8.3

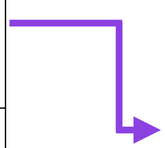
Introduction to Recurrent Neural Networks

Part 1: Modeling Sequence Data

Sebastian Raschka and the Lightning AI Team

A big limitation of Bag-of-Words model:
the word sequence order is lost

Text
The ghost pepper is so spicy, it is hauntingly hot
I tried to hug the sun today, but it was too hot to handle
I cannot handle spicy food



but	cannot	food	ghost	handle	hauntingly	hot	hug	i	is	it	pepper	so	spicy	sun	the	to	today	too	tried	was
0	0	0	1	0	1	1	0	0	1	1	1	1	1	0	1	0	0	0	0	0

BoW vectors

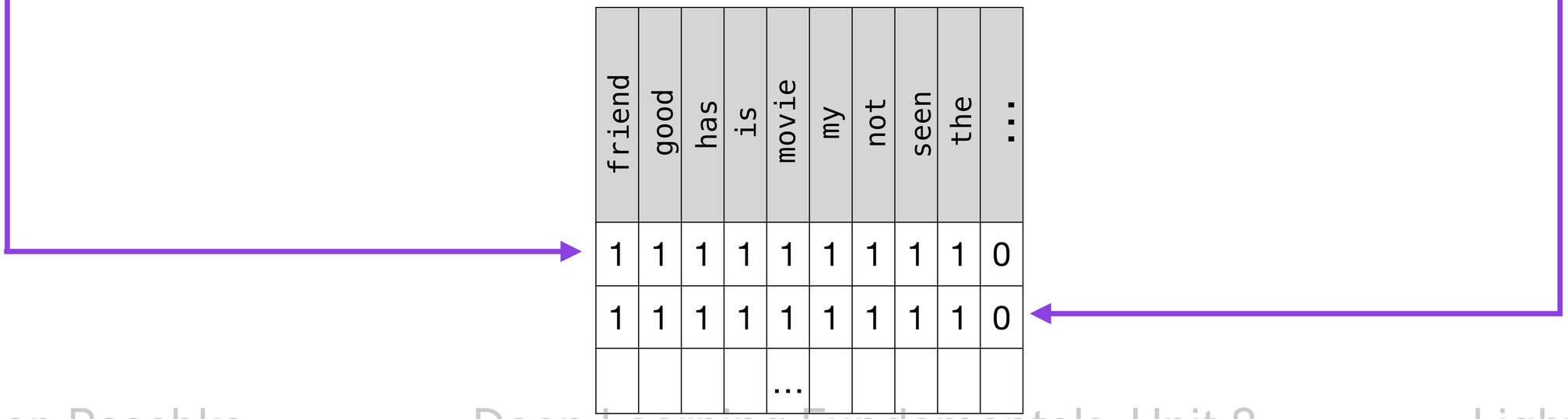
The movie my friend has not seen is good

The movie my friend has seen is not good

BoW maps these sentences to identical vectors

The movie my friend has not seen is good

The movie my friend has seen is not good



friend	good	has	is	movie	my	not	seen	the	...
1	1	1	1	1	1	1	1	1	0
1	1	1	1	1	1	1	1	1	0
				...					

