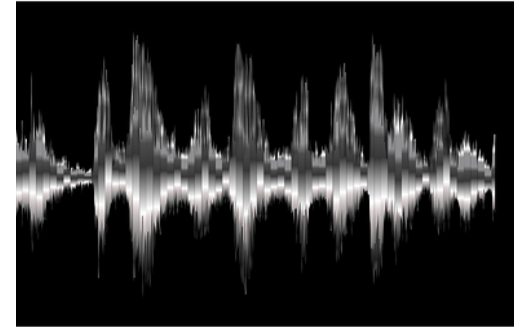


Deep Learning for sequential data

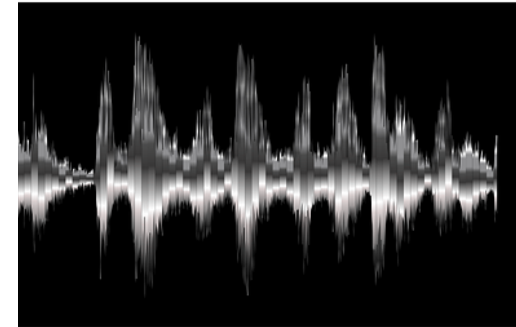
Sequential data

Text, Video, and Audio



Sequential data

Text, Video, and Audio



Time series: finance, industry, medicine...

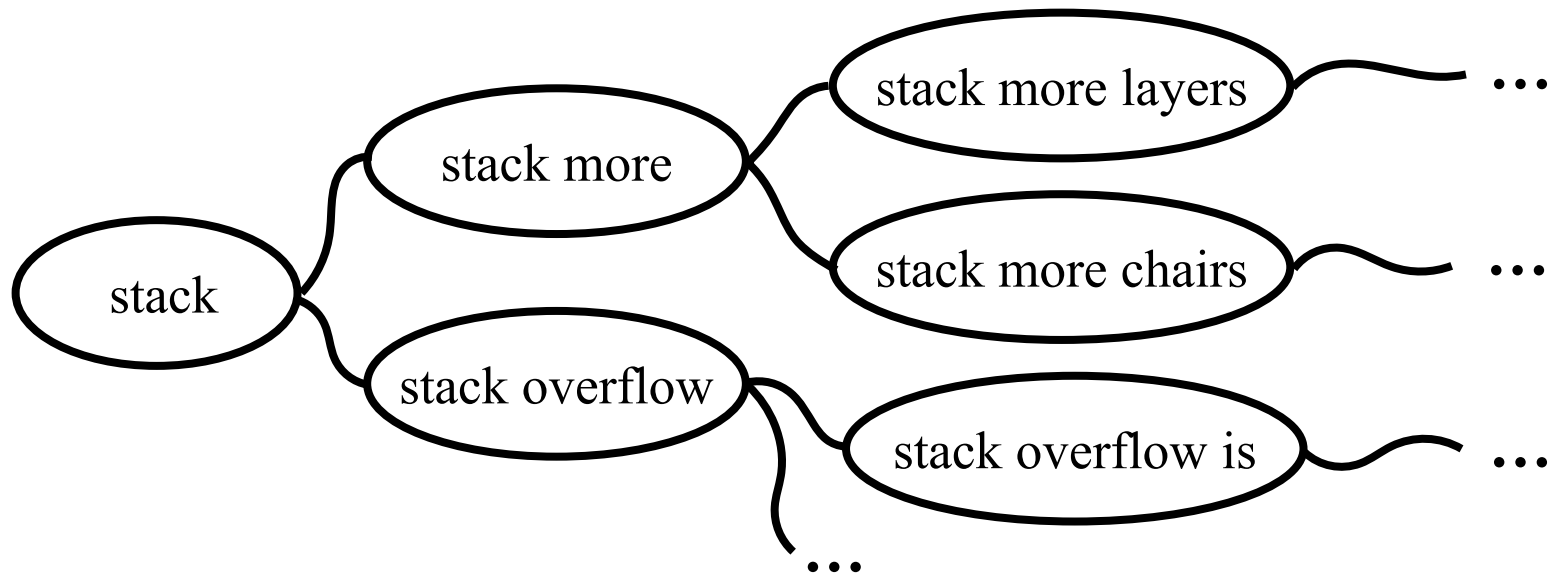


Sequences are everywhere!

