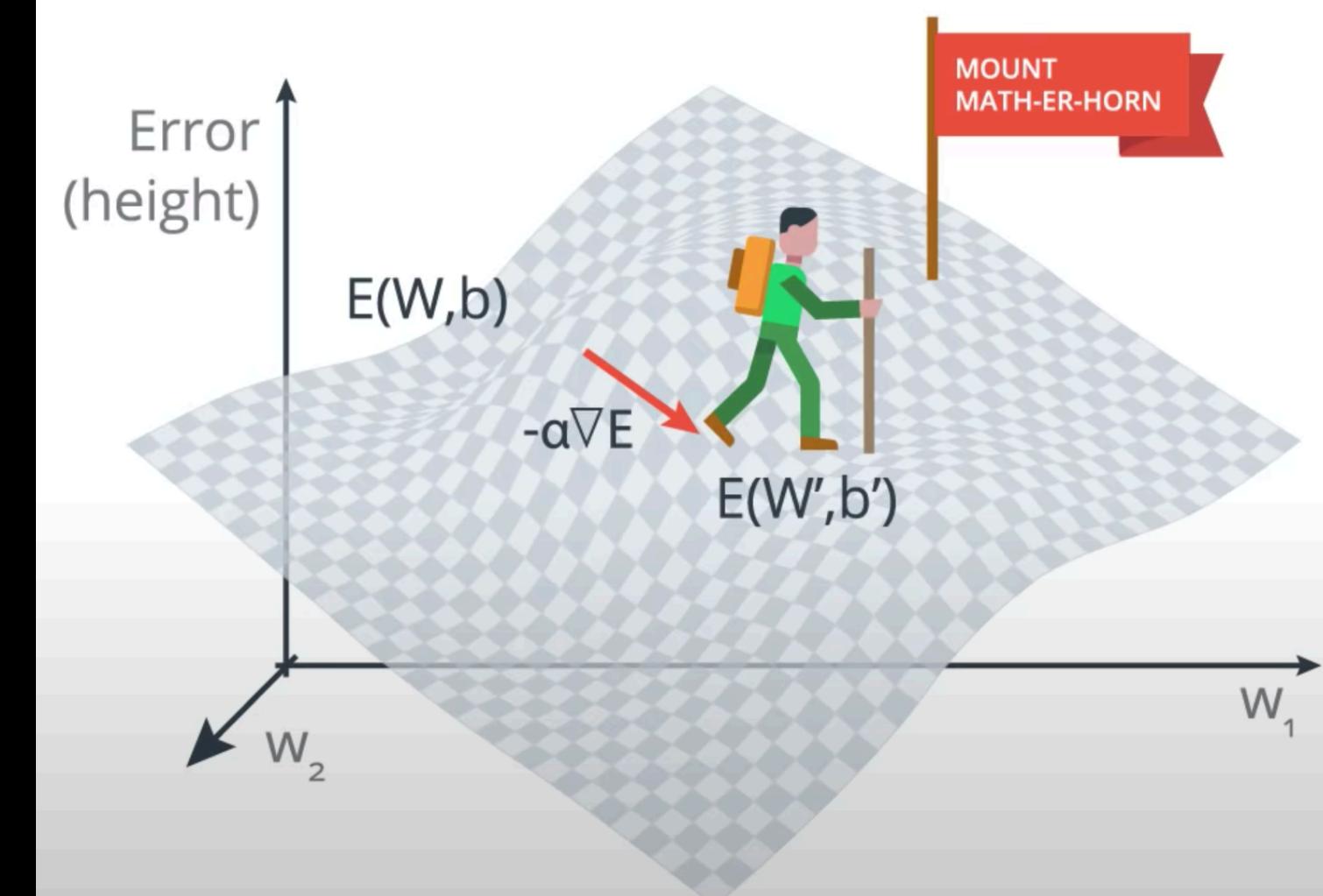
Training neural nets

- **Loss function**-models the distance between the system output and the gold output (common to use the loss used for logistic regression, the cross-entropy loss)
- **Gradient descent**-to find the parameters that minimizes this loss function

