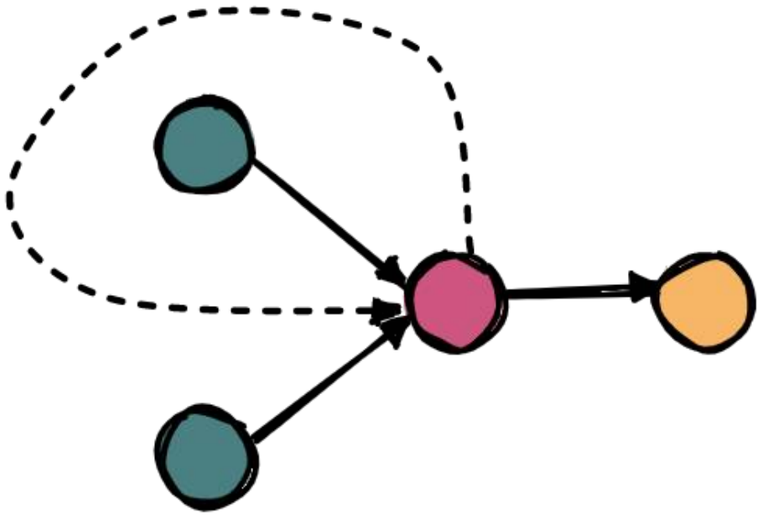
Recurrent neural networks

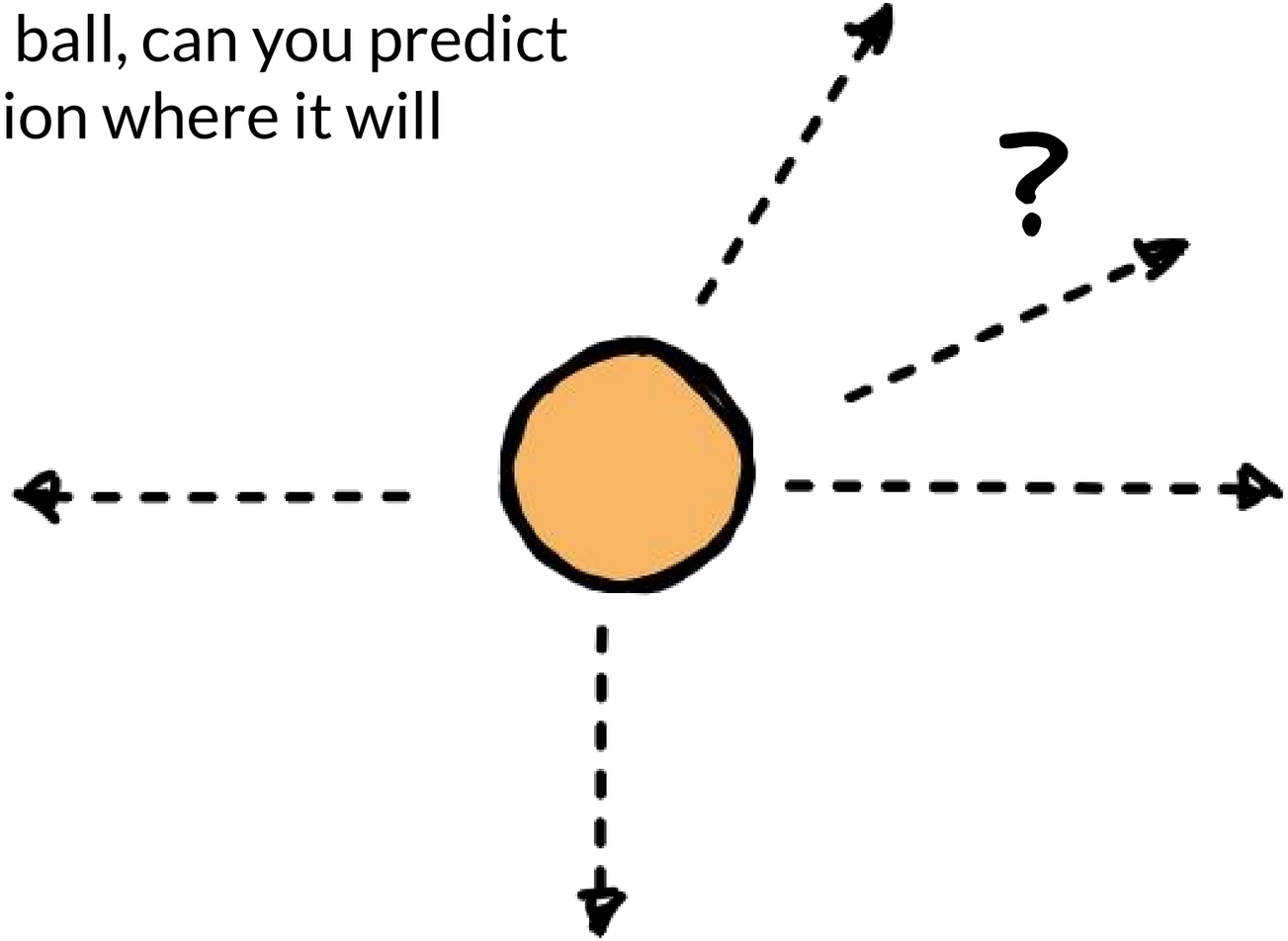


Week 19

Middlesex University Dubai; CST4050 Fall21;
Instructor: Dr. Ivan Reznikov

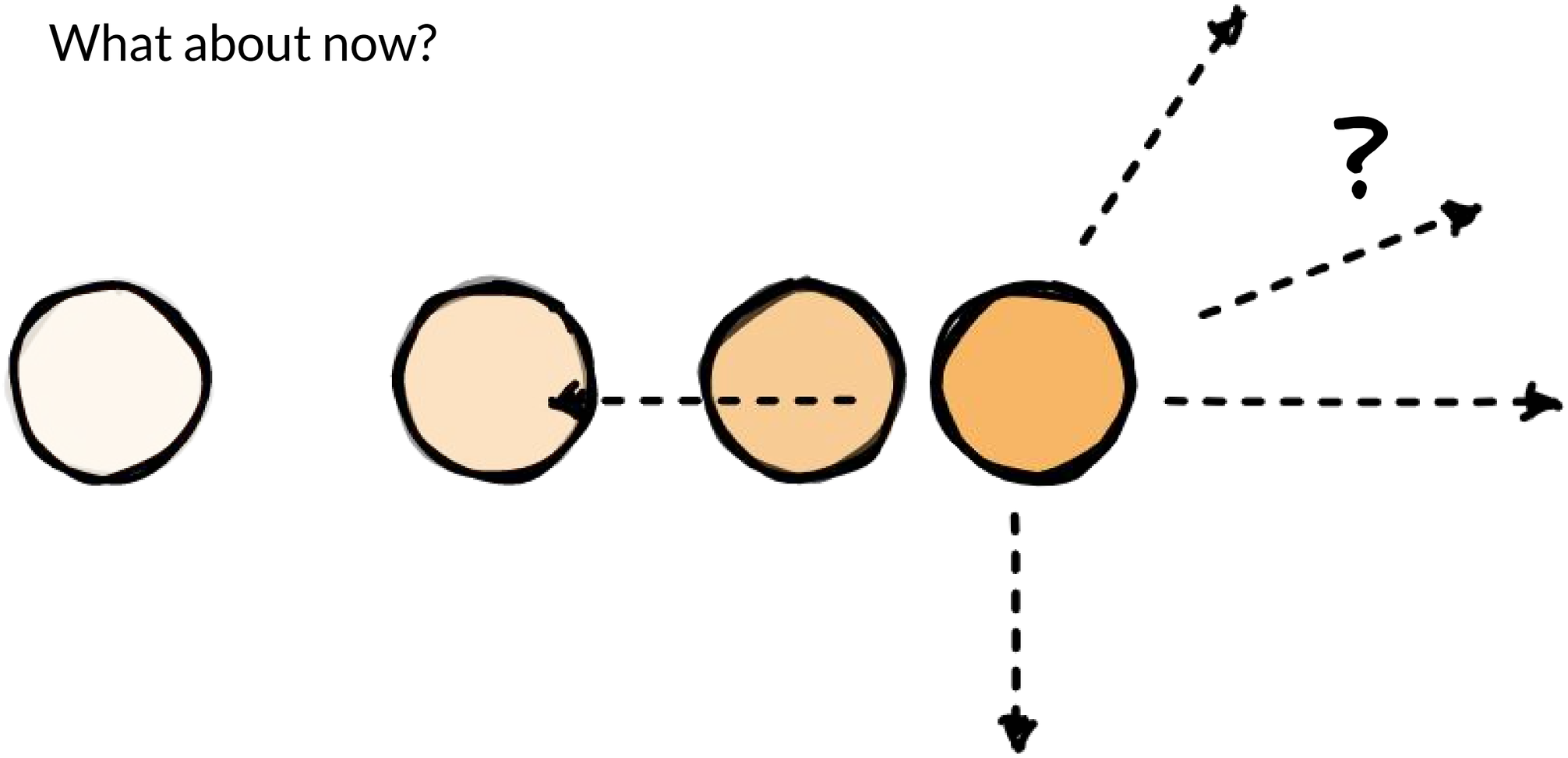
Ball position prediction

Given the image of the ball, can you predict the direction and location where it will move next?



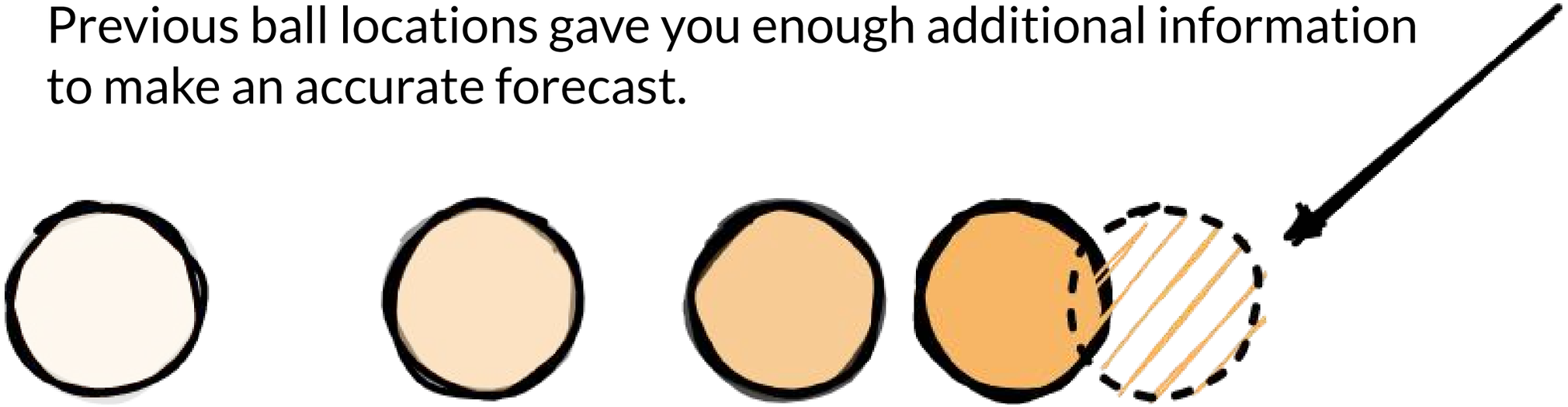
Ball position prediction

What about now?



Ball position prediction

High chance that you correctly the exact location and direction.
Previous ball locations gave you enough additional information to make an accurate forecast.



Sequences

Try saying all the numbers in order from 0 to 11 as fast as you can.

0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11

Sequences

What if we randomize the order of the numbers?
A bit slower, right?
The order matters.