Language model

We want to train a generative model of natural language

$$\begin{aligned} P(\text{text}) &= P(x_0, \dots, x_n) = \\ &= P(x_0)P(x_1|x_0)P(x_2|x_0, x_1) \dots P(x_n|\dots) \end{aligned}$$



Language model

We want to train a generative model of natural language

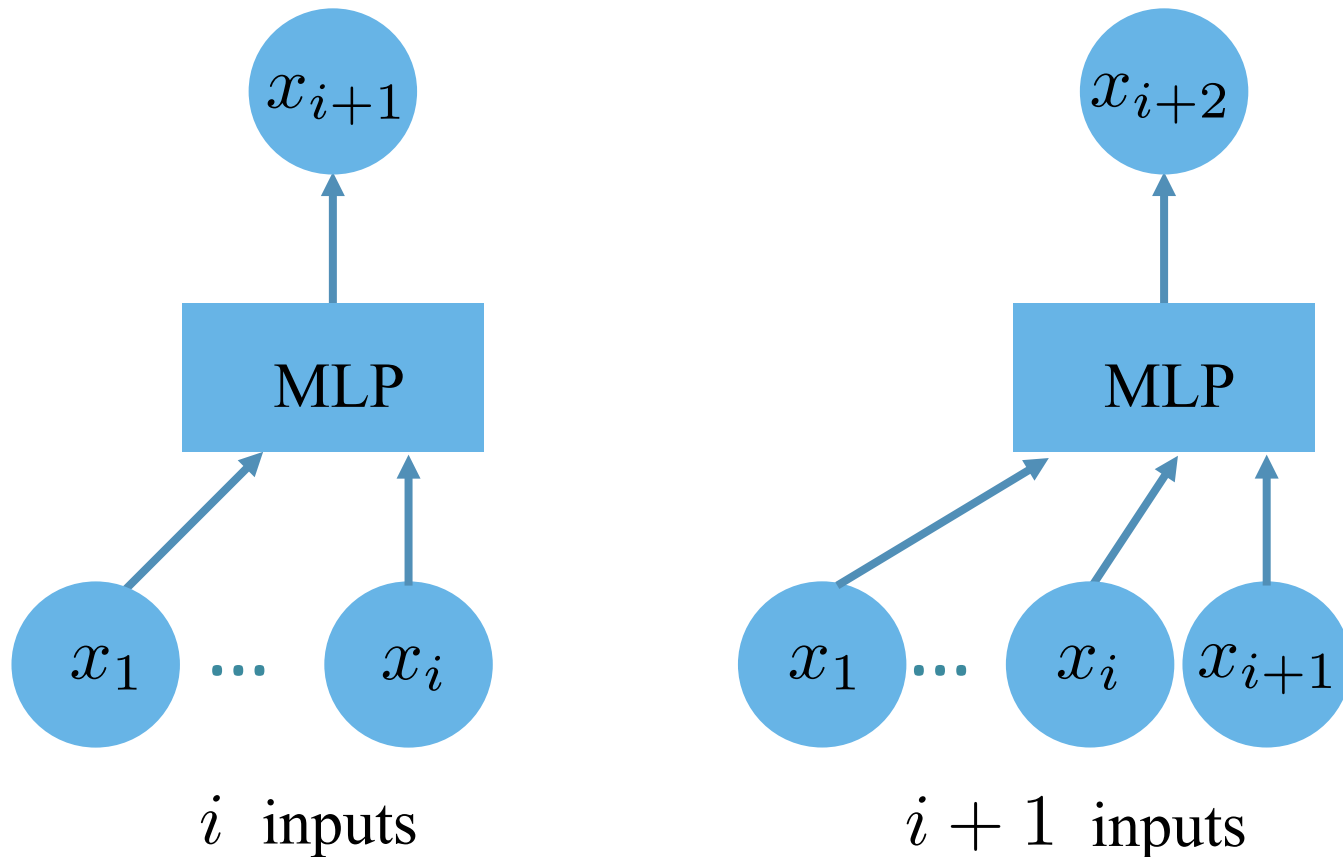
$$\begin{aligned} P(\textit{text}) &= P(x_0, \dots, x_n) = \\ &= P(x_0)P(x_1|x_0)P(x_2|x_0, x_1) \dots P(x_n|\dots) \end{aligned}$$