Gradient Descent



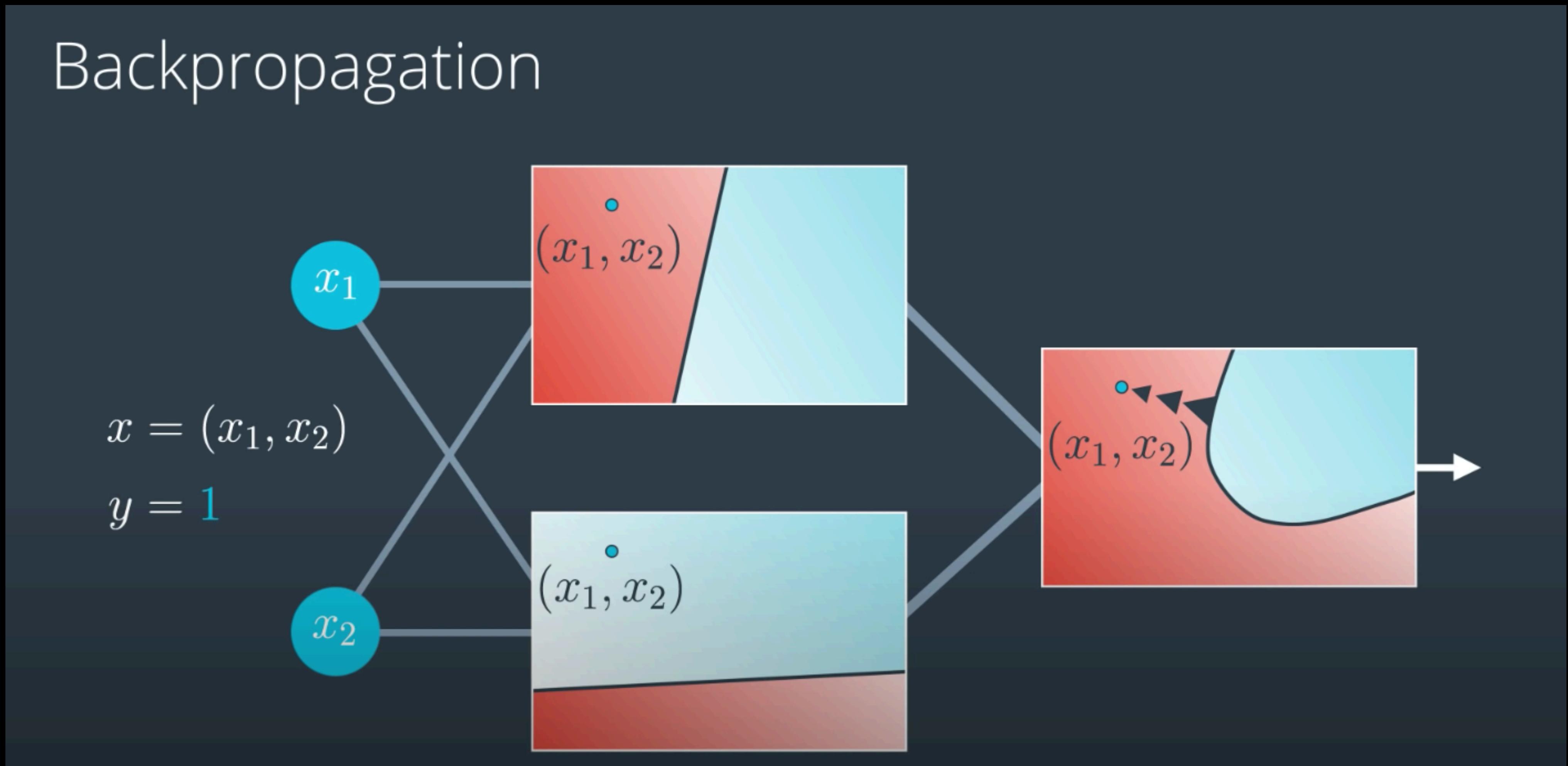
Gradient Descent



$\hat{y} = \sigma(Wx+b)$ ← Bad
 $\hat{y} = \sigma(w_1x_1 + \dots + w_nx_n + b)$
 $\nabla E = (\frac{\partial E}{\partial w_1}, \dots, \frac{\partial E}{\partial w_n}, \frac{\partial E}{\partial b})$