6, 9, 4, 7, 5, 2, 11, 8, 10, 1, 0, 3

Sequences

What if we start from 4?
Pretty much fast as the first time

4, x, x, x, x, x, x, 11

Sequences

Now let’s do the same exercise with the alphabet

ABCDEFGHIJKLMNOPQRSTUVWXYZ

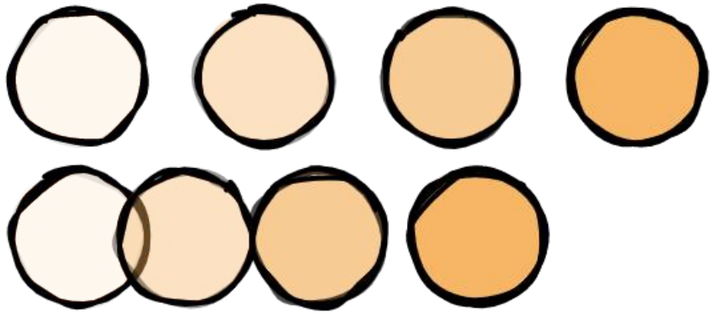
Sequences

Now start with the letter F.
It takes a while to pick up the pace.

F. . . . . Z

Why is that?

One of the reasons may be "more structured" sequenced:



?

?

5, 10, 15, ...
1, 3, 9, 27, ...

The official website for Game of Thrones on HBO, ___
Formula1 is the highest class of international racing ___

Predict next word

Problem 0: How to push text to a neural network, if the length of the sequence may vary?

Predict next word

The fact that Batman is so reliant on tech is his ...

Solution a: Fixed small window

[00001 00010] → prediction
is his

Problem a: Long-distance relationships
London is my home city. This is the reason I fluently speak English

Predict next word

The fact that Batman is so reliant on tech is his ...

Solution b: Fixed wide window

[00001 00010 01011 10110 11000 01110 11010 01110 10101 11000 11100]

The fact that Batman is so reliant on tech is his

→ prediction

Problem b: Lose of sense if met in different part of sequence

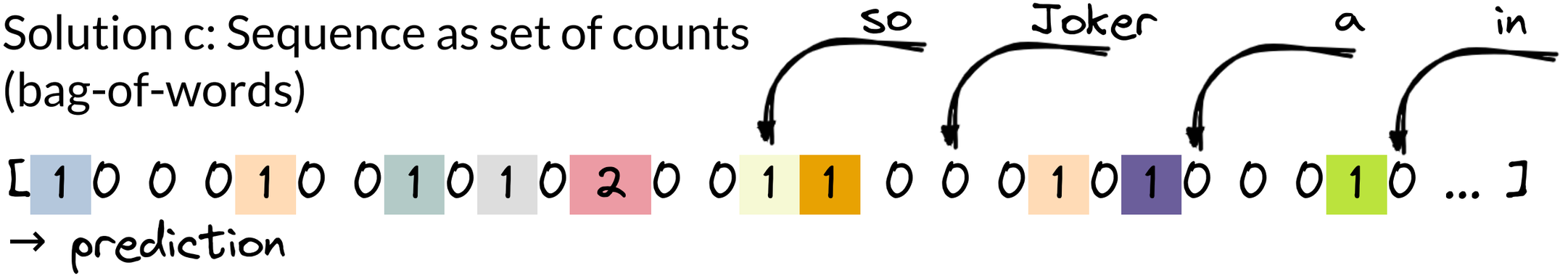
11010 01110 10101

reliant on tech

Predict next word

The fact that Batman is so reliant on tech is his ...

Solution c: Sequence as set of counts (bag-of-words)



Problem c: Lose of sense if met in different part of sequence

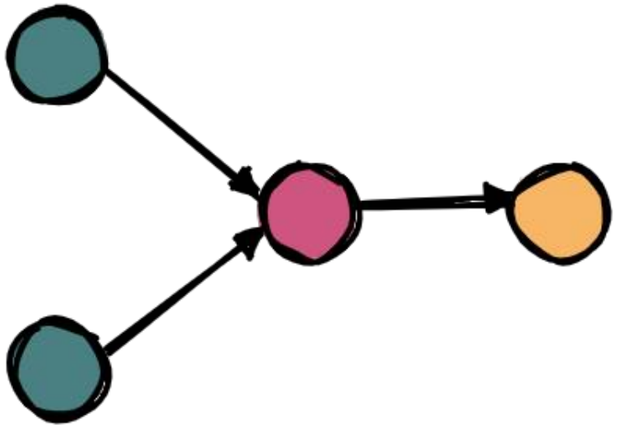
Batman is reliant on tech == tech is reliant on Batman

Model criteria

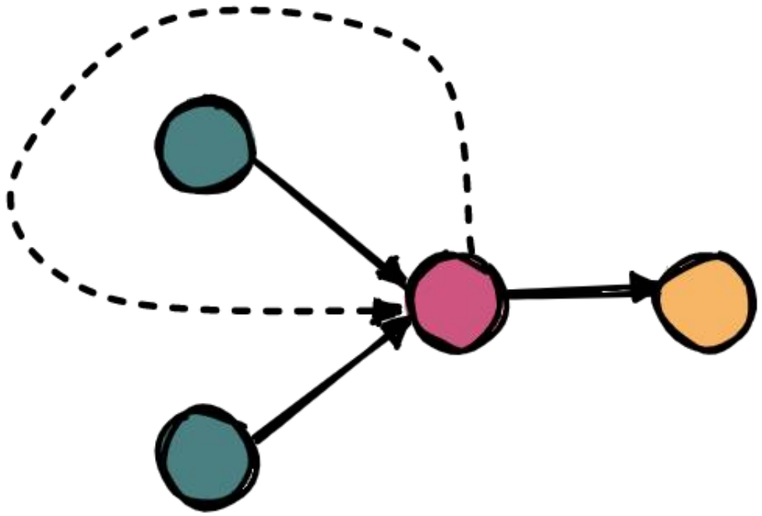
Requirements:
1. Handle sequences of different length
2. Track long-term dependencies
3. Share parameters across sequence
4. Save order information

Saving memories

Having memory may be pretty important when we deal with time-series events. A regular perceptron has two inputs from the input or previous layer. But what if we could pass the "memory" – the last value of the neuron? This is called a **recurrent** perceptron. The "memory" is called **hidden state**.



