
A4 homework submission

David Halpern

Deep Learning 2015, Spring

David Halpern
Department of Psychology
New York University
david.halpern@nyu.edu

1 Questions

1.1 Q2

$i = i$
 $\text{prev_c} = c_{t-1}^l$
 $\text{pre_h} = h_{t-1}^l$

1.2 Q3

Returns an rolled? version of the core LSTM module with parameters initialized uniformly within an interval.

1.3 Q4

model.s is the trained of the sequence. model.ds is the gradients of the model for the backward pass. model.start_s is the final hidden states of the current minibatch and gets reset when the sequence ends

1.4 Q5

Clipped at $\text{params.max_grad_norm}$ (shrink factor?)

1.5 Q6

Batch SGD

2 Experiments

2.1 1

Network parameters:

```
{  
  max_grad_norm : 7  
  seq_length : 50  
  batch_size : 50  
  lr : 1
```

```
max_max_epoch : 13
rnn_size : 400
init_weight : 0.1
decay : 2
dropout : 0.2
layers : 2
vocab_size : 50
max_epoch : 4
}
```

References

- [1] Pennington, Jeffrey, Socher, Richard, & Manning, Christopher D. (2014). Glove: Global vectors for word representation. Proceedings of the Empirical Methods in Natural Language Processing.
- [2] Xiang Zhang and (2015). Text Understanding from Scratch. CoRR, abs/1502.01710, .