Machine learning models for text

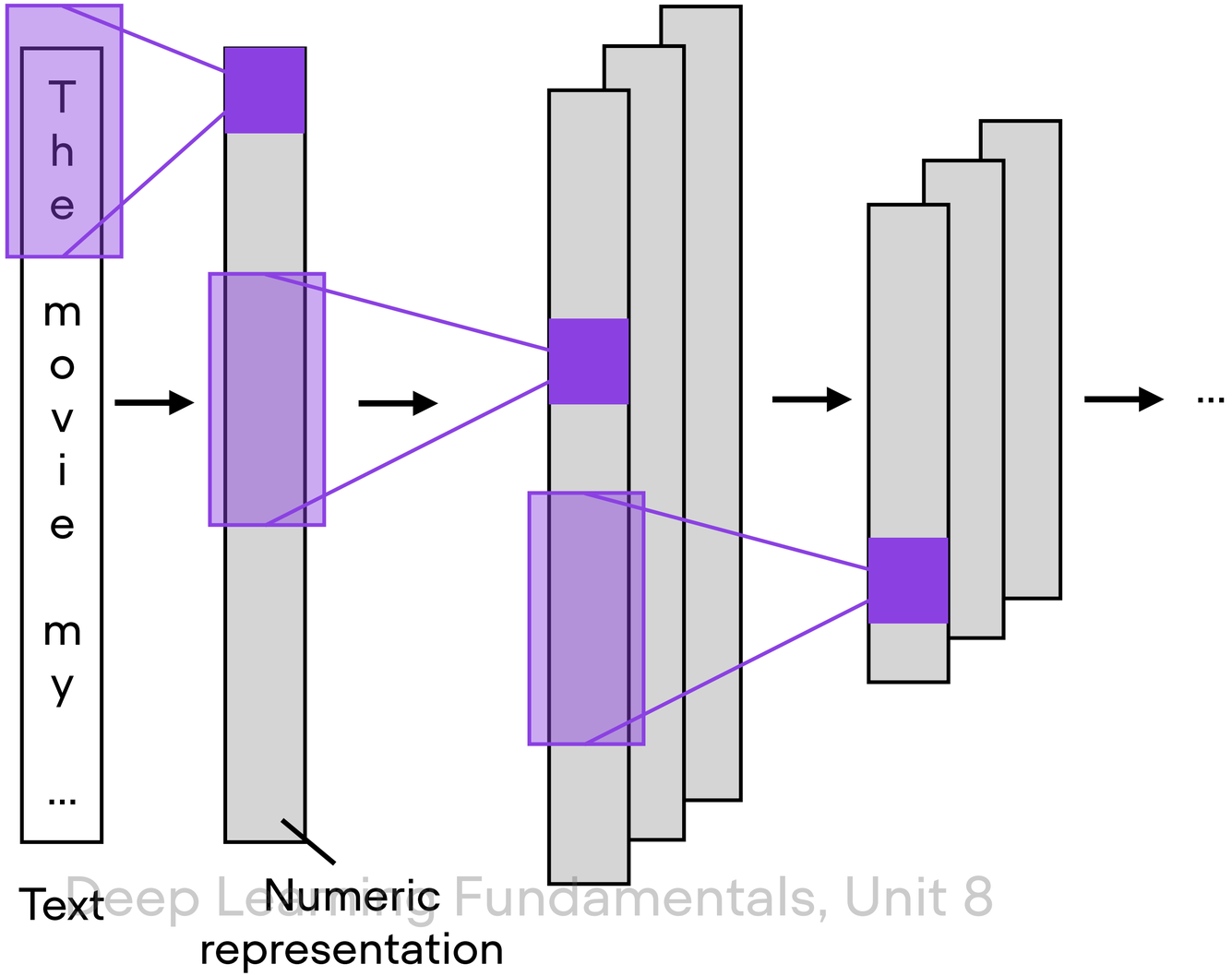
Classic methods for tabular data:
- Logistic regression
- Multilayer perceptron