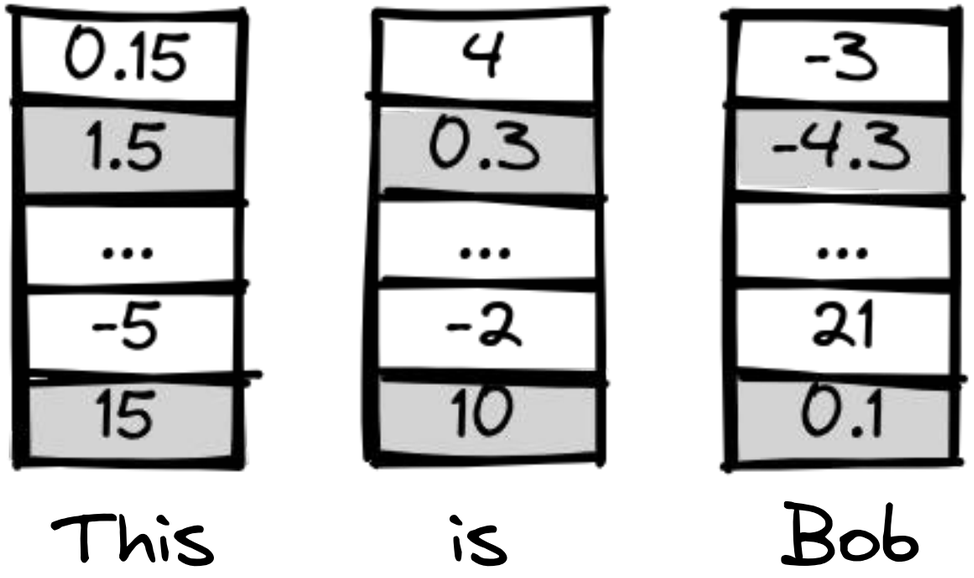
regular perceptron

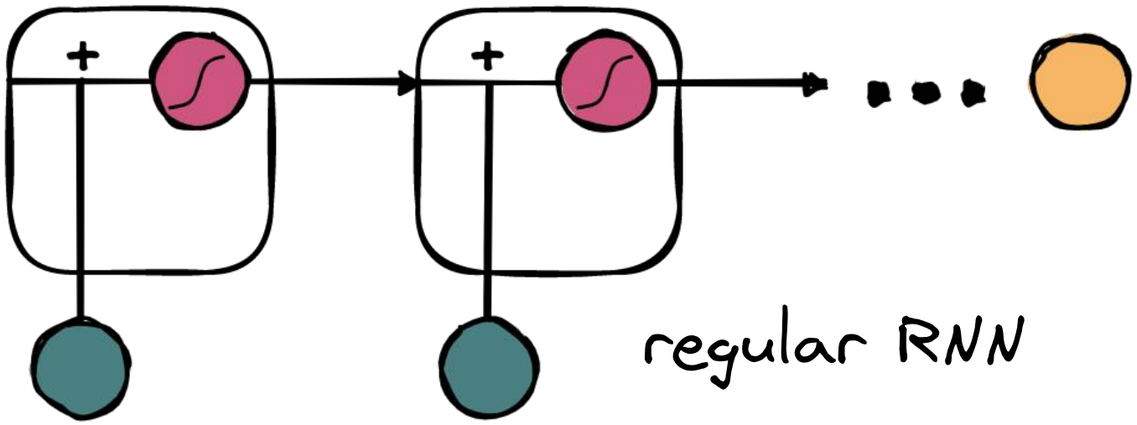


recurrent perceptron

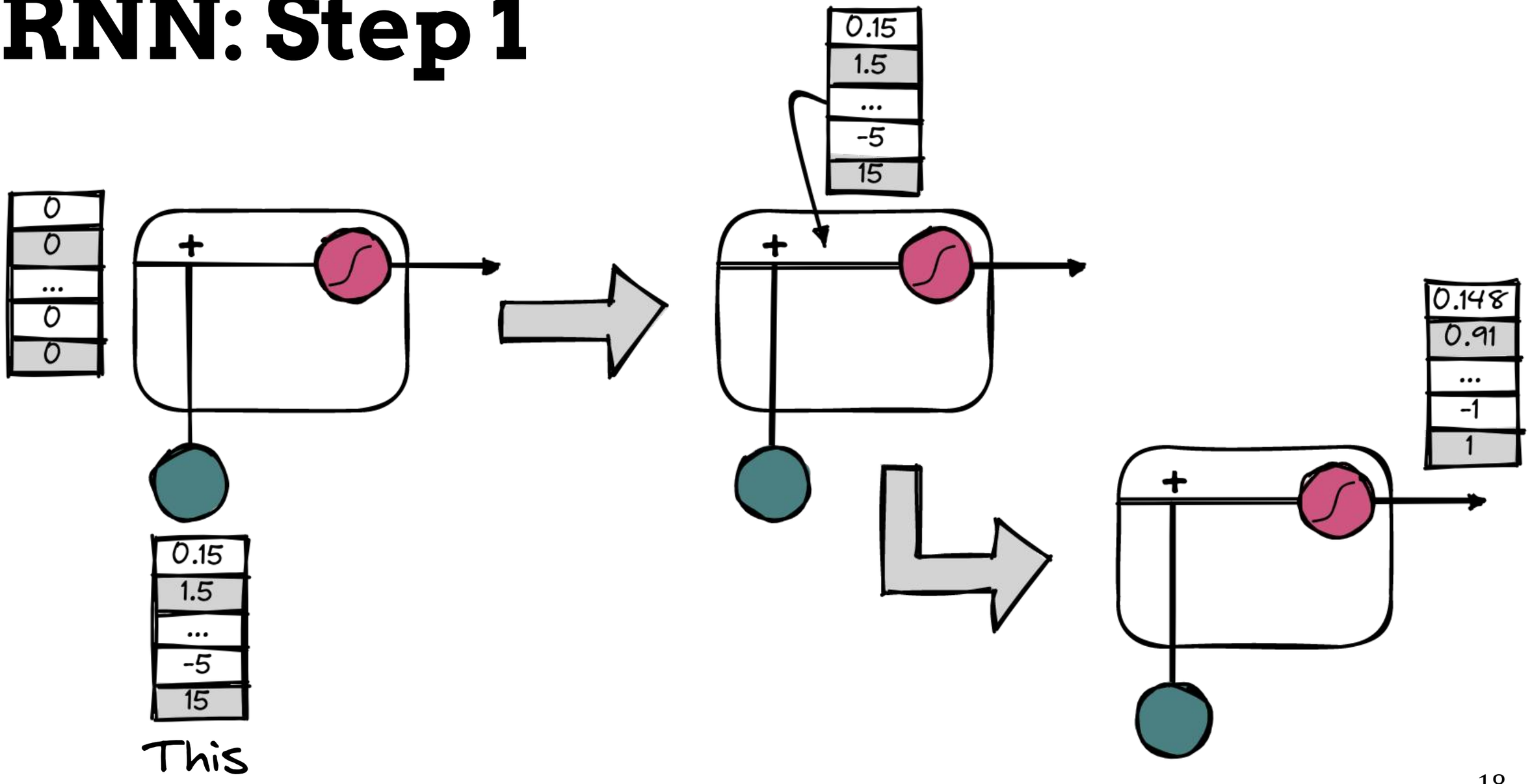
RNN: Case1

"This is Bob"