 $\alpha = 0.1$ (learning rate)
 $w_i' \leftarrow w_i - \alpha \frac{\partial E}{\partial w_i}$
 $b' \leftarrow b - \alpha \frac{\partial E}{\partial b}$
 $\hat{y} = \sigma(W'x+b')$ ← Better

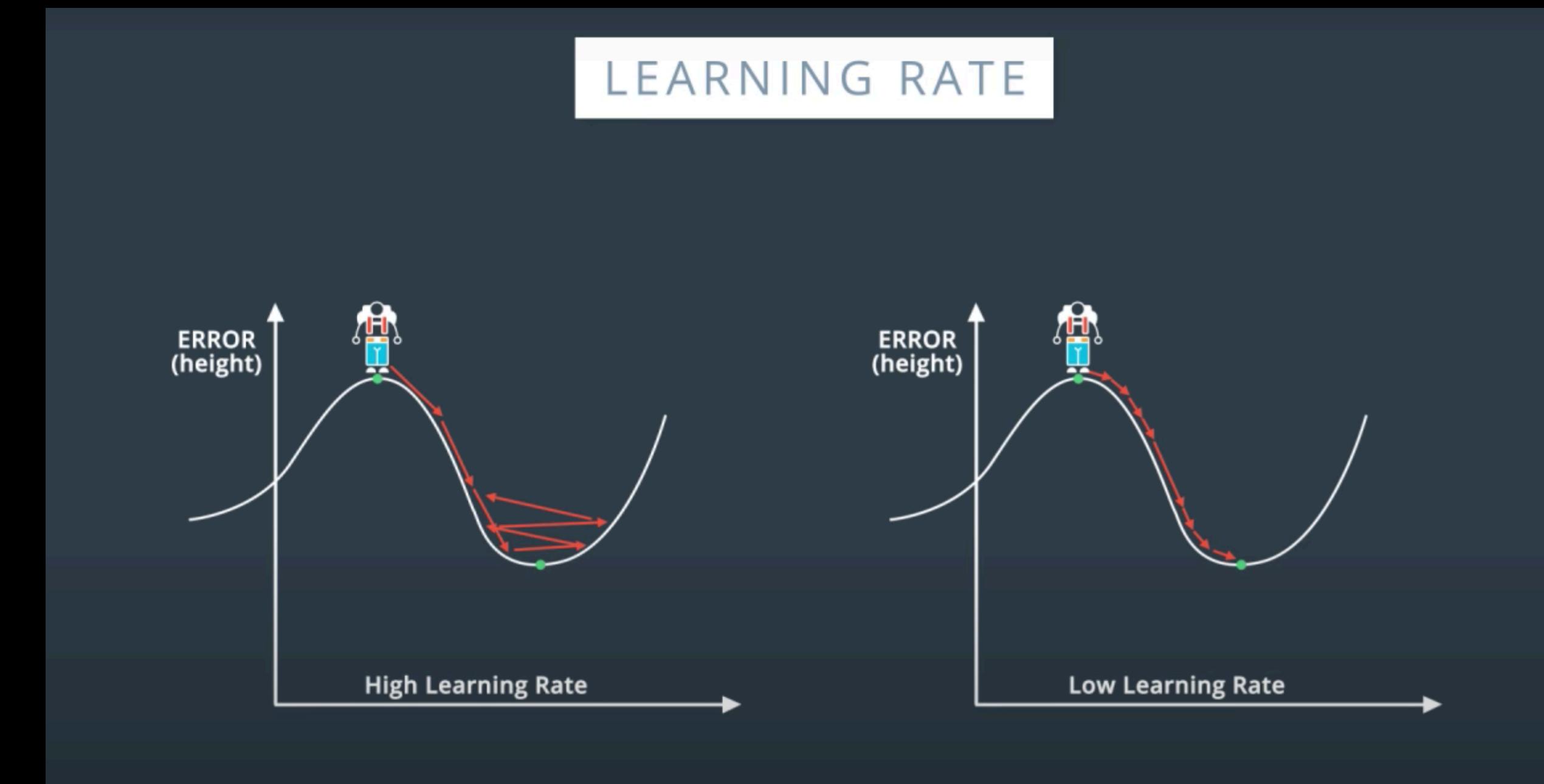
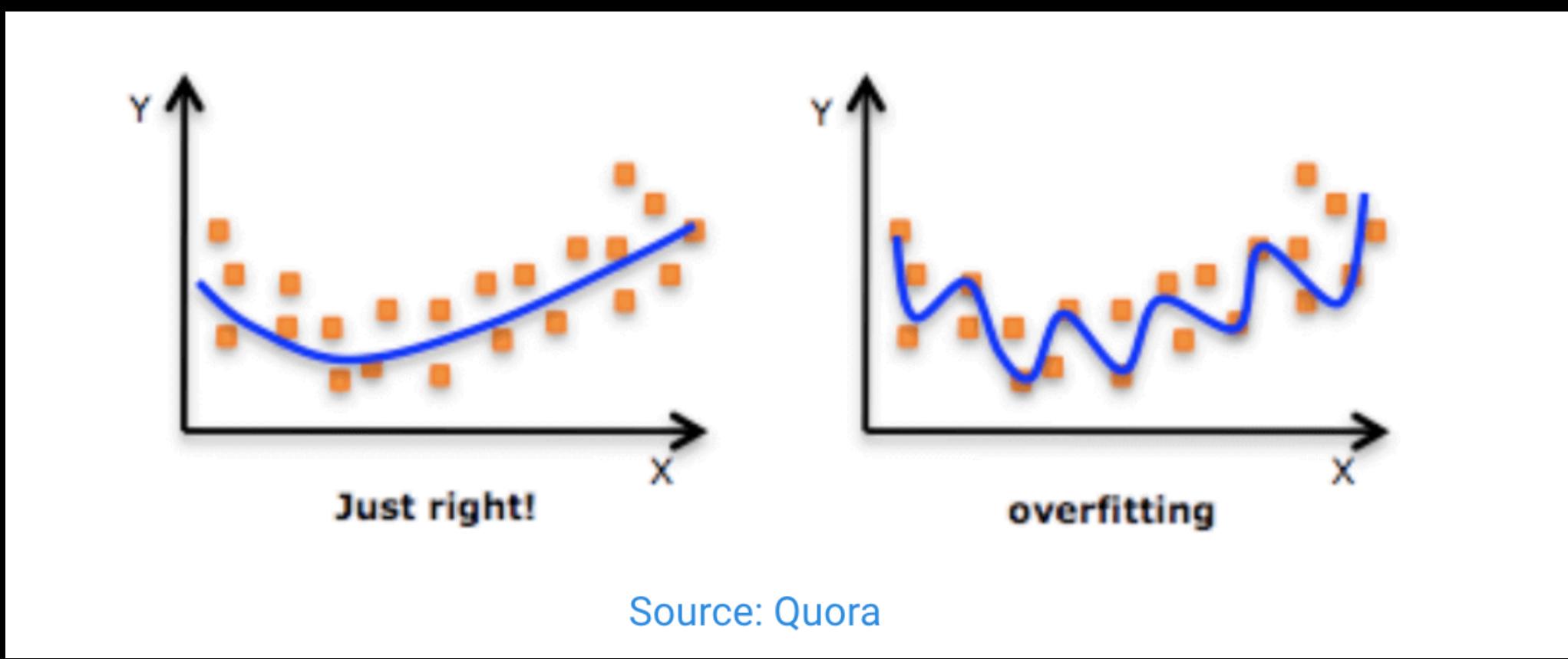
Backpropagation