Sequence models:
- 1D convolutional network
- Recurrent neural network

Transformers

A 1D convolutional network for text

Analogous to
image inputs
but 1D instead of 2D



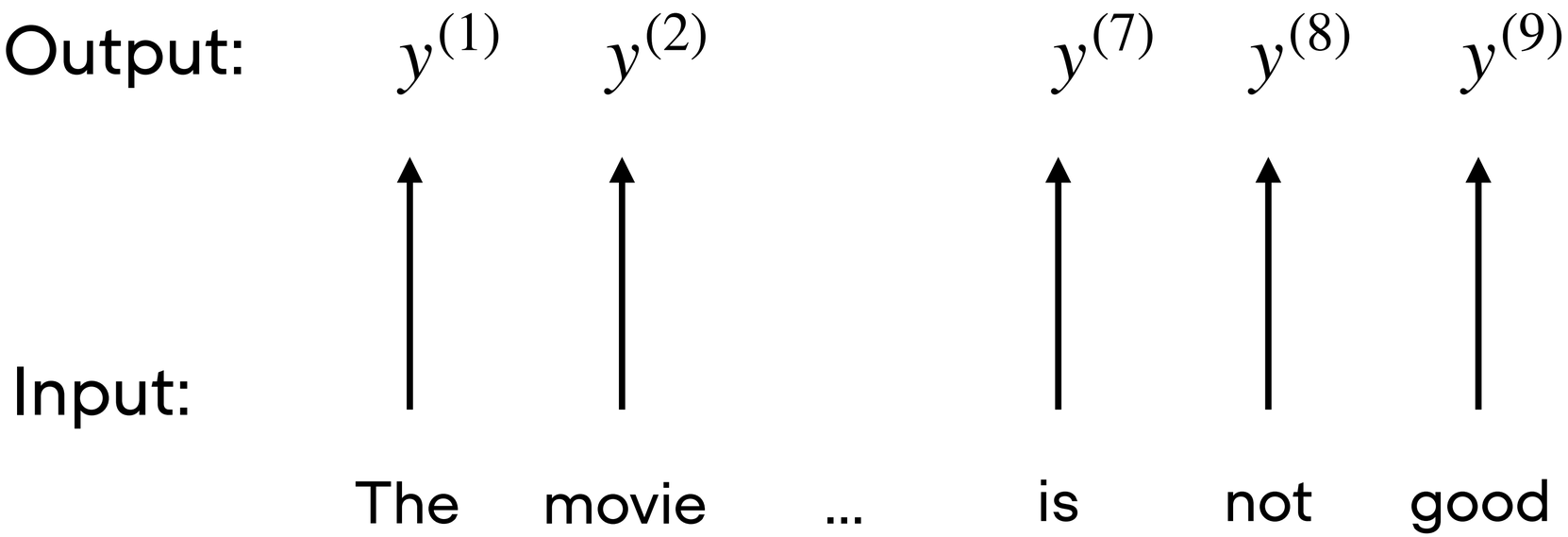
Machine learning models for text

Classic methods for tabular data:
- Logistic regression
- Multilayer perceptron

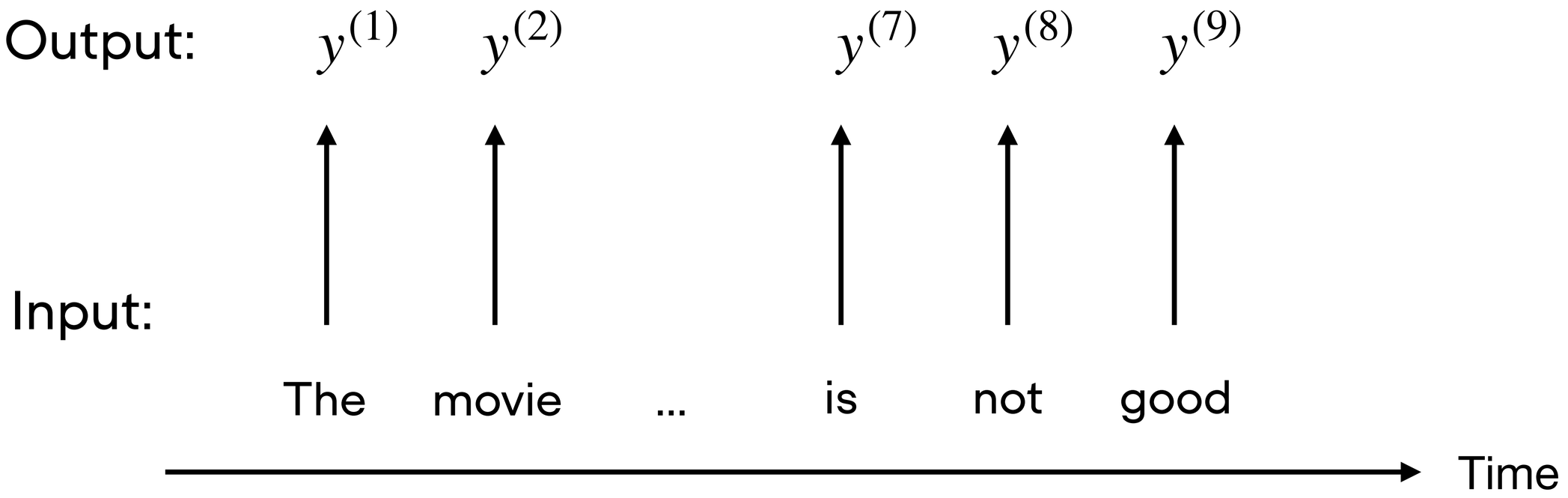
Sequence models:
- 1D convolutional network
- Recurrent neural network