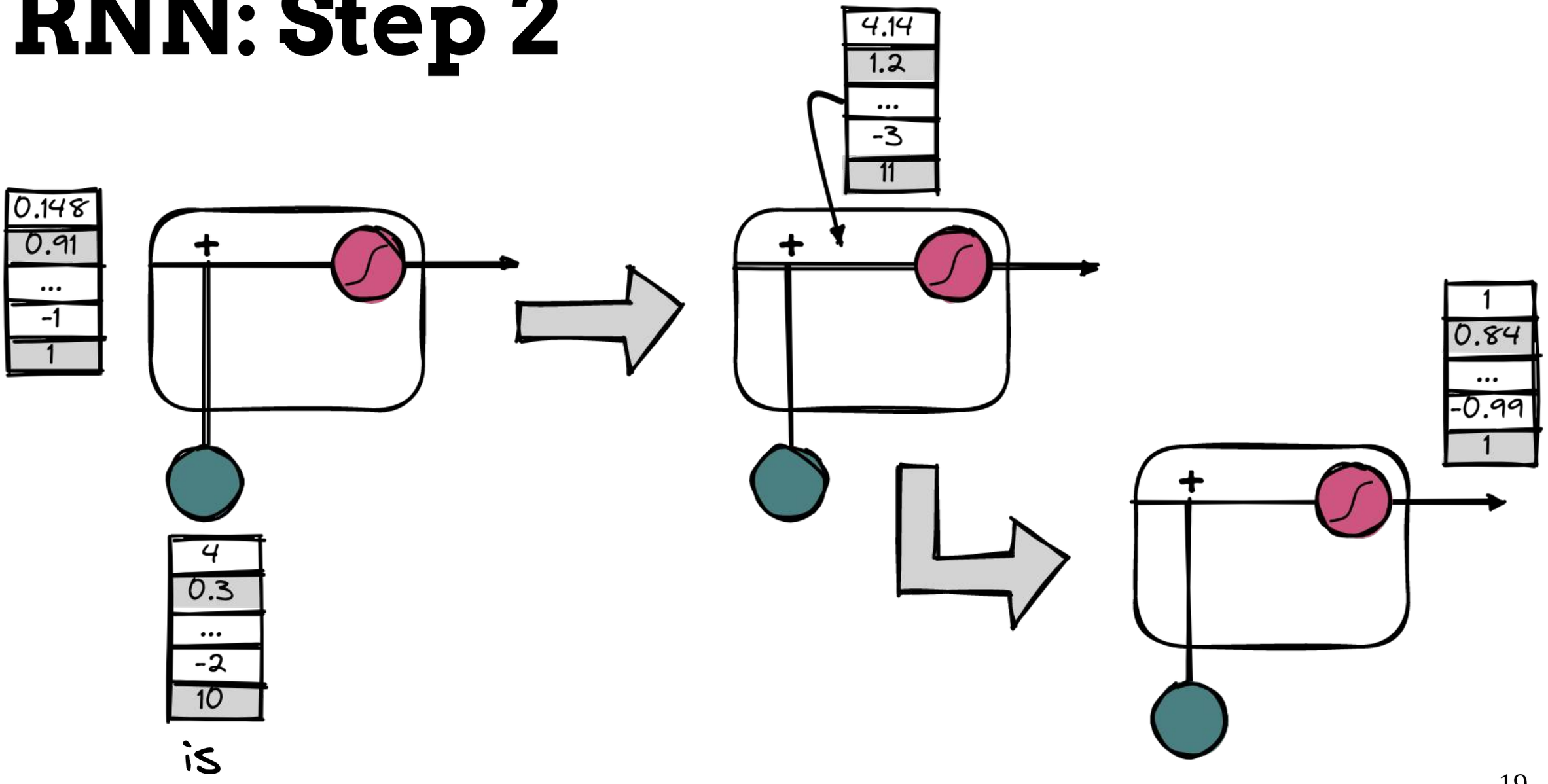




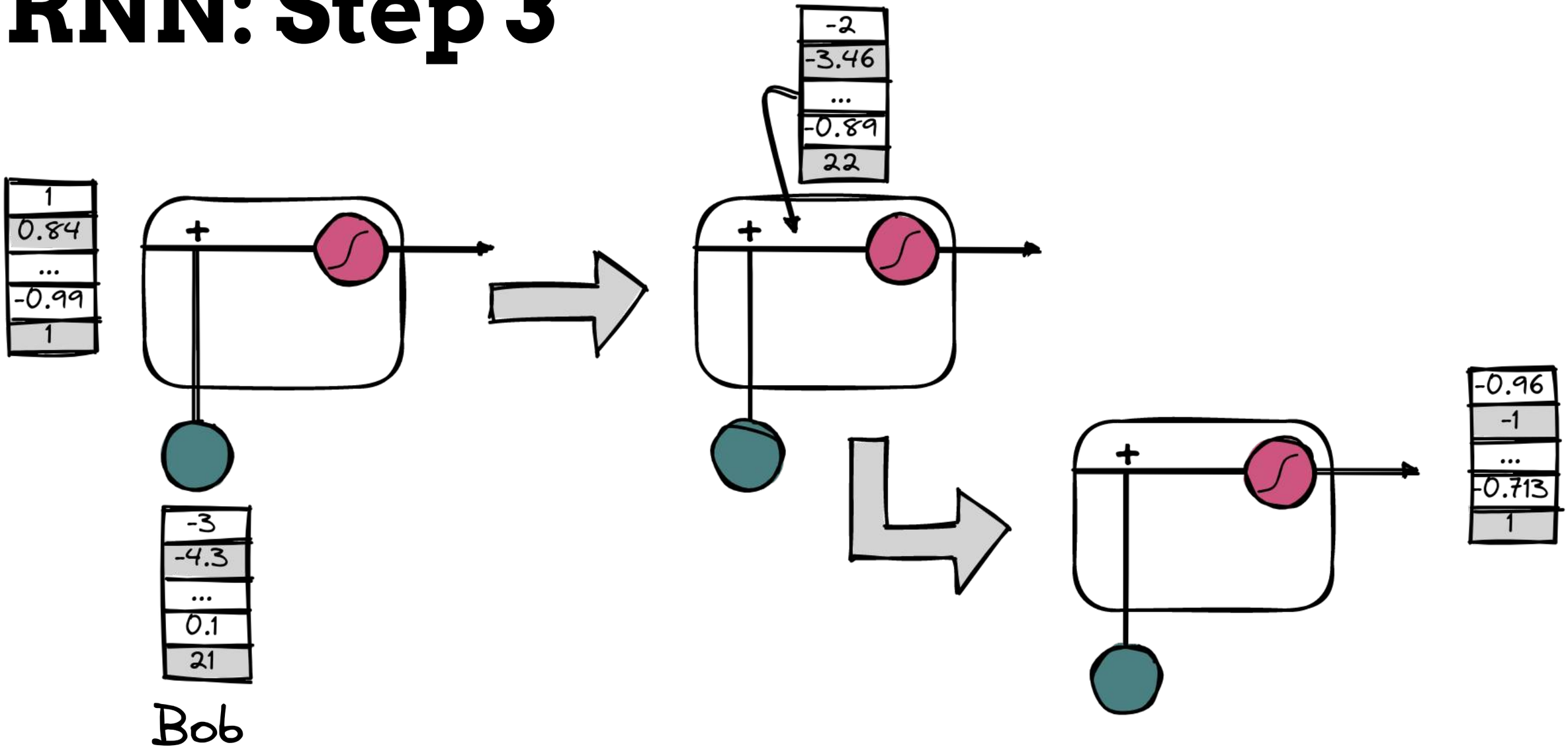
RNN: Step 1



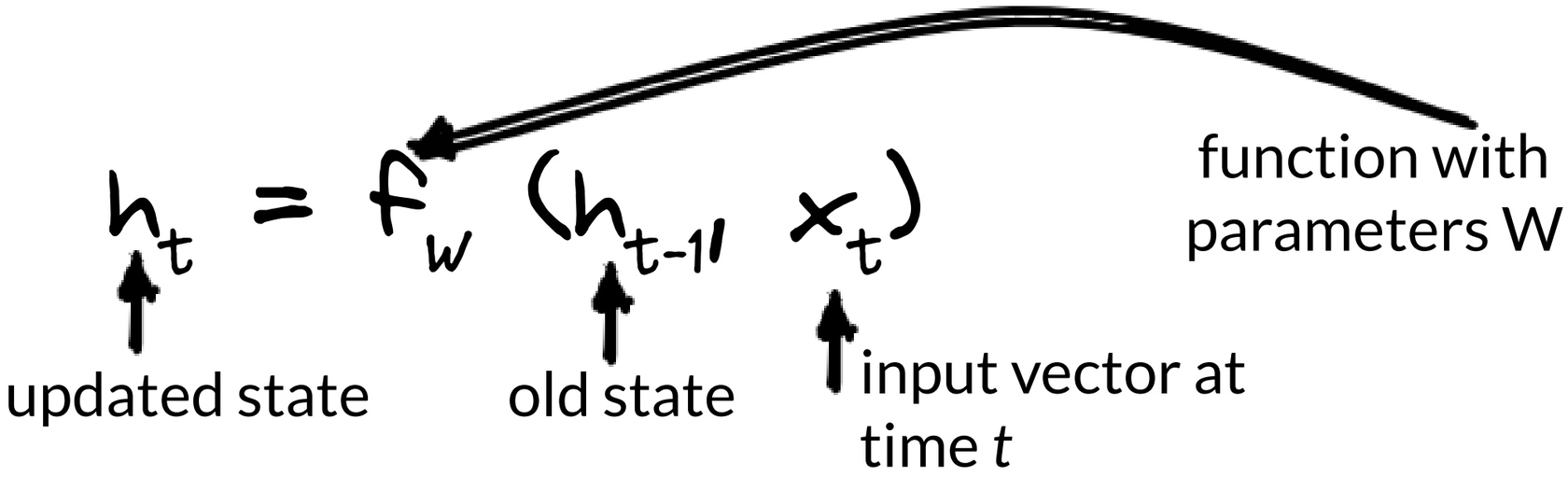
RNN: Step 2



RNN: Step 3



RNN: Formula