- Aims to minimize the cost function by adjusting the network's weights and biases



More details on training optimization

- Overfitting
- Dropout
- Learning rate



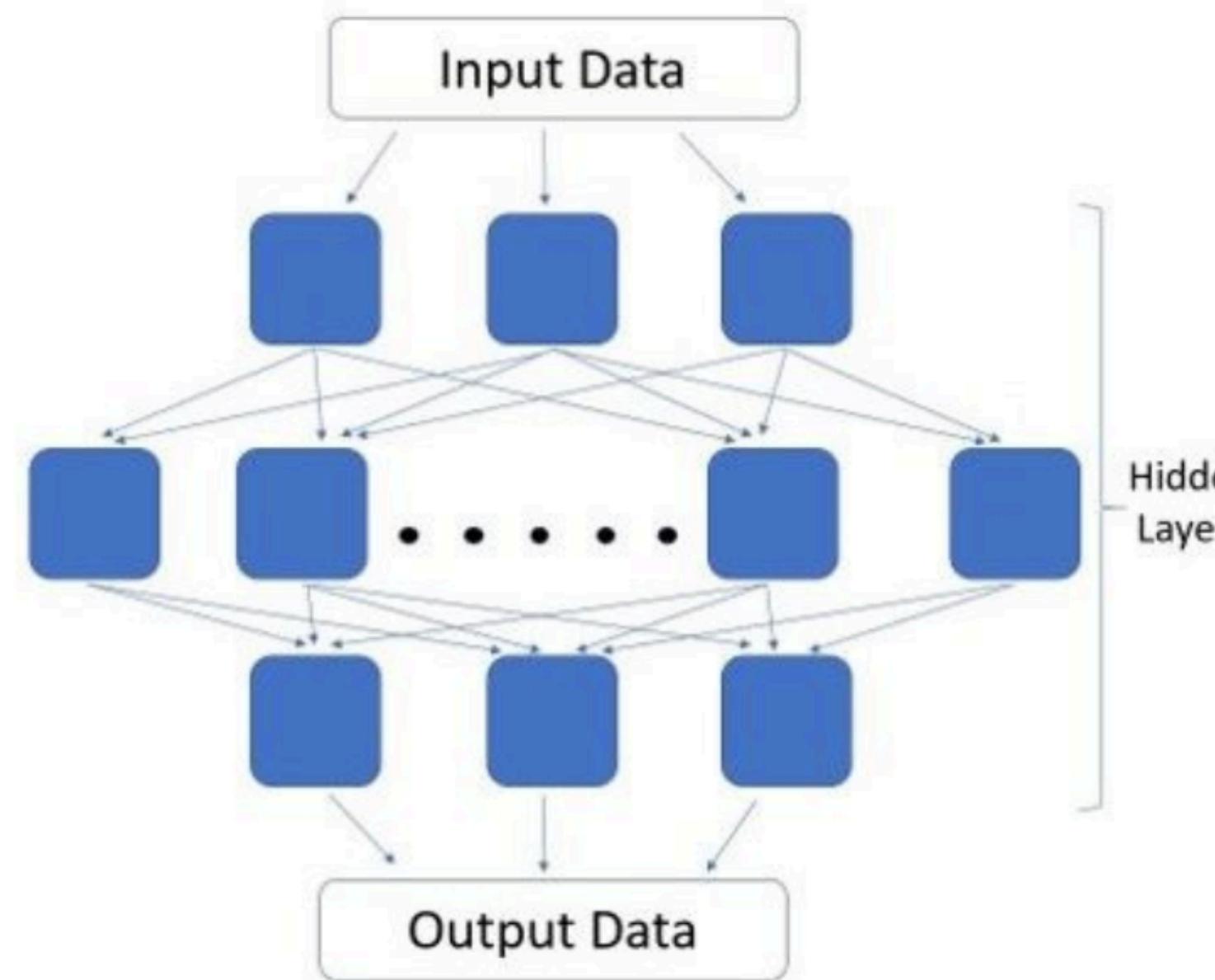
Recipe of Deep Learning

- Step1: Define a set of function
- Step2: Find the goodness of function
- Step 3: Pick the best function

Recurrent Neural Network

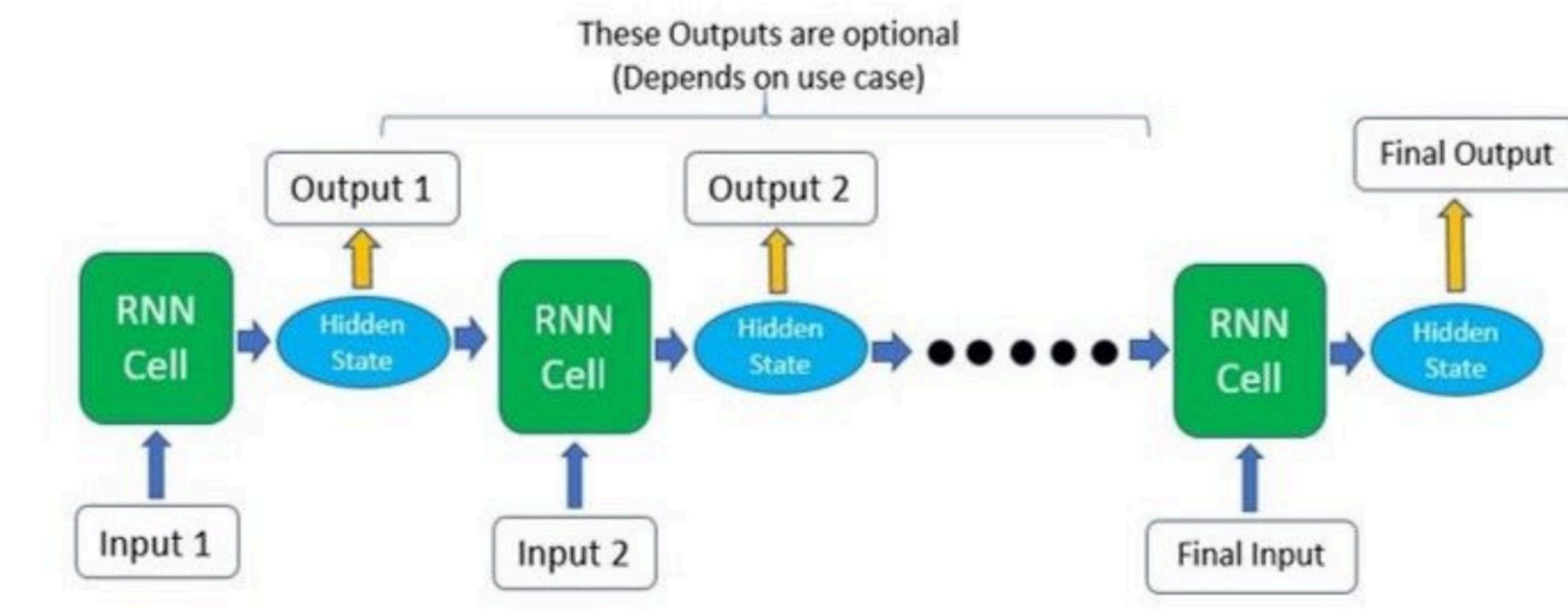
Comparison Between:

Traditional Feed-Forward Network



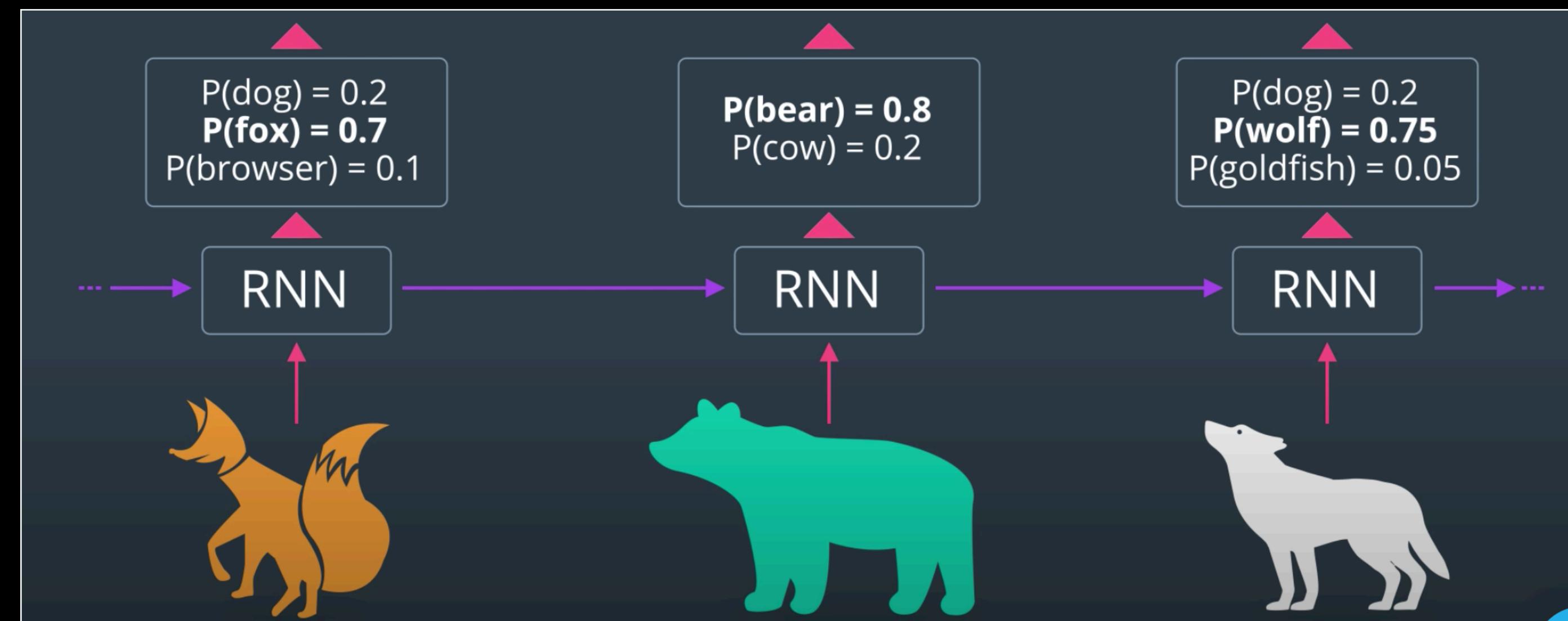
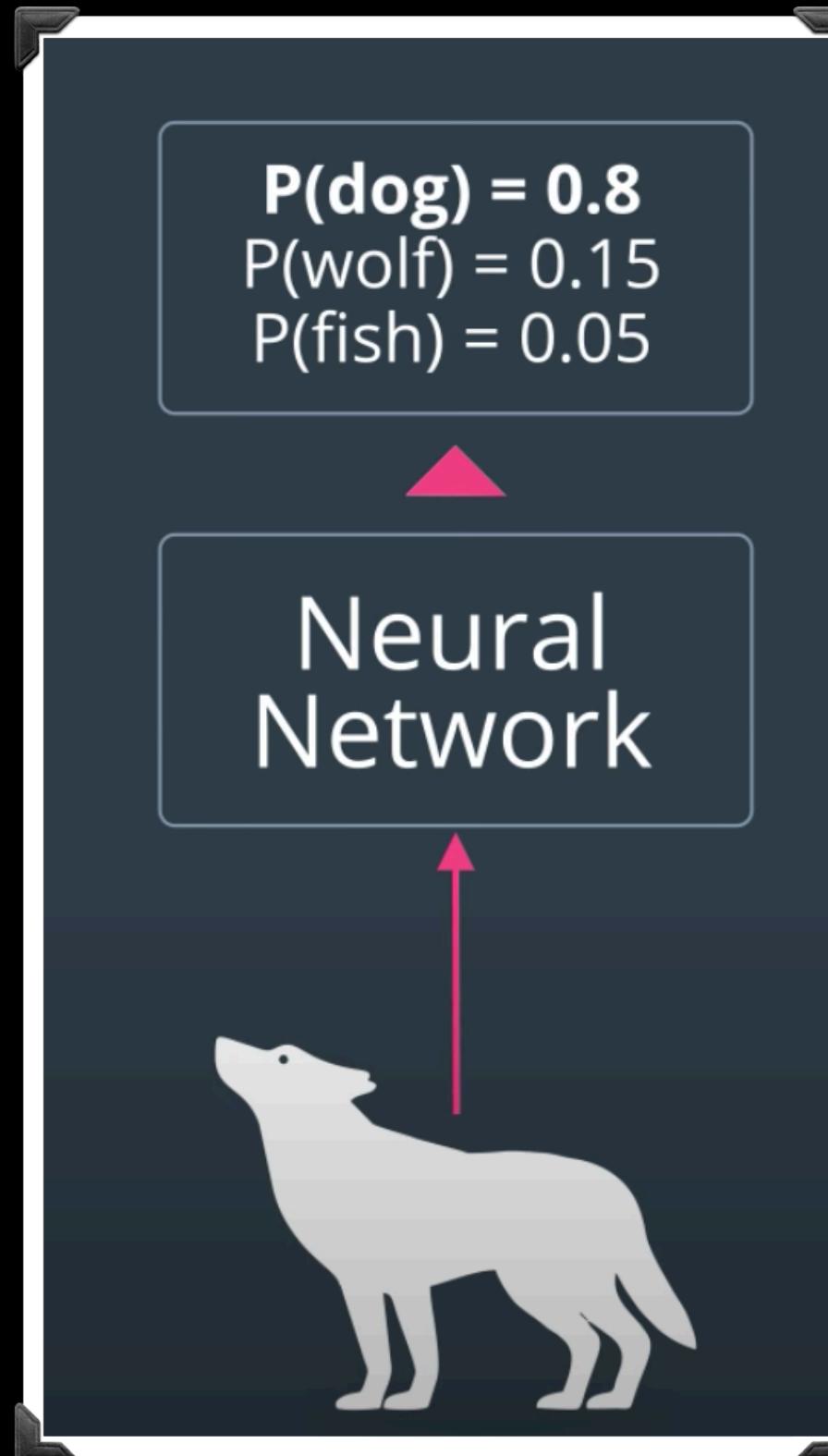
VS