Transformers

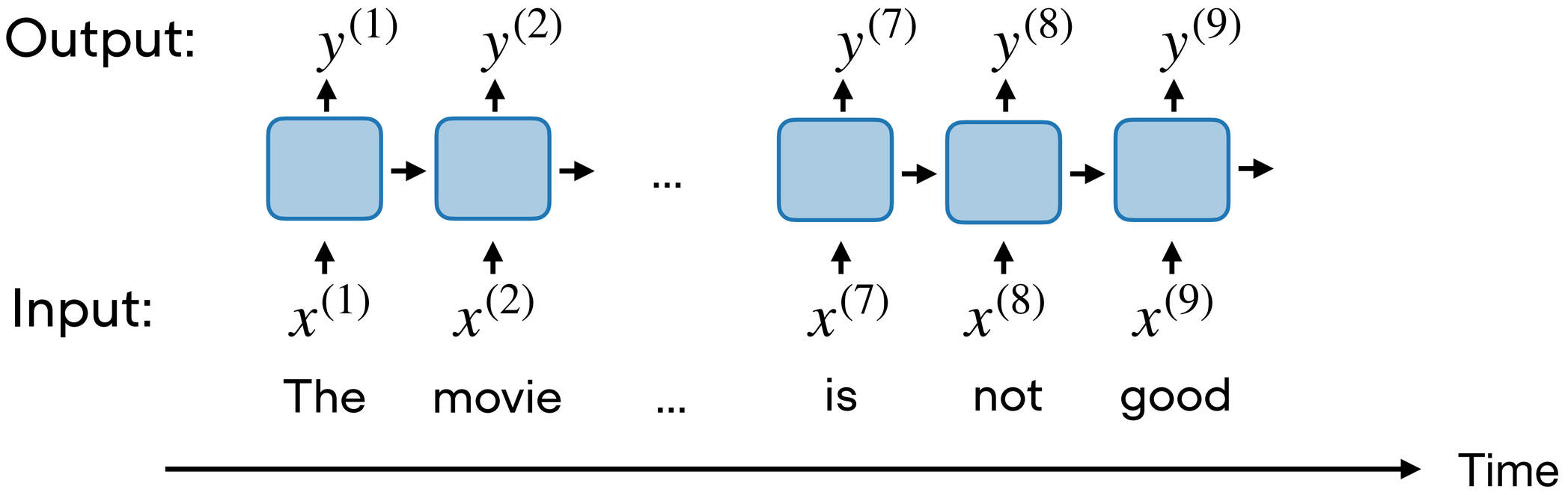
Recurrent neural networks (RNNS) for modeling sequences



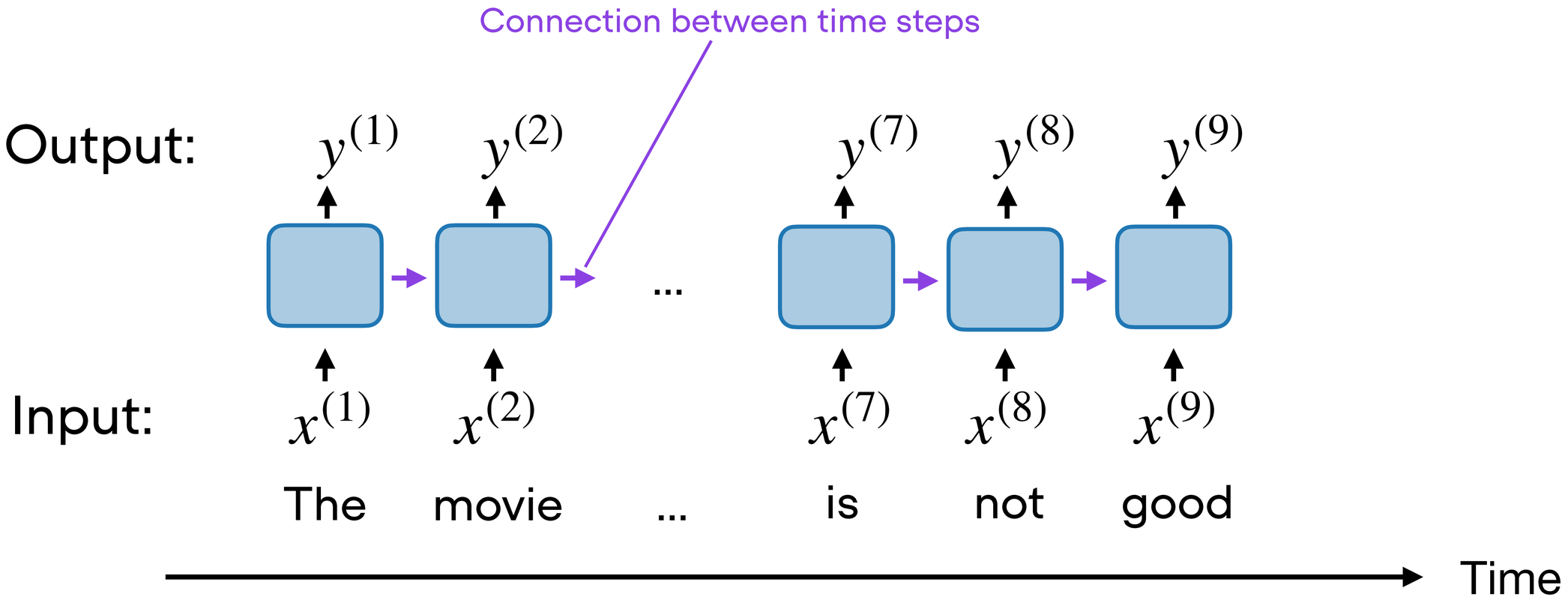
Recurrent neural networks (RNNS) for modeling sequences