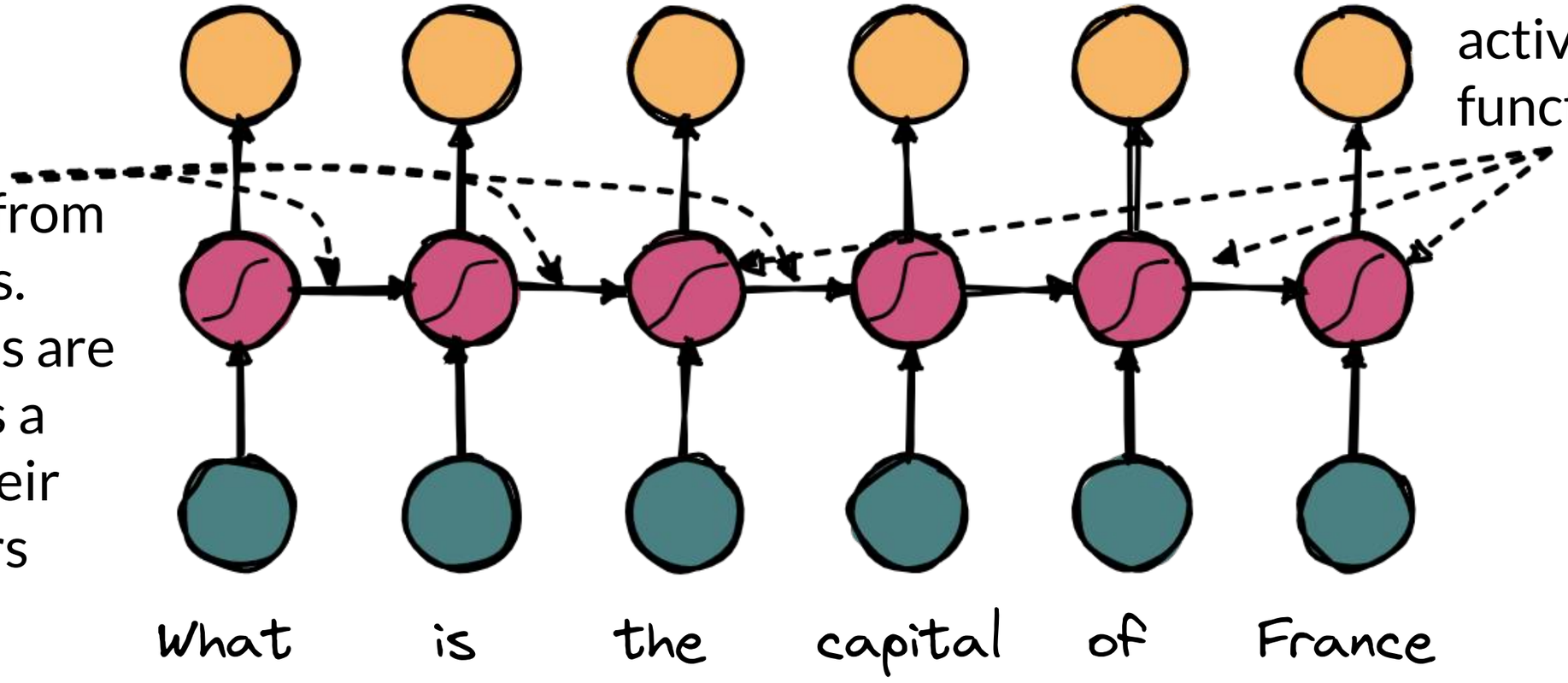

$h_t = f_w(h_{t-1}, x_t)$
function with parameters W
updated state
old state
input vector at time *t*

RNN: Case2

Let's take another query example:
"What is the capital of France?"