Why do we need it?

- Chatbots, question answering
- Machine translation
- Speech recognition
- Any text analysis you can imagine

Why not MLP?

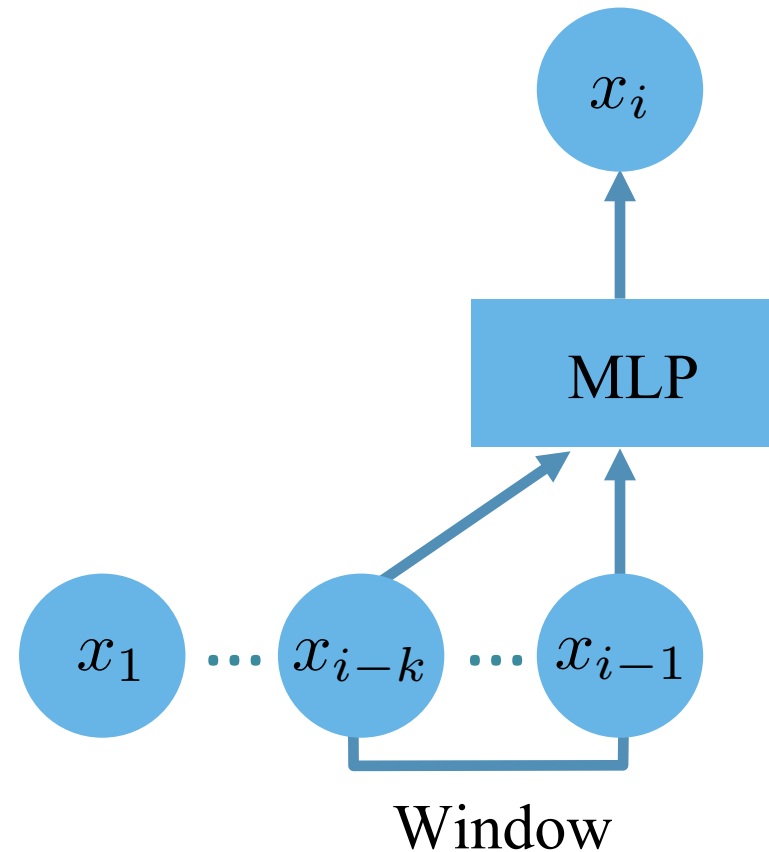
The main problem is arbitrary length of sequences:



How can we overcome it?

Why not MLP?