Recurrent neural networks (RNNS) for modeling sequences



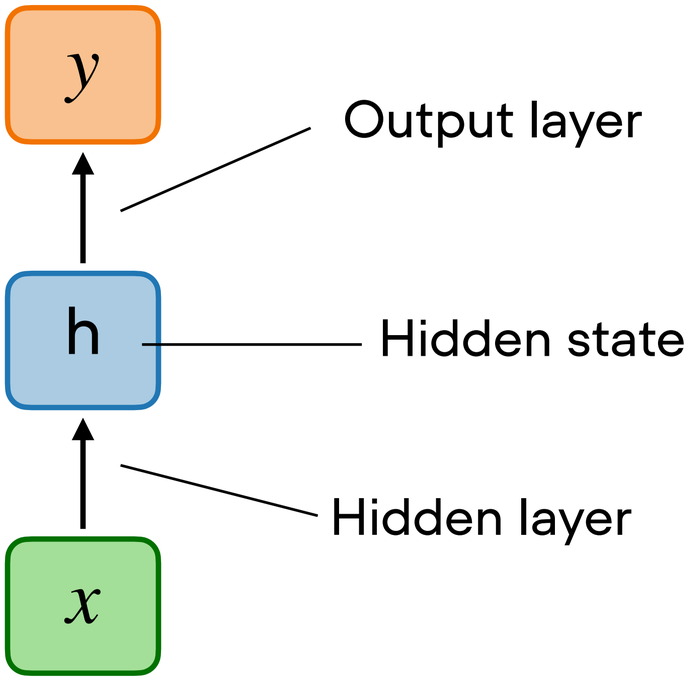
Recurrent neural networks (RNNS) for modeling sequences