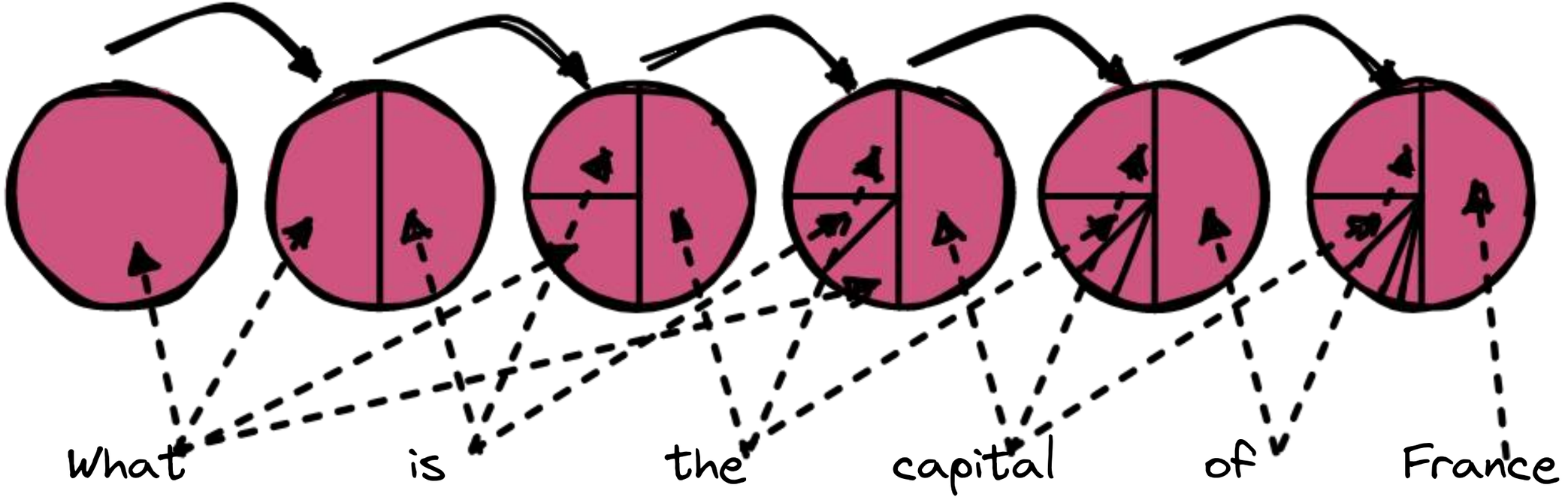
tanh activation function

Passing information from hidden states. Hidden states are connected as a chain, and their impact lowers across time