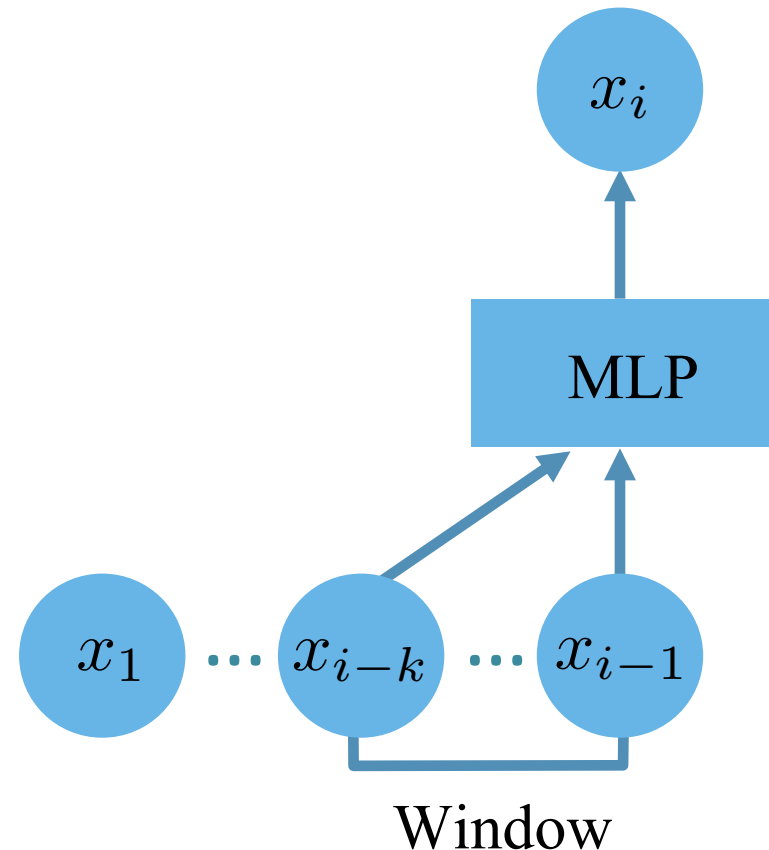
We can use a window of a fixed size as an input.



Why not MLP?

We can use a window of a **fixed size** as an input.

- This is just a heuristic and it is not clear how to choose the width of the window
- In some tasks we need very wide window therefore there is a problem with the large number of parameters



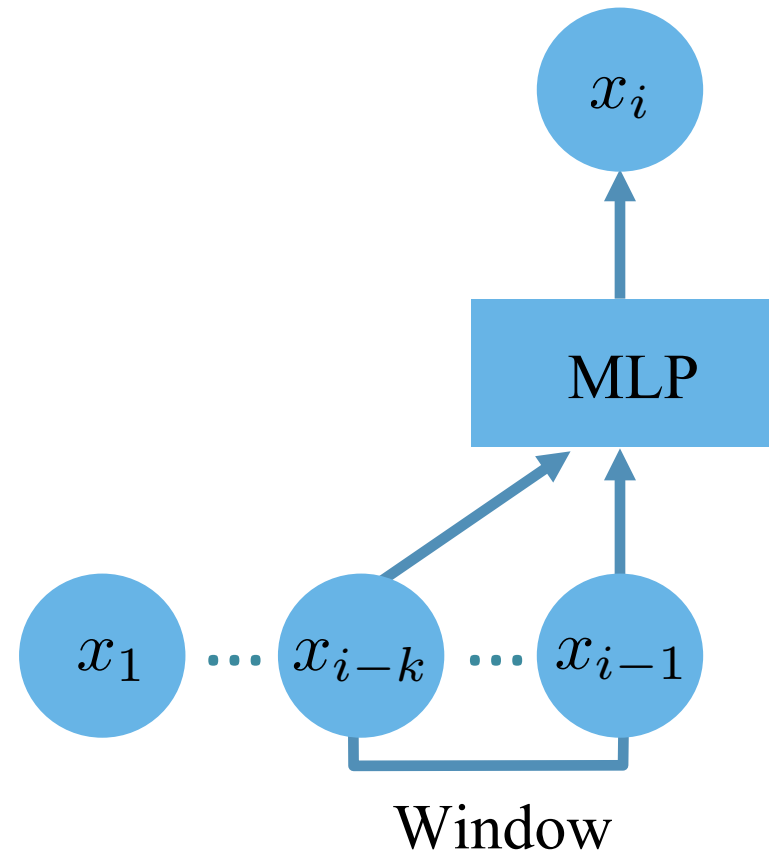
Why not MLP?

We can use a window of a **fixed size** as an input.