Recurrent Neural Network

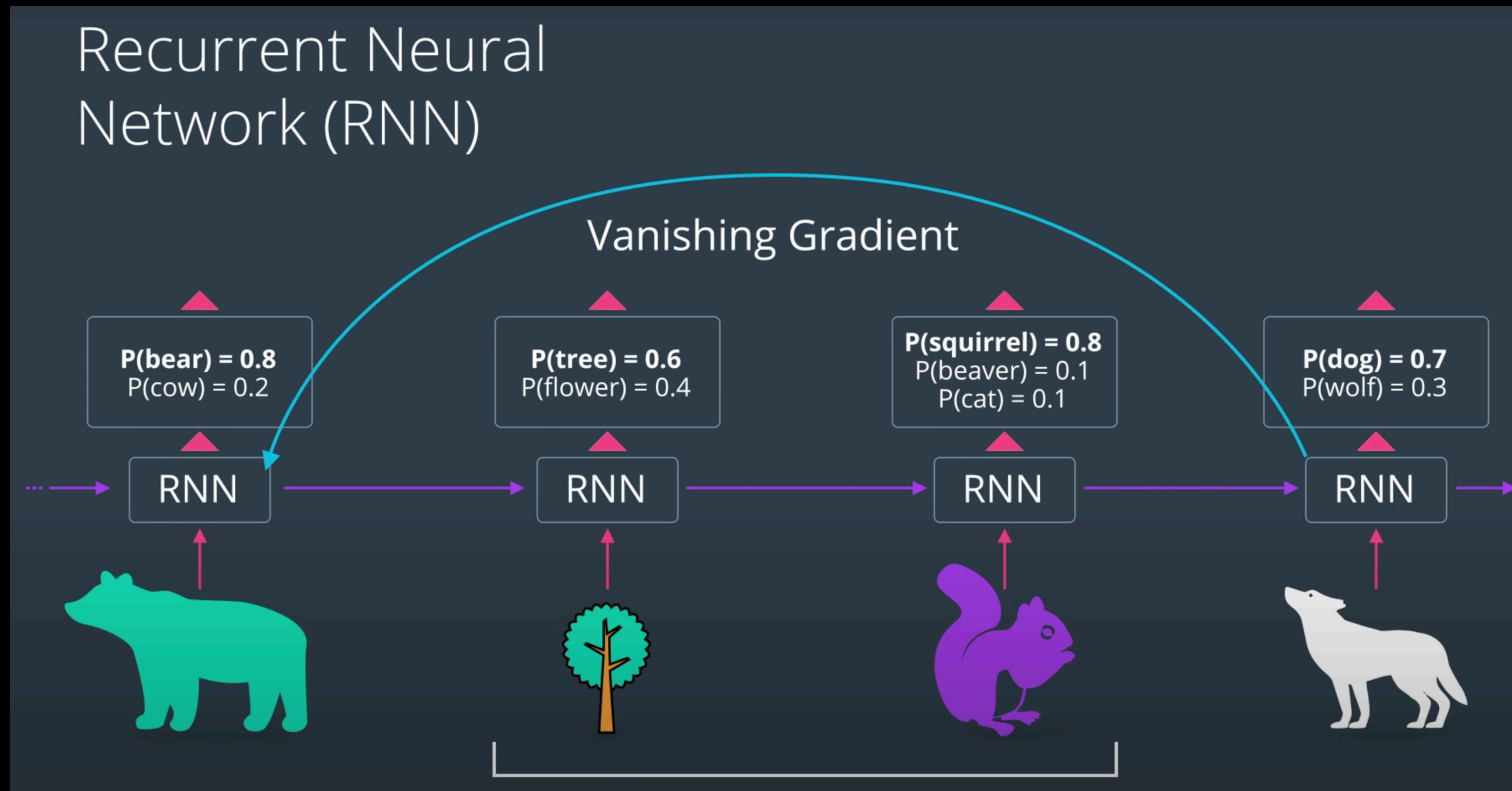


Overview of the feed-forward neural network and RNN structures

Recurrent Neural Network

