

# Unit 8 - Week 7

## Course outline

How to access the portal?

Week 1

Week 2

Week 3

Week 4

Week 5

Week 6

Week 7

Week 7 Lecture Materials

Introduction to Recurrent Neural Network

Unrolled RNN

RNN - Based Language Model

BPTT - Forward Pass

BPTT - Derivatives for W,V and U

BPTT - Exploding and vanishing gradient

LSTM

Truncated BPTT

GRU

Quiz : Assignment 7

Assignment 7 : Programming Exercise Self Assessment

Week 7 Feedback

Week 8

Week 9

Week 10

Week 11

Week 12

DOWNLOAD VIDEOS

## Assignment 7

The due date for submitting this assignment has passed. As per our records you have not submitted this assignment.

**Due on 2019-09-18, 23:59 IST.**

1) An RNN is fed with words,  $w_1, w_2, \dots, w_5$ , as input at every time step where  $t = 1, 2, \dots, 5$ . At  $t = 3$ , what does the RNN predict? 1 point

- $P(w_4)$   
  $P(w_3|w_1, w_2)$   
  $P(w_4|w_1, w_2, w_3)$   
 None of the above

No, the answer is incorrect. Score: 0

Accepted Answers:  $P(w_4|w_1, w_2, w_3)$

2) What are the advantages when characters are used instead of words as input to RNN? 1 point

- Vocabulary is restricted to characters and punctuation marks and white space  
 Learns words  
 Learns spellings of words  
 Softmax computation is not required at the output layer  
 Does not require word embeddings

No, the answer is incorrect. Score: 0

Accepted Answers: Vocabulary is restricted to characters and punctuation marks and white space

Learns words

Learns spellings of words

Does not require word embeddings

3) Is the statement "Vanishing gradient problem is easy to address than the exploding gradient problem" true? 1 point

- True  
 False

No, the answer is incorrect. Score: 0

Accepted Answers: False

4) Consider the following dimensions used in an RNN to answer the question: 1 point

Embedding vector size =  $1 \times 50$   
 Number of time states = 1000  
 Hidden layer size =  $1 \times 100$   
 Vocabulary size = 10000

What is the size of the context matrix which is connecting the hidden layer and the output layer?

- $R^{50 \times 1000}$   
  $R^{100 \times 100}$   
  $R^{100 \times 10000}$   
  $R^{50 \times 10000}$

No, the answer is incorrect. Score: 0

Accepted Answers:  $R^{100 \times 10000}$

5) The cost function  $J(\theta)$  computed at the output layer in the RNN (consider the notations used in the lecture) is usually represented as a function of 1 point

- $U, V, h_t$   
  $U, V, s_t$   
  $U, V, W$   
 None of the above

No, the answer is incorrect. Score: 0

Accepted Answers:  $U, V, W$

6) In the BPTT, the  $\frac{\partial z_t}{\partial V}$  is \_\_\_\_\_ 1 point

- $\tanh(Uh_{t-1} + Wx_t)$   
  $Uh_{t-1} + Wx_t$   
  $1 - \tanh^2(h_t)$   
 None of the above

No, the answer is incorrect. Score: 0

Accepted Answers:  $\tanh(Uh_{t-1} + Wx_t)$

7) In the BPTT, the  $\frac{\partial s_t}{\partial h_t}$  is \_\_\_\_\_ 1 point

- $Uh_{t-1} + Wx_t$   
  $\tanh^2(h_t)$   
  $1 - \tanh^2(Uh_{t-1} + Wx_t)$   
 None of the above

No, the answer is incorrect. Score: 0

Accepted Answers:  $1 - \tanh^2(Uh_{t-1} + Wx_t)$

8) The matrix consisting of the derivative  $\frac{\partial h^{[r]}}{\partial h^{[r-1]}}$  is known as \_\_\_\_\_ matrix 1 point

- Laplacian  
 Jacobian  
 Hermitian  
 Ramaseshan

No, the answer is incorrect. Score: 0

Accepted Answers: Jacobian

9) Vanishing gradient problem occurs due to \_\_\_\_\_ 1 point

- Vocabulary size > 100K  
 Hidden layers of size > 10  
 A very long time series as input or long term dependencies  
 None of the above

No, the answer is incorrect. Score: 0

Accepted Answers: A very long time series as input or long term dependencies

10) LSTM introduces a \_\_\_\_\_ at the hidden layer 1 point

- State vector  
 Memory vector  
 Word Vector  
 None of the above

No, the answer is incorrect. Score: 0

Accepted Answers: Memory vector