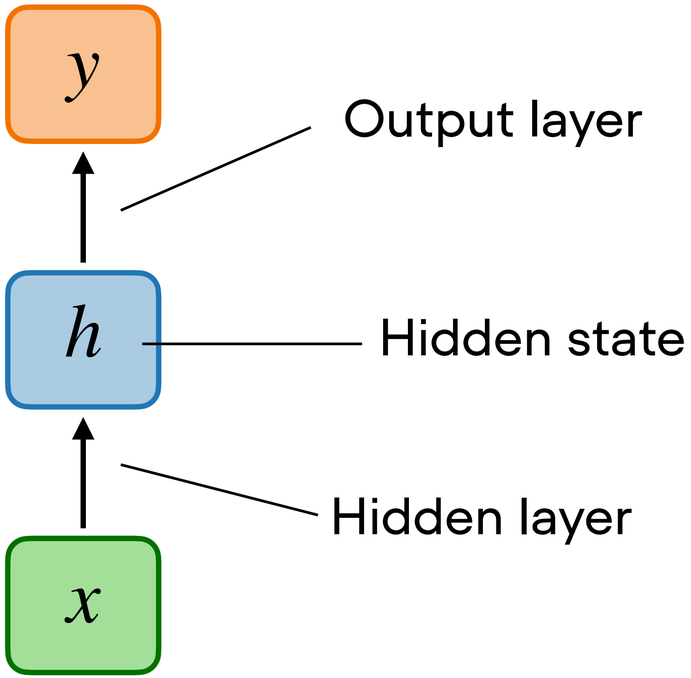
RNNs vs MLPS

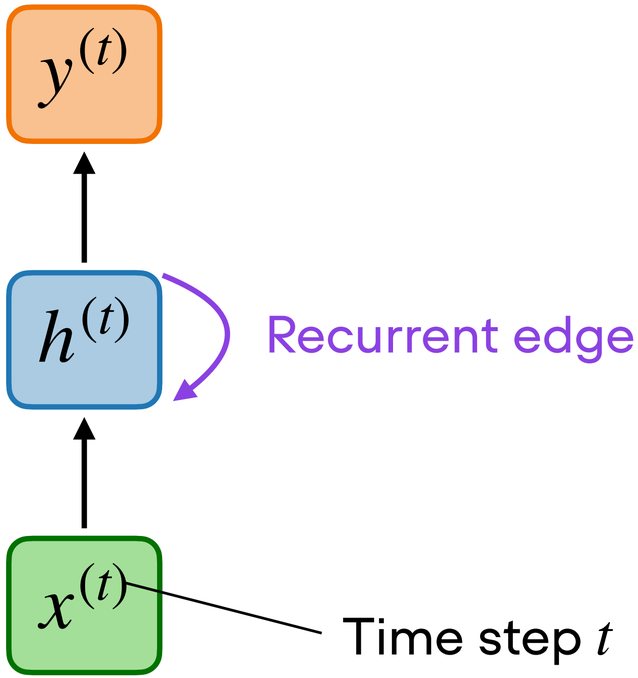
Recurrent neural networks (RNNS) for modeling sequences

Multilayer perceptron (unit 4)



Recurrent neural networks (RNNS) for modeling sequences

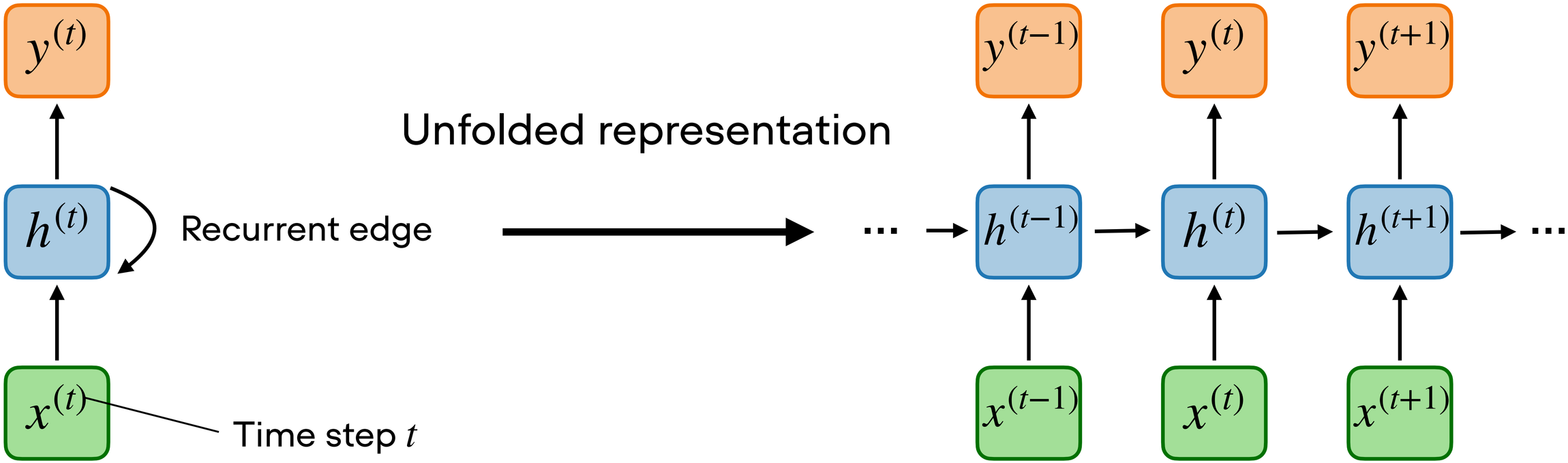




Multilayer perceptron (unit 4)