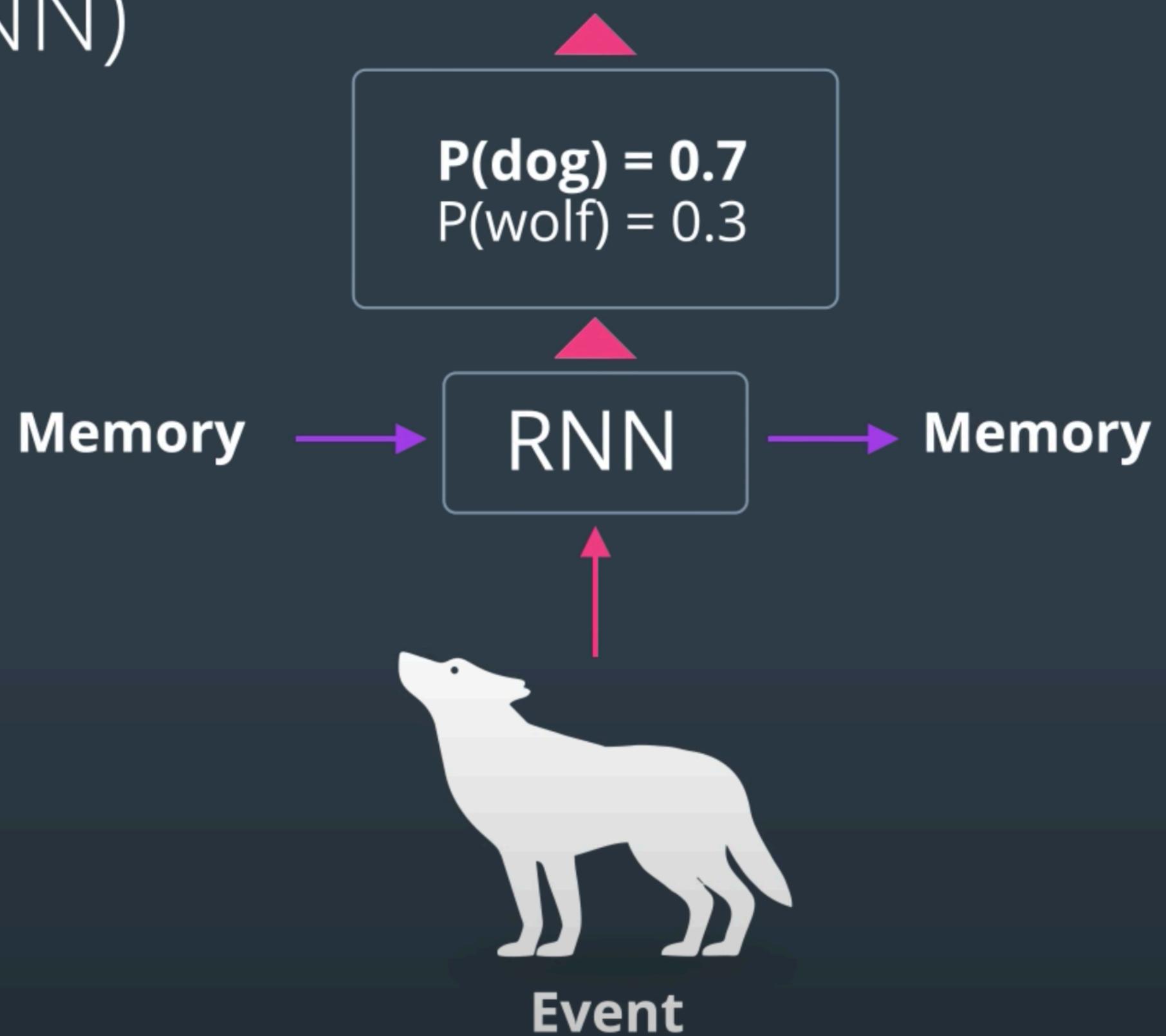


Problems of RNN



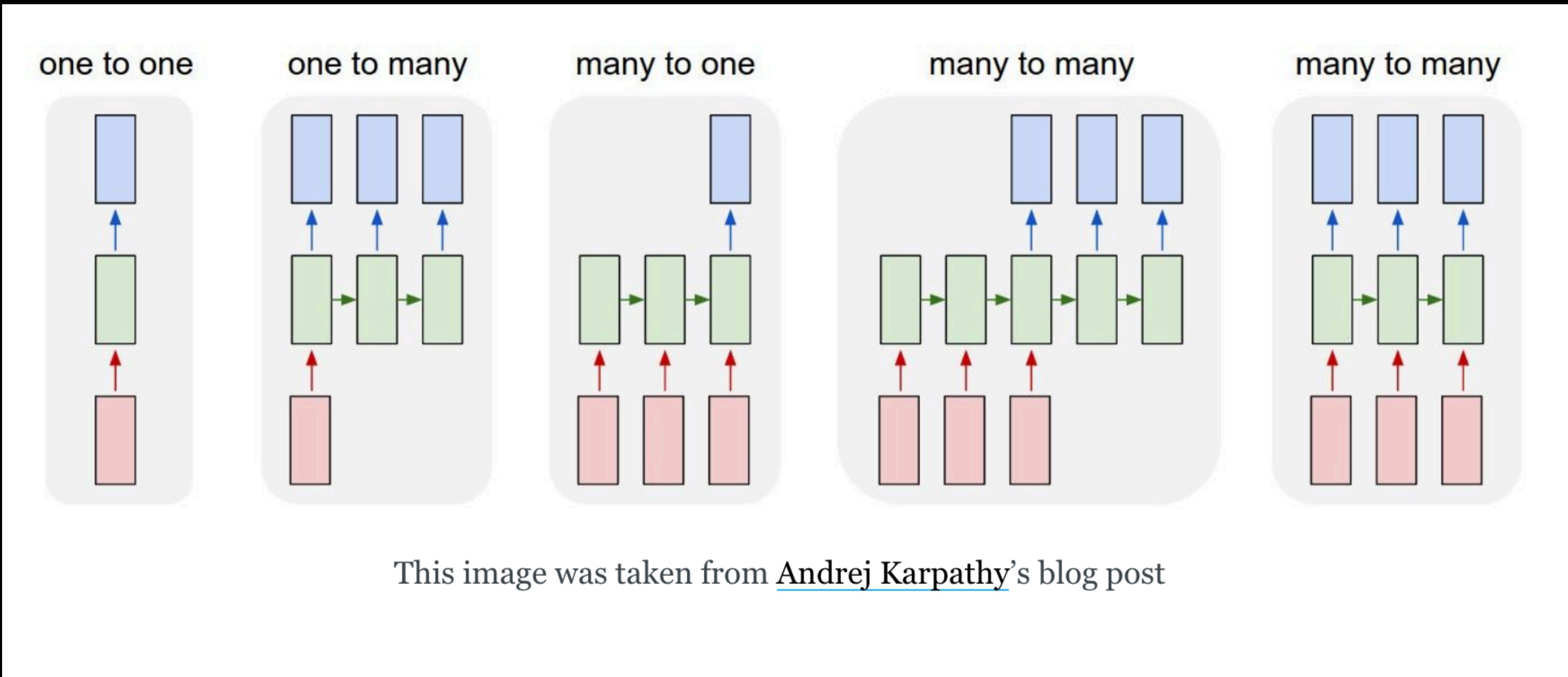
Recurrent Neural Network

Recurrent Neural
Network (RNN)



