Quiz Submissions - Quiz 8 - Attempt 1

Nhat Le (username: hung.le@mail.mcgill.ca)

Attempt 1

Written: Mar 21, 2019 7:36 PM - Mar 21, 2019 7:45 PM

Submission View

Released: Mar 20, 2019 11:59 PM

View the quiz answers.

Question 1 **1 / 1 point**

[Select the correct response to fill in the blank]

LSTMs solve the [BLANK] problem that plagues standard Elman RNNs.

- ✓ vanishing and exploding gradients
- computational efficiency
- multi-output prediction
- local minima

Hide Feedback

As discussed in Lecture 17, LSTMs help to solve the vanishing and exploding gradients problem.

Question 2 **0 / 1 point**

Elman and Jordan-type recurrent connections can be used simultaneously for **sequence classification** tasks. In other words, it is possible to combine both the recurrent connections used in Elman and Jordan RNNs for tasks, such as sentiment classification, where there is a single label associated with the entire sequence (i.e., where there are **not** individual outputs associated with each timepoint).

- ✗ True
- ➡ False

Hide Feedback

Jordan RNNs are only well-defined for sequence labeling tasks (i.e., tasks where these is an output at each timestep), since in a Jordan RNN the input at timestep *t* is the predicted output from timestep *t-1*.

Question 3 **0 / 1 point**

The "teacher forcing" technique is an alternative to stochastic gradient descent, used to train encoder-decoder models.

- ✗ ○ True
- ➡ ○ False

▼ Hide Feedback

Teaching forcing does not replace SGD; it is used to improve SGD-based training. I.e., it is complimentary to SGD.

Question 4 **0 / 1 point**

Which of the following approaches would **not** be helpful for dealing with exploding gradients during stochastic gradient descent:

- ○ Gradient clipping
- ○ Replacing Elman RNNs with LSTMs
- ✗ ○ Applying L2 regularization
- ➡ ○ None of the above

▼ Hide Feedback

All of the proposed approaches would be helpful for preventing exploding gradients. Both gradient clipping and LSTMs were discussed in lecture. L2 regularization will reduce the magnitude of the weights in the RNN parameter matrix, which will also help to reduce gradient explosion (since the weights will tend to be smaller in magnitude). However, L2 regularization could lead to vanishing gradients for this reason,

(Recall that L2 regularization penalizes according to the Frobenius norm when we are working with parameter matrices. To build on the above intuition, you can actually prove that the Frobenius norm gives an upper bound on the sum of the magnitudes of the eigenvalues, which means that constraining the Frobenius norm will generally lead to smaller eigenvalues. This proof is outside the scope of this course, however)

Question 5 **1 / 1 point**

Which of the following statements is false:

- ○ Elman RNNs are generally preferred over Jordan RNNs.
- ○ Teacher forcing can only be applied during training.

- [ ] None of the above

- [x] LSTMs help to solve the vanishing gradient problem because they use multiplicative instead of additive updates on the cell state.

Hide Feedback

LSTMs help to solve the vanishing gradient problem because they use **additive,** not multiplicative interactions.

Attempt Score: 2 / 5 - 40 %

Overall Grade (highest attempt)**:** 2 / 5 - 40 %

Done