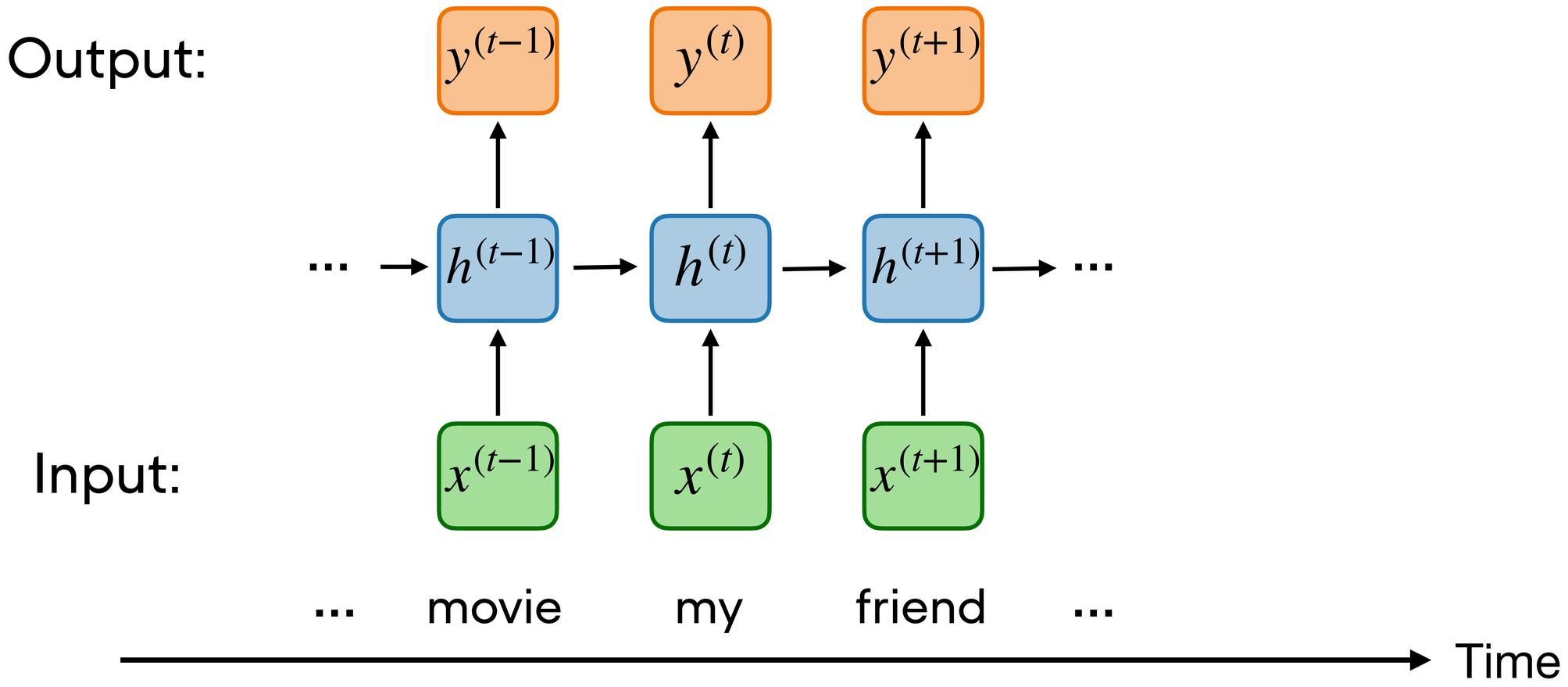
Recurrent neural network

Recurrent neural networks (RNNS) for modeling sequences

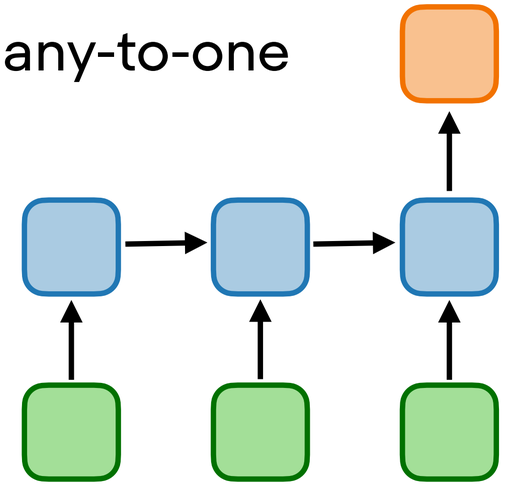


Recurrent neural network (RNN)