Question

How many weights are there in the first layer of the MLP?

- hidden neurons: 100
- window width: 100
- word embeddings size: 100



Why not MLP?

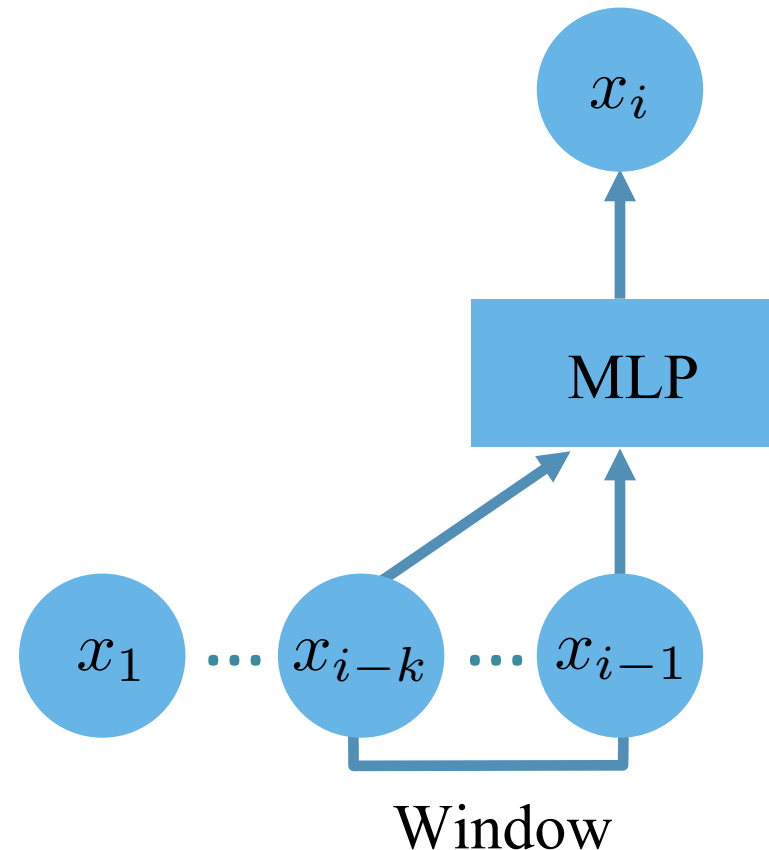
We can use a window of a **fixed size** as an input.

