



[Lecture 10. Recurrent Neural](#)

[Course](#) > [Unit 3 Neural networks \(2.5 weeks\)](#) > [Networks 1](#)

&gt;

2. Introduction

### **Audit Access Expires May 11, 2020**

You lose all access to this course, including your progress, on May 11, 2020.

Upgrade by Mar 25, 2020 to get unlimited access to the course as long as it exists on the site. **[Upgrade now](#)**

## 2. Introduction

### Introduction to Recurrent Neural Networks



## Video

[Download video file](#)

## Transcripts

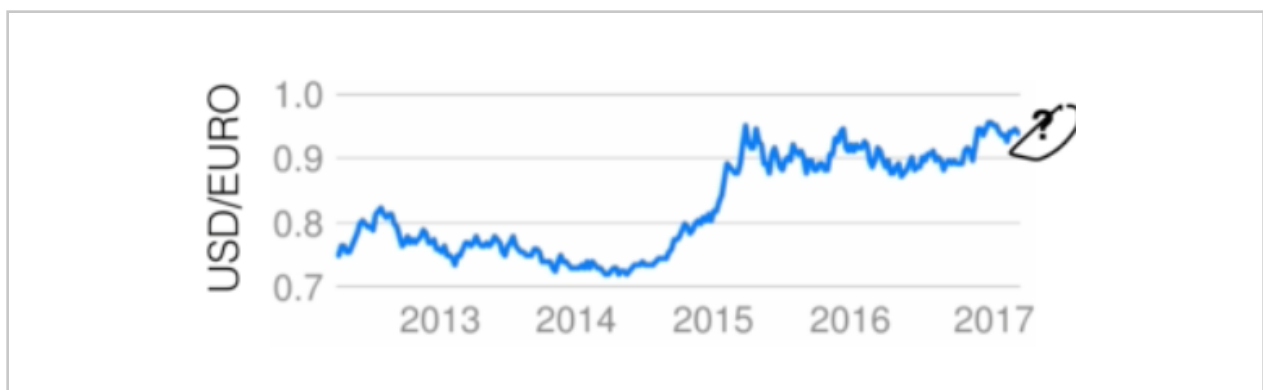
[Download SubRip \(.srt\) file](#)

[Download Text \(.txt\) file](#)

# Encoding Sequences with Feed-Forward Neural Networks

1/1 point (graded)

We have a temporal dataset of USD/EURO conversion rate from late 2012 to early 2017. Our goal is to predict the value of USD/EURO at the next timestep of early 2017.



If we are trying to encode the data into feature vectors for a feed-forward neural network, which of the following is the most viable strategy?

- ☒ slide a window of size 10 and use the most recent 10 points as a feature vector
- ☐ calculate the mean and the standard deviation of the entire sequence, and use them as a feature vector
- ☐ Use the length of the sequence and the standard deviation as a feature vector

**Solution:**

As discussed in the lecture, a common scheme to encode sequences is to use sliding windows and use data inside the most recent sliding window.

**Submit**

You have used 1 of 2 attempts

---

**i** Answers are displayed within the problem

---

## Context for Predicting Next Word

0/1 point (graded)

What is the issue with predicting the next word in the sentence using the previous **three** words as context?

(Choose all that apply.)

☐ Some words might need more context to predict ✓☐ Some words might need less context to predict, and additional words could be inefficient ✓☒ Some words might be closely related to words far away in the sentence ✓☐ Longer words are harder to predict because they have more letters

### Solution:

The amount of context we need to predict a word depends on the word. Therefore, some words could need more or less than 3 previous words as context. In addition, words could be related to other words far away in the sentence. Lastly, since we one-hot encode each of our words, the length of the word is irrelevant. All words, regardless of word length, have the same length one-hot encoding.

You have used 2 of 2 attempts

 Answers are displayed within the problem

## Discussion

**Topic:** Unit 3 Neural networks (2.5 weeks):Lecture 10. Recurrent Neural Networks 1 / 2. Introduction

by recent activity

There are no posts in this topic yet.



Learn About Verified Certificates

© All Rights Reserved