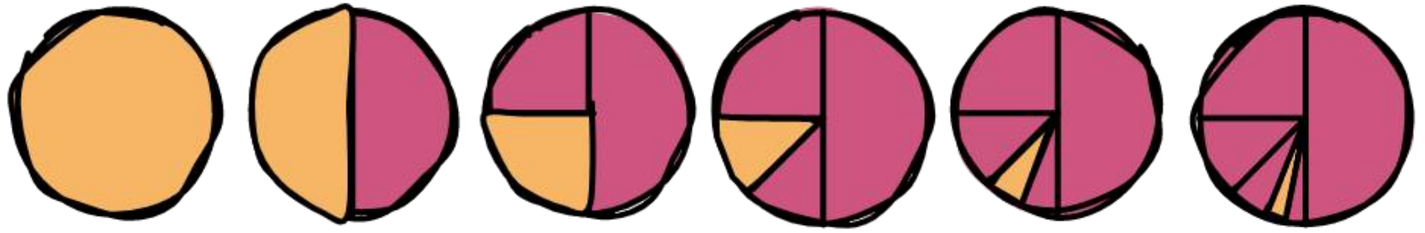


Vanishing gradient

Let's take a look at how importance of the word changes along the chain:



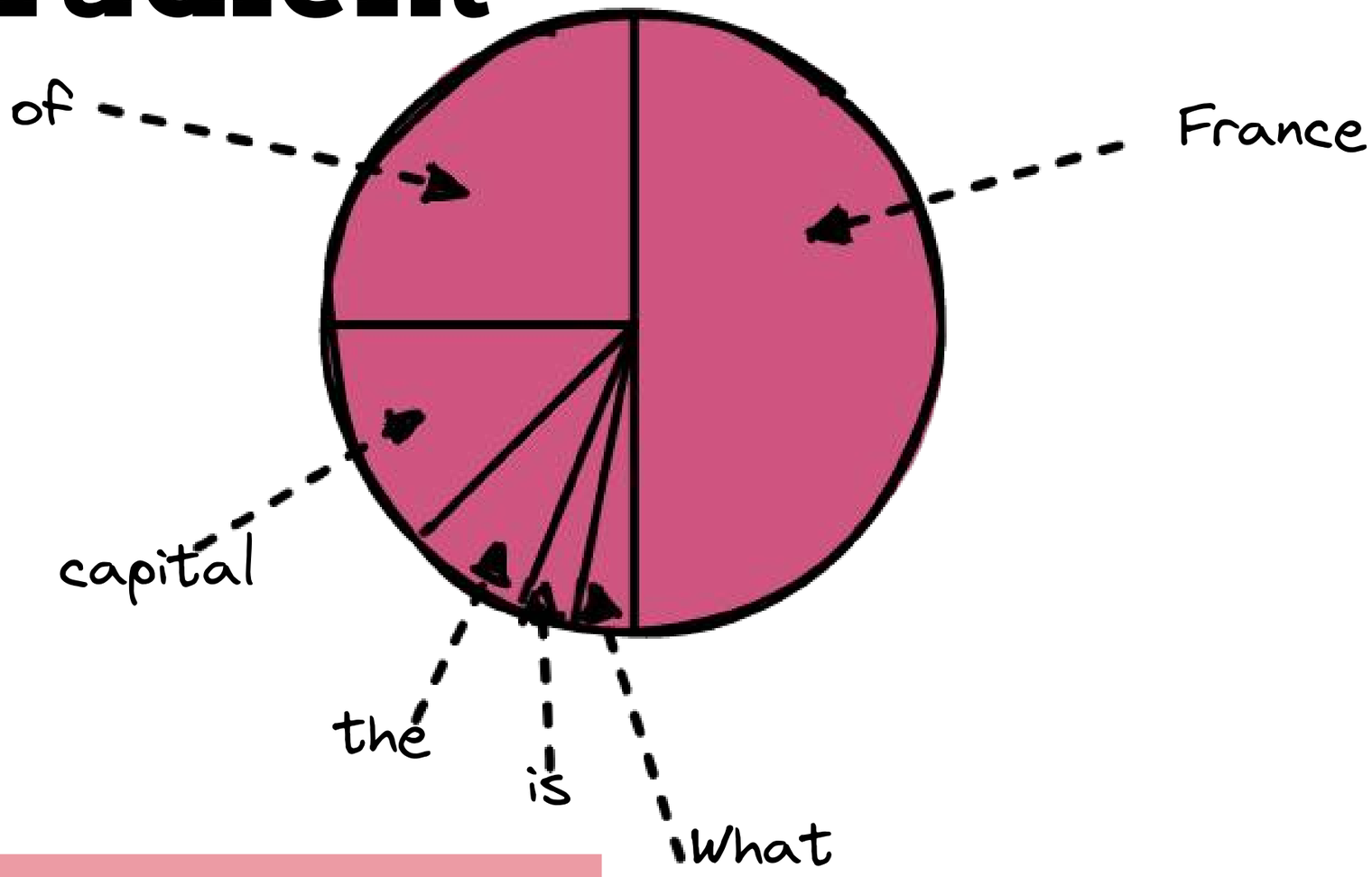
Impact of the word "What":



Vanishing gradient

As one can notice, the impact of recently met words is much higher than the ones from the beginning of the sequence.

This is an example of the vanishing gradient problem.



What is the capital of France