Question

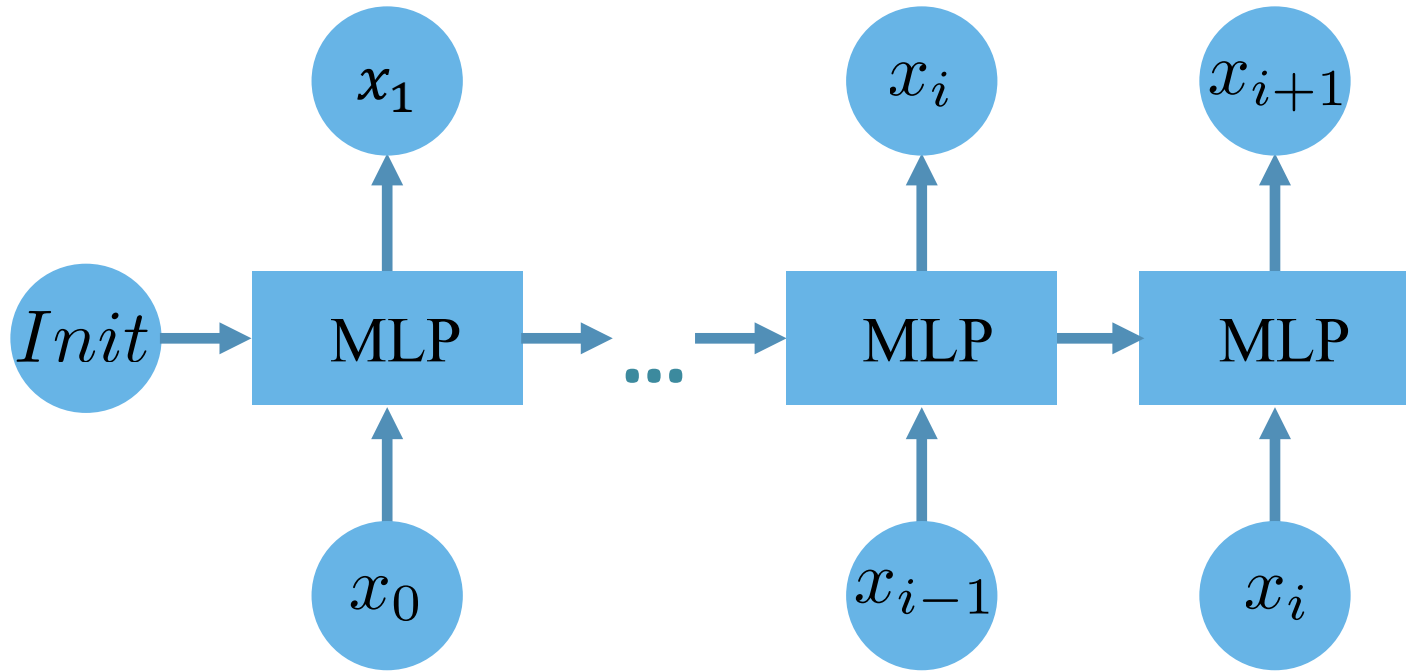
How many weights are there in the first layer of the MLP?

- hidden neurons: 100
- window width: 100
- word embeddings size: 100

More than a million!