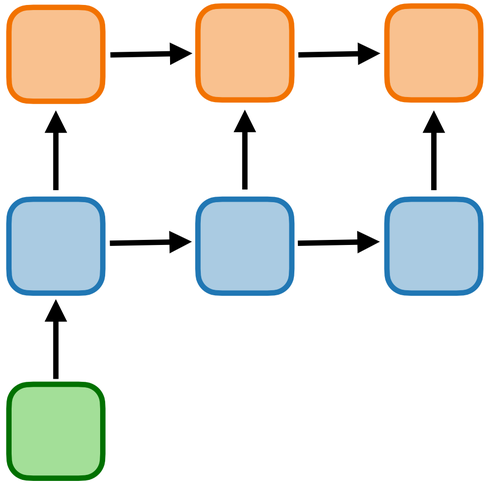
The same RNN


Output:
$y^{(t-1)}$
$y^{(t)}$
$y^{(t+1)}$
...
$h^{(t-1)}$
$h^{(t)}$
$h^{(t+1)}$
...
Input:
$x^{(t-1)}$
$x^{(t)}$
$x^{(t+1)}$
... movie my friend ...
Time

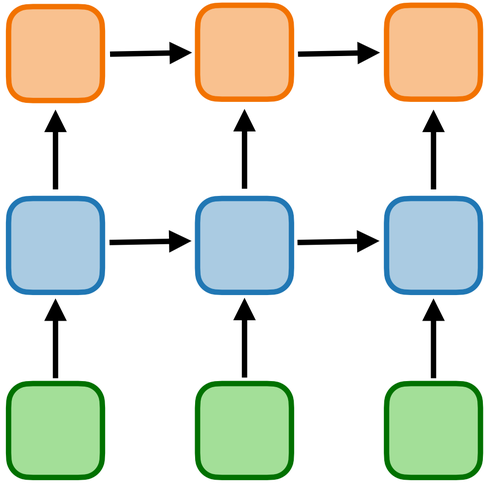
Many-to-one



One-to-many



Many-to-many



Many-to-many

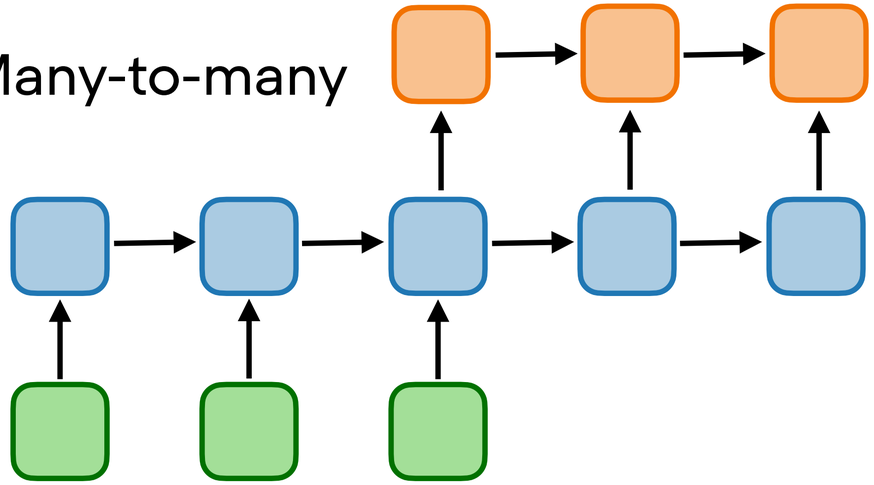


Figure inspired by
The Unreasonable Effectiveness of Recurrent Neural Networks by Andrej Karpathy (http://karpathy.github.io/2015/05/21/rnn-effectiveness/)

Next: The different modeling tasks