Applications of RNN

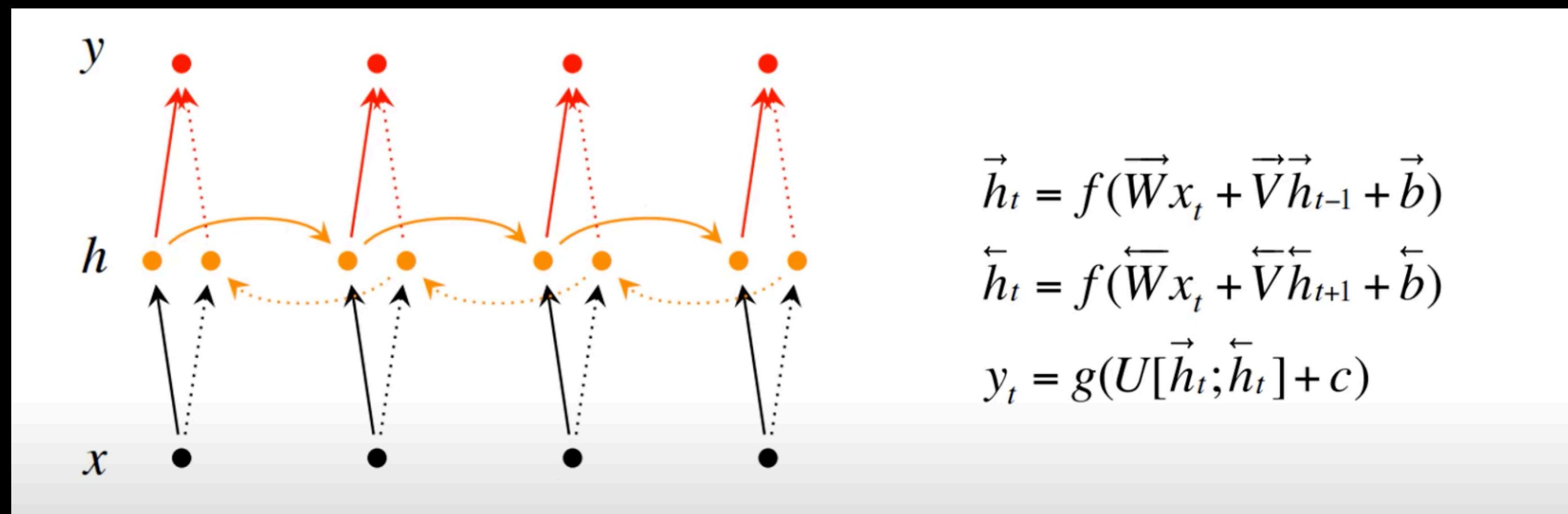
Sequence classification



Extensions

Bidirectional RNNs

- A Bi-RNN consists of two independent recurrent network, one where the input is processed from the start to end, and the other from the end to start.

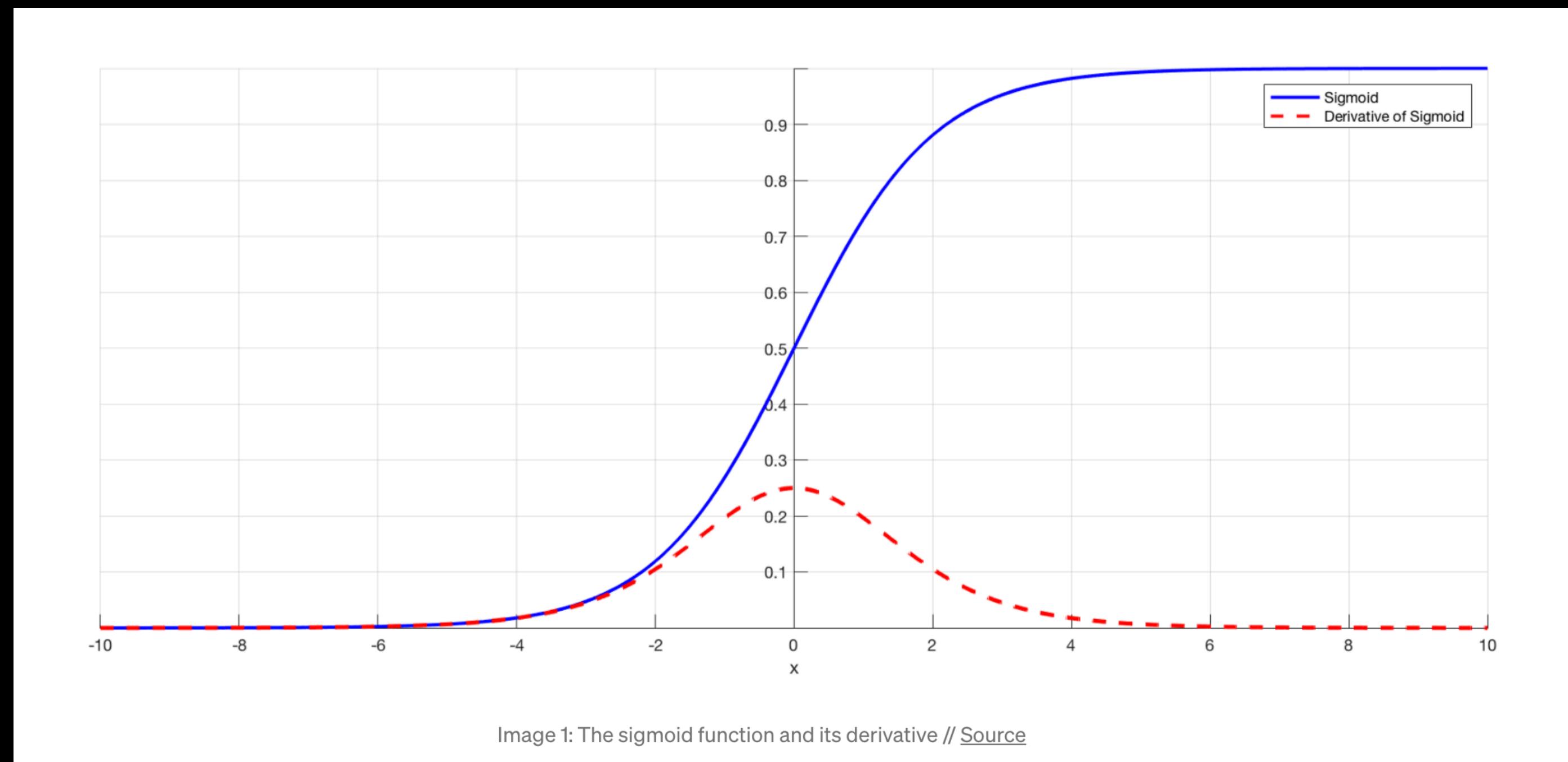


- h summarizes the past and future around a single token

- I am ____.
- I am ____ hungry.
- I am____hungry, and I can eat a horse.

Issues of RNN

- Gradient vanishing and exploding problem.
- It cannot process very long sequences



Reference

- Chi-Feng Wang, *The vanishing gradient problem*, 2019. (<https://medium.com/@reina.wang>)
- Lee,Hung-yi . *Applied Deep Learning course*
- Purnasai Gudikandula. *A Beginner Intro to Neural Networks* (<https://medium.com/@purnasaigudikandula/a-beginner-intro-to-neural-networks-543267bda3c8>)
- *Deep Learning From Scratch: Training Criterion* (<https://www.deeppideas.net/deep-learning-from-scratch-iii-training-criterion/>)