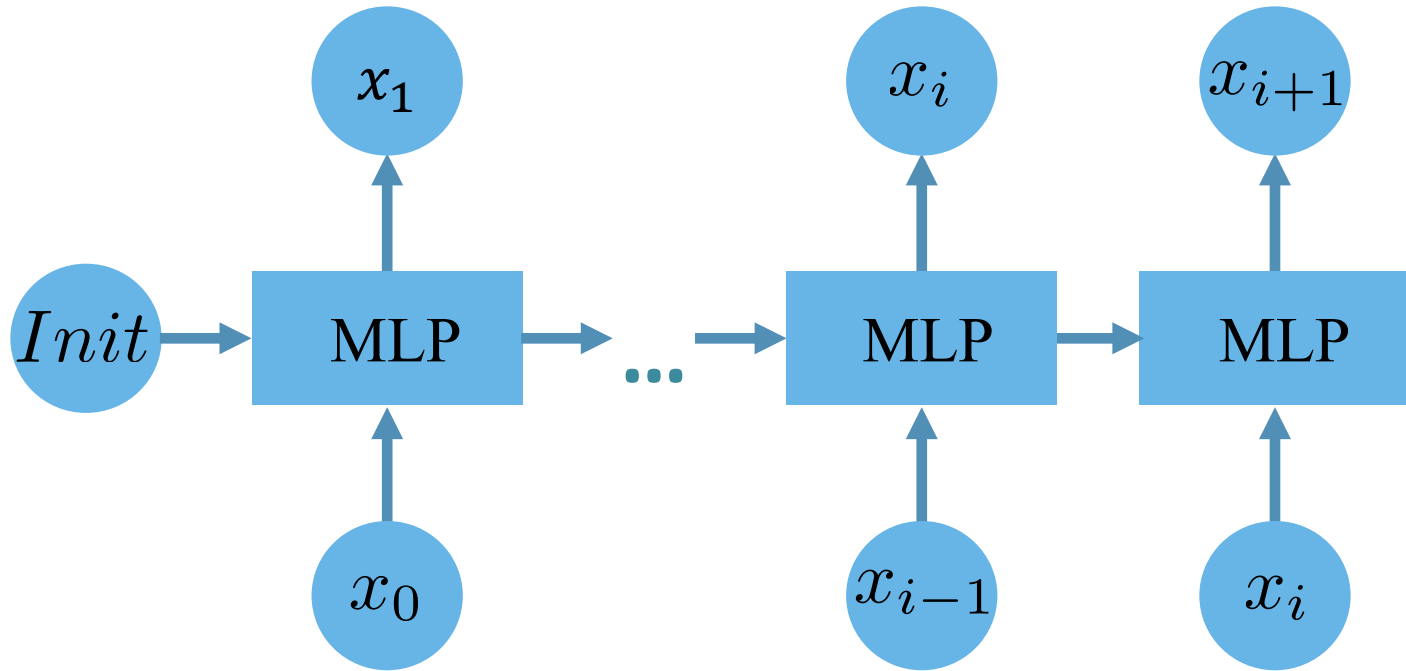
Recurrent Architecture



Problem #1: Arbitrary sequence length

Here: Fixed number of inputs at each time step.
At the first step we use some initial vector as an input from previous time step.

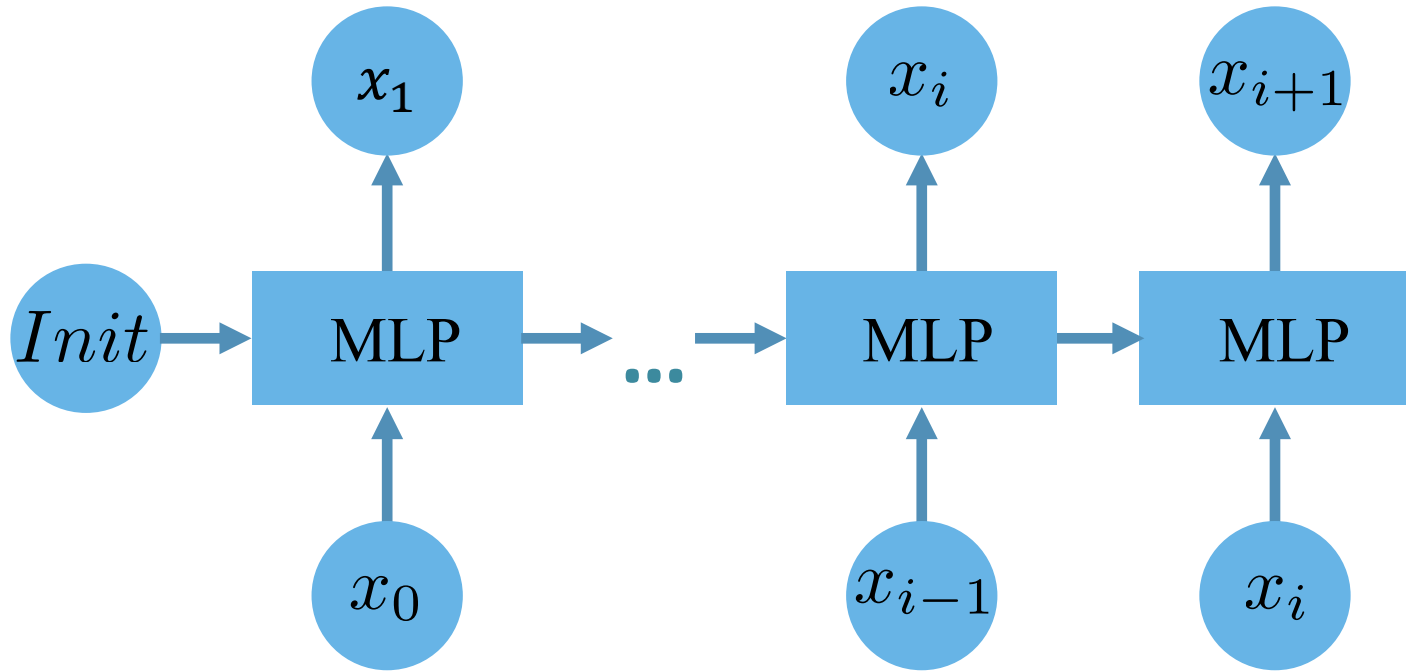
Recurrent Architecture



Problem #2: Large number of parameters

Here: All the parameters of an MLP are shared across the different time steps so we need a much smaller number of parameters.

