

Theoretical Tasks

November 20, 2017

1 Learning Sequences

Just as a fun task, speak out your phone number in your head. Pretty simple right! Now try to speak out the same phone number in reverse, it is not so easy now, is it. This is just one simple example to show that we might be learning things in a sequence. It is still a debatable hypothesis but is quite interesting to think about.

To learn and predict any kind of sequential information using Neural Networks, simple Feed Forward Networks are not enough, but, Recurrent Neural Networks(RNN) are specialized networks that we use for learning patterns in sequences.

Assuming that you have understood how to use Recurrent Neural Networks and some variants of it from the lecture, this task will help you understand them a little better.

1.1 Task 1

Consider a simple RNN with one hidden layer (Figure 1) and the following assumptions:

1. Hidden Layer Size: 128
2. Output Layer Size: 10
3. Input Vector Size at each Time Step: 20

1.1.1 Question 1

Calculate the number of parameters required to unroll this vanilla RNN for:

1. 5 time steps
2. 10 time steps

1.1.2 Question 2

Do the number of parameters increase as you increase the number of time steps to unroll? Explain your answer.

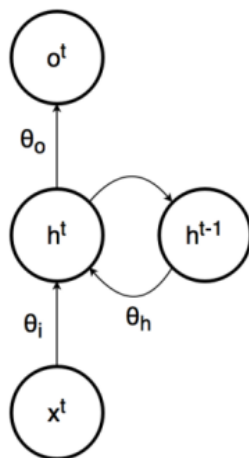


Figure 1: A vanilla RNN with a single hidden layer

1.2 Task 2

1.2.1 Question 1

Ideally you would expect an RNN model to learn long term dependencies in sequences. But in reality, that does not happen. Explain the reasons objectively. Try to be very precise here.

1.2.2 Question 2

State the variants of an RNN that you know of and give a single line explanation for each of them.

1.3 References

1. The Unreasonable Effectiveness of Recurrent Neural Networks: <http://karpathy.github.io/2015/05/21/rnn-effectiveness/>

2 LSTM

This task will help you in understanding LSTMs better by understanding the internal structure of the LSTM cell.

Consider an LSTM cell

2.1 Task 1

2.1.1 Question 1

What was the motivation behind using gates in an LSTM?

2.1.2 Question 2

What will happen if you manually set the output of the forget gate to 0. Try to give a formal explanation using the equations from the question above.

2.1.3 Question 3

How does an LSTM solve the vanishing gradient problem. Give a precise answer.

2.1.4 Question 4

How can you reduce the number of parameters of an LSTM cell?

2.1.5 Question 5

What is a Bidirectional LSTM and why is it useful? Try to be as formal as possible while giving an answer to this question.

2.2 References

1. Understanding LSTM Networks: <http://colah.github.io/posts/2015-08-Understanding-LSTMs/>

3 Acknowledgement

We would like to thank Saurabh Varshneya for being an active member of MindGarage and for us helping in making this exercise.