EECS545 Lecture 13 Quiz Solutions

1. Which of the following statements are INCORRECT about RNN/LSTM/GRU? (**Choose all options that apply**)

   (a) Recurrent neural networks can handle a sequence of arbitrary length, while feedforward neural networks can not.

   (b) Training recurrent neural networks is hard because of vanishing and exploding gradient problems.

   (c) Gradient clipping is an effective way of solving the vanishing gradient problem.

   (d) GRU is computationally more efficient than LSTMs if the hidden dimension size for LSTM and GRU are the same.

   Solution: (c). Gradient clipping is only a solution for solving exploding gradient problems, not vanishing gradient problems.

2. What are the benefits of using an RNN over applying CNN over time (i.e., 1D CNN)? (**Choose all options that apply**)

   (a) RNNs require less inference time than CNNs in Language modeling.

   (b) RNNs can better understand the sequential dependencies.

   (c) RNNs can better handle sequences with unknown lengths.

   (d) Training RNNs are easier than CNNs as RNNs are less likely to have gradients explode/vanish issue.

   Solution: (b) and (c)
   (a): RNNs need to infer each word one by one during inference time.
   (d): RNNs are more likely to suffer gradient explode/vanish issues.

3. Which tricks can help address at least one of the exploding/vanishing gradient problem? (**Choose all options that apply**)

   (a) Use LSTM instead of vanilla RNN.

   (b) Use sigmoid activation instead of tanh in vanilla RNN.

   (c) Use orthogonal initialization of model weights.

   (d) Use gradient clipping.

Solution: (a) (c) (d). Please revisit the slides from page 44). (b) would still suffer exploding/vanishing gradient issues; tanh can be written as $2\sigma(2x)$-1 if sigmoid is $\sigma(x)$.

4. (True/False) The vanishing gradient in RNN could cause the parameters to be biased to capture short-term dependencies.

Solution: True. Long-term dependencies in the sequence become affected as the gradient vanishes in the middle (Slide 47)

5. (True/False) Unlike GRU, LSTM does not use a memory unit to handle sequential data.

Solution: False. LSTM does have the memory unit, but GRU does not.