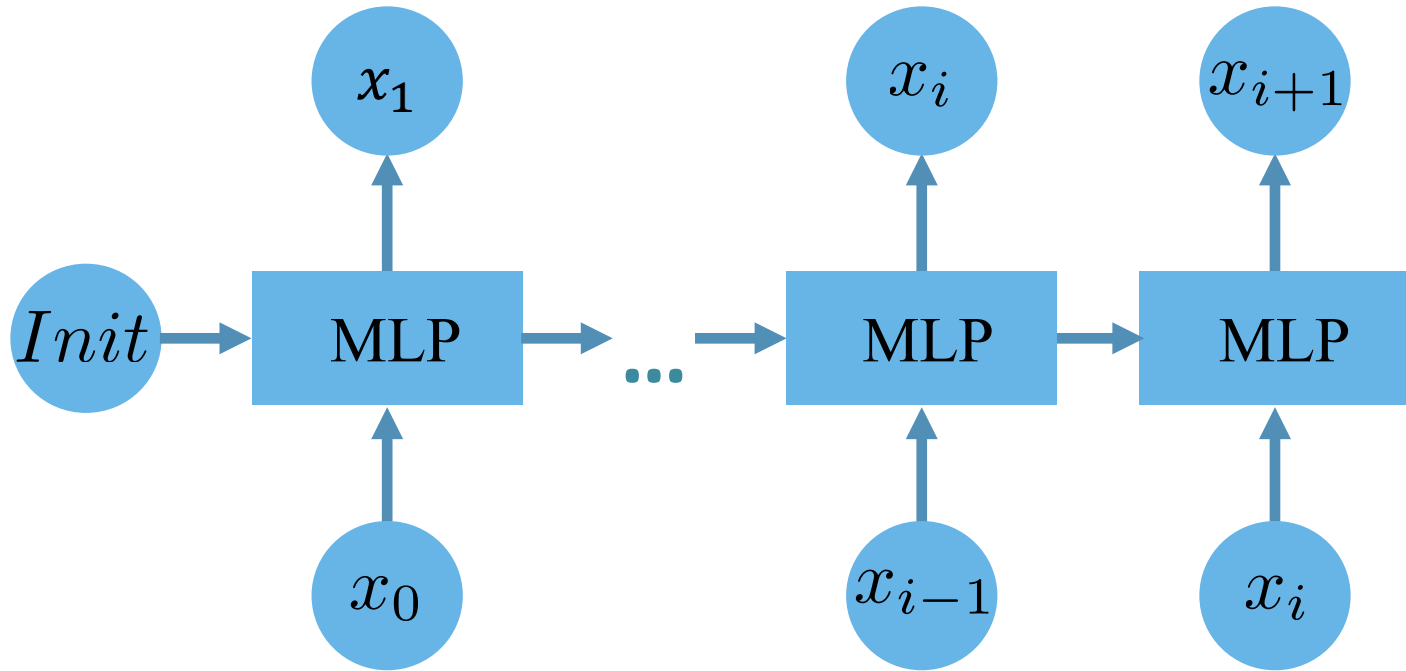
Recurrent Architecture



Question: How many weights are there in the first layer of the MLP?

- hidden neurons: 100
- word embeddings size: 100

Recurrent Architecture



Question: How many weights are there in the first layer of the MLP?

- hidden neurons: 100
- word embeddings size: 100

Only 20100!

Summary

- Sequential data is everywhere!
- Feedforward neural network isn't a very natural choice for such data because of arbitrary sequence length and large number of parameters
- Recurrent architecture is much more useful

In the next video:

Simple Recurrent Neural Network:
what is it and how to train it